

ACTES DU COLLOQUE

*Les textes des présentations au colloque sont publiés dans ces actes en anglais
ou en français.*

Deux papiers ont été traduits en français. Les versions originales sont en annexe.

TABLE DES MATIÈRES

Ouverture du Colloque	6
Allocution d'ouverture - Omar Azziman - Président du CSEFRS	7
Introduction - Rahma Bourqia - Directrice de l'INE - CSEFRS	9
1. L'enjeu de l'évaluation	13
■ Cognitive skills and Economic Growth: Heterogeneity Matters, <i>Nadir Altinok (France)</i>	14
■ Review of International and Regional Assessment Programs, <i>Marguerite Clarke (USA)</i>	25
■ Summary of the OECD Reviews on Evaluation and Assessment in Education, <i>Elizabeth Fordham, Paulo Santiago, Soumaya Maghnouj (OCDE)</i>	33
■ L'évaluation des politiques scolaires : un instrument au service de la construction des politiques scolaires de long terme ? <i>Nathalie Mons (France)</i>	50
2. Instances et Standards de l'évaluation	53
■ Ensuring the Relevancy of Accreditation Standards in the U.S.: A case study, <i>Paula Harbecke (USA)</i>	54
■ The new ANVUR Accreditation and QA framework, ongoing experience and caveats for its effective implementation, <i>Muzio Gola (Italie)</i>	62
■ AlmaLaurea's Tools for the Improvement of Quality Assurance in the Higher Education Systems, <i>Francesco Ferrante (Italie)</i>	72
■ L'examen des compétences universitaires et son impact sur les établissements de l'enseignement supérieur en Jordanie, <i>Bashir Al Zu'bi (Jordanie)</i>	79
■ L'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique du Maroc : une nouvelle expérience, <i>El Mokhtar Bakkour (Maroc)</i>	89
3. Techniques de mesures d'évaluation	99
■ L'évaluation formative en français : analyse métrologique, <i>Pierre Valois - Belkacem Abdous (Canada)</i>	100
■ Analyse des items, <i>Amina Benbiga - Larbi Karani (Maroc)</i>	114
■ Evaluation des apprentissages et amélioration de l'éducation : l'expérience argentine: ONE et IMESA, <i>Marisa Diaz (Argentine)</i>	124
4. Evaluation de l'enseignement supérieur	134
■ Safeguarding Quality of Higher Education: the case of the Netherlands, <i>Joop Hartog (Pays-Bas)</i>	135
■ L'évaluation des institutions de l'enseignement supérieur : une description et un retour d'expériences de trois projets pilotes, <i>Abdelaali Kaouachi (Maroc)</i>	145
■ L'évaluation de l'insertion des diplômés du supérieur au Maroc : le rendement de l'université, <i>Saad-Ellah Berhili (Maroc)</i>	157
■ Enjeux et expériences de l'évaluation dans le supérieur, <i>Mustapha Bennouna (Maroc)</i>	170

5. Evaluations et politiques publiques	178
■ Les évaluations en éducation : leurs effets sur les politiques éducatives, <i>Georges Felouzis (Suisse)</i>	179
■ Evaluation d'impact des politiques de réforme du système éducatif marocain : problématiques de conception et de mise en œuvre, <i>Abdelhaq El Hayani (Maroc)</i>	184
■ Modes de gouvernance, pratiques d'évaluation et leur impact sur les politiques éducatives publiques, <i>Amina Debbagh (Maroc)</i>	201
■ La Cour des Comptes : méthodes d'évaluation des institutions éducatives, <i>Mohamed Hadhoudi (Maroc)</i>	212
■ Les approches adoptées par l'ONDH pour évaluer les politiques publiques en éducation, <i>Ikbal Sayah (Maroc)</i>	220
Conclusion	225
■ Albert Sasson (<i>Maroc</i>)	
Annexes	227
■ Papiers - Version originale	
- Evaluacion de aprendizajes y mejora de la educacion. La experiencia de Argentina: ONE e IMESA- Marisa Diaz	228
- امتحان الكفاءة الجامعية وأثره على مؤسسات التعليم العالي الأردنية - بشير الزعبي	237

OUVERTURE



ALLOCUTION D'OUVERTURE

Omar Azziman *

Mesdames et Messieurs les Ministres,

Mesdames et Messieurs les experts,

Mesdames et Messieurs,

C'est avec un grand plaisir que je souhaite la bienvenue à tous les participants au Colloque International organisé par l'Instance Nationale d'Évaluation auprès du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique, sur le thème de «**L'évaluation en éducation et en formation : approches, enjeux et défis**».

Ce colloque représente une réelle opportunité, car il permet :

- Premièrement, de rassembler des experts mondiaux, des spécialistes du domaine, des chercheurs universitaires, et des responsables publics, marocains et étrangers, qui partagent les mêmes préoccupations dans le domaine de l'éducation et de la formation en général et de l'évaluation en particulier, considérée comme un choix méthodologique et un outil pour faire évoluer les politiques publiques en matière éducative ;
- Deuxièmement, d'assurer la participation du Maroc à l'année internationale de l'évaluation dans le but de soutenir le développement d'un environnement favorable à l'évaluation aux niveaux international, national et local.

A cette occasion, je souhaite adresser mes vifs remerciements à nos invités, marocains et étrangers. Je remercie également l'Instance Nationale d'Évaluation auprès du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique à travers sa Directrice, Madame Rahma Bourqia, ainsi que ses équipes pour leurs efforts afin de mettre à la disposition du Conseil des évaluations thématiques crédibles, permettant ainsi à ce dernier de remplir ses missions et à notre système éducatif de s'engager dans la voie du progrès.

Mesdames, Messieurs,

Cette rencontre internationale constitue un événement important pour notre Conseil, à l'heure où le Maroc se trouve engagé dans des chantiers structurants. Je citerai en premier lieu :

- La nouvelle Constitution de 2011 qui a apporté plusieurs nouveautés, parmi lesquelles l'intégration de l'évaluation des politiques publiques dans la liste des prérogatives du Parlement, tout en accordant la possibilité à d'autres institutions constitutionnelles d'intégrer l'évaluation dans leur domaine d'activité.

Ce tournant important que connaît le Maroc contribuera certainement à l'augmentation de la demande des travaux d'évaluation, ce qui entraînera assurément le renforcement du rôle de notre Conseil en tant qu'institution constitutionnelle dont la vocation est de contribuer à l'évaluation des politiques publiques dans le domaine de l'éducation, de la formation et de la recherche scientifique, dans le but d'augmenter l'efficacité et la performance du service public.

* Président du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique

Quant au second chantier, il est plus récent et concerne l'évolution de notre système éducatif. En effet, le Maroc possède désormais une **Vision Stratégique** pour la réforme du système éducatif qui intègre l'évaluation comme choix méthodologique décisif pour assurer la réussite de la réforme de l'école marocaine.

Ce choix méthodologique a été celui qui a présidé à l'élaboration de l'évaluation de la Charte Nationale d'Education-Formation (2000 et 2013). Ceci a permis d'identifier les dysfonctionnements et les contraintes et a également permis de trouver les solutions, de dessiner les grands choix pour réhabiliter l'école marocaine et améliorer ses performances.

Il est donc clair que l'évaluation doit désormais devenir le centre de nos préoccupations, en mettant en application la Vision Stratégique afin de donner à ce chantier toutes les chances de succès.

Notre conception de l'évaluation se fonde sur les éléments suivants :

- Premièrement, l'obligation collective d'accepter l'évaluation comme un outil qui vise l'amélioration, le développement et l'évolution. Plus que cela, l'évaluation est aussi une culture dont la finalité première est l'adaptation aux nouveautés et la création, notamment dans les domaines de l'éducation et de la formation.
- Le deuxième élément consiste à adopter les mécanismes, outils et indicateurs d'évaluation, fondés scientifiquement et méthodologiquement, et reconnus, au niveau international, pour leur transparence, objectivité et crédibilité, en prenant en considération les spécificités de notre système éducatif et les attentes de la société ainsi que les défis actuels ;
- Le troisième élément se définit par l'institutionnalisation des échanges réguliers et de l'entente interactive. Ceux-ci sont fondés sur la réalisation d'une valeur ajoutée entre l'instance qui se charge de l'évaluation et les secteurs et institutions relatives à l'éducation. Partant du principe que l'évaluation n'est pas seulement un processus d'identification de dysfonctionnements, mais son rôle se doit de dépasser cela, vers une veille prospective des difficultés, d'anticipation des contraintes, la valorisation des bonnes pratiques, le renforcement et le développement des acquis et le soutien de la recherche scientifique.
- Quant au quatrième élément, il concerne la conviction à partager selon laquelle l'évaluation peut contribuer à la mobilisation autour des questions de l'école. L'évaluation scientifique et objective met les décideurs, les acteurs éducatifs, les familles des apprenants devant la réalité de l'éducation et de la formation, et de là, permet l'ouverture d'un débat plus large, et la construction des voies possibles pour le développement des performances des établissements d'éducation et de formation.

Mesdames et Messieurs,

Ce ne sont là que quelques idées préliminaires que je souhaitais partager avec vous à l'occasion de l'ouverture de ce colloque qui constitue un forum à pérenniser.

J'ai la ferme conviction que les travaux de cette rencontre permettront un échange d'idées, d'expertises, d'expériences et de bonnes pratiques et que les conclusions et recommandations serviront à renforcer et à consolider la place de l'évaluation de l'éducation dans notre pays, en même temps qu'elles serviront à intensifier les relations de coopération entre les experts internationaux et marocains.

Je tiens à renouveler mes remerciements et ma reconnaissance à nos partenaires des organismes internationaux, ainsi qu'à tous les intervenants et participants à ce colloque auquel je souhaite un grand succès.

Je vous remercie pour votre attention.

NOTE INTRODUCTIVE

Rahma Bourqia *

Monsieur le Président du Conseil,

Madame et Messieurs les ministres,

Mesdames et Messieurs les participants et intervenants au colloque,

Les représentants des organismes internationaux partenaires et des institutions nationales,

Honorables invités.

C'est avec un grand plaisir que nous accueillons les participantes et participants à ce Colloque international sur « **L'Évaluation en Éducation et en Formation : Approches, enjeux et défis** ». Je tiens à vous remercier d'avoir répondu à notre invitation.

L'Instance Nationale d'Évaluation auprès du CSEFRS organise ce colloque à un moment déterminant dans l'histoire de notre système éducatif. En effet, le CSEFRS, après une évaluation approfondie des 13 dernières années de mise en œuvre de la Charte Nationale d'Éducation de Formation, adoptée en 2000, vient de publier sa Vision Stratégique 2015-2030 qui trace les grandes orientations, et recommande les principaux leviers pour une réforme profonde du système marocain d'éducation, de formation, et de recherche scientifique pour les 15 prochaines années.

Parmi les recommandations de cette Vision, figure la mise en place des dispositifs d'évaluation pour accompagner la mise en œuvre d'une réforme, à même de mettre notre système éducatif sur la voie de l'équité, de l'égalité des chances, de la qualité, de l'émancipation de l'individu et du développement du pays, tout en se dotant de la maîtrise nécessaire de la conduite du changement.

L'Instance Nationale d'Évaluation, conformément aux missions que lui confère la loi organisant le Conseil, et qui lui attribue l'évaluation du système d'éducation, de formation et de recherche scientifique, est engagée dans la prise en charge des recommandations du Conseil.

Pour contribuer à ce grand chantier de réforme de l'éducation, l'INE consolide ainsi son dispositif d'évaluation, en capitalisant sur ses propres acquis et sur les structures d'évaluations qui existent déjà dans le paysage national, et en mettant en place les instruments nécessaires à une évaluation externe et indépendante.

Cet engagement, au profit d'une évaluation externe, crédible et indépendante, implique que notre institution s'ouvre sur les différents systèmes nationaux et internationaux d'évaluation, dont beaucoup sont représentés par les participants et participantes à ce colloque.

L'organisation de ce colloque affiche, de notre part, une volonté d'échanger avec nos collègues sur les différentes méthodes d'évaluation et d'accréditation dans un esprit de partage des bonnes pratiques et d'identification des démarches innovantes de l'évaluation. Le rassemblement dans ce colloque, des chercheurs éminents, praticiens expérimentés, représentants de plusieurs institutions d'évaluation et provenant de différents pays et du Maroc, nous offre l'opportunité, pendant ces deux jours, d'un échange, qui s'annonce riche et intéressant, autour des approches, des enjeux et des défis de l'évaluation.

Ce colloque requiert pour nous au Maroc une importance capitale pour plusieurs raisons:

Premièrement, derrière l'évaluation, il y a une accumulation de savoirs, de théories ainsi qu'une communauté de chercheurs en évaluation. Nous savons que l'histoire et l'évolution de la pratique de l'évaluation ont été encadrées par toute une réflexion qui les a faites évoluer dans une démarche qui maintient le dialogue entre terrain et théorie. Ceci implique que le projet d'évaluation de l'éducation au Maroc, doit non seulement s'inscrire dans l'universalité du savoir international, mais également d'y contribuer, et autant que faire se peut, apporter les éléments de réflexion tirés du contexte de son système éducatif.

Ce colloque nous offre par conséquent l'occasion d'engager des échanges entre participants pour éclairer ce dialogue entre terrain et théories.

* Directrice de l'Instance Nationale d'Évaluation auprès du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique

La deuxième raison réside dans le fait que l'évaluation constitue un l'objet de multiples demandes, parce que l'école et l'université font face à des défis interpellant les systèmes éducatifs. Le système éducatif marocain est amené à se plier aux exigences de l'imputabilité, de la reddition des comptes, préconisées par la Constitution marocaine de 2011, de la transparence et de la performance requises par "le management public" qui s'impose de plus en plus dans le domaine de l'éducation.

Cette demande, exigeant la performance des institutions éducatives, provient de la société, des parents d'élèves, des médias, des organismes internationaux, et du monde économique... Toutes ces parties prenantes exercent une pression sur l'école et sur l'université pour améliorer la qualité des apprentissages, des formations, et de la performance de l'école et de l'université : imposant et instaurant ainsi l'évaluation et l'accréditation comme des mécanismes incontournables de régulation.

L'ancrage d'un système d'évaluation raisonnée et réfléchi dans notre système scolaire et universitaire, doit s'inscrire dans l'internationalisation de l'éducation et de ses standards, et viser l'amélioration de la performance de notre système d'éducation et de formation. Ce colloque nous offrira certainement l'occasion de nous informer et de nous inspirer de l'expérience des pays qui ont pu inscrire l'évaluation dans les systèmes éducatifs avec des effets positifs sur le rendement de l'éducation.

La troisième raison, qui rend ce colloque important pour le contexte marocain, réside dans le fait que le Maroc a opté, constitutionnellement, pour une régionalisation avancée qui va certainement consolider l'autonomie des Académies et des Universités. Ceci va entraîner un changement dans le mode de pilotage de l'éducation. Ce changement fait que l'Administration Centrale doit s'accommoder, non seulement avec la pression qu'exerce la globalisation sur l'éducation, mais également avec la nouvelle autonomisation/décentralisation dans le cadre de la régionalisation avancée.

Dans ce contexte, l'auto-évaluation (ou l'évaluation interne), adoptée et réalisée à des niveaux intermédiaires des systèmes scolaire et universitaire : aux niveaux des écoles et des établissements, contribue à une appropriation, par les territoires et par leurs acteurs, de la pratique et de la culture d'évaluation. C'est dans ce sens que le couple évaluation interne et évaluation externe indépendante, ainsi que les accréditations, doivent fonctionner alternativement comme de nouveaux modes de régulation et de guidage, entraînant ainsi de nouveaux mécanismes de guidage et de pilotage de l'éducation. Ceci nous interpelle pour se préparer à une nouvelle distribution de responsabilités entre le national, régional et local dans le pilotage de l'éducation.

La quatrième raison de l'importance de ce colloque se rapporte au fait que l'évaluation n'a d'intérêt que si elle mène les politiques publiques à prendre des mesures concrètes et à agir pour mettre l'école et l'université sur la voie de la qualité. Le sursaut d'une Nation suite à des évaluations est connu dans les cas des pays comme les Etats-Unis et l'Allemagne. La publication aux Etats Unis, en 1983, du rapport sur l'éducation "Nation at Risk" (Une Nation face au risque) fut suivie de grandes réformes du système éducatif. En Allemagne, les résultats de l'évaluation des enquêtes PISA en l'an 2000, révélant une faiblesse relative des acquis des élèves allemands, a provoqué ce qui est connu sous l'expression "le choc de PISA" (the PISA shock).

En se référant à ces cas, l'évaluation doit contribuer à créer un sursaut national et amener un pays à repenser son école et son avenir. Depuis 1999, le Maroc a adhéré aux enquêtes internationales d'IEA, (the International Association for the Evaluation of Educational Achievement), les enquêtes TIMSS et PIRLS sur l'évaluation des acquis des élèves, et a accumulé des données sur les résultats des vagues successives de ces enquêtes. Si les résultats ont permis une comparabilité internationale, ils n'ont pas toujours créé un débat public après leur publication.

En effet, si les résultats de l'évaluation n'entraînent pas un débat qui pousse les politiques publiques à améliorer la situation de l'éducation, les évaluations des enquêtes internationales TIMSS, PIRLS et PISA de l'OCDE, deviennent des machines à produire des données et un classement de pays. Ce classement, qui suscite l'intérêt médiatique, reconforte les pays "premiers de la classe" et renvoie aux moins performants l'image de crise de leurs systèmes éducatifs.

A un moment où nous aspirons à une refondation de notre école, les évaluations des enquêtes internationales doivent susciter une réflexion et un débat parmi les acteurs, et insuffler une dynamique dans les politiques publiques pour améliorer les acquis des élèves. Dans un pays comme le nôtre, la raison d'être légitime de l'adhésion à l'évaluation des acquis des élèves est de contribuer à améliorer les apprentissages et de relever le niveau de notre éducation.

Nous savons que ces enquêtes internationales ont évolué à travers le temps et se sont étendues à plusieurs domaines de l'éducation. On pourrait citer le programme PIAC (Programme for International Assessment of

Adults Competencies), destiné à évaluer le niveau d'alphabétisation des adultes, ou encore le programme TALIS (Teaching and Learning International Survey) de l'OCDE pour évaluer la performance des enseignants en relation avec les apprentissages, et d'autres... Cette évolution doit amener les pays émergents à intervenir, en amont, c'est-à-dire à contribuer dès la conception et l'élaboration des instruments et des standards, et en aval, à s'impliquer dans l'analyse des résultats. Ce qui permettrait une appropriation de l'évaluation.

Il est évident que cette appropriation du système d'évaluation, fait progressivement son chemin dans les pays émergents et se renforce avec la mise en place des dispositifs d'enquêtes nationales dans plusieurs pays. C'est dans ce sens que l'Instance Nationale d'Évaluation a mis en place un programme national de l'évaluation des acquis des élèves (dit PNEA), à réaliser tous les quatre ans, dont le modèle méthodologique sera présenté lors de ce colloque.

En soulignant l'importance de ce colloque, nous sommes conscients qu'une réforme profonde de notre éducation, ici au Maroc, qui met en œuvre la Vision Stratégique du Conseil pour les 15 années à venir, se doit de mobiliser tous les dispositifs nationaux et internationaux d'évaluation dont nous disposons pour les mettre au service de ce grand chantier de réforme.

Tout en insistant sur le rôle que doit jouer l'évaluation comme mécanisme de régulation, on ne doit pas oublier que la crise de l'éducation est tributaire, non seulement d'un fonctionnement interne au système d'éducation et de formation, mais également des facteurs externes qui doivent être pris en compte dans notre réflexion sur l'évaluation. Dans ce déficit de l'éducation, interviennent la culture des acteurs, les politiques éducatives et leur histoire, le système de gouvernance, les conflits autour des valeurs, les contextes socio-économiques des écoles et des universités, ainsi que le degré de mobilisation autour d'un projet éducatif qui nous projette dans l'avenir. C'est pour cette raison que toute évaluation de l'éducation ne doit pas être réduite à une lecture faite à partir d'une comparabilité internationale uniquement, mais également à partir des études évaluatives des contextes de l'éducation, menées avec des approches multidisciplinaires, qui expliquent les résultats des enquêtes internationales.

Portée par des organismes de la communauté internationale, la comparabilité des enquêtes internationales sur les acquis des élèves, bien que crédible et légitime, suscite des discussions et des débats. L'arrière fond philosophique, sur lequel se basent ces enquêtes, est l'universalité des normes de qualité qui doivent régir les systèmes éducatifs dans ce «grand laboratoire mondial de l'éducation». Mais cette conception de l'universalité de la qualité provoque parfois des débats de la part de tous ceux qui mettent l'accent sur les contextes de l'éducation comme facteurs explicatifs de la situation de l'éducation.

Dans notre démarche de l'évaluation, il ne s'agit pas de se ranger du côté de la tendance qui se focalise sur l'internationalisation et la standardisation de l'évaluation en éducation, ou du côté de son opposé, qui se concentre sur les particularités des contextes, mais de mettre les deux tendances au service de la réforme continue de l'éducation et la formation. C'est ainsi que l'évaluation doit être constamment soumise à une approche vigilante, réflexive et constructive pour lui éviter d'être réduite à une pratique sommative qui donne un certain pouvoir aux agents d'évaluation. Etant une pratique normative, stratégique et politique, interpellant l'action, l'évaluation est formative et formatrice qui vise à faire d'une éducation moyennement performante, une éducation qui performe plus et mieux.

Nous sommes conscients que l'évaluation fait partie, aujourd'hui, de la modernisation du management public (public management), qui transfère la culture de l'assurance qualité de l'entreprise à d'autres milieux et terrains. Mais l'effet de ce transfert doit être retravaillé et repensé dans le domaine de l'éducation : une éducation, dont la finalité ultime, est la production de la richesse humaine, la construction de têtes bien faites, la formation des individus qui ont acquis des compétences et munis de valeurs pour vivre dans une société inclusive équilibrée et développée, et pour qu'ils soient citoyens de leur pays et du monde.

L'évaluation n'a donc d'intérêt que si elle fonctionne comme un catalyseur qui incite le système éducatif à innover, à améliorer de manière continue et constante les apprentissages, la gestion de l'école et le bon fonctionnement de l'université.

Ce Colloque international sur "L'Évaluation en Éducation et en Formation : Approches, enjeux et défis" nous permettra certainement d'échanger sur toutes ces questions, entre chercheurs, praticiens de l'éducation, représentants des organismes internationaux et parties prenantes du système éducatif.

Ce colloque nous offre également une opportunité d'échange avec nos collègues issus de traditions évaluatives différentes, et de débattre des enjeux de l'évaluation et de l'accréditation ainsi que leurs rôles respectifs dans l'accompagnement des réformes.

Le colloque est organisé en six séances, réparties sur 2 jours :

- **Le premier jour :**

1. La première séance de cette matinée, qui commencera sous peu, porte sur les cadres théoriques et démarches de l'évaluation,
2. La deuxième séance, qui se tiendra cet après-midi, porte sur les évaluations internationales
3. La troisième séance de cet après-midi discutera de l'Assurance qualité et de l'accréditation ;

- **Le second jour :**

4. La quatrième séance, lors de la matinée du vendredi a pour titre **Normes et instruments d'évaluations**
5. La cinquième séance de la matinée du vendredi portera sur **l'évaluation de l'enseignement supérieur**
6. La sixième et dernière séance, a pour titre **l'évaluation et les politiques publiques** avec une intervention introductive qui sera suivi **d'une table ronde sur l'évaluation au Maroc.**

Je tiens à remercier tous les participants et participantes qui ont répondu à notre invitation et ont fait le déplacement pour partager avec nous leurs expériences, leurs réflexions et leurs pratiques en matière d'évaluation. Je remercie vivement, les représentants des institutions nationales et les partenaires nationaux et internationaux pour leur contribution à ce colloque et tous les acteurs de la communauté de l'éducation qui vont en suivre les travaux.

Je souhaite aux participantes et participants, nationaux et internationaux, la bienvenue et un agréable séjour à Rabat, et un plein succès à notre colloque.

Merci pour votre attention

1

L'ENJEU DE L'ÉVALUATION



COGNITIVE SKILLS AND ECONOMIC GROWTH: HETEROGENEITY MATTERS¹

Nadir Altinok - Abdurrahman Aydemir *

Abstract

Our paper reassesses the question of the impact of cognitive skills on economic growth using new indicators for cognitive skills, based on a pool of international assessments for pupils in primary and secondary schools. These new data extend measures of cognitive skills substantively in terms of both time and the number of countries involved. In particular, our data extends the coverage of less developed countries, among them adding 27 countries of sub-Saharan Africa, a continent that was largely missing from the analysis of the effects of learning outcomes on economic growth. Using this extended dataset and employing several identification strategies, cognitive skills are found to have a positive impact on economic growth. We also address the heterogeneity in the causal effect of cognitive skills on growth and show that the effect of skills on growth differs across regions and by the economic level of countries. While the effect is the largest in low-income countries, a lower impact of cognitive skills is found for high-income countries.

Introduction

The question of which factors determine economic growth has been a major topic in economic research. Levine and Renelt (1992) reported initial income level and the share of investment in GDP as the only variables that consistently had statistically significant impacts on economic growth across a wide array of specifications. They find that neither primary nor secondary enrolment rates consistently had positive, statistically significant effects on economic growth. The importance of human capital for economic growth has been called into question by a large number of studies that failed to find a positive relationship between the quantity of education and economic growth in cross-country analysis. Recent studies, however, have pointed out the importance of school quality as opposed to quantity (Barro, 1991, Hanushek and Kimko, 2000, Hanushek and Woessmann, 2012) and have provided evidence of the positive effect of school quality on the rate of economic growth.

This paper extends this recent literature in a number of substantive dimensions. We use a recent dataset that covers many countries, particularly less developed ones, that could not be included in growth regressions by previous studies. For example, among the newly added countries, our database includes 27 countries of sub-Saharan Africa, a continent that was largely missing from the analysis of the effects of learning outcomes on economic growth. The study also updates the period of analysis by including the most recent data on schooling quality (between 1965 and 2012). Using these data, we pay particular attention to the heterogeneity in the estimated impact of the relationship between education and growth according to level of development. We also assess the importance of minimum and advanced skills on growth experiences of countries with different levels of development. In addition to reporting cross-sectional regression results, we also present results from recently proposed alternative estimation methods for identifying the causal effect of education quality on growth using a panel data framework and new set of instruments. This allows to test the robustness of the estimated impact of cognitive skills on economic growth based on different estimation strategies and subsamples and provides several novel results.

An important number of research papers analyzed the education-growth relationship. In this paper, we only focus on studies that have included a qualitative dimension to education². Following Barro's analysis that underlined the importance of school quality, the work of Hanushek and Kimko (2000) was the first to

1 - This paper is a short version of a larger paper, which will be published with the title «Does one size fit all: The impact of cognitive skills on economic growth».

2 - For further details, see the comprehensive review by Durlauf, Johnson, and Temple (2005).

* Nadir Altinok: BETA/CNRS, Department of Economics, University of Lorraine (France), UFR Droit, Sciences Economiques et Gestion, Abdurrahman Aydemir: Sabanci University, Faculty of Arts and Social Sciences, Orhanli, Tuzla 34956, Istanbul.

include measures of educational quality using data from international student achievement tests (hereafter ISATs). Using data from ISATs that administered similar mathematics and science tests to students in 31 countries, they construct normalized test scores and include these as measures of the human capital stock (called «Labor Force Quality») in their analysis of economic growth (from 1960 to 1990). They find that the coefficient estimate for this variable is positive and statistically significant, and its inclusion in the analysis causes the years of schooling variable to lose statistical significance. Their estimated impact is large, suggesting that a one standard deviation increase in schooling quality (measured by test performance) increases the rate of economic growth by 1.4 percentage points. An important shortcoming of this analysis is the inclusion of the learning outcomes variable for only one point in time, hence making it impossible to analyze economic growth rates as a function of changes in learning outcomes.

A recent paper by Hanushek and Woessmann (2012) (hereafter HW) aimed at improving the work of Hanushek and Kimko (2000). Using the Cohen and Soto (2007) years of education instead of the Barro and Lee dataset, they update test score data that include more developing countries (50 countries, of which 27 are developing countries) and the period analyzed is extended to cover the 1960-2000 period. Another important change is related to the methodology used: HW used an approach that assumes stability over time of the variance of quality of student achievement in a restricted number of OECD countries, which serves as a standardization benchmark for performance variation over time. Confirming results from Hanushek and Kimko (2000), HW find that years of schooling have no impact on economic growth when the test score measure is included. Moreover, the test score variable is highly significant and adding it increases the adjusted R squared from 0.25 to 0.73. Overall, a one standard deviation increase in school quality is associated with a 1.3-2.0 percentage point higher rate of economic growth.

Following the pioneering analysis of Hanushek and Kimko (2000), we start from the idea that one year of education in country *i* does not confer the same rate of return as one year of education in country *j*. Our paper aims at improving and extending the literature in a number of ways. Firstly, we use a larger dataset for our test score variable, by including more developing countries: our dataset provides an increase in the number of countries of about 60%. A recent database (Altinok et al., 2014) compiled the results of countries in the international assessments of pupils for each skill (mathematics, sciences and reading) between 1960 and 2007. While this original dataset provides data until 2007, we replicate the same methodology in order to update it until 2012. Compared to the Altinok et al. (2014) dataset, while our updated dataset includes only 3 additional countries, by including more recent testing it achieves a substantive extension of 30% in the number of observations overall. This increase permits us to improve the precision of our learning outcomes indicators, since we have more observations for each country. The resulting updated database includes comparable cognitive skills for 125 countries, as compared to HW who take into account 77 countries. For instance, our database includes 27 countries of sub-Saharan Africa which were not captured by previous research.

Secondly, the methodology used for our test score dataset is exclusively based on datasets available at the student level which helps to reduce errors in measures of school quality. For instance, we improved the measure of test scores between 1960 and 1980 by adjusting the test scores of pioneering achievement tests with original raw data provided by the IEA (International Association for the Evaluation of Educational Achievement)³. In most previous studies, only mean scores published in international reports were used. In order to reduce measurement error bias, we also chose not to include data from an assessment for which there is no raw data available. This is the reason why we excluded International Assessment of Educational Progress (IAEP)⁴ and Monitoring Learning Achievement (MLA) assessments. Indeed, specific corrections are often needed for mean scores from student achievement tests in order to obtain comparable scores between countries and over time, even for the very recent student achievement tests such as TIMSS 2003. For instance, some countries took part in several student achievement tests, but with different sampling procedures. With original raw data, it is possible to exclude specific populations in order to obtain more comparable data on learning outcomes.

Thirdly, since our data include countries from several continents, in addition to estimating an average effect of education on economic growth, we can also test for heterogeneity in these effects. This paper is the first one in literature that assesses within a causal framework the differences in the amplitude of cognitive skills

3 - A major improvement of the Altinok et al. (2014) dataset deals with the information relative to trends between the first waves of IEA studies. Indeed, the trends between FIMS and SIMS and between FISS and SISS are only available for a common number of items. The list of these items and the anchoring process is described in IEA Reports (see Keeves, 1992; Robitaille and Garden, 1989). This additional information permitted Altinok et al. (2014) to simply adjust assessments on NAEP anchoring, but also to reproduce the trends found in these reports. Since, a significant number of items were used in both assessments, the trends found in these reports may reduce the bias which occurs when we only use NAEP anchoring.

4 - Unfortunately, raw data for IAEP are not available. We would like to thank ETS and NCES for their support.

and growth relationship⁵. In particular, we exploit the availability of more than 80 countries in our data and provide estimates separately by (i) income level of countries, (ii) regions, and (iii) total factor productivity. This analysis provides novel evidence on the cognitive skills and economic growth relationship. For example, among other regions, our analysis provides results for Arab countries and Sub-Saharan African countries, a region the growth experience of which received little attention by previous studies due to data constraints.

Lack of comparable data on cognitive skills is a general issue for growth empirics research. After improving the data by including more developing countries, our analysis yields three main results.

i) While we cannot find a robust effect of the quantity of schooling (measured as initial years of education), the coefficient associated with our updated cognitive skills variable is quite strong over most estimations. These results confirm those reported by HW.

ii) Our results show that including more developing countries increases the overall impact of cognitive skills on economic growth by about 25%.

iii) Moreover, we find that the magnitude of the effect is highest for low-income countries, followed by high-income and middle-income countries.

I. A Simple Growth Model

Following HW, we use a simple growth model: a country's growth rate (g) is a function of the skills of workers (H) and other factors (X). These factors include initial levels of income and technology, specific institutional dimensions, and other factors that are used in the growth empirics. Skills are often referred to simply as the workers' human capital stock. Our specification assumes that H is a one-dimensional index and that growth rates are linear in these inputs:

$$g = \gamma H + \beta X + \varepsilon \quad 1$$

The most important specification issue in this framework is the nature of the skills (H) and where they might come from. In the educational production function literature (Hanushek, 2002), skills are explained by many factors such as family inputs (F), the quantity and quality of inputs provided by schools (qS), individual ability (A), and other relevant factors (Z) which include labor market experience, health, and other specific characteristics:

$$H = \alpha F + \beta(qS) + \gamma A + \delta Z + v \quad 2$$

In our specification, the schooling term (qS) combines school attainment (S) and its quality (q). However, human capital is a latent variable that cannot be directly observed. Hence, we need a correct measure of human capital in order to test its impact on economic growth. The main existing theoretical and empirical work on growth begins by taking the quantity of schooling of workers (S) as a direct measure of H . Following the pioneering analysis of Hanushek and Kimko (2000), we focus directly on the cognitive skills component of human capital and evaluate H with test-score measures of mathematics, science, and reading achievement. There are many advantages of using measures of educational achievement (Hanushek and Woessmann, 2012). Firstly, they capture outputs of schooling by focusing on differences in the knowledge and ability generated by schools. Secondly, since they include all the general skills, they do not only rely on school skills but also skills from other sources (families and general ability). Another important advantage of using cognitive skills is the ability to assess the importance of different policies designed to affect the quality aspects of schools since cognitive skills allow for differences in performance among students with the same quantity of schooling.

II. Data and methodology

The dataset related to cognitive skills used in this paper builds upon the work of Altinok et al. (2014) and updates the 1960-2007 data to 1960-2012. Based on new data sources and the alternative method of

5 - Castelló-Climent and Hidalgo-Cabrillana (2012) develop a theoretical model of human capital investments distinguishing between low- and high-quality education. Using the Hanushek and Kimko (2000) dataset they show the education quality has a positive effect on growth only when quality is relatively high and conclude that quality may be not growth enhancing unless students achieve a minimum level of knowledge. Their main empirical exercise, however, does not control for potential endogeneity of cognitive skills and includes few developing countries compared to our data. Where they use an IV strategy to address endogeneity lagged values of education quality is used as an instrument but the authors caution against the validity of this instrument.

anchoring, there are several innovations in this dataset compared to previous research. The construction of this data benefits from international student achievement tests (ISATs) as well as regional student achievement tests (RSATs). ISATs include the well-known TIMSS, PIRLS and PISA tests⁶. Along with these international assessments, three major RSATs conducted in Africa and Latin America, such as LLECE, SACMEQ or PASEC⁷, which were never used in previous research on the effect of cognitive skills on economic growth. This helps us to extend the available data to a larger set of countries, in particular improving the representation of the developing world. The resulting updated database in this paper includes comparable cognitive skills for 125 countries, as compared to HW who take into account 77 countries between 1960 and 2000⁸.

The methodology to generate comparable achievement scores across countries used in Altinok et al. (2014) aims at improving the seminal work by Barro and Lee (1996) and Barro (2001), and consists of a major update of a previous work by Altinok and Murseli (2007). Hanushek and Kimko (2000) and Hanushek and Woessmann (2012) also use a method of anchoring for their database of cognitive skills across 77 countries. The alternative methodology for creating the data used in this paper differs from Hanushek and Woessmann (2012) in that it takes into account several improvements made by ISATs since 1995 and enables the inclusion of the main regional assessments that were absent in previous datasets. Details of this methodology can be found in Altinok et al. (2014).

III. Baseline results

In this section, we report cross-sectional estimates of the cognitive skills and economic growth relationship based on equation (1). Since we use an extended dataset based on a different methodology to HW, we first replicate results from HW using their own data as well as our dataset confined to the HW sample. We then report results from our new extended data, which show to what extent earlier results reported in the literature hold for a larger set of countries and with more recent data. In addition, we estimate the relationship between cognitive skills and economic growth for various subsamples of countries and using different sets of tests to construct education quality measures. Table 1 presents the baseline results. This table is divided into three parts. The first part (part A) replicates Table 1 from HW using the same dataset and sample of countries. In the second part (B) of Table 1, we use our dataset that expands the years used for calculation of test scores to 2012 but restrict the sample to the countries in HW. This allows us to check to what extent the expanded time span for the tests in our dataset provides additional information compared to previous research. Part (C) of Table 1 uses our updated dataset with the extended set of countries and aims to test the robustness of previous estimates to the inclusion of additional countries⁹.

The results of Table 1 show a strong positive relationship between cognitive skills and economic growth that remains significant across different specifications in a cross-country estimation. This confirms the findings of HW in a larger sample, including many developing countries. Secondly, the amplitude of the effect is quite similar to that in HW's paper, even after we include more countries in the estimation. Comparing results in column 8 in each panel (1-A; 1-B and 1-C) shows to what extent the amplitude of the effect changes between different indicators of cognitive skills and different samples of countries. In the HW estimates, an increase of one standard deviation of cognitive skills improves growth by about 1.2 percent (column 3). The effect is approximately similar with our data restricted to the HW sample of countries (1.3 percent). Including all countries with available data shows that the estimated effect (1.5 percent) is about 25% higher. While one-third of this increase can be attributed to the difference between our updated cognitive skills measures, the remaining part is due to the expansion of sample, and especially the inclusion of developing countries. The main finding is that for our full sample and with our updated cognitive skills measures, a move of one standard deviation of individual student performance translates into 1.5 percentage points difference in annual growth rates, other things being equal. It is also interesting to measure the level of one standard deviation in terms of score points. Since one standard deviation is equal to 100 points in our scale, this represents approximately the difference of performance between Greece (533 points) and South Korea (628 points). In addition, the difference between Turkey and the remaining OECD countries is approximately

6 - Respectively the Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS) and Programme for International Student Assessment (PISA).

7 - Respectively the Latin American Laboratory for Assessment of the Quality of Education (LLECE), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the Program on the Analysis of Education Systems (PASEC).

8 - It should be noted that the number of countries included inside estimations is always lower than the number of countries for which we have comparable data on cognitive skills. The main reason is the lack of data on other explanatory variables. For instance, whereas HW compiled comparable data on cognitive skills for 77 countries, only 50 of them were included in different estimations. In our case, whereas we have data for 120 countries, our sample is reduced to 80 countries.

9 - Because we need data for both economic growth and cognitive skills between 1960 and 2010, all former communist countries are eliminated even if they have test measures. This explains why our estimation does not include 125 countries.

equal to 0.5 standard deviation.

The above results show a strong positive relationship between cognitive skills and economic growth using cross-sectional variation. Therefore, our results confirm the results of HW, and partially the hypothesis highlighted in Pritchett (2001) where one explanation of the controversies in the lack of significant effect of education on economic growth was quality differences between countries. While the results are robust across various specifications and subsamples, reverse causality and endogeneity bias may potentially be driving the results. Reverse causality would arise if higher economic growth enables countries to develop better education systems that yield higher test performance. The presence of other factors that are correlated with cognitive skills such as cultural factors, institutions, and access to natural resources that affect growth will lead to an endogeneity bias in our estimations. The following section addresses these identification issues.

IV. Heterogeneity in the Impact of Cognitive Skills on Economic Growth

Countries place a high priority to investments in education and skills as a key driver of economic growth. The gains from these investments, however, depend on the interactions between skills, technology, and physical capital. For example, investments in skills may result in larger productivity gains in countries where skill supply is scarce compared to countries where skill supply is relatively abundant. Although there are many studies that assess the mean effect of cognitive skills on growth across countries there has been little research in the literature that addresses the heterogeneity of this relationship.

A second important issue regarding the heterogeneous effects of skills is which types of skills matter most for economic growth. Acemoglu and Zilibotti (2001) shows that a mismatch between supply of skills and the adopted technology leads to low productivity while Hanushek (2013) provides evidence that the impact of high performers on growth differs between OECD and non-OECD countries. Potential differences in the impact of different types of skills on growth has important policy implications since the countries that aim to improve cognitive skills face the choice of targeting improvements across the whole distribution or placing more emphasis on a specific part of the distribution, such as the bottom or the top.

In this section, we aim to extend the existing literature in a number of ways. We first provide further evidence of the heterogeneity of the relationship between cognitive skills and growth, presenting results for various subsamples. Secondly, we conduct an analysis that tests whether the effect of minimum and advanced level of cognitive skills differ between countries. Our third contribution is related to the estimation methodology. The previous literature that considers the heterogeneity of the relationship between cognitive skills and growth do not address endogeneity of cognitive skills. Using a larger sample of countries, we also address the endogeneity issue through a number of alternative instruments. Our analysis yields a rich set of results that indicate significant differences in the impact of cognitive skills on growth and to sheds further light on growth experiences of countries. Also, using a single data set that involves a consistently defined human capital measure and applying the same methodology for estimation provides comparable results across subsamples. This overcomes the challenge of synthesizing results over different studies that use different methodologies and measures of human capital in different country contexts and the resulting uncertainty in the policy arena as to the most effective type of education or skills for growth

We divide the sample into several parts and provide estimates separately by (i) income level of countries, (ii) total factor productivity, and (iii) regions. Higher income countries employ a higher level of capital stock and enjoy higher total factor productivity. Hence, the role of skills in growth for these countries may differ from those of low income countries. There are also marked differences by geography in the growth experiences of countries. The role skills play in these growth experiences has received little attention. For example, our analysis provides results for Arab countries and Sub-Saharan African countries, the continent that could not often be studied separately by previous studies due to data constraints.

The results are presented in Table 2, which is divided into 2 panels. The first panel reports results from OLS regressions (panel A) whereas the second panel (panel B) report results from the IV-GMM estimation. In order to test whether our results are driven by the use of specific instruments, we use different combinations of instruments presented by rows B1 through B6. In all of the IV-GMM estimates initial years of schooling is used as an instrument in combination with one or two other instrumental variables.¹⁰ In Panel B1, we use the government effectiveness as an instrument that proved to be a valid instrument for the whole sample in Table 4. In Panel B2, we use initial school drop-out rate for primary education as an instrument. Since pupils may leave schools because they do not receive a high standard of education, school drop-out rate

10 - Previous sections provided evidence for the validity of initial years of schooling as an instrument in the cognitive skills growth relationship. Nevertheless, we have also carried out estimations that does not use initial years of education as an instrument and obtained results that are very similar to those presented in Table 6. Due to space considerations these results are not presented.

for primary school may serve as a good instrument for education quality or cognitive skills. Hanushek et al. (2008), for example, show in a developing country context that a student is much less likely to remain in school if attending a low-quality school rather than a high-quality school. However, since growth rate of the economy could also impact on drop-out rate, in order to satisfy the exclusion restriction we use the initial level of school dropout as an instrument¹¹. A combination of government effectiveness, initial school dropout, and initial years of schooling is used as our instruments in panel B3. Primarily used by Islam et al. (2014), DALY was highly correlated with our cognitive skills variable in the first stage results of the IV estimation (see column 3 of Table 4) and it is included as an instrument in panels B4 and B5. Final specification uses the overall level of income inequality (measured with Gini index) as an instrument. A recent study shows a positive correlation between income inequality and inequality of education (Inter-American Bank, 1999) while Krueger (2012) and Corak (2013) show that countries with more inequality as measured by Gini coefficients have less intergenerational mobility. Overall level of inequality may, thus, capture disparities along the income distribution in access to education and its quality, hence lead to reductions in cognitive skills. Cingano (2014) provides support for this channel. The study finds that the main mechanism through which inequality affects growth is by undermining education opportunities for children from poor socio-economic backgrounds, lowering social mobility and hampering skills development. While we hypothesize an affect of inequality on growth only through its effect on cognitive skills, inequalities in education and income and growth may be jointly determined. In order to avoid reverse causality, we use the initial level of the Gini coefficient for each country as an instrument¹².

In order to obtain comparable effects in terms of standard deviations, we also standardize the cognitive skills variable in each sub-sample (with a mean equal to 0 and a standard deviation equal to 1). This allows us to directly compare the effect of cognitive skills expressed in terms of standard deviations between each sub-sample. Given the large set of results, we only report the coefficient estimate of the cognitive scores variable, the first-stage F-statistic, and the number of countries included in each subsample in brackets, but do not to present the first stage results.

In column 1 of Table 2, in the first two rows we reproduce baseline results where our cognitive skills variable has a positive and significant impact on economic growth, whether we consider the OLS or the IV estimations. In rows B2 to B6 of the first column IV estimation using different sets of instruments provide coefficient estimates that range between 1.7 and 2.1 and are all larger than the OLS estimate. According to the OLS estimation, an increase of one standard deviation in cognitive skills produces an increase in annual economic growth of about 1.5 percentage points. By using the IV estimation technique, the overall effect of cognitive skills is increased by about 25% (1.9%). One reason for the increased effect may be the downward bias due to measurement error.

The results in columns 2 to 8 that distinguish between the economic levels of countries provide important insights. Comparing columns 2 and 3 shows that while the effect of cognitive skills is positive and significant for both low and high income countries, both the OLS and IV results indicate that its amplitude is about 60% higher for the low-income countries. This emphasizes that the promotion of education policies that focus on the quality of education has especially large payoffs in least developed regions. These results are in contrast with Castelló-Ciment et al. (2012) that reports no effect of education quality on economic growth in countries at the bottom end of the quality distribution but a positive effect in countries with higher education quality. However, our results are in line with the results of HW that report a higher impact of cognitive skills for OECD countries. In countries with low levels of education quality, improvements in quality may lead to substantial improvements in productivity of workers. Higher estimated effects of quality on growth in low income countries may be due these productivity gains.

Estimations results by geographical region are presented in columns 4 to 6. IV estimates for each region shows a positive and significant impact of cognitive skills on economic growth. Compared to other regions, we find much larger effect of cognitive skills on economic growth for the Asian countries. It is possible that the early-period growth explosion of East Asia is mainly due to high level of cognitive skills in this region, compared to other regions like Latin America. The lowest coefficient estimates, on the other hand, are obtained for Latin America. As we focus on regions, with much smaller sample sizes, some of the F statistics are now lower than 10. Only in the case of Latin America, however, they are systematically below this threshold. Hence, we do not have a strong support for a positive effect of cognitive skills on growth for this region.

Besides the standard distinction of economic levels of countries and geographical locations, we also

11 - Since data availability differs greatly between countries, the year of the initial value of drop-out rate in primary education varies between countries. However, for most countries, the initial year is 1970.

12 - Similarly to drop-out rate, the initial level of Gini coefficient differs between countries. Due to data constraints, the initial year is often around 1980.

divided the sample into two parts, in the spirit of Nelson and Phelps (1966). It is possible that countries which are far from the technology frontier, i.e. with a low total factor productivity in 1960, will benefit more from higher cognitive skills levels than others countries. To test this possibility, we separate the sample by distinguishing low initial total factor productivity (TFP) countries and high initial TFP countries, using the median level of TFP in 1960 (columns 7 and 8). Results confirm that countries which are far from their technology frontier benefit more from cognitive skills than other countries. Comparing columns 7 and 8, the effect of cognitive skills is doubled for these countries in the standard OLS estimation. When we correct for possible endogeneity, measurement error and omitted variable bias by using the IV GMM estimation technique, the difference between the two groups becomes even larger. Another important finding is that the extent of bias between OLS and IV estimates is the largest for Arab States and Sub-Saharan Africa. This may be due to lower quality of assessment for this region.¹³ In conclusion, our cognitive skills variable is quite stable and in most subsamples has a positive and significant impact on economic growth. We find that the magnitude of the effect is higher for the low-income countries and across regions investing in the quality of education appears to be most rewarding for Arab States and Sub-Saharan African countries.

Conclusion

Among all the explanations for economic growth, one that is generally accepted concerns the level of human capital. The seeming obviousness of the idea, however, has met with inconsistencies in the existing literature. The most robust macro-economic analyses reveal contradictions in the relationship between education and growth. Pritchett (2001) showed that very often the impact of education on growth is negative and significant. However, the majority of these studies have ignored the qualitative dimension of human capital, recognizing only the purely quantitative indicators.

The use of national or international achievement tests in mathematics and sciences fills this gap in qualitative measurement. Hanushek and Kimko (2000), Barro (2001) and Hanushek and Woessmann (2012) used qualitative variables, but did not exploit all of the international assessments or all the countries surveyed (respectively 36, 43 and 50 countries included in their samples). In this research study, we used an updated dataset on cognitive skills for a larger number of countries. Consequently, our sample includes more developing countries than the previous studies.

The main advantage of using a larger dataset is probably the possibility of distinguishing between different subsamples. We look for potential differences in the amplitude of the effect of cognitive skills on economic growth between high-income and low-income countries, but also between different regions. While the overall effect of cognitive skills is higher for developing countries, it remains positive and significant in all subsamples. The highest amplitude is found for Sub-Saharan African countries, where an increase of one standard deviation of cognitive skills improves economic growth by about 2.5 percentage points.

13 - Contrary to other assessments where modern psychometric procedures were included, the PASEC assessment had no Rasch scaling of scores which may reduce survey quality and explain why the estimated IV coefficient is higher than the one found with OLS technique. See Wagner (2011).

Table 1. Standard estimates of the effect of cognitive skills on economic growth

(1)	(2)	(3)	(4) ^(a)	(5) ^(b)	(6) ^(c)	(7) ^(d)	(8) ^(e)	(9) ^(f)
(A) Data from Hanushek and Woessmann (2012), sample from Hanushek and Woessmann (2012)								
Cognitive skills	1.238 (8.62)	1.199 (7.38)	1.224 (6.88)	1.102 (8.13)	1.006 (3.33)	0.853 (5.02)	0.594 (5.18)	1.191 (6.04)
Years of schooling 1960	0.408 (4.41)	0.050 (0.79)	0.014 (0.18)	0.064 (0.76)	0.070 (0.79)	-0.003 (0.04)	-0.014 (0.23)	-0.067 (0.89)
GDP pc 1960	-0.399 (4.85)	-0.294 (9.21)	-0.321 (8.49)	-0.300 (7.48)	-0.317 (5.74)	-0.270 (5.10)	-0.334 (6.80)	-0.975 (4.38)
(B) Data from updated Altinok et al. (2014), Sample from Hanushek and Woessmann (2012)								
Cognitive skills	1.312 (8.86)	1.316 (7.24)	1.383 (6.84)	1.382 (11.13)	1.185 (4.44)	1.023 (4.57)	0.669 (3.56)	1.360 (6.44)
Years of schooling 1960	0.408 (4.41)	-0.004 (0.06)	-0.062 (0.76)	-0.045 (0.69)	0.006 (0.07)	-0.038 (0.52)	-0.018 (0.27)	-0.079 (1.11)
GDP pc 1960	-0.399 (4.85)	-0.319 (9.41)	-0.317 (8.18)	-0.294 (7.30)	-0.301 (6.29)	-0.243 (4.19)	-0.314 (6.49)	-1.171 (6.50)
(C) Data from updated Altinok et al. (2014), Sample from updated Altinok et al. (2014)								
Cognitive skills	1.629 (13.14)	1.510 (10.50)	1.382 (8.36)	1.587 (10.40)	1.101 (3.96)	1.531 (5.66)	0.720 (2.72)	1.610 (9.92)
Years of schooling 1960	0.465 (4.90)	0.115 (1.83)	0.168 (2.51)	0.099 (1.29)	0.149 (2.28)	0.055 (0.83)	0.003 (0.06)	0.036 (0.54)
GDP pc 1960	-0.220 (3.82)	-0.270 (9.43)	-0.310 (8.21)	-0.294 (7.29)	-0.314 (7.18)	-0.281 (6.37)	-0.308 (6.84)	-1.073 (7.41)
(A) Observations	50	50	50	52	50	47	45	50
(B) Observations	50	50	50	52	50	47	45	50
(C) Observations	84	84	80	85	80	68	68	80
(A) R-squared (adj.)	0.313	0.753	0.756	0.754	0.778	0.800	0.803	0.667
(B) R-squared (adj.)	0.313	0.792	0.792	0.794	0.823	0.820	0.791	0.760
(C) R-squared (adj.)	0.232	0.719	0.729	0.739	0.756	0.714	0.750	0.670

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2000 for sample from Hanushek and Woessmann (HW) (2012), 1960-2010 for sample from updated Altinok et al. (2014). Regressions include a constant. Test scores are average of math and science, primary through end of secondary school (for HW data) or through lower-secondary school (for Altinok et al. data), all years. Absolute t-statistics in parentheses

(a) Mean years of schooling refers to the average between 1960 and 2000 (HW data), 2010 (ADM data).

(b) Robust regression including the two outliers of Botswana and Nigeria (with reg robust estimation implemented in Stata).

(c) Specification includes dummies for the eight world regions taken in HW.

(d) Specification includes additional controls for openness and property rights

(e) Specification includes additional controls for openness, property rights, fertility, and tropical location.

(f) GDP per capita 1960 measured in logs

Table 2. Sensitivity of estimated effects of cognitive skills to the economic level of countries and regions

	(1) All countries	(2) High Income Countries ^(a)	(3) Low Income Countries ^(a)	(4) Arab States & Sub-Saharan Africa	(5) Asian Countries	(6) Latin American Countries	(7) High TFP countries	(8) Low TFP countries
A- OLS								
Cognitive skills	1.510 (10.50)	0.899 (7.68)	1.600 (8.18)	0.892 (2.96)	1.710 (16.22)	-0.062 (0.23)	0.917 (7.83)	1.836 (12.68)
Adj. R ² (Observations)	0.729 (80)	0.717 (40)	0.765 (40)	0.424 (25)	0.958 (14)	0.524 (17)	0.627 (36)	0.858 (36)
B1- IV-GMM								
Cognitive skills	1.875 (10.45)	1.306 (8.12)	1.833 (10.16)	2.500 (2.26)	1.236 (5.30)	0.542 (1.71)	0.956 (5.55)	2.483 (10.15)
F statistic (observations)	42.62 (79)	29.46 (39)	19.33 (40)	1.56 (25)	13.27 (14)	7.06 (17)	16.61 (35)	16.49 (36)
B2- IV-GMM								
Cognitive skills	2.051 (9.80)	1.225 (7.11)	2.074 (9.87)	1.886 (4.20)	1.650 (6.19)	1.497 (2.43)	0.740 (2.33)	2.214 (9.56)
F statistic (observations)	27.51 (74)	13.76 (35)	9.45 (35)	6.87 (25)	4.97 (11)	1.97 (17)	5.16 (30)	9.75 (36)
B3- IV-GMM								
Cognitive skills	1.994 (10.12)	1.254 (7.70)	2.064 (10.69)	1.965 (4.57)	1.743 (7.42)	0.423 (2.08)	0.932 (4.08)	2.410 (10.89)
F statistic (observations)	27.80 (73)	21.23 (34)	11.17 (39)	6.28 (25)	4.79 (11)	4.36 (17)	12.70 (29)	13.52 (36)
B4- IV-GMM								
Cognitive skills	1.938 (11.03)	1.258 (6.91)	1.963 (12.02)	1.843 (5.43)	1.630 (8.37)	1.128 (2.23)	0.839 (5.57)	2.352 (11.88)
F statistic (observations)	35.55 (78)	9.14 (39)	16.66 (39)	11.82 (25)	6.29 (12)	4.02 (17)	40.62 (34)	14.37 (36)
B5- IV-GMM								
Cognitive skills	1.912 (11.38)	1.209 (8.29)	1.966 (12.05)	1.883 (5.79)	1.635 (9.74)	0.534 (1.78)	0.821 (6.78)	2.439 (12.22)
F statistic (observations)	37.26 (77)	8.79 (38)	16.95 (39)	8.50 (25)	24.44 (12)	4.24 (17)	64.67 (33)	13.30 (36)
B6- IV-GMM								
Cognitive skills	1.712 (10.95)	0.954 (8.17)	1.779 (10.84)	0.773 (1.93)	0.912 (3.28)	1.476 (2.27)	1.085 (6.49)	2.073 (10.07)
F statistic (observations)	33.24 (78)	29.37 (38)	13.45 (40)	9.40 (25)	9.84 (14)	2.04 (17)	21.69 (36)	11.43 (34)

Notes: Dependent variable: average annual growth rate in GDP per capita, 1960-2010 for sample from updated Altinok et al. (2014). Regressions include a constant. Test scores are average of math and science, primary through lower secondary school, all years. Absolute t-statistics in parentheses. Each panel with IV estimations includes different instruments but always initial years of schooling (hereafter Yrs). B1: Yrs + governance effectiveness. B2: Yrs + drop out of primary education. B3: Yrs + governance effectiveness + drop out of primary education B4: Yrs + DALY. B5: Yrs + governance effectiveness + DALY. B6: Yrs + initial level of Gini index

(a) Countries above/below sample median of GDP per capita 1960

Traduction du Résumé**Abstract Translation**

Nadir Altinok - Abdurrahman Aydemir

COMPÉTENCES COGNITIVES ET CROISSANCE ÉCONOMIQUE : UNE QUESTION D'HÉTÉROGÉNÉITÉ

Notre papier évalue la question des compétences cognitives et leur impact sur la croissance économique, en utilisant de nouveaux indicateurs pour les acquis cognitifs et en se basant sur les évaluations internationales des élèves des cycles du primaire et du secondaire. Ces nouvelles données étendent la mesure des acquis en termes de temps et du nombre de pays concernés. De façon particulière, nos données couvrent moins de pays développés, et 27 pays d'Afrique Sub-Saharienne ont été ajoutés, un continent qui manquait dans l'analyse des effets des résultats d'apprentissage sur la croissance économique. En utilisant cette base de données étendue et en employant plusieurs stratégies d'identification, il apparaît que les acquis cognitifs ont un impact positif sur la croissance économique. Nous traitons également l'hétérogénéité dans l'effet de causalité des compétences cognitives sur la croissance et démontrons que l'effet des compétences sur la croissance diffère selon les régions et selon le niveau de développement économique des pays. Tandis que l'effet est le plus grand dans les pays à revenus bas, l'impact des compétences cognitives est réduit dans les pays à fort revenu.

Bibliography and References

- Acemoglu, D. and Zilibotti, F. 2001. Productivity Differences. *Quarterly Journal of Economics*. 116(2): 563-606.
- Altinok, N., Diebolt, C. & de Meulemeester, J.-L. 2014. A New International Database on Education Quality : 1960-2010. *Applied Economics* 46 (11), 1212-1247.
- Barro, R. 1991. Economic growth in a cross-section of countries. *Quarterly Journal of Economics* 106 (2), 407-443.
- Barro, R.J. 2001. Education and Economic Growth. in Helliwell, J.F. (Ed.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being* (pp. 14-41). Paris: OECD Press.
- Castelló-Climent, Amparo and Ana Hidalgo-Cabrillana 2012. The role of educational quality and quantity in the process of economic development, *Economics of Education Review*, vol. 31(4), pages 391-409.
- Cingano, F. (2014), "Trends in Income Inequality and its Impact on Economic Growth", OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing.
- Cohen, D., & Soto, M. 2007. Growth and human capital: good data, good results. *Journal of economic growth* 12 (1), 51-76.
- Corak, Miles, 2013. Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27, no. 3, 79-102.
- Durlauf, S.N., Johnson, P.A., & Temple, J.R. 2005. Growth econometrics. *Handbook of economic growth* 1, 555-677.
- Hanushek, E.A. 2002. Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics*. (pp. 2045-2141). Amsterdam: Elsevier.
- Hanushek, E.A. 2013. Economic growth in developing countries: The role of human capital. *Economics of Education Review* 37, 204-212.
- Hanushek, E.A., & Kimko, D.D. 2000. Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 1184-1208.
- Hanushek, E.A., & Woessmann, L. 2012. Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, And Causation. *Journal of Economic Growth* 17(4), 267-321.
- Inter-American Development Bank. 1999. Facing up to Inequality in Latin America: Economic and Social Progress in Latin America, 1998-99 Report. Johns Hopkins University Press.
- Keeves, J.P. 1992. *The IEA Science Study III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.
- Krueger, Alan, 2012. "The Rise and Consequences of Inequality in the United States," Speech at the Center for American Progress, Washington, DC, January 12, 2012.
- Levine, R., & Renelt, D. 1992. A sensitivity analysis of cross-country growth regressions. *The American economic review*, 942-963.
- Nelson R. and Phelps E., 1966. Investment in Humans, Technological Diffusion and Economic Growth. *American Economic Review*, 61, 69-75.
- Pritchett, L. 2001. Where has all the education gone?. *The World Bank Economic Review* 15 (3), 367-391.
- Robitaille, D.F. & Garden, R.A. (Eds.). 1989. *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics*. Oxford: Pergamon Press.
- Wagner, D. 2011. *Smaller, Quicker, Cheaper. Improving Learning Assessments for Developing Countries*. UNESCO: International Institute of Educational Planning. Paris.

REVIEW OF INTERNATIONAL AND REGIONAL ASSESSMENT PROGRAMS

Marguerite Clarke *

Abstract

Over the last 20 years, there has been a steady increase in the number of countries participating in international and regional assessments of educational achievement. One of the attractions of these assessments is that they allow countries to compare their students' achievement levels to those for other participating countries. But how similar are international and regional assessments to each other? And how useful are international assessments in particular for improving quality and learning in low- and middle-income countries? This paper discusses recent findings on this topic from work supported by the World Bank. It concludes with comments on recent initiatives that aim to further enhance the utility of these assessments both for the participating countries and for global efforts to track learning levels.

Introduction

This paper describes the main international and regional assessment programs that collect and report data on student achievement levels at the basic education level. The paper addresses two key questions: (i) How similar are international and regional assessments? (ii) How useful are international assessments for improving quality and learning in low- and middle-income countries?

International and regional assessment programs can be defined as those that (i) describe student achievement in one or more knowledge or skill areas in ways that estimate achievement levels in the education system as a whole; and that (ii) provide policymakers, educators, and others with information on their system's performance relative to others. All of these assessments seek cross-national answers to one or more of the following: How well are students learning? Does evidence indicate particular strengths and weaknesses in students' knowledge and skills? Do particular subgroups in the population perform poorly? What factors are associated with student achievement? Do the achievement levels of students change over time? In order to answer these questions, students in each country are assessed in a number of areas and background information is collected from students, teachers, parents, or principals.

* Experte en Éducation, Banque Mondiale

I. Main International and regional Assessment Programs at the Basic Education Level

The PIRLS, TIMSS, and PISA assessments are international in scope, involving the participation of countries from all regions of the world. PIRLS assesses reading at the equivalent of Grade 4 in participating countries while TIMSS assesses mathematics and science at the equivalent of Grades 4 and 8. PISA uses a sample of 15-year-olds in each country to measure achievement levels in reading, mathematics, and science.

LLECE, PASEC, and SACMEQ are more regional in nature, with each focusing primarily or exclusively on a particular geographic region or language group. LLECE measures achievement levels in reading, mathematics, science, and writing at the Grades 3 and 6 levels in countries in Latin America. PASEC measures student achievement in reading and mathematics in Grades 2 and 6 in Francophone countries while SACMEQ focuses on Grade 6 and assesses student achievement in reading, mathematics, and health in Anglophone countries in Africa.

Table 1 provides general information on six of the best-known international and regional assessment programs at the basic education level.

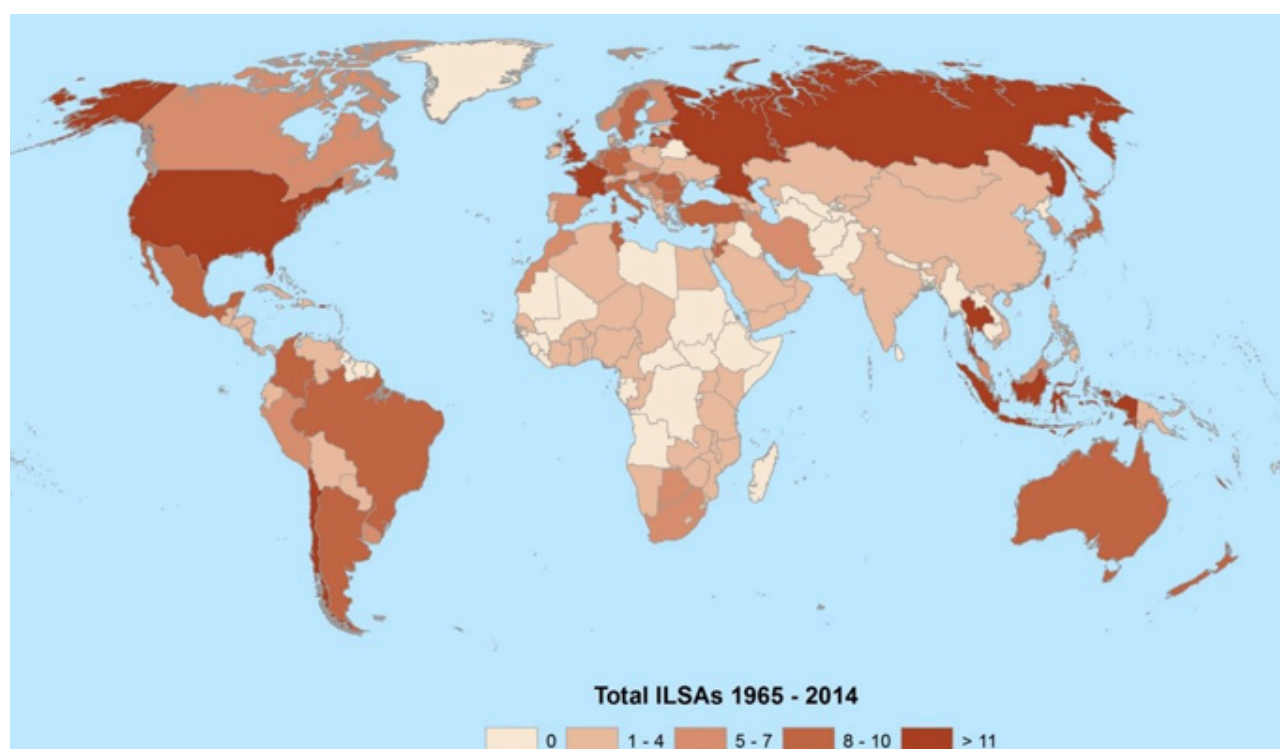
Table 1. Main International and Regional Assessment Programs at the Basic Education Level

Name	Target grades/ages	Main subject areas	Organization	Years	Participating regions
<i>Progress in International Reading Literacy Study (PIRLS)</i>	Grade 4	Reading	International Association for the Evaluation of Educational Achievement (IEA)	2001, 2006, 2011, 2016	Global
<i>Trends in International Mathematics and Science Study (TIMSS)</i>	Grades 4, 8*	Mathematics, science	IEA	1995, 1999, 2003, 2007, 2011, 2015	Global
<i>Program for International Student Assessment (PISA)</i>	15 year olds	Reading, mathematics, science	Organization for Economic Cooperation and Development (OECD)	2000, 2003, 2006, 2009, 2012, 2015	Global
<i>Latin American Laboratory for Assessment of the Quality of Education (LLECE)</i>	Grades 3, 6	Reading, mathematics, science, writing	UNESCO-(la Oficina Regional de Educación para América Latina y el Caribe) OREALC	1997, 2006, 2013	Latin America
<i>Program for the Analysis of Education Systems (PASEC)</i>	Grades 2, 6**	Reading, mathematics	Conference des Ministres de l'Éducation des États et Gouvernements de la Francophonie (CONFEMEN)	Every year between 1993 and 2010; 2014	Francophone Africa
<i>Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)</i>	Grade 6	Reading, mathematics, health	SACMEQ	1999, 2004, 2011, 2014	Anglophone Africa

*There is also a TIMSS Advanced module that is administered to students in Grade 11. This Grade 11 assessment does not follow the same cycle as the Grades 4 and 8 assessments and to date has only been administered in 1995, 2008, and 2015 and has involved the participation of fewer countries than the Grades 4 and 8 assessments.

**Until recently, PASEC was administered in Grade 5, not Grade 6.

Figure 1. Increase in countries participating in international and regional assessments



Source: Lockheed et al. (2015).

As shown in Figure 1, there has been a steady increase in countries participating in these international and regional assessments over the last five decades.

Approximately two-thirds of all countries with populations greater than 30,000 have participated in one or more of these assessment programs. While wealthier countries are more likely to participate in international assessment programs, participation rates are also growing among low- and middle-income countries.

II. How similar are international and regional assessments?

In 2014, the World Bank and the OECD commissioned a study to determine similarities and differences among existing international and regional assessment programs. The study involved analyzing the six assessments shown in Table 1 in the areas of population coverage, content and format, and approaches to analysis and reporting. Some of the report's (Creswell et al., 2015) key findings are summarized in Tables 2-4.

Table 2. How similar are international and regional assessments in coverage?

	Number of participating countries in most recent cycle	Main target population	Frequency
PISA	65	15 year olds	3-year cycle
TIMSS	77	Grades 4, 8	4-year cycle
PIRLS	49	Grade 4	5-year cycle
LLECE	15	Grades 3, 6	No fixed cycle
SACMEQ	15	Grade 6	No fixed cycle
PASEC	10	Grades 2, 6	No fixed cycle

Source: Based on Creswell et al. (2015).

Table 3. How similar are international and regional assessments in content and format?

	Curriculum-based	Main subject areas	Computer or paper delivery	Question formats
PISA	No	Reading, mathematics, science	Both	Selection, supply
TIMSS	Yes	Mathematics, science	Paper	Selection, supply
PIRLS	Yes	Reading	Paper	Selection, supply
LLECE	Yes	Reading, mathematics, science, writing	Paper	Selection, supply
SACMEQ	Yes	Reading, mathematics, health	Paper	Selection
PASEC	Yes	Reading, mathematics	Paper/Oral	Selection

Source: Based on Creswell et al. (2015).

Table 4. How similar are international and regional assessments in analysis and reporting?

	Use of Item Response Theory (IRT)	Proficiency levels	Background data on learning, language, home	Comparisons across systems, trends over time
PISA	IRT (1-parameter)	6	Yes	Both
TIMSS	IRT (3-parameter)	4	Yes	Both
PIRLS	IRT (3-parameter)	4	Yes	Both
LLECE	IRT (Rasch)	4	Yes	Partially
SACMEQ	IRT (Rasch)	8	Yes	Both
PASEC	2014 onwards	2014 onwards	Yes	2014 onwards

Source: Based on Creswell et al. (2015).

It is clear that the international assessments involve a far greater number of countries than the regional ones. The former also operate on a more predictable schedule, with PISA running on a standard three-year cycle and TIMSS and PIRLS on four- and five-year cycles respectively. The less-rigid schedule of the regional assessments is due to several factors, including responsiveness to the individual needs of participating countries and less predictable funding and human resources.

The majority of international and regional assessments test students in the primary grades and in the core subject areas of reading and mathematics. They also tend to rely on paper-based administration of tests. The exception to this is PISA, which focuses on an age group as opposed to a grade level, and offers both paper- and computer-based test administration. It also is the only assessment that claims to be 'curriculum-free' in that it focuses more on general competencies for success post-compulsory schooling. All of these assessments focus on the achievements of children in school. None attempt to measure the knowledge of out-of-school children although this may change given new international requirements for global tracking of achievement data.

There is a tendency for international assessments to use more sophisticated analytical techniques, such as item response theory (IRT), and to produce data that can be compared across countries and over time. The regional assessments tend to use less robust technical techniques, although this is changing as these programs develop more capacity. All of the assessments collect background data and use it in their analyses to determine factors associated with higher levels of student achievement.

Creswell et al. (2015) note that while these assessments differ in coverage, content, and analytical approaches, there is the possibility of getting them to 'talk' to each other under the right conditions. This objective has become more important since the emergence of the new United Nations Sustainable Development Goal for education, which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". The targets and indicators for this goal include a focus on global progress towards improved achievement levels in core subject areas, particularly reading and mathematics. This will inevitably

generate more demand for achievement data that can be compared across education systems and over time.

III. How useful are international assessments for improving quality and learning in low- and middle-income countries?

An increasing number of low- and middle-income countries are deciding to participate in one or more of the international assessments, and this demand is likely to grow. Participation fees vary from approximately US\$185,000 for PISA (three subjects, one age group) to US\$225,000 for PIRLS (one subject, one grade) to US\$450,000 for TIMSS (two subjects, two grades). The total direct costs that countries take on by participating in these assessments range from US\$400,000 to US\$2 million (this number includes the participation fee as well as implementation and translation costs), which is far less than one percent of annual expenditure on education in most countries (Wolff, 2007). These costs, while not huge, are still considerable in the case of low- and middle-income countries that have many demands on their education budgets.

While international assessments offer valuable opportunities for global benchmarking of system performance and for the building of sophisticated assessment capacity, researchers have noted certain problems that low- and middle-income countries in particular can face when participating in these assessments, including that (Greaney & Kellaghan, 2008):

- Many test questions can be too difficult for students from these countries, resulting in restricted test score variance and limited information on correlates of achievement;
- Poor performance on these high-profile assessments can carry political risks for key officials associated with delivery of education; and
- The demands of meeting strict international deadlines may prove difficult and can lead to countries dropping out or having their data cancelled.

Despite these concerns, a number of low- and middle-income countries have successfully participated in and made effective use of the results from international assessments. For example, Robitaille, Beaton, & Plomp (2000) note the following uses of TIMSS results by low- and middle-income countries:

- Kuwait: Committees were formed to revise the national mathematics and science curricula based on the country's results.
- Philippines: The President directed that a 'rescue package' be implemented to improve performance in science and mathematics in which teacher training would receive particular attention.
- Romania: New topics were added to the national mathematics curriculum.

A review of Latin American countries' experiences participating in PISA (Ferrer & Fiszbein, 2015) noted similar uses of PISA results by these countries.

A recent case study (Tyumeneva, 2013) of Russia's experience with participating in all three international assessments (PIRLS, TIMSS, and PISA) noted a general consensus in the country that the information generated had been effectively used to improve education quality. Specifically:

- the results created a general awareness among stakeholders of the need to reform the school system;
- access to the databases allowed for policy research and formulation of action plans to introduce reforms, and also helped build national capacity in assessment; and
- the results served as a reference for new national learning standards, curricula, and textbooks.

Despite such successes, a not uncommon theme in low- and middle-income countries' experiences with these assessments is difficulty in transforming the data into usable information. Bethell and Harutyunyan (2015) make this point in their case study on Armenia's participation in TIMSS, noting that while the country's results in the international TIMSS report allowed basic comparisons to be made, little else was done with the data, including no secondary analysis, no independent research, and no subject-specific reports for schools and teachers in Armenia. The authors attributed this situation to a lack of demand by national stakeholders for such data as well as limited national capacity to analyze the data. They identified several factors essential to establishing demand and capacity, including:

- educating policymakers, teachers, parents, civil society and the media as to the value of international

achievement data and how to interpret these data intelligently,

- building the analytical capacity of assessment specialists to supply the information demanded in the required form, and
- providing resources so secondary analysis and focused research can be conducted.

In 2014, the World Bank and the OECD commissioned a study on middle-income country experiences in PISA between 2000 and 2015¹. The findings (Lockheed et al. 2015) showed that while PISA results have informed some education policy in middle-income countries (particularly in the areas of curriculum and teacher training), the effects have been limited compared to higher-income countries. There tended to be less public discussion of results (including media coverage) and policy influence was mainly through private dialogue, including in the form of policy dialogue between countries and donors such as the World Bank.

This study also involved a multi-level analysis of 2012 PISA mathematics results for 18 middle-income countries: Argentina, Brazil, Bulgaria, Colombia, Costa Rica, Indonesia, Jordan, Kazakhstan, Mexico, Montenegro, Malaysia, Peru, Romania, Serbia, Thailand, Tunisia, Turkey, and Vietnam. The researchers found that the overall quality of education in many of these countries was much lower than that of most OECD countries, with a high share of students not reaching even a basic skill level in mathematics.

Tables 5 and 6 list the variables that best explained student achievement in the 18 countries. In all 18 countries, students who were in higher grades at the time of the test were more likely to score well. In all but one country, a student's socio-economic background was positively correlated with their achievement on PISA, with students from poorer backgrounds tending to perform less well. Boys also tended to perform better than girls. In most of the countries, students who had repeated a grade did less well than those who had never repeated a grade while those who had attended preschool tended to perform better than those who had not. At the school level (Table 6), in every country, the average socio-economic background of the children in a school predicted the school's performance on the 2012 PISA mathematics test, with schools with higher concentrations of poor children more likely to perform worse. Schools with more learning time did better in every country and schools that reported a positive disciplinary climate also did better in most countries.

These analyses also indicated, however, that existing PISA student background and school questionnaires may be more relevant for high-income countries than for low-income or lower-middle-income countries, since – in particular – the school questionnaire provided little information regarding the types of resources that often explain between- or within-school differences in student performance in low- and middle-income countries.

Table 5. Student-level variables important in more than 9 countries

Variable	Number of countries
Grade	18
Socio-economic background	17
Gender	17
Repeater	15
Preschool	11

Source: Based on Lockheed et al. (2015)

Table 6. School-level variables important in more than 9 countries

Variable	Number of countries
School average socio-economic background	18
More learning time	18
Positive disciplinary climate	13

Source: Based on Lockheed et al. (2015)

¹ - The methodology involved reviews of published academic literature, OECD documents and reports, and World Bank education projects and reports. It also included stakeholder interviews and case studies of 9 middle-income countries. The authors also conducted new empirical analyses of on-line media from 9 middle-income countries and PISA 2012 data from 18 middle-income countries.

Conclusion

This paper described the main international and regional assessment programs that collect and report data on student achievement levels in basic education. The paper addressed two key questions: (i) How similar are international and regional assessments? (ii) How useful are international assessments for improving quality and learning in low- and middle-income countries?

International and regional assessments are similar in that they all focus on measuring student achievement levels in the general areas of reading and mathematics. But they differ considerably in their scale and frequency, in the way in which students' knowledge is assessed, and in the types of approaches used to analyze and report the data. There is a need to better align these assessments so that the data can provide a clearer picture of how well students around the world are learning.

While international assessments offer valuable opportunities for global benchmarking of system performance and building sophisticated assessment capacity, low- and middle-income countries can face several issues when participating in these assessments, including test questions that are too difficult for their students, background questionnaires that lack appropriate questions for their contexts, and difficulty in analyzing the data in a way that national stakeholders will understand and use.

The various international and regional assessment organizations are already working to address some of these issues. The organizations that oversee PIRLS, TIMSS, and PISA have developed additional assessment options in the last few years – such as prePIRLS, TIMSS Numeracy, Literacy and Numeracy Assessment (LaNA), and PISA for Development – that are specifically designed to measure lower levels of the achievement continuum and to be more easily used in low- and middle-income country contexts, including with out-of-school children. These assessment programs are also creating more suitable background questionnaires and providing more support and capacity building at all stages of the assessment exercise. A new regional assessment program has emerged in East Asia and existing regional programs have made significant efforts to improve the quality and comparability of the data they collect. International and regional assessment programs are also working together to find ways to better align their programs so that the data can provide a clearer picture of global learning levels.

Traduction du Résumé Abstract Translation

Marguerite Clarke

EXAMEN DES PROGRAMMES D'ÉVALUATION RÉGIONAUX ET INTERNATIONAUX EN ÉDUCATION

Ces 20 dernières années, une constante augmentation du nombre de pays participants dans les évaluations internationales des acquis scolaires a été enregistrée. L'un des attraits de ces évaluations est qu'elles permettent aux pays de comparer les niveaux de leurs élèves à ceux des autres pays participants. Cependant, quel est le degré de similitude entre les évaluations régionales et internationales ? Quelle est l'utilité des évaluations internationales en particulier dans l'amélioration de la qualité et de l'apprentissage dans les pays à revenus bas et moyens ? Ce papier examine les conclusions récentes en la matière à partir d'un travail réalisé par la Banque Mondiale. Il conclue par des commentaires sur les initiatives récentes qui visent à renforcer davantage l'utilité de ces évaluations, à la fois pour les pays participants et pour des efforts globaux pour suivre et tracer les niveaux d'apprentissage.

Bibliography and References

Bethell, G., & K. Harutyunyan. (2015). The Development of the Student Assessment System in the Republic of Armenia: Achievements, Challenges, and Lessons Learned. Washington, DC: The World Bank.

<https://openknowledge.worldbank.org/handle/10986/22757>

Clarke, M. (2012). What Matters Most for Student Assessment Systems: A Framework Paper. Washington, DC: The World Bank.

<https://openknowledge.worldbank.org/handle/10986/17471>

Cresswell, J., U. Schwantner, & C. Waters. (2015). Review of Component Skills Assessed and Contextual Data Collection Used in Relevant International Assessments. OECD/World Bank.

<http://www.oecd.org/edu/a-review-of-international-large-scale-assessments-9789264248373-en.htm>

Ferrer, G., & A. Fiszbein. (2015). What Has Happened with Learning Assessment Systems in Latin America? Washington, DC: Inter-American Dialogue.

<http://www.thedialogue.org/resources/what-has-happened-with-learning-assessment-systems/>

Gardner, E. & M. Clarke (2015). Final READ Trust Fund Report, 2008-2015. Washington, DC: The World Bank.

<http://documents.worldbank.org/curated/en/2015/12/25515030/read-trust-fund-report-2008-2015>

Greaney, V., & T. Kellaghan. (2008). Assessing National Achievement Levels in Education. Washington, DC: The World Bank.

<https://openknowledge.worldbank.org/handle/10986/6904>

Lieberman, J., & M. Clarke (2012). Review of World Bank Support for Student Assessment Activities in Client Countries, 1998-2009. Washington, DC: The World Bank.

<https://openknowledge.worldbank.org/handle/10986/17476>

Lockheed, M., T. Prokic-Breuer, & A. Shadrova. (2015). The Experience of Middle-Income Countries Participating in PISA 2000-2015. OECD/World Bank.

<http://www.oecd.org/publications/the-experience-of-middle-income-countries-participating-in-pisa-2000-2015-9789264246195-en.htm>

Robitaille, D., A. Beaton, & T. Plomp (Eds.). (2000). The Impact of TIMSS on the Teaching and Learning of Mathematics and Science. Vancouver, BC: Pacific Educational Press.

Tyumeneva, J. (2013). Disseminating and Using Student Assessment Information in Russia. Washington, DC: The World Bank.

<https://openknowledge.worldbank.org/handle/10986/16276>

Wolff, L. (2007). The Costs of Student Assessments in Latin America, PREAL, Washington DC.

SUMMARY OF THE OECD REVIEWS ON EVALUATION AND ASSESSMENT IN EDUCATION

Abstract

*This article provides an overview of the main conclusions of *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*¹, the final report of an OECD 3-year in-depth investigation of evaluation and assessment policies in OECD countries. The report provides an international comparative analysis and policy advice to countries on how evaluation and assessment arrangements can be embedded within a consistent framework to improve the quality, equity and efficiency of school education.*

*This article is the OECD's contribution to the International Symposium on Evaluation and Assessment in Education and Training organised by the Moroccan National Authority of Evaluation (INE) in October 2015. It includes the executive summary and the overview chapters of the OECD report *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*².*

Introduction

Governments and education policy makers are increasingly focused on the evaluation and assessment of students, teachers, school leaders, schools and education systems. These are used as tools for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school, school leadership and teaching practices.

Results from assessment and evaluation are becoming critical to establishing how well school systems are performing and for providing feedback, all with the goal of helping students to do better.

This report compares the experience of 28 OECD countries, analyses the strengths and weaknesses of different approaches, and offers policy advice on using evaluation and assessment to improve the quality, equity and efficiency of education. It draws on a major study, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

1 - OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/en9789264190658/10.1787en>

2 - The authors of the OECD report *Synergies for Better Learning: An International Perspective on Evaluation and Assessment* are Deborah Nusche, Thomas Radinger, Paulo Santiago (co-ordinator) and Claire Shewbridge of the Directorate for Education and Skills.

Common themes

Several factors are driving the increased use of evaluation and assessment, including:

- An increased demand for effectiveness, equity and quality in education to meet economic and social challenges.
- A trend in education towards greater school autonomy, which is fuelling a need to monitor how schools are doing.
- Improvements in information technology, which allow for the development of both large-scale and individualised student assessment and facilitate the sharing and management of data.
- Greater reliance on evaluation results for evidence-based decision making.

The current state and use of evaluation and assessment varies greatly between OECD countries, but there are common themes:

Evaluation is expanding and becoming more diverse

Most OECD countries now see evaluation and assessment as playing a central strategic role, and are expanding their use. They are also taking a more comprehensive approach: Formerly, evaluation and assessment focused mainly on student assessment, but the focus is now broader and includes greater use of external school evaluation, appraisal of teachers and school leaders, and expanded use of performance data.

Indicators are rising in importance

Education systems are placing a stronger focus on measuring student outcomes, allowing comparisons of performance between schools and regions and over time. Most countries now have national databases on education, and issue education statistics and indicators. International benchmarking is also increasingly common and is informing national education debates.

Results are being put to varied uses

Results are being used to identify where schools are performing well and where they may need to improve. They are also being used to hold policy makers, school leaders and teachers accountable. For example, many countries now publish national tables of school results for use by, among others, parents, government officials and the media.

Rising reliance on educational standards

Many countries now set educational standards for what students should know and what they should be able to do at different stages of the learning process. This has encouraged monitoring to determine if students are meeting these standards.

Challenges and directions

Countries have different traditions in evaluation and assessment and take different approaches. Nevertheless, there are some clear policy priorities:

Take a holistic approach

To achieve its full potential, the various components of assessment and evaluation should form a coherent whole. This can generate synergies between components, avoid duplication and prevent inconsistency of objectives.

Align evaluation and assessment with educational goals

Evaluation and assessment should serve and advance educational goals and student learning objectives. This involves aspects such as the alignment with the principles embedded in educational goals, designing fit-for-purpose evaluations and assessments, and ensuring a clear understanding of educational goals by school agents.

Focus on improving classroom practices

The point of evaluation and assessment is to improve classroom practice and student learning. With this in mind, all types of evaluation and assessment should have educational value and should have practical benefits for those who participate in them, especially students and teachers.

Avoid distortions

Because of their role in providing accountability, evaluation and assessment systems can distort how and what students are taught. For example, if teachers are judged largely on results from standardised student tests, they may “teach to the test”, focusing solely on skills that are tested and giving less attention to students’ wider developmental and educational needs. It is important to minimise these unwanted side-effects by, for example, using a broader range of approaches to evaluate the performance of schools and teachers.

Put students at the centre

Because the fundamental purpose of evaluation and assessment is to improve student learning, students should be placed at the centre. They should be fully engaged with their learning and empowered to assess their own progress (which is also a key skill for lifelong learning). It is important, too, to monitor broader learning outcomes, including the development of critical thinking, social competencies, engagement with learning and overall well-being. These are not amenable to easy measurement, which is also true of the wide range of factors that shape student learning outcomes. Thus, performance measures should be broad, not narrow, drawing on both quantitative and qualitative data as well as high-quality analysis.

Build capacity at all levels

Creating an effective evaluation and assessment framework requires capacity development at all levels of the education system. For example, teachers may need training in the use of formative assessment, school officials may need to upgrade their skills in managing data, and principals – who often focus mainly on administrative tasks – may need to reinforce their pedagogical leadership skills. In addition, a centralised effort may be needed to develop a knowledge base, tools and guidelines to assist evaluation and assessment activities.

Manage local needs

Evaluation and assessment frameworks need to find the right balance between consistently implementing central education goals and adapting to the particular needs of regions, districts and schools. This can involve setting down national parameters, but allowing flexible approaches within these to meet local needs.

Design successfully, build consensus

To be designed successfully, evaluation and assessment frameworks should draw on informed policy diagnosis and best practice, which may require the use of pilots and experimentation. To be implemented successfully, a substantial effort should be made to build consensus among all stakeholders, who are more likely to accept change if they understand its rationale and potential usefulness.

I. The growing focus on evaluation and assessment

There is widespread recognition that evaluation and assessment arrangements are key to both improvement and accountability in school systems. This is reflected in their increasing importance in national education agendas. As countries strive to transform their educational systems to prepare all young people with the knowledge and skills needed to function in rapidly changing societies, some common policy trends can be observed in one form or another in most OECD countries, including decentralisation, school autonomy, greater accountability for outcomes and a greater knowledge management capacity. Decentralisation and school autonomy are creating a greater need for the evaluation of schools, school leaders and teachers while greater IT capacity allows for the development and analysis of large-scale student assessments as well as individualised assessment approaches. Results from evaluation and assessment are becoming critical to knowing whether the school system is delivering good performance and to providing feedback for further development. Evaluation and assessment are instrumental in defining strategies for improving practices within school systems with the ultimate goal of enhancing student outcomes. These developments are having a strong influence in the way in which policy makers monitor system, school, school leader, teacher and student performance.

Countries increasingly use a range of approaches for the evaluation and assessment of students, teachers, school leaders, schools and education systems. These are used as tools for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school, school leadership and teaching practices. Strong emphasis is being placed on better equipping and encouraging teachers to carry out self-appraisal and student formative assessment, on providing the incentives and means for school self-evaluation, on encouraging “value-added” evaluation and

on more regular standardised testing of students and national monitoring of the overall system. However, countries often face difficulties in implementing evaluation and assessment procedures. These may arise as a result of poor policy design, lack of analysis of unintended consequences, little capacity for educational agents to put procedures into practice, lack of an evaluation culture, or deficient use of evaluation results.

This report is concerned with evaluation and assessment policies in school systems that can help countries achieve their educational goals and student learning objectives. It draws on a major study, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

II. Main trends within evaluation and assessment

Although not all countries are in the same position, a number of trends within evaluation and assessment emerge.

Educational evaluation in school systems is expanding

It is apparent that education policy is increasingly conferring a central strategic role to evaluation and assessment as indispensable tools for improvement, accountability, educational planning and policy development. In the last two decades, most countries have introduced a wide range of measures intended to improve evaluation and assessment at all levels from the student to the school system itself. The expansion of educational evaluation results from increased demands for effectiveness, equity and quality in education so new economic and social needs are met. The greater importance of evaluation and assessment in education policy has involved the creation of specifically dedicated agencies which assume a central role in the governance of the evaluation and assessment framework. This recognises the need for specialised expertise, the imperative of building adequate capacity to deliver evaluation and assessment policies and the necessity of introducing some independence vis-à-vis education authorities.

There is a greater variety of evaluation and assessment activities

The expansion of educational evaluation has been accompanied by considerable diversification of evaluation and assessment activities. Although educational evaluation within school systems is not a recent concern, it has traditionally focussed mostly on the assessment of students. In recent years, countries are increasingly developing more comprehensive evaluation and assessment frameworks. These involve more responsibility given to the school itself, through greater emphasis on school self-evaluation; greater importance of external school evaluation as accountability requirements increase; more emphasis on school leadership and its appraisal as the pedagogical role of school leaders is consolidated; the emergence of formal systems of teacher appraisal; the expansion of student standardised assessment to monitor learning outcomes; the growing importance of performance data, particularly relating to student outcomes, to inform school and classroom practices as well as system-level policies; and the growing emphasis on the use of data for formative assessment.

Educational measurement and indicators development are rising in importance

The introduction of national standardised assessments for students in a large number of countries reflects the stronger focus on measuring student outcomes. These make data on student learning outcomes available, providing a picture of the extent to which student learning objectives are being achieved, and they grant the opportunity to compare student learning outcomes across individual schools, regions of the country and over time. Also, for the purpose of monitoring education systems and evaluating school performance, data are increasingly complemented by a wide range of education indicators based on demographic, administrative and contextual data collected from individual schools. Most countries have developed comprehensive national indicator frameworks. It is now common practice to report statistics and indicators in education in an annual publication. International benchmarking is also increasingly common.

Larger and more varied uses are given to evaluation and assessment results

Countries are giving a more varied use to evaluation and assessment results. There is a growing interest in using evaluation results for formative purposes. School leaders, teachers and policy makers are more and more using evaluation results to identify areas where schools are performing well, and where they may need to improve. These data may help shape policy and/or school management decisions on resource

distribution, curriculum development and definition of standards, or strategies for professional development. Another increasingly marked focus is the use of evaluation and assessment results to hold policy makers, school leaders and teachers accountable.

Accountability as a purpose of evaluation and assessment is gaining in importance

Countries are increasingly using evaluation and assessment for accountability purposes. This can take a variety of forms. First, there is a growing trend of public reporting, including the publication of standardised student assessment results at the school level for use by parents, government officials, the media and other stakeholders, the publication of school inspection reports, school annual reports, and system level reports providing an assessment of the state of education. Second, evaluation and assessment results are increasingly used to reward or sanction the performance of individual school agents. This goes alongside the expansion of school external evaluation and teacher appraisal procedures. A number of countries have instituted systems whereby either schools, school leaders or teachers receive rewards for their good performance or are the subject of sanctions for underperformance.

There is greater reliance on educational standards

The focus on student learning outcomes has, in many countries, driven the establishment or underlined the importance of educational standards for the quality of the work of schools and school agents, and encouraged means for monitoring progress towards those standards. Educational standards refer to descriptions of what students should know (content standards) and be able to do (performance standards) at different stages of the learning process. In many countries, there is growing emphasis on the development and use of ambitious educational standards as the basis of assessment and accountability. By creating a set of standards against which student performance can be measured, countries aim to assess students against a desired measurable outcome.

Assessment is becoming more international

National education debates are increasingly shaped by international comparisons, particularly of student performance in international student surveys. The growing availability of internationally comparable data on student performance has, in important ways, influenced national discussions about education and fostered education policy reforms in countries. International comparative data put countries under pressure to attain higher levels of performance building on policies identified as potentially effective in high-performing countries. The expansion of international assessment has also significantly contributed for some countries to introduce national standardised assessments.

Assessment involves greater technological sophistication

The expansion of assessment, particularly the spreading out of standardised student assessment, as well as the management of the data it generates has greatly benefited from greater capacity of information and communication technologies. Improvements include more individualised assessment approaches, better assessment of cognitive skills such as problem solving, capacity for rapidly marking large-scale assessments, reliability in marking and reduced cost to administer student assessment. Other examples include the development of rapid-assessment – a computer-facilitated approach to frequent, brief formative student assessment, more sophisticated value-added models to determine a school's or a teacher's contribution to student learning, and data information systems providing new opportunities for information sharing across school agents.

III. Main policy challenges

In all countries, there is widespread recognition that evaluation and assessment frameworks are key to building stronger and fairer school systems. Countries also emphasise the importance of seeing evaluation and assessment not as ends in themselves, but instead as important tools for achieving improved student outcomes. However, there are a range of challenges in ensuring that evaluation and assessment reach such ultimate objective. Although each country context is unique, some common policy challenges emerge. These are listed in Table 1.

Table 1. Main challenges in evaluation and assessment

Domain	Main challenges
The evaluation and assessment framework	<ul style="list-style-type: none"> Building a coherent and integrated evaluation and assessment framework Balancing the accountability and development functions of evaluation and assessment Ensuring articulations within the evaluation and assessment framework Securing links with classroom practice Finding a desirable measure of national consistency as against local diversity Developing competencies for evaluation and assessment and for using feedback Overcoming the challenge of implementation
Student assessment	<ul style="list-style-type: none"> Aligning educational standards and student assessment Finding a balance between summative and formative assessment Balancing external assessments and teacher-based assessments in the assessment of learning Developing fair assessments to all student groups Designing large-scale assessments that are instructionally useful Ensuring fairness in assessment and marking across schools Securing informative reporting of student assessment results
Teacher appraisal	<ul style="list-style-type: none"> Developing a shared understanding of high-quality teaching Balancing the developmental and accountability functions of teacher appraisal Accounting for student results in the appraisal of teachers Developing adequate skills for teacher appraisal Using teacher appraisal results to shape incentives for teachers

<p>School evaluation</p>	<p>Aligning the external evaluation of schools with school self-evaluation</p> <p>Ensuring the centrality of the quality of teaching and learning</p> <p>Balancing information to parents with fair and reasonable public reporting on schools</p> <p>Building competence in the techniques of self-evaluation and external school evaluation</p> <p>Improving the data handling skills of school agents</p>
<p>The appraisal of school leaders</p>	<p>Developing school leadership appraisal as an integral part of the evaluation and assessment framework</p> <p>Developing a clear understanding of effective school leadership</p> <p>Placing pedagogical/learning-centred leadership at the heart of school leadership appraisal</p> <p>Combining the improvement and accountability functions of school leadership appraisal</p> <p>Ensuring that all school leaders have opportunities for professional feedback</p> <p>Using appraisal results to shape incentives for school leaders</p>
<p>Education system evaluation</p>	<p>Meeting information needs at the system level</p> <p>Monitoring key student learning outcomes</p> <p>Securing comparability over time and across schools</p> <p>Developing analytical capacity to use education system evaluation results for improvement</p> <p>Communicating education system evaluation results clearly and comprehensively</p> <p>Maximising use of system-level information</p>

IV. Main policy directions

To meet the challenges outlined above, a number of policy options are suggested across the areas analysed – the evaluation and assessment framework, student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation. Table 2 summarises the main policy directions (see Annex B for a complete list of policy directions). Not all of the policy directions apply equally to all countries. In a number of cases many, or most, of the policy suggestions are already in place, while for other countries they may have less relevance because of different social, economic and educational structures and traditions. This is a challenging agenda, but tackling one area without appropriate policy attention to inter-related aspects will lead to only partial results. Nevertheless, it is difficult to address all areas simultaneously, and resource constraints mean that trade-offs are inevitable.

Table 2. Main policy directions

Policy objective	Main policy directions
The evaluation and assessment framework Embracing a holistic approach	<p>Integrate the evaluation and assessment framework</p> <p>Align the evaluation and assessment framework with educational goals and student learning objectives</p> <p>Secure links to the classroom and draw on teacher professionalism</p> <p>Promote national consistency while giving room for local diversity</p> <p>Establish articulations between components of the evaluation and assessment framework</p> <p>Place the students at the centre of the evaluation and assessment framework</p> <p>Sustain efforts to improve capacity for evaluation and assessment</p> <p>Engage stakeholders and practitioners in the design and implementation of evaluation and assessment policies</p>
Student assessment Putting the learner at the centre	<p>Ensure a good balance between formative and summative assessment</p> <p>Establish safeguards against an overreliance on standardised assessments</p> <p>Draw on a variety of assessment types to obtain a rounded picture of student learning</p> <p>Support effective formative assessment processes</p> <p>Ensure the consistency of assessment and marking across schools</p> <p>Ensure that student assessment is inclusive and responsive to different learner needs</p> <p>Put the learner at the centre and build students' capacity to engage in their own assessment</p> <p>Maintain the centrality of teacher-based assessment and promote teacher professionalism</p> <p>Engage parents in education through adequate reporting and communication</p>

<p>Teacher appraisal Enhancing teacher professionalism</p>	<p>Resolve tensions between the developmental and accountability functions of teacher appraisal</p> <p>Consolidate regular developmental appraisal at the school level</p> <p>Establish periodic career-progression appraisal involving external evaluators</p> <p>Establish teaching standards to guide teacher appraisal and professional development</p> <p>Prepare teachers for appraisal processes and strengthen the capacity of school leaders for teacher appraisal</p> <p>Ensure that teacher appraisal feeds into professional development and school development</p> <p>Establish links between teacher appraisal and career advancement decisions</p>
<p>School evaluation From compliancy to quality</p>	<p>Ensure the focus for school evaluation is the improvement of teaching, learning and student outcomes</p> <p>Evaluate and adapt external school evaluation to reflect the maturity of the school evaluation culture</p> <p>Raise the profile of school self-evaluation and align external school evaluation with school self-evaluation</p> <p>Develop nationally agreed criteria for school quality to guide school evaluation</p> <p>Strengthen school principals' capacity to stimulate an effective school self-evaluation culture</p> <p>Promote the engagement of all school staff and students in school self-evaluation</p> <p>Promote the wider use of the results of external school evaluation</p> <p>Report a broad set of school performance measures with adequate contextual information</p>
<p>The appraisal of school leaders Fostering pedagogical leadership in schools</p>	<p>Promote the effective appraisal of school leaders within the broader assessment and evaluation framework</p> <p>Develop a common leadership framework or set of professional standards for school leaders</p> <p>Promote the appraisal of pedagogical leadership together with scope for local adaptation</p> <p>Build capacity for effective school leader appraisal</p> <p>Ensure school leader appraisal informs professional development</p> <p>Consider career advancement opportunities to reward successful school leaders</p>

<p>Education system evaluation Informing policies for system improvement</p>	<p>Ensure a broad concept of education system evaluation within the evaluation and assessment framework</p> <p>Ensure policy making is informed by high-quality measures, but not driven by their availability</p> <p>Develop a national education indicator framework and design a strategy to monitor student learning standards</p> <p>Ensure the collection of qualitative information on the education system</p> <p>Ensure collection of adequate contextual information to effectively monitor equity</p> <p>Establish and secure capacity for education system evaluation</p> <p>Strengthen analysis of education system evaluation results for planning and policy development</p>
---	---

Common policy themes

Despite the major differences and traditions across countries, they share some common policy priorities.

Fostering synergies within the evaluation and assessment framework

The full potential of evaluation and assessment will not be realised until the framework is fully integrated and is perceived as a coherent whole. This requires a holistic approach to building a complete evaluation and assessment framework in view of generating synergies between its components, avoiding duplication of procedures and preventing inconsistency of objectives.

At the outset, it might prove useful to develop a strategy or framework document that conceptualises a complete evaluation and assessment framework and articulates ways to achieve the coherence between its different components. The strategy should establish a clear rationale for evaluation and assessment and a compelling narrative about how evaluation and assessment align with the different elements in the education reform programme. It should describe how each component of the evaluation and assessment framework can produce results that are useful for classroom practice and school development activities. The strategy could also contribute to clarifying responsibilities of different actors for the different components and allow for better networking and connections between the people working on evaluation and assessment activities. As such, it should also create the conditions for a better articulation between the different levels of educational governance, including evaluation agencies and local education authorities.

Furthermore, the process of developing an effective evaluation and assessment framework should give due attention to: achieving proper articulation between the different evaluation components (e.g. school evaluation and teacher appraisal); warranting the several elements within an evaluation component are sufficiently linked (e.g. teaching standards and teacher appraisal; external school evaluation and school self-evaluation); and ensuring processes are in place to guarantee the consistent application of evaluation and assessment procedures (e.g. consistency of teachers' marks). This is in addition to proper links to initial teacher education and strategies for professional development; situating education system evaluation in the broader context of public sector performance requirements; and ensuring references for evaluation and assessment are well aligned with student learning objectives.

Aligning student learning goals with evaluation and assessment

A critical aspect in the effectiveness of the evaluation and assessment framework is its proper alignment with educational goals and student learning objectives. This involves a range of aspects. First, it requires evaluation and assessment procedures to align with the main principles embedded in educational goals and student learning objectives. For instance, if educational goals are based on principles such as student-

centred learning, collaborative work, achievement of competencies and assessment for learning then there should be greater emphasis on the developmental function of evaluation and assessment, involving more attention to student formative assessment, greater emphasis on self-reflection for all the school agents, greater focus on continuous improvement in teacher appraisal, and better use of results for feedback.

Second, evaluation and assessment procedures require direct alignment with student learning objectives. This implies designing fit-for-purpose student assessments which focus on the competencies promoted in student learning objectives, ensuring the overall evaluation and assessment framework captures the whole range of student learning objectives, and developing teaching and school management standards which are aligned with student learning objectives. Third, it is essential that all school agents have a clear understanding of education goals. This requires goals to be clearly articulated; the development of clear learning expectations and criteria to assess achievement of learning objectives; room for schools to exercise some autonomy in adapting learning objectives to their local needs; and collaboration among teachers and schools to ensure moderation processes which enhance the consistency with which learning goals are achieved. Fourth, it is essential to evaluate the impact of evaluation and assessment against student learning objectives on the quality of the teaching and learning. Particular attention should be given to identifying unintended effects as evaluation and assessment activities have considerable potential to determine the behaviour of school agents.

Focussing on the improvement of classroom practices and building on teacher professionalism

To optimise the potential of evaluation and assessment to improve what is at the heart of education – student learning – policy makers should promote the regular use of evaluation and assessment results for improvements in the classroom. All types of evaluation and assessment should have educational value, and be meaningful to those who participate in the evaluation or assessment. To this end, it is important that all those involved in evaluation and assessment at the central, local and school level have a broad vision of evaluation and assessment and of the need to bring together results from different types of evaluation and assessment activities to form rounded judgements about student learning, performance of school agents and practices within the school system and use evaluation and assessment information for further improvement.

This calls for an articulation of ways for the evaluation and assessment framework to generate improvements in classroom practice through the assessment and evaluation procedures which are closer to the place of learning. Evaluation and assessment have no value if they do not lead to the improvement of classroom practice and student learning. An important step in this direction is a national reflection about the nature and purpose of evaluation components such as school evaluation, school leader appraisal, teacher appraisal and student formative assessment within the overall education reform strategy and the best approaches for these evaluation components to improve classroom practices. Impacting classroom practice is likely to require the evaluation and assessment framework to place considerable emphasis on its developmental function. Channels which are likely to reinforce links to classroom practice include: an emphasis on teacher appraisal for the continuous improvement of teaching practices; ensuring teaching standards are aligned with student learning objectives; involving teachers in school evaluation, in particular through conceiving school self-evaluation as a collective process with responsibilities for teachers; focussing school evaluation on the quality of teaching and learning and their relationship to student learning experiences and outcomes; promoting the appraisal of the pedagogical leadership of school leaders; ensuring that teachers are seen as the main experts not only in instructing but also in assessing their students, so teachers feel the ownership of student assessment and accept it as an integral part of teaching and learning; building teacher capacity for student formative assessment; and building teachers' ability to assess against educational standards.

The central agent in securing links between the evaluation and assessment framework and the classroom is the teacher. This highlights the importance for evaluation and assessment frameworks to draw on the professionalism of teachers in ensuring evaluation and assessment activities result in authentic improvement of classroom practices and student learning. In addition, establishing links between evaluation and assessment and classroom learning requires establishing clear roles for local structures – school management, school supervision, local education authorities – in the implementation of evaluation and assessment policies. The point is that the fulfilment of the developmental function of evaluation and assessment requires articulation at the local level.

Effectively conceiving the accountability uses of evaluation and assessment results

Evaluation and assessment provide a basis for monitoring how effectively education is being delivered to students and for assessing the performance of systems, schools, school leaders, teachers and students, among others. They can serve as an instrument for the accountability of school agents when the results of an evaluation or assessment have stakes for school agents such as linkages to career advancement or salary progression, one-off rewards, sanctions, or simply information to parents in systems based on parental school choice. By measuring student outcomes and holding teachers, school leaders and schools responsible for results, accountability systems intend to create incentives for improved performance and identify “underperforming” schools and school agents.

At the same time, high-stakes uses of evaluation and assessment results might lead to distortions in the education process as a result of school agents concentrating on the measures used to hold them accountable. For instance, if those measures are based on student standardised tests, this might include excessive focus on teaching students the specific skills that are tested, narrowing the curriculum, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and potentially even manipulation of results. Also, when the framework tends to stress the accountability function there is a risk that evaluation and assessment are perceived mostly as instruments to hold school agents accountable, to “control” and assess compliance with regulations. An additional challenge is that the developmental function of evaluation and assessment might be hindered in processes which stress the accountability function as the high stakes involved will limit the ability of school agents to openly reveal their weaknesses in view of receiving feedback to improve their practices.

As a result, it is important to design the accountability uses of evaluation and assessment results in such a way these undesired effects are minimised. This involves safeguards against excessive emphasis on particular measures, such as student standardised tests, to hold school agents accountable and drawing on a broad range of assessment information to make judgements about performance; communicating that the ultimate objective of evaluation and assessment is to enhance student outcomes through the improvement of practices at the different levels of the school system; building on a variety of evaluation and assessment procedures achieving each a well-identified distinct function; ensuring that the publication of quantitative data is perceived as fair by schools and set in a wider set of evidence; and conceiving individual performance-based rewards for school personnel as career advancement opportunities and non-monetary rewards.

Placing the student at the centre

Given that the fundamental purpose of evaluation and assessment is to improve the learning of the students, a key principle is to place the students at the centre of the framework. This translates into teaching, learning and assessment approaches which focus on students’ authentic learning. Students should be fully engaged with their learning, contributing to the planning and organisation of lessons, having learning expectations communicated to them, assessing their learning and that of their peers, and benefitting from individualised support and differentiated learning. To become lifelong learners, students need to be able to assess their own progress, make adjustments to their understandings and take control of their own learning. Student feedback to teachers can also be used for teacher formative appraisal. In addition, it is important to build community and parental involvement and an acceptance of learning and teaching as a shared responsibility. A particularly important priority for some countries is to reduce the high rates of grade repetition. There are alternative ways of supporting those with learning difficulties in the classroom.

In addition, evaluation and assessment should focus on improving student outcomes and achieving student learning objectives. This should be reflected in the priorities for national monitoring, the importance of evidence on student performance for school evaluation and teacher appraisal, the value of clear reporting on student results, and the emphasis on feedback for improving student learning strategies. There is also the increasing recognition that the monitoring of student outcomes must extend beyond knowledge skills in key subject areas and include broader learning outcomes, including students’ critical thinking skills, social competencies, engagement with learning and overall well-being.

Going beyond measurement in educational evaluation

As described earlier, measures of student learning are becoming increasingly available (in particular through

national standardised assessments) and most countries have developed education indicator frameworks. Performance in schools is increasingly judged on the basis of effective student learning outcomes. This is part of the general shift to outcome measures in the public sector. The advantage is that student outcomes become the focal point for analysis.

An imperative is that measures of performance are broad enough to capture the whole range of student learning objectives. However, it is not always possible to devise indicators and measures of good quality across all the objectives of the education system. Hence, it needs to be recognised that policy making at the system level needs to be informed by high-quality data and evidence, but not driven by the availability of such information. Qualitative studies as well as secondary analysis of the available measures and indicators are essential information to take into account in policy development and implementation. Qualitative approaches include the narrative provided by external school evaluation reports, key stakeholder feedback on broader outcomes (e.g. school climate, student engagement), and qualitative appraisal of teachers and school leaders. The qualitative aspects can feed into the policy debate by providing evidence on a broader set of student learning outcomes, as well as help shed light on some of the factors associated with student learning outcomes.

Building capacity for evaluation and assessment

The development of an effective evaluation and assessment framework involves considerable investment in developing competencies and skills for evaluation and assessment at all levels. Hence, an area of policy priority is sustaining efforts to improve the capacity for evaluation and assessment. Depending on country specific circumstances, areas of priority might be: developing teachers' capacity to assess against student learning objectives; improving the skills of teachers for formative assessment; improving the data handling skills of school agents; or developing expertise for teacher appraisal and school evaluation, including ensuring that designated evaluators are qualified for their role. Capacity building through adequate provision of initial teacher education and professional development should be a priority making sure provision is well aligned with the national education agenda. This should go alongside the development of training and competency descriptions for key people within the evaluation and assessment framework.

There is also a need to reinforce the pedagogical leadership skills of school directors as their role in many countries still retains a more traditional focus on administrative tasks. The objective is that school leaders operate effective feedback, coaching and appraisal arrangements for their staff and effectively lead whole-school evaluation processes. Peer learning among schools should also be promoted. In addition, there needs to be strong capability at the national level to steer evaluation and assessment. This can be ensured through the establishment of agencies with high levels of expertise which have the capacity to foster the development of skills for evaluation and assessment across the system. Such agencies could provide important leadership in modelling and disseminating good practice within the evaluation and assessment framework.

A further strategy involves initiatives at the central level to build up a knowledge base, tools and guidelines to assist evaluation and assessment activities. These typically include detailed plans to implement student learning objectives, including guidelines for schools and teachers to develop student assessment criteria. Other examples are tools for teachers to use in the assessment of their students (e.g. test items banks), Internet platforms proposing formative teaching and learning strategies, tools for the self-appraisal of teachers, instruments for school leaders to undertake teacher appraisal, and resources for school self-evaluation.

Designing evaluation and assessment procedures which are fit for purpose

Establishing clarity about the purposes and appropriate uses of different evaluations and assessments is important to ensure that evaluation and assessment frameworks optimally contribute to improvements at the classroom, school and system level. A key goal for countries is to develop, for each component of the evaluation and assessment framework, a clear vision and strategy where different approaches developed nationally and locally each serve a clearly defined purpose and the format of the evaluation or assessment is aligned to that particular purpose. For evaluation or assessment to be meaningful, it must be well-aligned to the type of skills and competencies that are valued. Coherent evaluation and assessment frameworks should aim to align student learning objectives, practices in the school system, and evaluation and assessment around key learning goals, and include a range of different evaluation and assessment approaches and formats, along with opportunities for capacity building at all levels.

More specifically, because standardised central student assessment is a relatively new phenomenon in many OECD countries, it is important to be clear about its purposes, to develop large-scale assessments over time to be able to accommodate the purposes that are reasonable, point out inappropriate uses and provide guidance for the way in which these assessments can be used as part of a broader assessment framework. Also, to build a systematic and coherent system of teacher appraisal, it is important that the aspects it seeks to monitor and improve are clear, that the approaches to appraisal are adapted to the different stages of a teachers' career and in line with the purposes they are aiming to achieve. Similarly, the fundamental purpose of school evaluation needs to be clearly and consistently understood across the school system. For instance, external school evaluation can be part of the strategy to bring about general improvement across all schools or, more narrowly, it can focus on "underperforming schools".

Evaluation and assessment systems also need to underline the importance of responding to individual needs and school community contexts, and design evaluation and assessment strategies that suit the needs of different learner groups or distinct schools agents.

Balancing national consistency with meeting local needs

In order to contribute to national reform agendas, a certain degree of national consistency of approaches to evaluation and assessment is desirable. This is likely to provide greater guarantees that evaluation and assessment practices are aligned with national student learning objectives. However, in certain countries, there are strong traditions of local ownership – at the jurisdiction level (federal systems), local level (region or municipality), or school level. In these cases, a high degree of autonomy is granted in school policies, curriculum development and evaluation and assessment. There is an understanding that shared or autonomous decision making and buy-in from those concerned are essential for the successful implementation of evaluation and assessment policy. It is also clear that local actors are in a better position to adapt evaluation and assessment policies to local needs.

Hence, the evaluation and assessment framework will need to find the right balance between national consistency and local diversity. A possible approach is to agree general principles for the operation of procedures such as school evaluation, teacher appraisal, school leader appraisal and student assessment while allowing flexibility of approach within the agreed parameters to better meet local needs. The principles agreed should come along with clear goals, a range of tools and guidelines for implementation. In decentralised systems, it is also important to encourage the different actors to co-operate, share and spread good practice and thereby facilitate system learning, development and improvement.

Implementing evaluation and assessment policy successfully

The process of evaluation and assessment policy design and implementation involves a number of challenges to yield sound results. Ideally, evaluation and assessment policy would need to be based upon informed policy diagnosis, drawn on best practice, backed up by adequate research evidence, and consistent – both intrinsically and with other education policies. Of equal importance is consensus-building among the various stakeholders involved – or with an interest – in educational evaluation. This should go alongside the involvement of practitioners such as school leaders and teachers in the design, management and analysis of evaluation and assessment policies.

In order to build consensus, it is important that all stakeholders see proposed evaluation and assessment policies within the broader policy framework and strategy. Indeed, individuals and groups are more likely to accept changes that are not necessarily in their own best interests if they understand the rationale for these changes and can see the role they should play within the broad evaluation and assessment framework. There is therefore much scope for government authorities to foster the chances of successful policy implementation, by improving communication on the long-term vision of what is to be accomplished for evaluation and assessment as the rationale for proposed reform packages.

Other approaches for successful policy implementation include the use of pilots and policy experimentation when needed, opportunities for education practitioners to express their views and concerns as evaluation and assessment policies are implemented, the communication of key evaluation and assessment results to stakeholders, developing expertise and capacity for evaluation and assessment across the system, reducing excessive bureaucratic demands on schools, and ensuring sufficient resources are provided for implementation.

Traduction du Résumé**Abstract Translation**

OCDE

OCDE : RÉSUMÉ DES EXAMENS D'ÉVALUATION DE L'OCDE DANS LE DOMAINE DE L'ÉDUCATION

Ce papier présente une vue globale des conclusions principales du rapport « Des synergies pour un meilleur apprentissage : une perspective internationale sur l'évaluation¹ » sur l'enquête de l'OCDE, menée sur trois années des politiques d'évaluation dans les pays de l'OCDE. Le rapport fournit une analyse internationale comparative et des recommandations sur la politique à adopter pour mener des évaluations dans un cadre cohérent afin d'améliorer la qualité, l'équité et l'efficacité de l'éducation.

Ce papier constitue la contribution de l'OCDE au Colloque International sur l'Évaluation en Éducation et en Formation, organisé en octobre 2015, par l'Instance Nationale d'Évaluation (INE) au Maroc. Ce papier comporte un résumé et des chapitres du rapport de l'OCDE « Des synergies pour un meilleur apprentissage : une perspective internationale sur l'évaluation² ».

1 - OCDE (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, examen de l'OCDE de l'évaluation en Éducation, OCDE, Publishing, Paris, DOI.

2 - Les auteurs du rapport de l'OCDE *Synergies for Better Learning: An International Perspective on Evaluation and Assessment* sont Déborah Nusche, Thomas Radinger, Paulo Santiago (coordonnateur) et Claire Shewbridge de la Direction pour l'Éducation et Compétences.

Bibliography and References

Campbell, C. and B. Levin (2008), "Using data to support educational improvement", *Educational Assessment, Evaluation and Accountability*, No. 21, pp. 47-65.

Clarke, M. (2012), "What matters most for student assessment systems: A framework paper", *SABER – Systems Approach for Better Education Results*, Working Paper No. 1, The World Bank, Washington, DC.

Crooks, T.J. (2004), "Tensions between assessment for learning and assessment for qualifications", paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies (ACEAB), Fiji, March 2004.

Darling-Hammond, L., A. Amrein-Beardsley, E. Haertel and J. Rothstein (2012), "Evaluating teacher evaluation", *Phi Delta Kappan*, Vol. 93, No. 6, pp. 8-15.

Davis, S., K. Kearney, N. Sanders, C. Thomas and R. Leon (2011), *The Policies and Practices of Principal Evaluation: A Review of the Literature*, WestEd, San Francisco, California,

www.wested.org/online_pubs/resource1104.pdf.

Earl, L. (2003), "Assessment as learning: Using classroom assessment to maximize student learning", in T.R. Guskey and R.J. Marzano (eds.), *Experts in Assessment Series*, Corwin Press, Thousand Oaks, California.

Fazekas, M. and T. Burns (2012), "Exploring the complex interaction between governance and knowledge in education", *OECD Education Working Papers*, No. 67, OECD Publishing, Paris,

www.oecd.org/edu/workingpapers.

Goldring, E., A.C. Porter, J. Murphy, S.N. Elliott and X. Cravens (2007), *Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices*, a report prepared for the Wallace Foundation grant on Leadership Assessment, March 2007,

[www.wallacefoundation.org/knowledgecenter/school-leadership/principal-evaluation DocumentsAssessing-Learning-Centered-Leadership.pdf](http://www.wallacefoundation.org/knowledgecenter/school-leadership/principal-evaluation/DocumentsAssessing-Learning-Centered-Leadership.pdf).

Hooge, E., T. Burns and H. Wilkoszewski (2012), "Looking Beyond the Numbers: Stakeholders and Multiple School Accountability", *OECD Education Working Papers*, No. 85, OECD Publishing, Paris,

www.oecd.org/edu/workingpapers.

Looney, J. (2011a), "Alignment in complex systems, achieving balance and coherence", *OECD Education Working Papers*, No. 64, OECD Publishing, Paris,

www.oecd.org/edu/workingpapers.

Linn, R.L. (2000), "Assessments and accountability", *Educational Researcher*, Vol. 29, No. 2, pp. 4-16.

Looney, J. (2011b), "Integrating formative and summative assessment: Progress toward a seamless system?", *OECD Education Working Papers*, No. 58, OECD Publishing, Paris,

www.oecd.org/edu/workingpapers.

Looney, J. (2009), "Assessment and innovation in education", *OECD Education Working Papers*, No. 24, OECD Publishing, Paris,

www.oecd.org/edu/workingpapers.

OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, *OECD Reviews of Evaluation and Assessment in Education*, OECD Publishing, Paris.

DOI: <http://dx.doi.org/10.1787/9789264190658-en>

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in*

Education: Norway, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011b), OECD Reviews of Evaluation and Assessment in Education: Sweden, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Santiago, P., F. Benavides, C. Danielson, L. Goe and D. Nusche (2013), OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile, OECD Publishing, Paris, at

www.oecd.org/edu/evaluationpolicy.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), OECD Reviews of Evaluation and Assessment in Education: Portugal, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012b), OECD Reviews of Evaluation and Assessment in Education: Mexico, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012c), OECD Reviews of Evaluation and Assessment in Education: Czech Republic, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), OECD Reviews of Evaluation and Assessment in Education: Australia, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Santiago, P. and F. Benavides (2009), Teacher Evaluation: A Conceptual Framework and Examples of Country Practices, OECD, Paris,

www.oecd.org/edu/evaluationpolicy.

Shewbridge, C., M. Hulshof, D. Nusche and L. Stenius Stæhr (2014), OECD Reviews of Evaluation and Assessment in Education: Northern Ireland, United Kingdom, OECD Publishing, Paris, forthcoming at

www.oecd.org/edu/evaluationpolicy.

Shewbridge, C., J. van Bruggen, D. Nusche and P. Wright (2014), OECD Reviews of Evaluation and Assessment in Education: Slovak Republic, OECD Publishing, Paris, forthcoming at

www.oecd.org/edu/evaluationpolicy.

Shewbridge, C., M. Ehren, P. Santiago and C. Tamassia (2012), OECD Reviews of Evaluation and Assessment in Education: Luxembourg, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011a), OECD Reviews of Evaluation and Assessment in Education: Denmark, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Shewbridge, C., M. Hulshof, D. Nusche and L. Stoll (2011b), OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium, OECD Publishing, Paris,

www.oecd.org/edu/evaluationpolicy.

Taylor, E.S. and J.H. Tyler (2011), "The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers", NBER Working Paper Series, No. 16877, National Bureau of Economic Research, Cambridge, Massachusetts.

L'ÉVALUATION DES POLITIQUES SCOLAIRES : UN INSTRUMENT AU SERVICE DE LA CONSTRUCTION DES POLITIQUES SCOLAIRES DE LONG TERME ?

Nathalie Mons *

Résumé

Depuis les années 1990, dans de nombreux pays, notamment de l'OCDE, les systèmes scolaires ont connu un raz-de-marée réformiste (Mons, 2007). L'évaluation y occupe une place de choix, du moins dans les discours politiques. Elle serait devenue un instrument crucial d'amélioration de la qualité des systèmes scolaires. Mais la réalité est en décalage avec ces discours politico-administratifs volontaristes. L'OCDE (2015) a montré que, seules 11% des politiques scolaires menées entre 2008 et 2014 dans les pays de l'OCDE intégraient une évaluation de l'impact de la réforme. De plus, quand elle existe, cette évaluation des politiques éducatives se cantonne souvent à la publication de rapports qui circulent peu auprès des praticiens et décideurs et ne produisent donc que des effets limités sur les décisions des politiques et les pratiques des acteurs de terrain.

Les discours, plus ou moins fondés, développés autour d'une crise de l'éducation ont conduit à un raz-de-marée réformiste dans de nombreux pays depuis les années 1990 (Mons, 2007). Les grandes enquêtes internationales, comme PISA, ont pu montrer qu'à budget identique, selon les politiques menées, les résultats des élèves - les scores nationaux moyens mais aussi l'étendue des inégalités scolaires - peuvent varier, interrogeant les choix politiques historiques.

Dans ce raz-de-marée réformiste, l'évaluation occupe une place de choix, dans les discours politiques. Elle serait devenue un instrument crucial d'amélioration de la qualité des systèmes scolaires. Mais la réalité est en décalage avec ces discours politico-administratifs volontaristes. Si l'évaluation des acquis des élèves s'est bien généralisée, notamment dans les pays de l'OCDE, l'évaluation des politiques éducatives reste en retrait. Elle souffre aujourd'hui de trois faiblesses. Une enquête récente (OCDE, 2015) a montré qu'entre 2008 et 2014, seules 12% des réformes mises en oeuvre dans l'OCDE relevaient du champ de l'évaluation contre, par exemple près d'un quart de réformes portant sur l'amélioration du fonctionnement des écoles (school improvement reforms). Toujours selon cette étude, l'évaluation demeure de plus en plus peu intégrée au cycle de fabrication des politiques éducatives. Seules 10% des politiques scolaires analysées dans l'enquête intègrent une évaluation de leur impact. Troisième faiblesse : quand elle existe cette évaluation des politiques éducatives se cantonne souvent à la publication de rapports qui circulent peu auprès des praticiens et décideurs et ne produisent donc que des effets limités sur les décisions des politiques et sur les pratiques des acteurs de terrain. En résumé, malgré des discours volontaristes, l'évaluation en éducation est peu présente et quand elle est développée, peu incluse dans un cercle vertueux de fabrication de politiques publiques, ce qui interroge au final son efficacité.

Or la faible intégration de l'évaluation dans le cycle de fabrication des politiques éducatives a des effets négatifs d'ampleur. Trois écueils apparaissent notamment : faute d'accrochage à une base scientifique, les réformes sont souvent fondées sur des a priori idéologiques - chaque camp politique a sa perception des « bonnes solutions » en éducation. Il en découle une vision court-termiste des politiques éducatives, qui changent sans évaluation à chaque alternance politique. Ces revirements continus et cette absence de légitimité scientifique produisent à leur tour des effets négatifs lors de la mise en œuvre des réformes : les praticiens, peu convaincus par les discours politiques idéologiques adoptent bien souvent la stratégie de l'autruche qui consiste à attendre la prochaine réforme sans mettre en œuvre la politique en cours. Il en

* Présidente du Conseil national d'évaluation du système scolaire (Cnesco), France
Professeure de sociologie, Université de Cergy-Pontoise, France

résulte une perte d'autorité, de crédibilité et de légitimité de l'action publique en éducation.

Inverser cette tendance n'est pas aisée. Décréter la production d'évaluation ne suffit pas, car, comme nous l'avons précisé en amont, le lien entre évaluation et évolution des décisions ou des pratiques des acteurs de terrain n'a rien d'automatique. Transformer l'évaluation en levier du changement institutionnel en éducation, en faire un instrument au service de la construction de politiques scolaires de long terme efficaces doit correspondre à une stratégie réfléchie.

Cette stratégie doit intégrer au moins certaines des caractéristiques suivantes :

1. L'évaluation doit être scientifiquement développée pour être légitime auprès de tous les acteurs parties prenantes de la communauté éducative, seule une évaluation scientifique, pluridisciplinaire peut entraîner l'adhésion des acteurs ;
2. L'évaluation doit être cohérente et complète (depuis l'analyse des acquis des élèves jusqu'à l'étude des effets des politiques nationales en passant par les résultats des établissements scolaires, des dispositifs intermédiaires ou politiques locales) ;
3. Donner une perspective internationale à l'évaluation permet de renforcer sa richesse - davantage de résultats d'évaluation et de recherche sont à disposition - et sa légitimité - les évaluateurs externes au pays apportent une crédibilité complémentaire à l'activité évaluative ;
4. L'évaluation doit être participative : les acteurs qui devront s'emparer des résultats des évaluations doivent dès le début être associés aux enquêtes pour pouvoir y adhérer. Ce modèle de l'« évaluation participative » (qui s'oppose par exemple au « modèle d'évaluation gestionnaire ») permet d'éviter les écueils du rejet de l'« évaluation-sanction » et de l'indifférence face à l'« évaluation hors-sol », évaluation qui ne répond pas aux questions que les praticiens ou les politiques se posent pour pouvoir progresser ;
5. La diffusion des résultats des évaluations doit être pensée comme un objectif en soi, une phase de vie des évaluations n'allant pas de soi (« ideas don't fly » expliquent les politistes). Il faut donc penser une stratégie particulière de circulation des idées qui s'incarnent dans des véhicules d'information concrets et s'adressent à l'ensemble des parties prenantes concernées sans en omettre (professionnels de l'éducation, politiques mais aussi grand public et notamment parents).

Cette politique d'évaluation active au service de la construction de politiques éducatives de long terme efficaces a commencé à être mise en œuvre en France au Conseil national d'évaluation du système scolaire (Cnesco), jeune institution créée par la loi en 2013. Ainsi en complément des rapports d'évaluation qui associent systématiquement les praticiens et un réseau d'une centaine de chercheurs-associés français et internationaux, le Cnesco a développé des formats d'activité originaux qui visent à diffuser les résultats des évaluations et des recherches en direction des publics multiples de l'éducation. Ainsi, les conférences de consensus permettent de créer des passerelles entre le monde de l'évaluation et de la recherche et celui des praticiens, représentés par un jury qui a pour mission d'élaborer des recommandations au vu des résultats des évaluations (par exemple sur les pratiques pédagogiques efficaces pour les premiers apprentissages en mathématiques). Les conférences de comparaisons internationales visent, elles, un public de décideurs et leur permet d'interagir avec des décideurs et des chercheurs étrangers sur des politiques scolaires précises (par exemple les bénéfices de la mixité sociale à l'école et les stratégies pour la construire dans le système scolaire). Les forums en région, enfin, s'adressent au grand public, et aux parents en particulier (par exemple sur la thématique du manque d'efficacité du redoublement et des solutions alternatives).

Traduction du Résumé

Abstract Translation

Nathalie Mons

EVALUATION OF SCHOOL SYSTEMS : AN INSTRUMENT IN THE SERVICE OF BUILDING LONG TERM EDUCATIONAL POLICIES

Since the 1990s, in main countries, namely OECD ones, education systems have known a reform tidal wave (Mons, 2007). Evaluation occupies a key place, at least in political speeches. It has become a crucial instrument for education system quality improvement. However, reality is sometimes at odds from these administrative speeches. OECD (2015) has shown that only 11% from school policies between 2008 and 2014 in OECD countries do include evaluation of impact of reforms. Moreover, when education policies evaluation exists, it is often limited to reports publication, which produces only limited efforts in decision-making and practices for actors on the ground.

-2

INSTANCES ET STANDARDS DE L'ÉVALUATION



ENSURING THE RELEVANCY OF ACCREDITATION STANDARDS IN THE U.S.: A CASE STUDY

Paula Harbecke *

Abstract

The Commission on Institutions of Higher Education (CIHE) is the higher education accreditation body of the New England Association of Schools and Colleges (NEASC), one of six regional accrediting agencies in the United States. Currently, NEASC accredits approximately 240 institutions of higher education across the six New England states and overseas. The NEASC Standards of Accreditation establish the criteria of quality that institutions are required to demonstrate in order to achieve accreditation.

Because NEASC is a membership organization, the Standards are developed and regularly updated through a participatory process involving member and candidate institutions in the articulation of the dimensions of quality required for accreditation. Recognizing the importance of higher education to the citizenry and for the economy, the public, as well, is invited to participate in the process. Thus, the Standards represent the accrued wisdom of over 200 colleges and universities and interested others about the essential elements of institutional quality. In other words, the NEASC Standards can be said to represent the articulation by the New England higher education community of the characteristics and behavior of those colleges and universities that – as accredited institutions – deserve the public trust.

The Commission continually evaluates the relevancy and effectiveness of its Standards and its processes for applying them, and makes such changes as conditions warrant. The current NEASC Standards went into effect January 2006, with midterm revisions effective July 2011. In 2014, the Commission again began the process of updating its Standards at a time of great change in American higher education, striving to articulate Standards for Accreditation for the next decade.

In October and November 2014, representatives of member institutions were invited to a series of eight meetings around the region and one in London for international institutions to review the current standards and the context for standards revision. Many useful comments were submitted regarding the current Standards (what works well, what doesn't) and the membership's desires for the revised Standards (what should be included and what should not be). The discussion and written feedback from individuals and groups, along with the experience of the Commission with changes in higher education, were key to identifying five issues meriting further consideration in the revision of the Standards for Accreditation:

- 1. Student Achievement and Success.*
- 2. Faculty: Roles and Composition.*
- 3. Partnerships: with institutions, for-profit entities, and others.*
- 4. Competency-Based Education, Prior Learning Assessment, and "Credits from Elsewhere."*
- 5. Technology, including the library and distance education.*

Using the New England region as a case study, this session will discuss the process periodically undertaken by accrediting bodies in the United States to update and keep current Standards of Accreditation that establish the criteria used to assess institutional quality. The five issues identified by the membership during the current revision process and the impact of each on the Standards will be examined.

I. Background Information

As specified by its policy, and in light of changes in higher education and public expectations for accreditation, every ten years the New England Association of Schools and Colleges' Commission on Institutions of Higher Education undertakes a comprehensive process to review and update its Standards for Accreditation. The Commission's current Standards were adopted in 2005; in 2011 they underwent modest revisions. The most recent version of the NEASC-CIHE Standards, effective July 1, 2011, identifies eleven dimensions of quality:

* Vice President, New England Association of Schools and Colleges Commission on Institutions of Higher Education, USA

- Standard 1: Mission and Purposes
- Standard 2: Planning and Evaluation
- Standard 3: Organization and Governance
- Standard 4: The Academic Program
- Standard 5: Faculty
- Standard 6: Students
- Standard 7: Library and Other Information Resources
- Standard 8: Physical and Technological Resources
- Standard 9: Financial Resources
- Standard 10: Public Disclosure
- Standard 11: Integrity

In 2014, the Commission again began the process of updating its Standards at a time of great change in American higher education, striving to articulate Standards for Accreditation for the next decade. In addition to a general overview of the Commission's approach to revising the standards, the principal focus of this paper is on the larger issues, each of which cuts across two or more of the existing Standards.

Commission Mission: The Commission is the higher education accreditation body of the New England Association of Schools and Colleges (NEASC). The stated mission of the Commission is:

The Commission develops, makes public and applies criteria for the assessment of educational effectiveness among institutions of higher education leading to actions on their institutional accreditation. By this means, the Commission assures the education community, the public and interested agencies that accredited institutions have clearly defined objectives which meet with criteria published by the Commission; that they have the organization, staffing and resources to accomplish, are accomplishing and can continue to accomplish these objectives. In addition, through its process of assessment, the Commission encourages and assists in the improvement, effectiveness and excellence of affiliated educational institutions.

This statement reflects the historical dual purposes of regional accreditation: to provide quality assurance through a program of periodic institutional evaluation, and to promote institutional improvement.

Because NEASC is a membership organization, the Standards are responsive to the expressed ideas of the Commission's accredited and candidate institutions. And because regional accreditation is a means of self-regulation that works in the public interest, the Standards may also be said to represent the articulation by the New England higher education community of the characteristics and behavior of those institutions of higher education that deserve the public trust. In addition, because the Commission is recognized by the U.S. Department of Education as a "reliable authority on the quality of education," the Commission's Standards for Accreditation must reflect the regulatory requirements that accompany its status as a gatekeeper for federal financial aid for students at accredited and candidate institutions.

NEASC's Commission on Institutions of Higher Education is undertaking a review of its Standards at a time of great change in American higher education. As it is striving to articulate Standards for Accreditation for the next decade, the Commission is paying particular attention to changes in technology, finance, student demographics, and public expectations for higher education, as well as to appropriately addressing institutional capacity and enterprise risk management.

Preliminary Work. At its June 2014 retreat, the Commission heard from a panel of experts on the external environment for higher education and the changing public expectations for higher education. Commission members reviewed a series of documents, including Reflective Essays on what students have gained as a result of their education, and the standards and processes used by the other U.S. regional accrediting bodies.

The Commission then articulated some broad outlines for the revision:

- The Standards should: 1. continue to be mission-centric; 2. emphasize outcomes increasingly, much more than inputs, especially student learning; 3. leave plenty of room for innovation (which should be mission-compatible); 4. increase the expectation for accountability; and 5. increase the expectation for transparency.

- The Commission should: 6. consider having fewer standards – by combining and revising current standards; 7. keep a simple, clear process so that institutions focus on the content and not the process; 8. consider whether to separate compliance from improvement; 9. ensure the standards allow for a differentiated process with institutions; and 10. reflect expectations for quality in competency-based education.

These commitments clarified that the Commission is responsive to the increasing expectation that accreditation serve the public interest through a greater focus on student learning and that it ensures a system of accountability and transparency. At the same time, the Commission seeks to continue to ensure that accreditation serves each institution through mission-centric Standards, and a process that promotes institutional improvement as well as assuring quality. The Commission also made a commitment that the revised Standards would not be longer than the current Standards.

In October and November 2014, representatives of member institutions were invited to a series of eight meetings around the region and one in London for international institutions to review the current standards and the context for standards revision. Many useful comments were submitted regarding the current Standards (what works well, what doesn't) and the membership's desires for the revised Standards (what should be included, what should not be). The discussion and written feedback from individuals and groups, along with the experience of the Commission with changes in higher education, were key to identifying the issues below for further consideration before any draft of a revised set of Standards for Accreditation was prepared.

II. Thematic Areas for Consideration

The titles given to these thematic areas are suggestive of the underlying issues, each of which cuts across organizational boundaries of institutions and arguably, at least, is related to two or more of the current Standards.

1. Student Achievement and Success

The 2007 mid-course review of the New England Standards resulted in this addition to the standard on Planning and Evaluation: "Based on verifiable information, the institution understands what its students have gained as a result of their education and has useful evidence about the success of its recent graduates. This information is used for planning and resource allocation and to inform the public about the institution." (2.7) Beginning in 2011, institutions completing their fifth year interim report were asked to include a Reflective Essay of 15-20 pages addressing one or more of the following:

1. What students gain as a result of their education: "Based on verifiable information, the institution understands what its students have gained as a result of their education and has useful evidence about the success of its recent graduates."
2. Assessment of student learning: "What and how students are learning."
3. Measures of student success, including retention and graduation.

In general, the Commission has found that the Reflective Essays represent a stronger way for institutions to use the accreditation process to focus on their understanding of achievement and success of their students. Because institutional missions and the compositions of the student bodies vary greatly among the membership, the particular findings in the Reflective Essays also vary greatly. Because the Commission is committed to increased attention to student learning, accountability, and transparency in the revised Standards, it will be useful to consider how these topics should be represented in the revised Standards for Accreditation.

Questions considered were:

- Are the three topics outlined above – what students gain, assessment of student learning, and measures of student success including retention and graduation – the appropriate way to group an increased emphasis on student achievement?
- The Standards are designed to apply to all students, all locations, all modalities of instruction (including competency-based education). How should that be clarified in the Standards?
- The Standards frame "assessment" as essentially a private activity for institutional improvement, and the Commission is more interested in the thoroughness and validity of the process and the extent to which institutions have developed systems that lead to changes in curriculum, instruction,

or services for students than in the specific findings of the assessments. Is this the right approach? Do the Standards and processes convey this expectation clearly?

- How should retention and graduation rates be considered for graduate students? Students in competency-based programs? Part-time students? Students in online programs?
- What are reasonable expectations – and aspirational expectations – that institutions know about their recent graduates?
- How should institutions be expected to make information on student success available to the public?

2. Faculty: Roles and Composition

In the classical model of a college or university, faculty are full-time, tenured or on the tenure track. They work together in organized structures such as departments and curriculum committees as part of shared governance to construct and oversee the curriculum. They develop and teach the courses, assess student learning, and stay in touch with many of their graduates long after they leave the institution. These full-time faculty members are responsible to evaluate and improve the curriculum and instruction. Their interactions are formal (e.g., through committees) and informal (e.g., hallway conversations). Together they constitute a group that has all of the experience needed to oversee the curriculum and evaluate its effectiveness with respect to student achievement. An external perspective – consultants, specialized accreditation, academic and professional societies, the identification of peer institutions and benchmarks – supports these faculty members in overseeing and ensuring quality.

In New England, some institutions continue to operate this way. Some never have, relying from the first on part-time faculty. Others once relied on the classical model, but for reasons primarily financial do so less and less as the proportion of part-time and adjunct faculty increase. Still other institutions have, at least in part, deconstructed the faculty roles so that different individuals are responsible for curriculum development, course development, instruction, coaching, advising, and assessment.

Questions considered were:

- Is the Commission's current standard on Faculty sufficiently open to multiple models of faculty appointment?
- In institutions with large proportions of part-time or adjunct faculty, what responsibility, if any, does the accredited institution have for faculty development and evaluation for these appointments? For including part-time or adjunct faculty in governance?
- Where there are large numbers of part-time or adjunct faculty, how should the Commission articulate its expectation for faculty ownership of and responsibility for the quality of the curriculum, as improved through feedback from an effective system to assess student learning outcomes?
- When faculty roles are de-constructed, what should the Commission's expectation be for feedback loops connecting those who construct the curriculum with those who teach/mentor/coach students and those who assess student learning?
- Should this Standard address "instructional staff" and academic support personnel as well as faculty?

3. Partnerships: with institutions, for-profit entities, and others

In the classical model, colleges and universities operate as stand-alone entities. They develop and offer an academic program, relying on outside entities for a prescribed set of functions that might include the production of textbooks and library resources and possibly off-campus locations for internships and service-learning opportunities. While some institutions – or parts of institutions – still operate in this mode, others are more cooperative – some would say entangled – with other entities: sister institutions in a system, parents in a corporation, higher education partners for dual or joint degree programs. A few have contractual relationships with non-accredited entities involving courses and degrees. The development of on-line learning saw the rise of commercial operations eager to share the wealth of new enrollments by offering services that ranged from generating student leads to virtually offering the curriculum. Another example of a partnership involves organizations that recruit international students and then partner with accredited institutions on ensuring those students are fully prepared for the academic work that counts for a degree.

Because accreditation ultimately places responsibility for the academic program with the accredited institution, with respect to partnerships, the Commission has drawn the line so that the accredited entity must be responsible for the content of advertising and promotion; the admission of students; the oversight of the curriculum; the hiring, supervision and retention of instructors; and the assessment of student learning. The clearest explication of the Commission's expectations is here:

The institution demonstrates its clear and ongoing authority and administrative oversight for the academic elements of all courses for which it awards institutional credit or credentials. These responsibilities include course content and the delivery of the instructional program; selection, approval, professional development, and evaluation of faculty; admission, registration, and retention of students; evaluation of prior learning; and evaluation of student progress, including the awarding and recording of credit. The institution retains, even with contractual or other arrangements, responsibility for the design, content, and delivery of courses for which academic credit or degrees are awarded. The institution awarding a joint, dual, or concurrent degree demonstrates that the program is consistent with Commission policy, and that the student learning outcomes meet the institution's own standards and those of the Commission.

Questions considered were:

- Are the Commission's Standards and expectations clear with respect to institutional partnerships?
- Are there consortial arrangements that should also be acknowledged in the Standards?
- Are there other types of partnerships now or on the horizon that should be addressed by the Standards?
- Should the standard on Organization and Governance be clearer about partnerships? The standard on Public Disclosure? Integrity? Financial Resources?
- Here and in the Standards and Commission policy, the term partnership is used to address any formal arrangement (e.g., with an MOU, MOA, or contract) between the accredited institution and a non-regionally accredited entity. Is that distinction appropriate and clear? How would it apply to public institutions in a system or institutions that are part of a corporation?
- Should the self-study or the interim report ask institutions specifically to address the academic quality of any partnership arrangements?

4. Competency-Based Education and Prior Learning Assessment (PLA)

In the traditional model, students enter college as first-time full-time students and earn all of their credits from the institution awarding their degree. Many colleges – and many students – still follow this model. Broadened enrollment to 'non- traditional' students and ways for students to enter college with credits in hand that complicated the model – for example, AP and CLEP exams – have been around for decades. As the adult proportion of the student body has increased and the national expectation arose that more Americans have a college certificate or degree, there has been more interest in recognizing learning and achievement that happens beyond the campus. Recently interest has grown in developing programs that let students earn a degree through demonstrating competencies, in some cases in programs untethered to semesters or credits (i.e., programs offered through direct assessment). Also, there exist ways for students to earn credits or have learning turned into credit recommendations through commercial enterprises (e.g., StraighterLine), higher education consortia offering MOOCs (e.g., edX), and credit recommendations from the American Council on Education (i.e., CREDIT) for experiences gained through military service or employment. CAEL, which oversees prior learning assessment, claims research shows that students who earn credits through PLA are more likely to graduate than students who do not.

Questions considered were:

- How should these paths be represented in the Commission's standards, and where should they be placed?
- The Commission expects that student achievement for someone who earns a degree through a competency-based program is equivalent to that of a student who earns a degree through a conventional credit-based program. How should institutions be expected to validate the quality of their competency-based degree in regards to this assumption?
- When institutions accept a large number of credits from a non-accredited entity, what should be the Commission's expectation for validating that the decision was a good one educationally? Does this

expectation differ in any way from credits accepted through transfer agreements? Dual enrollment programs? What are the best practices now and what should they be going forward?

- What data should the institution be expected to present on the number of credits accepted or awarded from outside entities, and how should those data be presented and analyzed?
- What aspects of the Commission's Policy on Dual Enrollment should find their way into the Standards for Accreditation?
- In part because a master's degree can be earned through as few as 30 credits, the Commission's standards currently do not accept credit for prior learning at the graduate level. Some fields – e.g., business and social work – have developed dual tracks in a graduate program, one for students with prior study or significant work experience and a longer one for newcomers to the field. What should the Commission's standards expect in this regard?

5. Technology, including the library and distance education

The Fall 2014 regional meetings to discuss the Standards revision revealed that large numbers of institutions feel that the current Standards are arguably most in need of updating when it comes to matters related to the library and technology, often expressed as a concern over how Standard 7 (Library and Other Information Resources) and Standard 8 (Physical and Technological Resources) work together. Or don't. Also mentioned frequently was the observation that the language used to address programs offered online seemed outdated. Technology now undergirds so much of what an institution does that it is a fundamental resource, in its importance not unlike the buildings, classrooms, and laboratories. Data security has become a major concern of institutions. Some institutions no longer have libraries – they have a Learning Commons or some other arrangement to offer learning resources to students. And despite a shift in the standards to encourage a greater emphasis on students gaining skills of information literacy, many self-studies still concentrate on the input side of the ledger when it comes to information resources.

Questions considered were:

- Should there continue to be a separate standard on the Library? If not, what should be kept from the current standard and where should it go?
- Should matters related to technology be included in a revised Standard 8 or should they be disbursed to other standards, and if so, which ones? Should the technology for teaching and learning be addressed within a Standard separate from administrative technology?
- Would it be helpful to have one standard on financial, physical, and technological resources? Should human resources be included? Or is that too much? Are financial resources so key to the well-being of institutions that the area should continue to have its own standard?
- Cyber-security is a big issue on campuses today. How big a deal should it be in the Standards for Accreditation?
- What should be said in 2015 about online education in the Standards for Accreditation? Should online education matters be woven into the Standards or addressed together in one section?
- How should the Standards reflect that many institutions have two or more distinct student bodies (e.g., those on campus and those studying at a distance)? How should the programs and services for these student bodies be addressed in the self-study, and how can the Standards for Accreditation be articulated in a way that supports the institution looking at all students, all locations, all modalities?

III. Initial Conclusions and Outcomes

In August 2015, after an analysis of the considerable input received through the Fall 2014 meetings, discussion and feedback on the Standards Revision Discussion Paper that presented the five thematic areas presented above, as well as the Commission's own discussions, a first draft version of the revised Standards for Accreditation was released. Primary revisions included:

- Less emphasis on inputs/resources: The three current standards focused on resources – 7: Library and Other Information Resources; 8: Physical and Technological Resources; and 9: Financial Resources – were combined into one new standard titled Institutional Resources.
- More emphasis on student achievement/outcomes: The sections that address assessment and

student achievement that are currently located in three different standards – 2: Planning and Evaluation; 4: The Academic Program; and 6: Students – were brought together for heightened emphasis in a new standard titled Educational Effectiveness.

- More emphasis on what is owed to the public: To highlight the public role the standards continue to have, two of the current standards – 10: Public Disclosure and 11: Integrity – were combined, and a new section on transparency added, to form a revised standard titled Integrity, Transparency, and Public Disclosure.
- Recognition of the different models institutions now use to support teaching, learning, and scholarship: In addition to faculty, other academic staff who may be involved in these endeavors are now explicitly included. The new standard is titled Teaching, Learning, and Scholarship.

Other modifications incorporated into the first draft of the NEASC's new 2016 Standards for Accreditation were the inclusion of competency-based education, the recognition of multiple student bodies, mention of student safety and security, and the addition of human resources as a subsection in the Standard on Institutional Resources. As a result, the current eleven dimensions of quality used in New England to define the characteristics and behavior of those colleges and universities that – as accredited institutions – deserve the public trust, now are nine:

- Standard 1: Mission and Purposes
- Standard 2: Planning and Evaluation
- Standard 3: Organization and Governance
- Standard 4: The Academic Program
- Standard 5: Students
- Standard 6: Support for Teaching and Learning
- Standard 7: Institutional Resources
- Standard 8: Educational Effectiveness
- Standard 9: Integrity, Transparency, and Public Disclosure

IV. Next Steps

At its November 2015 meeting, the Commission will consider the input from the Fall 2015 meetings held to discuss the first draft of the revised Standards for Accreditation, as well as all comments submitted by email and post. A second draft version will then be developed for discussion by the membership at the NEASC Annual meeting in December 2015. Approval and adoption of the final version of the new Standards for Accreditation by the Commission is expected in January 2016.

The above is a summary of the inputs and the outcomes to date of the New England region's process to review – during a time of great change in U.S. higher education – the relevancy and effectiveness of its Standards of Accreditation and thereby ensure the criteria in place to assess institutional quality over the coming decade are current and up to date, reflect the needs and views of its membership, and are responsive to the public interests in higher education. Reflecting its status as a membership organization, the process was participatory and inclusive, involving member and candidate institutions, as well as interested members of the public. As a result of this process, the Standards of Accreditation will continue to represent the articulation by the New England higher education community of the characteristics and behavior of those colleges and universities that – as accredited institutions – deserve the public trust.

Traduction du Résumé

Abstract Translation

Paula Harbecke

GARANTIR LA PERTINENCE DES STANDARDS D'ACCREDITATION AUX USA : UNE ÉTUDE DE CAS

La Commission des institutions de l'Enseignement supérieur¹ est l'instance d'accréditation de l'enseignement supérieur de l'Association de la Nouvelle Angleterre des écoles et lycées (NEASC)², l'une des six agences régionales d'accréditation aux États Unis. Actuellement, l'Association (NEASC) accrédite environ 240 institutions de l'enseignement supérieur dans les six états de la Nouvelle Angleterre et outre-mer. Les standards d'accréditation de l'association instaurent le critère de qualité que les institutions doivent adopter pour obtenir l'accréditation.

Vu que la NEASC est une organisation de membres, les Standards sont développés et régulièrement actualisés par le biais d'un processus participatif impliquant les institutions membres et candidates dans l'élaboration des critères de la qualité requise pour l'accréditation. Au vu de la reconnaissance de l'importance de l'enseignement supérieur pour les citoyens et pour l'économie, le public est également invité à participer dans le processus. Les Standards représentent l'expérience accumulée de plus de 200 universités.

La Commission évalue régulièrement la pertinence et l'efficacité de ses Standards et de ses processus. Les Standards actuels de la NEASC sont entrés en vigueur en janvier 2006, avec des révisions à mi-parcours en juillet 2011. En 2014, la Commission a de nouveau commencé le processus d'actualisation de ses Standards à un moment qui a connu des changements majeurs dans l'enseignement supérieur américain, tentant ainsi de maintenir l'articulation des Standards pour la décennie suivante.

En octobre et novembre 2014, les représentants des institutions membres ont été invités à une série de 8 réunions dans la région, et une réunion à Londres pour les institutions internationales, afin de passer en revue les standards actuels et étudier le contexte pour les étudier de nouveau. Plusieurs commentaires très utiles ont été soumis concernant les Standards actuels (ce qui marche le mieux, et ce qui ne marche pas très bien) ainsi que les volontés des membres concernant les standards révisés (ce qui doit être inclus et ce qui ne le doit pas). La discussion et les retours écrits des individus et des groupes, avec l'expérience de la Commission, étaient la clé pour identifier les cinq problématiques qui méritaient d'être étudiées pour la révision des Standards de l'Accréditation :

1. Les résultats de l'étudiant et la réussite ;
2. Faculté : Rôle et composition
3. Partenariats : avec les institutions, les entités à but non lucratif, et autres ;
4. L'éducation basée sur la compétence, l'évaluation des acquis,
5. Technologie, y compris la bibliothèque et l'apprentissage à distance.

En se basant sur la région de la Nouvelle Angleterre en étude de cas, la présentation examinera le processus mis en œuvre de façon périodique en accréditant les instances des USA pour actualiser et maintenir les Standards d'accréditation qui établissent les critères utilisés pour évaluer la qualité institutionnelle. Les cinq problématiques identifiées par les membres pendant la dernière révision du processus ainsi que l'impact de chacune d'elles seront examinés dans le cadre de cette étude de cas.

1 - Commission on Institutions of Higher Education (CIHE)

2 - New England Association of Schools and Colleges

THE NEW ANVUR ACCREDITATION AND QA FRAMEWORK, ONGOING EXPERIENCE AND CAVEATS FOR ITS EFFECTIVE IMPLEMENTATION

Muzio M. Gola *

Abstract

*The **ANVUR** accreditation and QA FRAMEWORK is the result of a long developmental work done in Italy which took into account the international practice and ensured full compliance with the ENQA ESG's.*

A feature of the ANVUR FRAMEWORK is that strictly connects external accreditation practices and internal Quality Assurance practices and binds them under a common set of documents. A winning idea was to use the same set of documents, presented exclusively on-line, to inform the public, the internal QA agents and the external accreditation assessors, making them at the same time an instrument for communication to the public, then for transparency/accountability, and finally for evaluation/accreditation.

*The **ANVUR** FRAMEWORK is a ICT based tool, and is structured in a modular manner designed for usage in a web medium (screen Tables and formats).*

*The **ANVUR** FRAMEWORK is based on a triad of fundamental documents (the annual Programme information model, the related Review Report and the annual report from the Peer Student-Teacher Committee) which are repeated every year in a cross-linked feedback format; when collected and examined over a typical three year span their collection allows the dynamics of the Programme to be monitored by internal and external assessors. This documentation system proves to be systematic, continuously adjourned, yet sustainable and reliable. It reflects everyday's working conditions of the Programme. It is maintained on an ongoing basis. Moreover it suppresses the need of the traditional self-evaluation report.*

The availability of a tool that is at the same time a Programme design & review tool and a documented checklist for Programme evaluation & accreditation makes the accreditation process a more straightforward process, because assessors are provided with a special password which allows them to access not only the said fundamental document, but also all the uploaded supporting documents retrievable with a click of the mouse. This allows a very effective desk review to be performed before the on-site visit by the Evaluation Committee. Experience says that the external assessors have a much easier life and can do a better job.

Even with such a robust document structure, experience garnered during the first evaluations proves that several problems may arise which require the utmost caution.

The Evaluation Committee has a formidable task to perform, since an evaluation combines top-down elements at the University level with bottom-up elements at each Programme level, plus some essential research aspects at Department level.

*The transition from the internal QA application of the **ANVUR** FRAMEWORK to the external evaluation and accreditation has put into the limelight problems related to the capability of external assessors to capture the essential of the whole picture and to evaluate it consistently allowing useful comparisons to be made across different institutions.*

The paper describes and discusses in detail measures which have been devised to keep under strict control the behaviour of the visiting committee in order to ensure a uniform quality of the individual contribution and their integration into one coherent set of judgements.

* Vice-rector for Quality Accreditation and Evaluation, Politecnico di Torino, Italy

I. A quick cold start and a faster than expected development

The so called “University reform law n.240” issued on Dec. 30 2010 in Article 1 sub.4 decreed the establishment of the new Italian agency “ANVUR”, with the following statement: “The Ministry, in compliance with the freedom of education and the autonomy of universities, indicating objectives and strategic guidelines for the system and its components, and through the National Evaluation Agency of the University and Research system (ANVUR) as within its competence, verifies and evaluates the results according to quality criteria, transparency and promotion of merit, on the basis of the best experiences spread internationally, guaranteeing the consistent public resources distribution with the objectives, addresses and activities carried out by each university, in accordance with the principle of national cohesion as well as the evaluation of the results achieved.”

In effect, the seven members of ANVUR “Board of Directors” were nominated may 1st 2011. From then on ANVUR moved very fast, facing a huge amount of different tasks which may be examined at <http://www.anvur.org/index.php?lang=it>.

We shall consider here only the setting up of ideas and procedures for Quality Assurance and Accreditation of Universities and Programmes, therefore excluding the parallel impressive activities on the Research Evaluation Exercise and the Professors Habilitation. In short, the steps taken and their timing were as follows:

- January 2012: the regulations regarding the operation of ANVUR (modes and timing) were defined with the ministerial decree “19/2012”
- May 2012: publication of the first AVA prospective document for sharing with stakeholders; this document sets the cultural background, the main organizational and procedural elements of the QA system of teaching and research to be introduced in universities, “key QA documents” and “key QA actors” as described later
- May to July 2012: stakeholder comments are received
- July 2012: the AVA document is defined and delivered by ANVUR to the Ministry of higher Education
- October 2013 through December 2014: training and information meetings delivered by ANVUR experts at the universities, about 25 meetings at universities in different geographical areas and different types
- January 2013: the Ministerial Decree 47/2013 is issued, adoption of indicators and procedures proposed in the AVA-ANVUR document
- February to May 2013: SUA-CdS’s (base document for the QA of a Programme, see later) are filled by 90 universities and overall about 4300 Programmes
- March to May 2013: first call and training of “disciplinary” experts to be included in the Committee of Evaluation Experts (CEV) both for “initial” and the “periodic” accreditation (on-site visits); further ones have followed in time
- June 2013: completion of the procedures of Initial Accreditation of new Programmes
- December 2013: a revision of the Ministerial Decree 47/2013 is issued under the name 1059/2013, adjustment of some indicators and criteria
- January 2014: the call for University accreditation is open, Institutions submit their applications to be accredited (first “periodic accreditation”) on a voluntary basis
- January to June 2014: final development of the Guidelines for the Committee of Evaluation Experts (accreditation), after studies already carried out already in 2013; submission of the draft to the university community
- April 2014: the Guidelines for the SUA-RD (base document for the QA of Departmental research, see later) is published
- June 2014: first call and training of “system” experts to be included in the CEV for periodic accreditation (on-site visits)
- June 2014: first periodic accreditation test on two “pilot” voluntary Universities, one traditional (public, Lecce) and one telematics (private, Mercatorum Univ. in Rome); full scale assessment including the on-site visit, with final assessment of critical issues of the method

- July 2014: first call and training of student experts to be included in the CEV for periodic accreditation (on-site visits)
- July to December 2014: Guidelines for the Evaluation of Doctoral Programmes are approved by the Board of Directors then revised and finally published
- October 2014: the final version of the Guidelines for the Committee of Evaluation Experts (accreditation) is published
- November 2014: accreditation of the first two Universities and Programmes (Universities of Perugia and L'Aquila), historical event for the Italian university system
- November 2014 to May 2015: universities draw up the newly developed SUA-RD (Quality Assurance of Departmental Research)
- January 2015: first accreditation (on-site visit) of a telematic university (UNINETTUNO).

In summary, notable is the inception of a completely new system for Quality Assurance, Evaluation and Accreditation (a loose translation of the original Autovalutazione, Valutazione, Accreditamento, acronym "AVA") which was developed at an incredible pace, being finally ready for the first application in fall 2014, with the accreditation of the two first "test" higher education institutions, i.e., the Universities of Perugia and L'Aquila.

II. The core values

Tenets are:

- first of all, it has no meaning think about any form of external evaluation of accreditation if inside the institution Quality Assurance has not already been successfully applied in-house,
- second, the end purpose of all these efforts is neither Quality Assurance for its own sake, nor accreditation: it is the quality of education made available to students.

While it is not possible to expose in this concise presentation all the details of the "key" QA and accreditation documents, it may suffice to state that for the "education" part the AVA system has its foundations on three "interfaces":

- interface 1: labour & professions market/ Programme
- interface 2: teaching modules & teachers/ students
- interface 3: student services/students

Their focus points are as follows:

<p>Interface 1:</p> <p>Labour & professions market/ Programme</p>	<ul style="list-style-type: none"> • Which are the professional roles and the cultures to which that particular type of student is being prepared? • Which are the roles or functions that degree holder will cover when at work, at least in her/his first five to ten years of career? • Which are the competencies typical of those roles and functions? (which means knowledge, skills, etc. used in a working context) <p>The Programme must be designed around roles / functions / competencies, in order not to create a mismatch between exit outcomes and exit requirements.</p> <p>It is difficult, and would be very risky, to believe that all the above can be defined without consultation with external stakeholders or without prospective studies from the employers' and professional side.</p> <p>Self reference is lurking!</p>
---	---

<p>Interface 2: Teaching modules & teachers/ students</p>	<p>The borderline between the world outside and the world inside the academy is quite fuzzy. There is no mechanic one-to-one relation between the tools we transmit to the student and the task she/he will be asked to solve.</p> <p>Then, to the best of our knowledge we have to make an hypothesis, and build subject specific knowledge / understanding / skills / know-how, plus transversal skills towards that aim. Then expected learning outcomes have to be specified.</p> <p>This knowledge building, however, has an external effectiveness: the capability of rapidly grasping those competencies at work which were taken as the far objective. This happens outside the University, and success will be checked only with a follow up of alumni performance over a certain time span.</p> <p>But it has also an internal effectiveness (not to be confused with efficiency, which tells us how much effort was necessary): the success of the single steps though which pieces of knowledge are gradually built. This is fully in the hands of ourselves, the professors i.e. the professionals who know how Learning Outcomes must be transmitted and, last but by not least, how to verify whether they were effectively reached by the students.</p> <p>Finally, student assessment or examination as Learning Outcome certification is crucial. Through assessment expected learning outcomes are validated as achieved learning outcomes.</p> <p>This all what professional teaching is about. A first class university professor should be excellent in this professionalism, besides of course striving to be an excellent researcher with high level papers. Otherwise, why she/he does not move to a pure research institution?</p> <p>The teaching & learning activity must be supported by adequate methods and infrastructure: persons, methods, rooms, laboratories, technical staff, services.</p>
<p>Interface 3: Student services / students</p>	<p>The ENQA-ESG mention this point in the standard <i>"1.6 Learning resources and student support"</i>.</p> <p><i>Institutions should have appropriate funding for learning and teaching activities and ensure that adequate and readily accessible learning resources and student support are provided.</i></p> <p>The structure of the SUA-CdS takes into account in very fine detail all the requirements explained by the ESG-ENQA 2015 guidelines attached to this standard.</p>

III. Key documents and key actors

The very intense work on the part of ANVUR on this part of its mandate has found a symmetric counterpart in the required of universities. From the point of view of universities, their records show up to the date of this conference (October 2015):

- **SUA-CdS** (Scheda Unica Annuale – Corso di Studio, which might be freely translated into “Annual Integrated Framework – Programme), in the academic years 2013/14, 2014/15, 2015/16; in short, they are public documents, available on the net, revised annually, where every Programme/Department formulate commitments and work plans, used for communication with all parties involved, basis of all self-assessment and accreditation activities
- **Annual Reviews**, in the academic years 2012/13, 2013/14, 2014/15 (a test Review was run in a limited number of forerunner Programmes at each institutions already for the academic year 2011/12); they are compulsory for all Programmes and must be issued every year, as the name says; they are basically an analysis of ongoing monitoring results based both on indicators (included, but not limited to, the student opinion survey) and on direct observation of Programme’s performance in its various aspects, from methods to infrastructure, being divided in three parts named:
- **Cyclic Review** (the first) starting from the academic year 2015, however to be written by Programmes in the preparation of an accreditation visit and limited to those Programmes selected for accreditation by the institutions themselves; they must be issued every three to five years according to conditions;

In short, Annual and Cyclic Reviews are periodic self-assessments and plan pursuing the aim to verify the adequacy and effectiveness of the study program, research the causes of dissatisfaction and finds solutions to remove them.

- **SUA-RD** (Scheda Unica Annuale – Ricerca Dipartimentale, which might be freely translated into “Annual Integrated Framework – Departmental Research): spring 2014, pilot testing (2013 data) on a reduced number of institution and departments before the issue the ANVUR SUA-RD guidelines; spring 2015, SUA-RD’s are filled by all departments in all institutions (2014 data); to be repeated every year
- To these we must add the “**Guidelines for the Committee of Evaluation Experts**” (shortly named “**Accreditation Guidelines**”) where seven Quality Assurance requirements (four, from AQ1 through AQ4 concern the top-down teaching and learning AQ measures as set by the Governing Bodies, to be compared with the Requirement AQ5 which takes into account the bottom-up rules as applied by the Programme; then AQ6 which concerns the Departmental research and finally AQ7 which concerns some “dimensional” limits to number of instructors, students, class hours and credits). Each AQ requirement is split into more “indicators”, from two to five, and each indicator is composed of “points of attention, or facets. Algorithmic rules are given to compose the partial judgements on “facets” into judgements on indicators.

These documents are taken care by the key actors involved in AQ operations inside each university, which are (according to regulations set by the Governing Bodies at the top-down level) :

- **The Programme Review Committee**, which is charged in the first place with the follow-up of Programme’s performance and with the writing-up of Annual Review Reports (each year) and of the Cyclic Review report every three to five years, depending on conditions.
- **Peer Student-Teacher Committees** (Italian acronym CPDS): first report in winter 2013 (academic year 2012/13) according to ANVUR-AVA regulations, then in fall 2014, etc. yearly (The CPDS existed in some universities from a variable number of years, in the earliest case since 1992; their mandate was adapted according to new AVA regulations); in short, CPDS Reports are a source of independent information and analysis that evaluates teaching/learning processes from the point of view of students, to be taken onto account by the Programme Review Board in their Annual Reviews
- **Evaluation Board** (translation of the Italian “Nucleo di Valutazione”), annual Report issued by end of spring; the Report covers an evaluation of the Quality Assurance system of the University, according to ANVUR-AVA regulations, in years (The Evaluation Board existed in all universities before AVA already from 1999; their mandate was enlarged according to the new AVA regulations); in short, with the annual report of the Evaluation Board Document the overall effectiveness of the Quality Assurance of teaching and research is evaluated

- **Quality Promotion Unit** (a loose translation from “Presidio della Qualità di Ateneo), with the following “minimum” tasks:
 - to proper and uniform supervise QA procedures in the entire university
 - to propose common tools for QA and training activities for the purposes of its effective application
 - to support Programmes, their staff, their design and Review committees, and Department Directors for common tasks
 - to prepare an annual report for the Academic Authorities and such other reports as the Rector may require.

and in a broader and more mature vision:

- to promote and encourage principles which will enhance the quality of the educational and work experience of students and staff
- to promote and update guidelines for departments and units to carry out self-evaluation
- to liaise with departments or units in their self-evaluation
- to organise peer review procedures for departments or units
- to liaise with university initiatives and offices regarding staff development
- to promote the integration of policies and processes aimed at enhancing the quality of outcomes across the University.

Such combination of key documents and key actors gives a common “minimum” structure to the AQ of all universities subjected to ANVUR periodical evaluation and accreditation. Such uniformity is, in the balance, an advantage more than a constraint in a situation of growing awareness: it provides common rules whose practice and assimilation can grow by sharing common experiences, it allows attention to be concentrated on results rather than procedures, it cross fertilizes the quality culture of internal AQ actors with the one of external evaluators / accreditors, who after all do belong to the same pool of academics and, to a minor extent, technical staff.

IV. The weaknesses, the threats, and the lessons learned

The system is novel, experience on the part of universities is taking shape. The amount of visited universities after one-and-a-half year from the first accreditations amounts to around one-sixth of the total.

The time to grow a solid AQ inside universities has been still very limited, the consequence is that Programmes, Departments, and Governing bodies start really understanding all the needs and implications of AQ, at least partly, while preparing for accreditation. This is not “according to the theory”, but it is human. It is expected that the next run will see more awareness and sense of ownership.

However, criticism against the AVA system, but also against the other ANVUR initiatives, have been voiced by a number of organised opposers. This is quite normal – accreditation and QA are not popular at first – but more and more people holding responsibilities in Programmes and Departments are captured by the system as time goes by and the rationale is understood.

However, supporters of a revision have good reasons. Which, I maintain, if not well thought out, might come too early, violate cornerstone principles, and produce more damage than repair. Here we examine some “critical” complaints and revision proposals.

1. The system needs some urgent repairs

- The Annual Review Report is considered a burden, and a frequent criticism is against the “corrective actions” which are repeated every year even if they take more than one year to take effect;

Comment: in the opinion of AVA developers, this may be partly true, but the criticism stems largely from the fact that the meaning of the system has not really been understood. Proposed corrective action are quite frequently of a minor momentum, and seem to be far fetched. Difficult to say whether the criticism comes from a real irrelevance of the tool, rather than the expression of a latent general dissatisfaction or an “evaluation fatigue”.

- The key documents ought to be strictly aligned, looking at the same facts from different perspectives

and with different purposes. However it is true that documents are affected by a number of semantic misalignments which some universities have difficulties to assimilate and interpret.

Comment: some questions, especially in the “Accreditation guidelines”, are redundant, others are expressed in a way which lead to interpretation problems.

- The algorithmic rules to compose the accreditation judgements from “facets” to “indicators” then to the final assessment of the “requirements” are deemed to be over-and-unnecessarily complicated, really far-fetched. Some believe that the remedy is to formulate the statements expressing indicators more concisely, grouping together all those aspects which at the moment are split into several facets, in this way they expect that the judgment composition will be done automatically as a balance between the observations on single facets.

Comment: it is true that an overly complicated system can lead to unexpected results, and that the consideration of a large number of facets may result in a lower “mark” because every fault will sum up score reductions. The more the facets, the higher the probability of a reduction. Overall the system is considered from many as far fetched.

It is quite easy to predict that the combination of a growing pressure and the pending renovation of the whole ANVUR Board of Directors in a span of about one year (the first non-renewable mandate has expired) will facilitate an early revision of the system.

However, one thing is a clarification of obscure aspects, to prune crooked branches, another thing is to throw away the Annual Review altogether. In a moment when Programmes are slowly catching up with a better understanding of their task in a QA logic.

One thing is to eliminate undeniable redundancies and inconsistencies, something else is to revolutionize the total architecture, on a system which has not yet completed its first round.

One thing is to revise the rules for judgement composition, something else is to transform an indicator composed of several “analytically separated facets” into a big and long question amassing several partial aspects. The analytically separated facets allow to single out the aspects which are faulty and need a remedy, the amassed indicator does not solve the problem of judgment and makes and may lead to “average and flatten” the perception of problems.

Finally, a problem remains open: is it better to start from a revision of the accreditation guidelines or from a combined realignment of the QA “key documents”? A subject for mindful reflection.

2. The “self-evaluation report” controversy

It is worth stressing here that the combination of the SUA-CdS and the Annual Review reports of Programmes supplemented by those of the CPDS, plus the Cyclic Review Report (every 3 to 5 years) replace, in an innovative way, the traditional self-assessment report which has no longer a reason to be present.

Let us further underline this point in the light of the AVA official statements:

- all documents must reflect and represent the daily work of Programmes
- it is not conforming to this principle to craft documents purposely intended for the accreditation
- as a general rule the online documentation should be available online without prior notice, or, in the case of a visit, be accessible by an evaluator with a quite short notice for organizational purposes and courtesy and with the only purpose to make sure that the last and more adjoined version is available.

This point is not well understood by all. There are those who state that it ought to be re-introduced on the basis that the main national QA accreditation schemes base their accreditation schemes – endorsed by ENQA-ESG’s - on the triad: self-evaluation report, peer external visit, final evaluation report. However, I maintain that this would undermine the philosophy of the system.

Of course, in the contexts before AVA the meaning and the function of the self-evaluation report was that of an occasional document which:

1. was (politically and psychologically) a desirable instrument to underline the active participation of the institution to the evaluation process
2. was, in addition, an effective instrument to gather and organise all the existing material on the internal activities and their results, documents that an external committee would have no means to collect

3. was, finally, an opportunity to the institution under evaluation to “read and better understand” its own QA system.

It is well known that one practical consequence of this approach could be the write-up of such report by a very limited number of experts, while the majority of staff would not feel the essentials of the QA and might not change their behavior accordingly.

Today, instead, the prevailing focus of AVA is on the day-by-day mode of operation of the Programmes and Departments, the prime actors of internal QA, continuously supported and monitored respectively by the Quality Promotion Unit, the Review Committee, the Peer Teacher-Student Committee, and the Evaluation Board.

This AVA concept is the most advanced interpretation of the ESG 2009 statement (which was rephrased in the 2015 revision): The confidence of students and other stakeholders in higher education is more likely to be established and maintained through effective quality assurance activities which ensure that Programmes are well-designed (the main instrument here is the SUA-CdS), regularly monitored (the AVA student questionnaires at the lowest level going up through the already mentioned committees up to the Evaluation Board), and periodically reviewed (the Cyclic Review), thereby securing their continuing relevance and currency.

It is important to stress once more that the AVA approach brilliantly solves the problem of preparing for the accreditation, because replacing the occasional self-assessment report with a regularly distributed control ensures a much more effective internal quality assurance system. The self-evaluation report, in fact, does not accompany the processes and is a kind of final-check before the external evaluation. In this sense it is not formative at all, whereas formative intentions are evident in the AVA analytical framework.

Also note that the self-assessment report induces adaptive behaviors and expresses a quality system that is basically put into play just on the occasion of external evaluations: a disconnected structure management. And this is not consistent with the guidelines ENQA. When a university learns and applies properly AVA, its operations develop simultaneously to the their AQ management and, ultimately, it is less costly than having to draw up a self-assessment report and then fix (necessarily ex-post) «non-conformities».

This point is a litmus test for deep understanding of the meaning of Quality Assurance in the AVA system. Therefore it is a tenet deserving a defense to the death, even though defenders of the faith at the end are burnt at stake!

3. The limited availability of evaluation experts

A bottleneck to the system is the lack of experts, mainly of disciplinary experts. It is, as a matter of fact, troublesome to receive sufficient candidatures of academics to the function of members of the evaluation committees. It is difficult to compel the candidates to follow a training course of adequate length with theory and simulations. It would be an excellent idea to select only those candidates who have a satisfactory scientific pedigree, a good teaching record, and had a significant experience in academic management (Programme director, Dean, Department director etc.).

This reduces the number of evaluators of a sufficiently quality, but it is a critical point, since using the same experts over and over for more than a couple of accreditation per year may jam them, and result in huge delays in the delivery of final reports. Take note of the fact that starting from the preliminary desk review, through the preparation to the visit, the visit itself, the writing of the final report, to its definitive approval, the duration of an accreditation may span easily over four-five months.

In summary, the selection and the training of experts ideally ought to start well before the beginning of accreditations. Like an army that should be trained well before starting a war.

4. How analytical the final accreditation report should be

Here too there is a debate. The “principle” instructions delivered to the Committees of Evaluation Experts are:

- the report must contain explicit, precise, and traceable references of documents, websites, interviews etc. on which the final judgments are based
- all results of the Report must be based on clear and solid evidence collected by the Committee and, if possible, also on benchmarking elements; CEV members will avoid subjective opinions i.e. not

based on factual findings

- the process on which decisions are based must be transparent, show «what» and «why» the CEV has used come to a certain conclusion, be well motivated - both positive and negative.

Individual experts have different opinions about how analytical a report should be. Can we accept very synthetic judgements, and fully trust the authority and expertise of evaluators? Or should every judgement be analytical, whatever the reputation of experts? The debate is not whimsically theoretical. A long, analytical report is very demanding, and overstress the experts. Some believe that a professional diagnosis/verdict should not require too much explanation, i.e., believe that the final report has a summative function. That the era of analytical reports is over. Others believe that a very analytical report, admittedly a burden, is indispensable because it fulfils what is one of the main function of external evaluation: the formative function.

The author of this paper believes that a final accreditation report has to be analytical, which does not mean rambling and long winded. Since a positive judgement indicates a best practice which other institutions may like to imitate, and a negative judgement must help the institution to recognize and overcome its deficiencies, in both cases the verdict must be extremely well justified.

The agency should prepare clear instructions on this point in order to guide the CEV, experts and president, towards uniform behaviours not only about the level of detail and motivation of the final report but also on the metric adopted for the final scores assigned to individual facets of the accreditation guidelines.

Traduction du Résumé

Abstract Translation

Muzio M. Gola

LA NOUVELLE ACCRÉDITATION DE ANVUR ET LE CADRE QA, EXPÉRIENCE EN COURS ET MISES EN GARDE POUR SA MISE EN ŒUVRE EFFECTIVE

Le nouveau cadre d'accréditation et de contrôle qualité Anvur est le résultat d'un long travail de développement qui a été effectué en Italie, qui a pris en compte la pratique internationale et garantit une totale conformité avec les ENQA ESG.

Le cadre ANVUR est basé sur des documents fondamentaux (le programme annuel du modèle d'information, les Review reports ainsi que les rapports annuels du Comité pair élève-enseignant) qui sont répétés chaque année dans des formulaires croisés, et collectés et examinés sur trois années, ce qui permet le contrôle des dynamiques du Programme par des évaluateurs internes et externes. Ce système documenté démontre qu'il est systémique, constamment ajourné, mais durable et fiable. Il reflète les conditions quotidiennes du Programmes. Il est maintenu sur une base régulière, et ne nécessite pas un rapport d'autoévaluation traditionnel.

Avoir un outil qui est en même temps un Programme et un outil de contrôle et également une checklist documentée pour le Programme d'évaluation et d'accréditation rend le processus d'accréditation plus direct, car les évaluateurs ont en leur possession un mot de passe spécial qui leur permet d'accéder, non seulement au document de base, mais également à tous les documents de support avec un seul clic. Ceci permet donc la réalisation d'une étude documentaire efficace avant la visite terrain de la part du Comité d'évaluation. L'expérience a montré que le travail des évaluateurs externes est ainsi facilité, et leur rendement meilleur.

Ce papier décrit et examine dans le détail les mesures qui ont été élaborées pour maintenir le contrôle du comportement du comité afin de garantir une qualité uniforme de la contribution individuelle, et leur intégration dans un seul ensemble de jugements.

Bibliography and References

<http://www.polito.it/>

ALMALAUREA'S TOOLS FOR THE IMPROVEMENT OF QUALITY ASSURANCE IN THE HIGHER EDUCATION SYSTEMS

Francesco Ferrante *

Abstract

AlmaLaurea is an interuniversity consortium, based on a bottom up approach, that represents almost 92% of Italian graduates from 72 Universities and counts on 20 years' experience carrying out surveys on Italian graduates' profile and job condition.

The AlmaLaurea's model is based on three pillars: an annual survey (census survey based on administrative data) on the profile of graduates (i.e. the internal effectiveness of the higher education institutions); an annual survey (census survey) on the occupational status of graduates 1, 3 and 5 years from graduation (i.e. the external effectiveness of the higher education institutions); an online databank with more than 2.000.000 CV, a tool aiming to improve the match between supply and demand of graduates.

Providing every year the Italian Ministry of Higher Education, the national quality assurance agency and Universities with reliable and updated information, the mission of AlmaLaurea is the improvement of quality assurance in the higher education systems.

With the financial support of the EU, AlmaLaurea is currently coordinating an international project for the implementation of its model in Morocco and Tunisia.

I. Almalaurea

AlmaLaurea is an Inter-university Consortium set up in Italy in 1994. Nowadays it involves 73 Universities and approximately 91% of Italian graduates. The Consortium is supported by the universities taking part in it, by the Italian Ministry of Education, University and Research (MIUR) and by all companies and institutions using the databank and the services offered by AlmaLaurea.

AlmaLaurea originates from an idea of a group of researchers of the University of Bologna, coordinated by Andrea Cammelli, professor of statistics and founder of the Consortium.

In 1988 two important events, the ninth centenary of the University of Bologna and the signing of the Magna Charta of Universities, in which the main values of the European academic tradition were stated, acted as a stimulus to reflect on the importance of some aspects of the academic world which had not been taken into consideration up to that point, namely: students' success in their academic studies, graduates' social and professional position and their access to the labour market.

After a project started by the Statistical Observatory of the University of Bologna, AlmaLaurea was officially set up with the creation of the databank in 1994. This was a small-scale "cultural revolution", with the establishment of a unifying overall logic capable of overcoming the academic particularism. Moreover, it was the beginning of a new culture based on evaluation and on the social role universities play in relation to youth employability.

AlmaLaurea's activities are based on three pillars:

- The graduates' profile: an annual survey and report on the internal efficiency of the higher education system;
- The graduates' employment conditions: an annual survey and report on the external efficiency of the higher education system;
- The online graduates' databank: a powerful tool that improves the match between supply and demand of graduates and their transnational mobility.

* *Consultant Scientifique, Almalaurea, Italie*

This complete, periodic, well-timed and updated documentation is provided to the Governing bodies of the universities that are part of the consortium, to Committees dealing with Teaching Activities and Career Guidance and made available to all the stakeholders involved in higher education issues – families, students, companies and policy-makers – as a solid basis for fostering all decision-making processes and activity planning.

II. SPECIFIC AIMS AND OBJECTIVES

AlmaLaurea is aimed at:

- Facilitating and improving the access of young people to the labour market and their placing. By doing this AlmaLaurea simplifies companies' research of personnel, reducing the time gap between qualified work demand and offer;
- Guaranteeing the optimisation of human resources through a steady update of data regarding the career of students holding a diploma or a degree;
- Looking after all monitoring processes concerning the academic programmes chosen by students and analysing graduates' characteristics and performances, so that a direct comparison between different courses, faculties and universities is made possible;
- Analysing the effectiveness of the study opportunities offered by each university;
- Taking into account the professional profile that is required by Italian and foreign companies.
- Analysing the effectiveness that the study opportunities offered by each university have in the labour market by monitoring the different occupational openings;
- Developing a synergic relationship with high schools which is aimed at guiding students towards university and the labour market;
- Promoting any activity aimed at the achievement of the above mentioned goals, both at a national and at a European level.

III. ENHANCING THE VALUE OF INFORMATION

"Is it useful to decide without knowing? Decisions should be indeed followed by actions."

Luigi Einaudi, economist, former President of the Italian Republic, *Prediche inutili*, 1959

The principle stated by Einaudi – to know in order to decide – is the driving force behind the two annual population surveys carried out by AlmaLaurea on the graduates' profile and employment conditions. The surveys are representative of the population of Italian graduates and the statistical documentation is freely available on the AlmaLaurea web site both in Italian and in English. Thanks to the high response rates, the collected data allows for robust statistical analysis.

Graduates' profile census survey (response rate: ≈ 86%, 2014 survey).

Since 1998 – at only 5 months from the end of every solar year – the different academic paths undertaken by the Italian graduates have been analysed at university, faculty and degree course level.

The collected information includes more than one hundred variables, among them: graduation mark, marks obtained in single exams, regular attendance, graduation age, social class, language skills, IT skills, courses attended abroad, work experiences made during the study period, reasons behind the choice of a specific degree course, personal aspirations and evaluation of the academic experience made (ranging from the relationship with the teachers to the services offered).

Graduates' employment condition census survey (response rate: 77% - 86%, 2014 survey)

The most significant and recent labour market trends are analysed every year on the basis of the occupational openings for graduates after one, three and five years from graduation. On an experimental basis, the survey is run 10 years after graduation. Surveys are released only 3 months after the end of the reference solar year. The variables taken into consideration for each university and faculty include the occupational activity carried out, how long it takes for graduates to access the labour market and how they manage to access it, earnings and the effectiveness of a degree for finding a job.

IV. THE CV DATABANK

“Compared to the students of universities that are not part of the Consortium, graduates coming from universities members of AlmaLaurea have more opportunities to find a job (+ 2%), have a higher average monthly income (+ 3%), are more satisfied with their jobs and have greater mobility (+ 2.5%)” .

The size, quality and timeliness of its services make the AlmaLaurea databank quite unique in Italy and an unequalled model on the international scale.

The databank contains:

- 2.2 million graduates' curricula (about 220,000 CVs are entered every year), certified by universities, translated into English and made available to companies looking for graduates to hire. Curricula are added to the databank yearly and updated by the graduates themselves;
- one hundred search variables available to Italian and foreign companies, both private and public for personnel search and selection purposes and for any information request.

It is a real register of student's skills which facilitates the access of young people (graduands and graduates) to the labour market giving to graduates having a longer work experience both career opportunities and improvement of their employment conditions.

AlmaLaurea is open towards foreign graduates and companies operating on an international level. This is made possible both by the translation of curricula and by the equivalence of qualifications at the European level.

In the last 10 years Italian and foreign companies used more than 3,5 million graduates' curricula for their personnel needs.

On the basis of the extended experience of AlmaLaurea, the same system and tools have been developed at secondary school level with **AlmaDiploma**, for the assessment of the secondary school system and career guidance.

The online AlmaDiploma database, based on the same model and functionalities of AlmaLaurea database, is aimed at facilitating secondary school leavers' job placement. More than 200,000 curricula of secondary school leavers (about 40,000 CVs are entered every year) are in AlmaDiploma database.

AlmaLaurea also provides higher education guidance tools to assess students' strengths, support university degree choices, and better understand the labour market through **AlmaOrientati**.

V. FURTHER SERVICES

Personnel search and recruitment are carried out by AlmaLaurea SRL, a company entirely owned by the AlmaLaurea Consortium. It is authorised to search and recruit personnel and it is enrolled on the register of employment agencies.

Recruiting online and job offers publishing

Using more than 100 search variables, Italian and foreign companies can search through the online databank the candidate that best fits their recruitment requirements.

Moreover, companies can publish job posting on AlmaLaurea web site; all graduates interested in the job offers can answer by sending their CV. Graduates can signed up for the e-mail alerts service and receive e-mail alerts about job offers, master's degree courses and postgraduate courses.

Personnel selection

AlmaLaurea supplies both to Italian and foreign companies specific services aimed at facilitating the selection of newly graduated students or graduates having a longer work experience. AlmaLaurea selects the candidates who comply with the requirements of the company and then suggests a small number of them, selected by AlmaLaurea experts in recruiting, assessment and selection.

Career guidance

Moreover, AlmaLaurea operates in the career guidance sector. It provides services targeted at newly graduated students entering the labour market for the first time and advanced career counselling services targeted at graduates having work experience.

Employer branding

Italian and foreign companies have the possibility to publish on the AlmaLaurea web site their company profiles, as to increase their visibility and so to attract graduates from the member universities.

Furthermore, AlmaLaurea provides additional services for companies interested in contacting the best new graduates of each graduation session.

Career days

AlmaLaurea supports companies in the planning and organization of ad hoc recruiting days, inviting graduates from all over Italy or from specific regions, accordingly to companies' needs and in concomitance with their vacancies.

Benefits for Companies and Graduates

AlmaLaurea recruiting and selection services provide the following guarantees both to companies and graduates:

- Effectiveness: AlmaLaurea supplies a lot of its services on line, in a short time and with great quality;
- Nearly 20 years experience of AlmaLaurea in recruiting and selection services for graduate profiles;
- Sound know-how of the university system and of the graduates' performances and characteristics;
- AlmaLaurea certifies the graduates' CVs on account of close links with the universities and guarantees to the companies the possibility of reaching graduates from almost all universities.

VI. ALMALAUREA ABROAD: INTERNATIONAL COOPERATION AND PROJECTS

The forthcoming challenges posed by globalisation regarding education, training and work offer AlmaLaurea steady growth prospects. If, as Andreas Schleicher, OECD's deputy director for Education and Skills put it: "Educational systems have to prepare students for jobs that have not yet been created, technologies that have not yet been invented and problems that we don't know will arise", the creation of a new knowledge-based society implies taking young people and their destinies - the future of each country - as a point of reference.

The AlmaLaurea internationalisation process is based on the commitment of enforcing provision of data and information to local governments for the governance of HE system and labour market; enhancing the linkages between HE, economy and society; innovating the HE system; cooperating in developing better conditions in labour markets and society.

Among the main goals of the internationalisation process:

- Diffusion of AlmaLaurea experiences and best practices in European and international context;
- Cooperation for graduates database's planning in foreign countries according to local socio-economic environment and needs;
- Promotion of international networks of bodies and institutions with similar competences on issues related to graduates' employability and education system performances monitoring.

Projects and activities in details

Since 2004, AlmaLaurea has been cooperating with societies, students, universities and firms on the international scene by sharing its experience and know-how.

The ongoing internationalisation program has been developed in different countries and areas, mainly through the participation in European grants.

AlmaLaurea is currently implementing in the following projects financed by the European Union through the Erasmus + programme:

- Coordination of the VOYAGE project (beneficiary country: Vietnam). Main goal: to set up, first time in Vietnam, the online database of Vietnamese graduates and to carry out surveys on the graduates' profile and design the methodology to run employment condition survey in Vietnam.
- THEQA+ project (beneficiary country: Turkey). Main goal: to set up, first time in Turkey, the online database of Turkish graduates and to carry out surveys on the graduates' profile
- RESUME project (beneficiary countries: Lebanon, Morocco, Tunisia). Main goal: to develop entrepreneurial spirit at university level as to enhance graduates' employability

The recently achieved projects are:

- Coordination of the TEMPUS ISLAH project (beneficiary countries: Morocco and Tunisia). Main goal: creation of a network of local and national observatories for enhancing graduates' employability .
- Coordination of the TEMPUS HEN-GEAR project (beneficiary country: Armenia). Main goal: improvement of the effectiveness of the higher education system and the employability of graduates (as project coordinator).
- IPA Adriatic ADRIA-HUB project (beneficiary countries: Bosnia Herzegovina, Croatia, Montenegro and Serbia). Main goal: promotion of employability of qualified graduates from the Adriatic labour market, by merging personal skills with industrial needs (as project partner).

Following a brief description of the main activities carried out in the last years.

Cooperation with European and Neighbouring countries

2005: Creation of a unified pilot database of European students' CVs, involving universities from France, Hungary, Netherlands, Poland and Italy (EAL-NET project).

2009: Support of high quality in education in European and Neighboring countries (QACEP project).

2009: development of a pilot database to facilitate graduates' assessment and employability, in cooperation with 4 universities in Morocco (TEMPUS GRINSA project).

2011: Action as knowledge-partner in the development of the higher education & research area in the Euro-Mediterranean Region (JISER-MED project).

2012: Transfer of know-how to the AlmaLaurea like initiative in Spain (PortaLaurea project).

Cooperation with Latin American Countries

2004: Participation in the set up of tools for a better link between university and labor market and for quality evaluation of education in Argentina, Brazil, Colombia, Chile, France, Germany, Guatemala, Italy, Peru, Mexico, Spain (EuropAid-Alfa program).

2011: Contribution to the implementation of web based services for enhancing graduates' job placement and employability in Argentina, Brazil, Bolivia, Chile, Colombia, Nicaragua, Paraguay, Spain (EmploiALCUE project).

2011: Collaboration to foster cooperation among higher education institutions in European and Caribbean countries in a structured and sustainable perspective (Cariberasmus project).

Cooperation in Africa

2008: Transfer of know-how to support the strengthening of research activities in the epidemiologic sector in Kenya, Tanzania, Uganda (EduLink project).

AlmaLaurea Networks and Research Groups

AlmaLaurea has strengthened collaborations with European networks in sectors close to its field of expertise. The INGRADNET network is coordinated by the University of Kassel and works in the field of tracking studies harmonization initiative at European level.

The EUNIS network (European University Information System) in 2011 awarded AlmaLaurea with the prestigious Eunis Elite Award for excellence in implementing Information Systems for Higher Education.

AlmaLaurea has also established tight cooperation with National and regional institutions developing similar graduates tracking surveys: with the CEREQ, the most important French observatory on employment and qualifications, producing statistics, research activity and providing support for policy implementation; with AQU (Catalan University Quality Assurance Agency) involved in the assessment, accreditation and certification of quality in the sphere of the universities and HE centers of Catalonia, and with INCHER (University of Kassel), carrying out the most extensive German Graduates tracking survey.

AlmaLaurea took part to the research group REFLEX (Flexible Professional in the Knowledge Society, New Demands on Higher Education in Europe) in the frame of the 6th Framework Programme, and Graduates Job Conditions survey (5 years on from graduation) was carried out in 10 countries.

AlmaLaurea has also started a collaboration with the Union for the Mediterranean, strongly supporting the implementation of its system in the MENA countries; contacts have been established also with higher education institutions active in Albania, Romania, China and Vietnam as future perspectives.

Furthermore, AlmaLaurea has been invited by the World Bank at its Headquarters in Washington D.C. to present its integrated model and tracing system to higher education experts, as best practice to be disseminated in the international context.

Bibliography and References

Manuel F. Bagues and Mauro Sylos Labini, 2007, "Do on-line labor market intermediaries matter? The impact of AlmaLaurea on the university-to-work transition", NBER Working Papers 12621, National Bureau of Economic Research, Inc.

Traduction du Résumé

Abstract Translation

Francesco Ferrante

LES INSTRUMENTS DE L'ALMALAUREA POUR L'AMÉLIORATION DE L'ASSURANCE QUALITÉ DANS L'ENSEIGNEMENT SUPÉRIEUR

L'AlmaLaurea est un consortium interuniversitaire, basé sur une approche bottom up (du bas vers le haut), qui représente environ 92% des lauréats italiens de 72 universités, et a plus de 20 années d'expérience dans les enquêtes menées sur les profils des lauréats italiens et le marché du travail.

Le modèle de l'AlmaLauréa est fondé sur trois piliers : une enquête annuelle (enquête de recensement fondée sur des données administratives) sur les profils des lauréats (autrement dit le rendement interne des établissements de l'enseignement supérieur) ; une enquête annuelle (de recensement) sur les préoccupations des lauréats diplômés après une, trois et cinq années avec plus de 2.000.000 CV, un outil qui vise à améliorer de la cohérence entre l'offre et la demande pour les lauréats.

En fournissant annuellement le Ministère de l'Enseignement Supérieur Italien, l'agence nationale de l'assurance qualité ainsi que les universités, avec des données actualisées, la mission de l'AlmaLauréa consiste à améliorer l'assurance qualité dans le système de l'enseignement supérieur.

Avec le support de l'Union Européenne, l'AlmaLauréa coordonne actuellement un projet international pour l'implémentation de ses modèles au Maroc et en Tunisie.

Bibliography and References

Manuel F. Bagues and Mauro Sylos Labini, 2007, "Do on-line labor market intermediaries matter? The impact of AlmaLaurea on the university-to-work transition", NBER Working Papers 12621, National Bureau of Economic Research, Inc.

L'EXAMEN DES COMPÉTENCES UNIVERSITAIRES ET SON IMPACT SUR LES ÉTABLISSEMENTS DE L'ENSEIGNEMENT SUPÉRIEUR EN JORDANIE

(Traduction - texte original en annexe)

Bashir Khalifeh Al Zu'bi *

Résumé

Vu l'augmentation continue des établissements de l'enseignement supérieur en Jordanie, privés et publics, il est devenu nécessaire de suivre et de contrôler ces institutions afin d'avoir un regard sur leurs outputs éducatifs, qui, à leur tour impactent les rendements de l'éducation, étant donné que l'objectif principal de ces établissements consiste à former des lauréats compétents pour les marchés du travail, local et régional. C'est la raison pour laquelle l'examen des compétences universitaires a été mis en place en Jordanie. Suivent les principaux aspects relatifs à cet examen, et qui seront présentés dans ce papier.

1. *Présentation de l'examen des compétences universitaires (définition, objectifs, niveaux ...)*
 2. *Passation de l'examen et restitution des résultats ;*
 3. *Impact de l'examen des compétences universitaires.*
-

I. Présentation de l'examen

L'examen est sous forme de plusieurs questions qui mesurent les compétences générales et moyennes à maîtriser par les étudiants en cycle de licence dans les établissements jordaniens, publics ou privés.

Objectifs de l'examen :

Cet examen vise à contrôler la qualité des outputs des programmes académiques dispensés dans les établissements de l'enseignement supérieur jordaniens, et non à juger la performance des étudiants et leur accorder un diplôme (les étudiants représentent le moyen et non la finalité), et ce pour atteindre les résultats suivants :

1. Élaborer les politiques relatives à l'enseignement supérieur de la part des autorités en charge de l'enseignement supérieur et de ses établissements en Jordanie. Il est donc possible, depuis le diagnostic, de déterminer les points forts et points faibles des résultats à obtenir par les étudiants. De plus, ceci permettra d'apporter l'aide nécessaire aux établissements pour prendre les mesures utiles pour dépasser les lacunes rencontrées.
2. Exploiter l'examen comme un outil de garantie de la qualité : les résultats seront des indicateurs importants pour établir les programmes et spécialités d'enseignement dans le cycle supérieur et augmenter la capacité des classes et prendre les décisions adéquates relativement aux accréditations à accorder ou à retirer.
3. Créer un esprit de compétition entre les différents programmes et filières afin de garantir la qualité dans les process et intrants, impactant ainsi les extrants (outputs) et au final, les performances des établissements de l'enseignement supérieur ;
4. L'exploiter comme un outil de reddition des comptes envers les établissements d'un côté, et du ministère de l'enseignement supérieur et de la Commission d'accréditation des établissements de l'enseignement supérieur d'un autre côté ;
5. Fournir des indicateurs quantitatifs et qualitatifs, précis et rigoureux, sur les performances des établissements de l'enseignement supérieur ;

* Président de la Commission Supérieure de l'Accréditation en Education, Jordanie

6. Contrôler le degré de développement et d'amélioration des outputs de l'université, en procédant à une comparaison de la performance de l'université sur plusieurs années.

Constitution de l'examen

L'examen, dans sa forme actuelle, est davantage un examen de compétences qu'un test de niveau. Celui-ci mesure les connaissances et les compétences acquises par l'étudiant, met l'accent sur les aptitudes que l'étudiant doit acquérir pendant sa formation universitaire (compréhension, analyse, créativité...) et sa capacité à en faire usage dans les différentes situations qui pourraient l'affronter dans le marché du travail.

Dans ce contexte, l'examen met l'accent sur 3 niveaux de compétences, tel que précisé ci-dessous :

- **Niveau macro** : ce niveau concerne les compétences générales dispensées dans toutes les filières de l'enseignement supérieur qui permettront à leurs lauréats de s'intégrer dans le marché du travail, et de jouer leur rôle de façon pleine et efficace dans la vie professionnelle. Les compétences sont constituées des compétences relatives au Liberal Arts, et comportent les compétences suivantes :

- La recherche scientifique et explication des résultats ;
- L'esprit critique ;
- La communication ;
- L'outil informatique ;
- Résolution problèmes et du travail en équipe ;
- La connaissance, des droits et des obligations ;
- Les comportements civiques et moraux ;
- Les Mathématiques ;
- Les traitements phonétiques ;

- **Niveau méso** : ce niveau de l'examen insiste sur les outputs à développer de la part de la faculté pour les lauréats. Ce sont des compétences appartenant à plusieurs catégories de filières :

- Les lettres, les langues et leur application ;
- L'information et le journalisme ;
- La loi islamique ;
- Les sciences sociales et de services ;
- Les sciences administratives et de commerce ;
- Les sciences de l'éducation et la psychologie ;
- Les sciences informatiques et techniques ;
- Les sciences mathématiques ;
- Les sciences agronomiques et leur application ;
- Les sciences médicales et paramédicales ;
- Les sciences naturelles essentielles ;
- Les sciences d'ingénierie et techniques ;
- L'audiovisuel ;
- Le droit.

- **Niveau micro** : L'examen concerne, à ce stade, toutes les compétences devant être acquises par les étudiants selon les filières étudiées. Lors de la dernière session du second semestre 2014-2015, il a concerné quelques filières (52), dont l'économie, la finance, la gestion des hôpitaux, le marketing, business administration, terres et environnement, économie agricole, la nutrition, ... Le reste des filières et spécialités sera intégré dans les prochaines éditions de cet examen.

II. Application de l'examen des compétences

L'examen de compétences, de niveau macro, est appliqué à tous les programmes et filières d'universités, publiques et privées, celui de niveau medium s'applique à tous les programmes, au nombre de 14. L'examen de niveau micro a été appliqué lors du second semestre de l'année 2014-2015 sur 52 spécialités. Il est également prévu, au début du premier semestre 2015-2016 de faire passer d'autres spécialités au niveau micro. Le tableau suivant montre les semestres d'études qui ont tenu l'examen de compétences universitaires, ainsi que les effectifs des étudiants qui y ont participé.

Semestres d'études où a été tenu l'examen de compétences universitaires et les effectifs des étudiants qui y ont participé, ainsi que le nombre de programmes et spécialités et programmes académiques pour chaque niveau

Semestre	Effectif des lauréats participant à l'examen	Nombre de niveaux de l'examen	Nombre d'examens comptabilisés	Nombre d'examens des programmes académiques et spécialités contenus dans l'examen
2ème semestre 2012-2013	8582	2 (général+medium)	17164	4
1er semestre 2013-2014	7251	2 (général+medium)	14502	8
2ème semestre 2013-2014	6542	1 (medium)	6542	12
1er semestre 2014-2015	19762	2 (général+medium)	39524	15
2ème semestre 2014-2015	42685	3 2 (général+medium-micro)	112748	77

Les étudiants ciblés

L'examen de la compétence universitaire concerne l'ensemble des étudiants en fin de cycle. Les universités sont informées des noms d'étudiants invités à passer cet examen la veille de la date de sa réalisation car il est estimé qu'aucune préparation n'est nécessaire. L'objectif étant de mesurer les compétences et habilités acquises depuis l'accès à l'université.

Dispositions de l'examen de la compétence universitaire

L'exécution de l'examen se fait selon l'architecture proposée par le Centre national des examens auprès de la Commission d'accréditation des établissements d'enseignement supérieur, en collaboration avec les établissements d'enseignement supérieur. Ainsi, on affecte ressources humaines et matérielles aux établissements, en plus des cadres représentant la Commission. L'examen est informatisé et a plusieurs versions afin de faciliter et fluidifier son usage d'une part, et pour palier aux risques de fraude d'autre part. Il est passé par l'ensemble des étudiants simultanément.

• Modalités du processus d'information sur les résultats de l'examen

1. Informer le Conseil de l'enseignement supérieur et de la recherche scientifique des résultats de l'examen de la compétence universitaire ;
2. Informer les Conseils des présidents des universités des résultats de l'examen de la compétence universitaire ;
3. Tenir une conférence de presse et publier les résultats de l'examen sur les sites officiels ;

4. Fournir pour chaque établissement un rapport détaillé sur les résultats de l'examen en termes de compétences et outputs (maîtrisés ou non) ;
5. Demander aux universités de mettre en place des plans d'amélioration sur la base des résultats de l'examen ;
6. Transmettre à la Commission les corrections proposées par les universités, qui modifient leurs plans éducatifs sur la base des résultats de l'examen et révisent, par conséquent, le contenu des matières enseignées en veillant à leur adéquation avec leurs descriptifs ;
7. Suivre les résultats de la compétence universitaire en collaboration avec les universités afin d'étudier les plans d'amélioration et d'évaluer le développement et le progrès des outputs sur la base de comparaisons longitudinales des résultats pour chaque université ;
8. S'appuyer sur les résultats de l'examen de la compétence universitaire comme principal indicateur dans l'accréditation, pour mesurer la qualité des outputs et pour décider de l'augmentation de la capacité des programmes et spécialités académiques.

Nouvelle vision proposée de la structure de l'examen de la compétence universitaire

L'architecture actuelle de l'examen de la compétence universitaire repose sur trois niveaux (général, moyen, approfondi) car il cible les compétences que chaque étudiant aurait dû acquérir au terme du cycle d'étude. Cet examen se tient sur trois semaines successives. En raison des autres examens qui coïncident avec l'examen de la compétence universitaire et de l'obligation des étudiants à assister aux cours magistraux, il faut revoir cette architecture. En effet, une nouvelle vision propose de faire une seule séance d'examen de la compétence universitaire couvrant tous les niveaux à la fois.

III. Impacts de l'examen de la compétence universitaire sur les systèmes éducatifs au sein des établissements de l'enseignement supérieur

Les nouvelles méthodes d'évaluation et d'accréditation, utilisées par plusieurs organismes internationaux, ont permis de se focaliser sur les outputs du système éducatif. L'accréditation stipule qu'il faut disposer d'un système d'évaluation interne continue qui sert à évaluer les outputs et de les améliorer en conséquence. En outre, la demande du marché de travail, ainsi que la forte compétition qui le caractérise, impose de mener des évaluations régulières des programmes académiques. Ces évaluations permettent d'améliorer la performance des programmes académiques en adéquation avec les nouveautés afin de permettre aux étudiants d'accéder au marché de travail au terme de leur cursus. L'évaluation ne concerne plus seulement les programmes académiques mais concerne également l'établissement de l'enseignement supérieur, de ses structures pédagogiques, de recherche et administratives et de ses activités estudiantines et celles au service de la société.

Ainsi, la Commission d'accréditation des établissements de l'enseignement supérieur veille sur la qualité de ces établissements et de leurs programmes académiques. Cet intérêt confirme la volonté de la Commission de donner de l'importance à la qualité de l'enseignement supérieur et de créer l'esprit de compétition entre les programmes et spécialités afin d'améliorer ses outputs, ce qui influence positivement la performance de l'ensemble des établissements de l'enseignement supérieur en Jordanie. En outre, les universités disposent des résultats de l'examen de compétence universitaire pour améliorer les outputs en comparant les performances de l'université sur plusieurs années, permettant ainsi de disposer d'indicateurs quantitatifs et qualitatifs continus et précis sur le rendement des superviseurs des programmes et spécialités. **L'examen de la compétence universitaire profite :**

- A l'étudiant : impact sur l'amélioration des enseignements ;
- A l'enseignant : mieux planifier les enseignements ;
- Au programme : amélioration et révision en vue d'en améliorer le contenu ;
- A l'université : améliorer la qualité et le développement des plans et des programmes.

La Commission a prévu de réaliser des études analytiques des résultats de l'examen de la compétence universitaire, parmi lesquelles une étude analytique des établissements et de leurs programmes et plans éducatifs, une étude de la réalité des besoins universitaires couvrant les compétences et résultats du niveau

général que les étudiants sont censés acquérir en vue de répondre aux attentes du marché de travail. Ainsi, selon l'étude analytique des établissements et de leurs programmes et plans éducatifs, une batterie d'indicateurs quantitatifs et qualitatifs a été définie sur la base des indicateurs suivants :

- Quelle est la composition du programme académique ?
- Quels sont les seuils des différentes compétences caractérisant le lauréat du programme ?
- Est-ce que le choix des programmes est basé sur des besoins réels de la société ?
- Existe-t-il une description accréditée du programme ?
- A-t-on préparé une matrice des connaissances, des habilités et des compétences ?
- Les outputs sont-ils en adéquation avec les standards académiques en vigueur ? Comment s'en assure-t-on ? Quel degré d'efficacité et d'avantage réalisé et quels sont les domaines de cette réalisation ?
- Existe-t-il une description des cours enseignés ?
- Est-ce que cette description contient la matrice des connaissances, habilités et compétences spécifiques à chaque programme universitaire ?
- Est-ce que la description des cours réalise les connaissances, habilités et compétences répondant aux standards académiques du programme ?
- Est-ce que les contenus sont conformes aux outputs espérés ?
- Quels sont les domaines d'amélioration et les champs d'application des programmes durant les cinq dernières années ?
- Est-ce que les programmes sont révisés périodiquement ?
- A quel moment cette révision s'opère-t-elle ?
- Est-ce que l'on profite des résultats de révision pour renouveler et améliorer le programme ?

Ci-après, un exposé du résumé de l'étude des compétences principales au niveau général et les cursus et besoins universitaires couvrant ces compétences ainsi que les résultats qui en découlent.

Compétences et outputs du niveau général des établissements de l'enseignement supérieur

Le niveau général englobe l'ensemble des compétences et résultats censé être acquis par n'importe quel lauréat des établissements de l'enseignement supérieur afin qu'il soit en adéquation avec les attentes de la vie professionnelle et sociale et celles du marché de travail. Ceci permet également au lauréat de s'inscrire dans le processus de développement personnel et de la formation tout au long de la vie. Ces compétences générales sont les :

- Compétences de la recherche scientifique et explication des résultats ;
- Compétences de l'esprit critique ;
- Compétences de la communication ;
- Compétences de l'outil informatique ;
- Compétences pour résoudre les problèmes et du travail en équipe ;
- Compétences de la connaissance, des droits et des obligations ;
- Compétences des comportements civiques et moraux ;
- Compétences mathématiques ;
- Compétences des traitements phonétiques.

Le tableau suivant indique les principales compétences du niveau général, ainsi que les cursus et besoins universitaires des facultés couvrant ces compétences :

Les compétences principales d'ordre général et les besoins des universités et facultés qui couvrent ces compétences

Compétences	Besoins de l'université	Besoins de la faculté	Remarques
Compétences de recherche scientifique et interprétation des résultats	Aucun besoin spécifique mentionné	Recherche scientifique, méthodologie de la recherche, développement des concepts scientifiques, principes de la recherche scientifique, les méthodes de la recherche scientifique, la pensée scientifique, les bases de la recherche scientifique, l'initiation à la recherche en éducation et statistiques, séminaire de recherche en traduction, méthodologies de la recherche et des statistiques sociales.....	La plupart des facultés dans les universités couvrent les compétences de recherche scientifique et l'interprétation des résultats.
La pensée critique	Aucun besoin spécifique mentionné. Les compétences en communication (langue arabe 101 et 102, et la langue anglaise 99,101,102), ces cours comprennent des compétences (dialogue, communication, assimilation, lecture, écriture).	Pensée analytique, scientifique, la pensée et la créativité, les méthodes de penser, l'enseignement de la pensée, la pensée créative.	Cours de compétences en communication couvrant la compétence de la pensée critique, et plusieurs autres exigences.
Compétence en communication	Aucun besoin spécifique mentionné. Les compétences en communication (langue arabe 101 et 102, et la langue anglaise 99,101,102), ces cours comprennent des compétences (dialogue, communication, assimilation, lecture, écriture).		Toutes les universités publiques et privées enseignent ces cours

Compétences en culture informatique	Education nationale		Toutes les universités publiques et privées enseignent ces cours
Résolution de problèmes et esprit d'équipe	Aucun besoin spécifique mentionné Les compétences en communication (langue arabe 101 et 102, et la langue anglaise 99,101,102), ces cours comprennent des compétences (dialogue, communication, assimilation, lecture, écriture).		Toutes les universités publiques et privées enseignent ces cours
La connaissance des droits et devoirs des compétences	Education nationale		Toutes les universités publiques et privées enseignent ces cours
La connaissance des règles et des comportements éthiques	Education nationale		Toutes les universités publiques et privées enseignent ces cours
Compétences du traitement mathématique	Aucun besoin spécifique mentionné		La plupart des exigences de facultés et des exigences interdisciplinaires des facultés scientifiques et humaines couvrent cette compétence.
Aptitudes de traitement verbal	Les compétences en communication (langue arabe 101 et 102, et la langue anglaise 99,101,102), ces cours comprennent des compétences (dialogue, communication, assimilation, lecture, écriture).		Toutes les universités publiques et privées enseignent ces cours

Résultats de l'examen d'aptitude universitaire pour des 3 dernières sessions

Le tableau suivant montre les moyennes des réponses correctes à l'examen d'aptitude universitaire niveau général, pour des 3 dernières sessions pour chaque compétence.

Comparaison des résultats de la maîtrise des compétences et des résultats du niveau général du premier examen semestriel de l'année universitaire 2013/2014 et le premier et le deuxième de l'année universitaire 2014/2015 semestre.			
Compétences	La moyenne arithmétique des réponses correctes Pour le premier semestre 2013/2014	La moyenne arithmétique des réponses correctes Pour le premier semestre 2014/2015	La moyenne arithmétique des réponses correctes Pour le deuxième semestre 2014/2015
Compétences de recherche scientifique et interprétation de résultats	46%	48.29%	43.74%
La pensée critique	42%	43.60%	52.50%
Compétence de communication	22%	21.88%	57.27%
Compétences en culture informatique	44%	43.70%	68.03%
Résolution de problèmes et esprit d'équipe	49%	47.00%	57.29%
La connaissance des droits et devoirs	53%	50.78%	62.91%
La connaissance des règles et des comportements éthiques	45%	43.00%	70.52%
Compétences de traitement mathématique	42%	42.57%	46.20%
Aptitudes de traitement verbal	45%	47.84%	51.03%
La proportion globale de compétence du niveau général	44%	45.00%	56.28%

Comparativement aux années précédentes, les résultats de l'examen de la compétence universitaire ont montré qu'il existe, en 2014-2015, une amélioration de la maîtrise des compétences et habilités. Cette amélioration est due aux changements qui ont été opérés sur la structure de cet examen obligatoire pour la fin de cycle. Ces changements de la structure de l'examen et non pas les compétences visées ont poussé les étudiants à devenir plus sérieux dans leur travail est plus présents lors des cours. Ainsi, sur la base de ces résultats, les universités ont dû adapter le contenu de leurs cours pour les mettre en adéquation avec les exigences de l'examen, notamment en révisant les cursus, l'introduction de matières optionnelles différentes ainsi que l'organisation de rencontres avec les étudiants afin de leur expliquer, pour leur vie professionnelle et scientifique, l'intérêt de l'examen de la compétence universitaire.

• Améliorations des programmes académiques proposées par les établissements de l'enseignement supérieur sur la base des résultats de l'examen de la compétence universitaire

L'organe d'accréditation des établissements d'enseignement supérieur transmet à chaque établissement un rapport détaillé sur les résultats de l'examen de la compétence universitaire. Cet organe suit ces résultats année après année et, sur la base des comparaisons longitudinales, étudie avec les établissements leurs plans d'amélioration. Les résultats de l'examen de la compétence universitaire est une référence dans la détermination de la qualité des outputs universitaires mais aussi dans la décision d'augmenter la capacité d'accueil académique. Parmi les plans d'amélioration des programmes académiques proposés on cite :

1. Développement des programmes et révision des cursus universitaires tout en introduisant des matières optionnelles à tous les niveaux ;
2. Introduction des compétences et habilités objets de l'examen de compétence universitaire dans les examens mensuels dans les disciplines en relation avec ces compétences et habilités ;
3. Organisation de rencontres avec les étudiants afin de leur expliquer l'intérêt de l'examen de la compétence universitaire ;
4. Révision du contenu des matières enseignées aux étudiants et s'assurer de leur conformité avec leurs descriptions ;
5. Garantir aux étudiants la liberté d'expression et de débat afin d'encourager la réflexion innovante et promouvoir les talents ;
6. Fournir un environnement universitaire riche en termes d'expérimentation et d'outils pédagogiques permettant davantage de créativité chez les étudiants. Il faut aussi multiplier les occasions aux étudiants pour organiser et participer à des activités, notamment en mettant à leur disposition les laboratoires, clubs et autres centres universitaires ;
7. Utilisation de plusieurs méthodes d'enseignement pour mieux orienter les étudiants et tenir compte des différences entre eux ;
8. Communication des travaux créatifs et innovants des étudiants à leurs pairs afin de consolider leur confiance en soi et leur fierté tout en les poussant à continuer les réussites en termes de créations et d'innovations ;
9. Amélioration et développement des cursus professionnels et révision de leurs contenus et des méthodes d'évaluation des étudiants ;
10. Encouragement des occasions d'échange d'expériences, recherches, projets communs et conférences entre les enseignants du même cycle ;
11. Organisation de cycles de formations au profit des enseignants sur les différentes compétences et habilités en lien avec l'activité d'enseignement et de son management ;
12. Communication avec les lauréats afin de s'assurer de l'adaptation de leurs compétences avec les exigences du marché de travail ;
13. Définition de plans de travail dans les cycles et décanats visant l'encouragement des activités hors curricula dans les cursus ;
14. Développement de stratégies d'évaluation dans les cursus en y intégrant des méthodes autres que les examens actuels ;
15. Consolidation de la notion de citoyenneté et du civisme tout en mettant en place des dispositions visant la protection des établissements de l'enseignement supérieur ;
16. Mise en place d'un système de standards garantissant la qualité de l'éducation dans l'enseignement supérieur conformément aux standards définis par l'union des universités arabes en matière de qualité de l'enseignement supérieur.

Recommandations issues de l'analyse des résultats de l'examen de la compétence universitaire

1. En se basant sur les résultats de l'examen de la compétence universitaire, il apparaît qu'une majorité d'étudiants des universités publiques et privées, a des faiblesses dans les compétences de recherche scientifique, explication des résultats, traitement numérique, réflexion critique, communication et traitement du langage. Cette situation est due à l'absence de programmes universitaires qui couvrent ces compétences. Ainsi, il est recommandé aux universités d'introduire un module universitaire

général intitulé compétences de réflexion et de recherche scientifique (incluant les méthodes de réflexion et de recherche scientifique, résolution de problèmes, communication et information) ;

2. Il est recommandé aux universités jordaniennes publiques et privées de développer les programmes et de réviser les cursus éducatifs tout en introduisant des matières optionnelles diverses dans l'ensemble des cycles tenant compte des compétences du niveau général ;
3. Il est recommandé d'organiser de rencontres avec les étudiants afin de leur expliquer l'intérêt, pour leur vie professionnelle concrète, des compétences et habilités liées à l'examen de la compétence universitaire ;

Il est recommandé de développer les plans éducatifs dans l'éducation nationale afin de tenir compte du besoin en termes de certaines compétences et habilités d'ordre général telles que : la culture informatique, la connaissance des droits et obligations, la connaissance des règles et des comportements indécents

Traduction du Résumé

Abstract Translation

Bashir Khalifeh Al Zu'bi

UNIVERSITY PROFICIENCY EXAM AND ITS IMPACT ON HIGHER EDUCATION INSTITUTIONS IN JORDAN

Given the continuous increase of the institutions of higher education in Jordan, both public and private, it becomes necessary to control and follow up these institutions in order to have a clear idea about their outputs. These outputs impact also the performance of education as a whole, given the fact that the main objective of the institutions is to develop skilled graduates for local and regional labour markets. This is the reason for the implementation of the Exam of university skills in Jordan. Here are its main aspects, which will be presented in this paper:

1. Presentation of the exam (definition, objectives, levels...);
2. Taking the exam and presenting the results,
3. The exam impacts.

L'AGENCE NATIONALE D'ÉVALUATION ET D'ASSURANCE QUALITÉ DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE DU MAROC. (UNE NOUVELLE EXPÉRIENCE)

El Mokhtar BAKKOUR *

Résumé

L'idée de la qualité de la formation remontait à 1999 avec l'adoption de la Charte nationale de l'éducation et de formation. Traduit par la loi 01-00 portant organisation de l'enseignement supérieur en exigeant la mise en place d'une instance nationale d'évaluation pour le système de l'enseignement supérieur et de la recherche scientifique.

Après plus de 14 ans, cette instance voit le jour sous l'appellation: Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique, créée par la loi 80-12, elle a le statut d'établissement public doté de la personnalité morale et de l'autonomie financière.

Les missions de l'Agence concernent l'évaluation des formations, de la recherche scientifique, les programmes de coopération et l'évaluation des établissements publics et privés et les établissements étrangers.

En tant qu'institution, l'Agence est administrée par un Conseil d'administration présidé par le chef de gouvernement ou l'autorité gouvernementale déléguée par lui. Ce Conseil d'administration est composé de membres de droit, de membres désignés et d'un membre élu. En outre, le conseil crée en son sein trois commissions spécialisées en plus des commissions ad-hoc.

Quant à la gestion de l'Agence elle est confiée à un Directeur nommé conformément à l'article 92 de la Constitution.

Pour permettre à l'Agence d'accomplir ses missions, elle doit disposer d'un personnel : Personnel recruté par l'Agence son statut du personnel, personnel détaché ou mise à la disposition par les administrations publiques, mais aussi elle peut recourir à des consultants ou des contractuels marocains ou étrangers.

Pour plus de transparence, de visibilité et d'information, l'agence établit des rapports annuels destinés au chef de gouvernement, au ministre de tutelle et aux établissements concernés.

La mise en œuvre de l'Agence paraît une tâche assez difficile, vu la diversité de ses missions et la consistance des évaluations.

I. Les fondements de la mise en place de l'Agence

1. La Charte nationale d'éducation et de formation. 1999

Cette Charte a fait un consensus de tous les représentants de la nation en 1999, et a ainsi tracé les grandes orientations de l'éducation et de formation pour les dix années à venir.

L'un des six espaces d'innovation qui sont identifiés par cette Charte c'est:

« L'amélioration de la qualité de l'éducation et de la formation. »¹

¹ - Royaume du Maroc : Commission spéciale d'éducation et de formation (COSEF) : La Charte nationale d'éducation et de formation, 1999. P : 46.

* Directeur de l'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique du Maroc

2. Les dispositions de la loi 01-00

La loi 01-00² portant organisation de l'enseignement supérieur de l'année 2000 qui vient mettre en œuvre les préconisations de cette charte nationale à travers les articles 77 et 79 :

L'article 77 : « Le système de l'enseignement supérieur est soumis, dans sa globalité, à une **évaluation régulière**, portant sur sa rentabilité interne et externe, touchant tous les aspects pédagogiques, administratifs et de recherche..... »

L'article 79 : « Pour la réalisation des audits et de l'**évaluation** requise à l'article 77 ci-dessus, il sera procédé à la création d'**instances spécialisées** de régulation bénéficient de l'autonomie et de l'indépendance nécessaires, notamment **une instance nationale d'évaluation** et un observatoire..... »

3. Le Discours Royal

Ce sont des orientations majeures pour une bonne qualité de l'enseignement que ce soit à l'école ou à l'université :

« Ce système, qui nous interpelle aujourd'hui, se doit non seulement d'assurer l'accès égale et équitable à l'école et à l'**université** pour tous nos enfants, mais également de leur **garantir** le droit à un **enseignement de qualité**, doté d'une forte attractivité et adapté à la vie qui les attend. »³

II. La Création de l'Agence : Le Ministère honore son engagement

1. Le plan d'action du Ministère de l'enseignement supérieur, de la recherche scientifique et de la formation des cadres 2013-2016.

Le plan d'action du Ministère pour les quatre années à venir a été structuré sur Six axes et 39 projets :

Le projet 7 de l'Axe 1 intitulé : « Mettre en place un système de régulation et de développement de la qualité ».

Deux résultats phare attendus de ce projet 7:

- **L'évaluation externe des universités** : Peut être ça été mal formulé, car la mise en place de ce système vise l'évaluation externe de tous les établissements de l'enseignement supérieur.
- **L'amélioration du rendement des universités** : Mais en fait Tous les établissements de l'enseignement supérieur.

2. La loi n°80-12

Plus de 14 ans d'attente, sa mise en place a été prévue par la loi 01-00 relative à l'organisation de l'enseignement supérieur.

L'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique a été créée par la Loi n° 80-12⁴. Cette loi comprend 15 articles répartis sur quatre chapitres :

- Dénomination et objet ;
- Administration et gestion ;
- Personnel de l'Agence ;
- Organisation financière.
- **Le décret n° 2-15-813 du 28 décembre 2015⁵** en application de la loi 12-80 visant à :
 - Déterminer le siège de l'Agence qui est fixé à Rabat ;
 - Déterminer l'autorité de tutelle de l'Agence qui est l'autorité gouvernementale chargée de l'enseignement supérieur et de la recherche scientifique ;

2 - Dahir n° 1-00-199 du 15 safar 1421 (19 mai 2000) portant promulgation de la loi n° 01-00 portant organisation de l'enseignement supérieur. B.O. n° 4800 du 1er juin 2000, p : 393

3 - <http://www.maroc.ma/fr/discours-royaux/texte-int%C3%A9gral-du-discours-adress%C3%A9-par-sm-le-roi-%C3%A0-la-nation-%C3%A0-l'occasion-du-59%C3%A8me>

4 - Dahir n° 1-14-130 du 3 chaoual 1435 (31 juillet 2014) portant promulgation de la loi 80-12. (B.O. n° 6284 du 21 Aout 2014) p : 3888.

5 - BO n° 6431 du 18 janvier 2015, édition générale (en langue arabe).

- Déterminer les modalités de désignation et d'élection de certains membres du conseil d'administration de l'Agence.

3. La vision stratégique 2015-2030 du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique

Cette vision stratégique qui a pour souci majeur, la qualité de l'enseignement et de la recherche, intitulée : « **Pour une école de l'équité, de la qualité et de la promotion.**»

Le projet 2 traduit clairement cette vision et cette ambition de la qualité par : « **La mise en place de L'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique et sa mise en œuvre.**»

III. Statut légal de l'Agence

Le choix du statut juridique d'une structure d'évaluation et d'assurance qualité est déterminant pour le degré d'autonomie de la structure et son influence sur le système de l'enseignement supérieur.

De manière générale, dans la pratique internationale, on peut identifier quatre principaux types de statut légal dans la création d'une structure d'assurance qualité :

- Création d'une structure intégrée dans les services de l'Etat, par exemple sous forme d'unité spécialisée au sein ministère de l'Education ou celui de l'enseignement supérieur (cas du Cameroun) ;
- Création d'un organisme ayant statut d'association : l'agence d'assurance qualité constituée par des établissements d'enseignement supérieur (cas des Philippines) ;
- Création d'un organe de régulation public ou d'une structure sous la responsabilité d'un organisme public. L'Etat peut prendre l'initiative de sa mise en place, mais la structure est indépendante (cas de l'Egypte).
- Création d'un organisme d'assurance qualité particulier pour les établissements et les diplômés d'un secteur professionnel (cas de la CTI en France)⁶.

La tendance internationale actuelle, dans la mise en place d'une structure d'assurance qualité, suggère le respect de certaines normes, à savoir :

- La structure d'assurance qualité doit être indépendante de l'Etat et des établissements d'enseignement supérieur ;
- La structure d'assurance qualité ne doit pas interférer dans la procédure d'évaluation par les pairs malgré son rôle essentiel dans le processus ;
- La structure d'assurance qualité doit établir la confiance, dans le sens de faire comprendre qu'évaluer c'est pour évoluer et non pas dans le sens d'inspecter ou de sanctionner.
- La création d'un système d'assurance qualité externe a besoin d'expertise ; la gestion du système nécessite des ressources considérables ;
- L'assurance qualité doit améliorer la qualité sur le long terme⁷.

Pour le cas marocain, le choix était clair, et d'après l'article premier de la loi 80-12 c'est la création d'une Agence sous forme d'un établissement public, doté de la personnalité morale et de l'autonomie financière et qui est soumise à la tutelle de l'Etat, cette tutelle de l'Etat a été définie dans le décret d'application décret n° 2-15-813 du 28 décembre 2015 qui est l'autorité gouvernementale chargée de l'enseignement supérieur et de la recherche scientifique.

L'Agence est également soumise au contrôle financier de l'Etat applicable aux entreprises publiques et aux autres organismes conformément à la loi 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes.⁸

6 - Institut International de planification de l'éducation (UNESCO) : Assurance qualité externe : option pour les gestionnaires de l'enseignement supérieur. Module 4 : Créer et organiser une structure d'assurance qualité. 2011. P : 8.

7 - Institut International de planification de l'éducation (UNESCO) : Op.cit. pp : 49-50.

8 - Dahir n° 1-03-195 du 16 ramadan 1424 (11/11/2003) portant promulgation de la loi n° 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes. BO. N° 5170 du 18/12/2003.

IV. Missions diversifiées de l'Agence

Pour la réalisation des missions d'une Agence d'évaluation, elle doit s'appuyer sur des valeurs connus et partagés.

De manière générale, on identifie deux grandes catégories de valeurs :

Valeurs fondamentales : la justice, l'égalité et l'équité ;

Valeurs instrumentales : la cohérence, la rigueur et la transparence.⁹

Dans cette logique et cette mouvance mondiale, que doit s'inscrire l'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique dans l'accomplissement de ses missions.

Les articles 3, 4 et 6 de la loi 80-12 relative à L'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique définissent ses missions :

1. L'évaluation pour le compte de l'État

D'après l'article 3, l'Agence effectue pour le compte de l'Etat l'évaluation système de l'enseignement supérieur et de la recherche scientifique dans le but de garantir la qualité. A cette fin elle est amenée à :

- Evaluer des établissements d'enseignement supérieur publics et privés et les établissements de recherche scientifique;
- Examiner et évaluer les filières de formation en vue de l'obtention ou du renouvellement de l'accréditation.
- Evaluer les activités des centres d'études doctorales et dresser le bilan des formations et des travaux de recherche réalisés dans ces centres.
- Evaluer la recherche scientifique et l'efficacité de ses structures.
- Evaluer les programmes et les projets de coopération universitaire dans le domaine de la formation et la recherche scientifique.¹⁰

2. L'évaluation à la demande

Selon les dispositions de l'article 4 de la loi 80-12, l'Agence peut effectuer des évaluations dans la limite de ses attributions au profit:

- Des départements ministériels qui entrent dans leurs intérêts l'enseignement supérieur, la recherche scientifique ou la formation des cadres.
- Du conseil supérieur de l'éducation, de la formation et de la recherche scientifique;
- De l'Académie Hassan II des sciences et technique;
- Du centre national de la recherche scientifique et technique;
- Des universités ou des établissements d'enseignement supérieur et de recherche.

3. Une vocation à l'international

Le législateur marocain a visé l'ouverture sur l'international. De ce fait, il a étendu les missions de l'agence pour effectuer des évaluations au profit des établissements de formation et de recherche scientifique étrangers :

- Soit dans un cadre de convention de coopération conclu par le Maroc;
- Soit sur demande ces établissements étrangers à leur demande et dans le cadre de conventions conclues à cet effet, sur autorisation de l'administration.

9 - Organisation internationale de la francophonie : Guide 5 : Conception et réalisation d'un référentiel d'évaluation p: 22

10 - Une proposition de loi vient d'être votée à la 1ère chambre du parlement qui vise l'élargissement des missions de l'Agence à « l'études et évaluation des dossiers des équivalences ».

V. Administration et gestion de l'Agence

L'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique est administrée par un conseil d'administration, et gérée par un directeur.

• Administration de l'Agence

Comme pour tout établissement public (à l'exception des universités), le conseil d'administration est présidé par le Chef du Gouvernement ou l'autorité gouvernementale déléguée par lui à cet effet.

Le conseil d'administration est l'organe suprême chargé de l'administration de l'Agence. A cet effet, il dispose de tous les pouvoirs et attributions nécessaires à l'administration de l'Agence.

Sur convocation de son président, le conseil d'administration se réunit, aussi souvent que les besoins l'exigent, mais au moins deux fois par an :

- avant le 30 juin pour arrêter les états de synthèse de l'exercice clos ;
- avant le 15 octobre pour examiner et arrêter le budget et le programme prévisionnel de l'exercice suivant.

Le conseil délibère valablement par la présence de la moitié au moins de ses membres, si ce quorum n'est pas atteint, il délibère valablement lors d'une deuxième réunion sans tenir compte de ce quorum.

Il prend ses décisions à la majorité des voix de ses membres présents, en cas de partage égal des voix, celle du président est prépondérante.

D'après l'article 7 de la loi 80-12, ce conseil est composé de membres de droit (5), de membre désignés (8) et un membre élu (1).

Membres de droit :

1. De deux représentants du ministère de tutelle ;
2. Du secrétaire perpétuel de l'Académie Hassan II des sciences et des techniques ou son représentant ;
3. Du président de la commission nationale de coordination de l'enseignement supérieur ou son représentant ;
4. Du président du Conseil supérieur de l'éducation, de la formation et de la recherche scientifique ou son représentant.

Membres désignés :

Les dispositions du Le décret n° 2-15-813 du 28 décembre 2015¹¹ en application de la loi n° 80-12, viennent préciser les modalités de désignation de ces membres comme suit :

1. Un représentant des établissements ne relevant pas des universités, désigné par l'autorité gouvernementale chargée de l'enseignement supérieur sur proposition du conseil de coordination prévue dans l'article 28 de la loi 01-00 ;
2. Un représentant de l'enseignement supérieur privé, désigné par l'autorité gouvernementale chargée de l'enseignement supérieur sur proposition de la commission de coordination de l'enseignement supérieur privé, prévue dans l'article 61 de la loi 01-00 ;
3. Deux anciens présidents d'université relevant de l'enseignement supérieur public, désignés par l'autorité gouvernementale chargée de l'enseignement supérieur ;
4. Quatre membres connus pour leur compétence scientifique et technique, désigné par l'autorité gouvernementale chargée de l'enseignement supérieur sur proposition du Directeur de l'Agence ;

Membre désigné :

1. Un représentant élu par et parmi le personnel de l'Agence.

En outre, le président du conseil d'administration, et pour faire bénéficier ce conseil des compétences des personnalités, dont leur présence peut être jugée utile, il peut inviter aux réunions de ce conseil et à titre consultatif, toute personne physique ou morale des secteurs public ou privé.

¹¹ - BO n° 6431 du 18 janvier 2015, édition générale (en langue arabe).

• Gestion de l'Agence:

D'après l'article 10 de la loi 80-12, l'Agence est dirigée par un directeur qui est nommé conformément aux dispositions de l'article 92 de la Constitution¹², c'est une nomination en conseil de gouvernement et qui se fait par décret du Chef de Gouvernement.

Le directeur de l'Agence dispose de tous les pouvoirs et attributions nécessaires à la gestion de l'Agence, c'est le représentant de l'Agence en interne et en externe.

Il peut recevoir délégation du conseil d'administration pour le règlement d'affaires déterminées qui sont de la compétence de ce conseil.

• Commissions spécialisées créées par le conseil d'administration:

La mission de l'Agence est l'évaluation, et pour lui permettre l'accomplissement de ses missions de l'article 11 de la loi 80-12 a prévu la création au sein du conseil d'administration trois commissions spécialisées.

Ces commissions spécialisées et qui sont permanentes qui traduisent les principales missions de l'Agence :

- La commission d'évaluation des établissements ;
- La commission d'évaluation des formations ;
- La commission d'évaluation de la recherche scientifique.

En plus de ces commissions spécialisées, le conseil d'administration peut décider la création de commission « ad-hoc » dont il fixe la composition et les modalités de fonctionnement et à laquelle il peut déléguer certaines de ses attributions.

La composition et les modalités de fonctionnement de ces commissions spécialisées et des commissions ad-hoc sont définies dans le règlement intérieur de l'Agence prévu à l'article 9 de ladite loi approuvé par le conseil d'administration.

VI. Personnel de l'Agence

Pour l'accomplissement de ses missions, l'Agence doit disposer d'un personnel.

D'après l'article 13 de loi 80-12 il est composé de:

- Personnel recruté par l'Agence conformément à son statut du personnel;
- Personnel détaché ou mis à sa disposition par les administrations publiques conformément à la législation et la réglementation en vigueur en la matière;
- Recours à des consultants et contractuels marocains ou étrangers pour des missions déterminées.

VII. Reddition des comptes

L'Agence est placée sous tutelle de l'Etat, divers mécanismes de contrôle : en tant qu'établissement public, un représentant du ministère des finance qui assiste au conseil d'administration de l'Agence ; et un contrôle d'Etat est ce conformément à la loi 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes.¹³

Rapports de l'Agence :

Le contrôle et la reddition des comptes se fait par le biais de rapports :

- Rapport annuel soumis au Chef de gouvernement comprenant : les activités de l'année, les recommandations visant l'amélioration de la qualité des établissements objet de l'évaluation;
- Rapport annuel soumis au Ministre de tutelle partant sur : l'état, les résultats et les perspectives qui se dégagent des opérations de ladite évaluation.
- Publication des rapports annuels ;
- Transmettre aux universités et aux établissements d'enseignement supérieur et de la

12 - Dahir n° 1-11-91 du 27 chaabane 1432 (29 juillet 2011) portant promulgation du texte de la constitution. B.O. n° 5964 bis du 30 juillet 2011, p : 1902.

13 - Dahir n° 1-03-195 du 16 ramadan 1424 (11/11/2003) portant promulgation de la loi n° 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes. BO. N° 5170 du 18/12/2003.

recherche scientifique les rapports des évaluations effectuées pour leur compte, afin de les présenter à leurs conseils.

Evaluation cyclique de l'agence :

On parle d'une Agence d'évaluation et d'assurance qualité, qui est mise en place pour évaluer les formations, les établissements d'enseignement et de recherche, dans le but d'améliorer de manière générale la qualité de l'enseignement et de la recherche.

L'Agence doit aussi obéir à cette logique d'évaluation et d'assurance qualité. S'inscrivant dans cette logique internationale, et selon les références et lignes directrices pour le management de la qualité dans l'espace européen de l'enseignement supérieur, « il est recommandé que chaque agence européenne se soumette à un examen externe cyclique de ses activités et de ses méthodes au moins tous les cinq ans. Les résultats devront être publiés dans un rapport qui fera état du degré de conformité de l'agence avec les références européennes pour le management de la qualité des agences d'évaluation. »¹⁴

VIII. La mise en œuvre de l'agence: Un immense chantier

C'est un long chemin à parcourir, c'est un chantier qui commence, d'abord sur le plan de la structuration de l'Agence, vue le nombre d'institutions, de filières de formation, de structures de recherche et de projets de coopération.

La mise en place des structures :

- Projet de décret fixant les critères de qualité sur proposition du conseil d'administration;
- L'élaboration de l'organigramme de l'Agence;
- L'élaboration du statut du personnel de l'Agence;
- L'élaboration des règlements intérieurs de l'Agence et du conseil d'administration;
- L'élaboration d'un règlement fixant les conditions et formes de passation des marchés;
- Mettre en place un barème des tarifs des prestations rendues par l'Agence;
- Sélection et formation des évaluateurs.

Diversifications et consistance des évaluations :

Selon le premier alinéa de l'article 5 de loi 80-12 « l'évaluation effectuée par l'Agence consiste en un audit des performances académiques et institutionnelles sur la base de critères de qualité fixés par voie réglementaire, sur proposition de l'Agence. »

Dans le cadre de ses missions, et pour bien mener sa tâche :

- L'Agence effectue des enquêtes sur le terrain ;
- Des visites aux établissements objet d'évaluation ;
- Consultation et examen de tous les documents et informations nécessaires pour bien mener sa mission ;
- Tenir des entretiens avec les responsables, le corps enseignants, le personnel administratif, les étudiants et les acteurs économiques entretenant des relations avec les établissements objet d'évaluation.

Pour bien accomplir ses missions d'évaluations, la loi impose à ce que tous les établissements concernés par l'évaluation doivent mettre à la disposition de l'Agence tous les documents et informations demandés.

A titre indicatif, et pour illustrer l'importance du chantier qui doit être initié par l'Agence :

Evaluation des établissements (institutionnelle)

- 125 établissements universitaires;
- 29 établissements ne relevant pas des universités;
- 212 établissements d'enseignement supérieur privé.

¹⁴ - ENQA : références et lignes directrices pour le management de la qualité dans l'espace européen de l'enseignement supérieur. Helsinki 2005, p 29.

Evaluation des filières de formation :

- 2000 filières des établissements publics;
- 400 filières des établissements privés.

Evaluation des CEDOC :

- 100 CEDOC.

Evaluation de la recherche :

- 1400 structures de recherche (équipes, laboratoires, centres...)

Conclusion

L'évaluation externe et l'assurance qualité demeure un nouveau métier au Maroc, surtout l'Agence nationale d'évaluation et d'assurance qualité de l'enseignement supérieur et de la recherche scientifique récemment créée et qui est entrain de se mettre en place a besoin de soutien financier, technique , ingénierie d'évaluation, formation des experts et encourager et facilite sa collaboration et son accès aux réseaux européens, arabes des Agences d'évaluations et d'assurance qualité.

L'Agence est là, mais la qualité reste une culture, qui reste un concept complexe est politisé, est il un concept idéologique¹⁵.

15 - Lee Harvey. Op.cit. p : 38

Traduction du Résumé**Abstract Translation**

El Mokhtar BAKKOUR

THE ESTABLISHMENT OF THE NATIONAL AGENCY FOR ASSESSMENT AND QUALITY ASSURANCE OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

The idea of the quality of training goes back to 1999, while adopting the National Charter of Education and Training. It was implemented through the law 01.00, organizing the Higher Education in Morocco, requiring the implementation of the National Agency for Evaluation of the Higher Education and Scientific Research.

After more than 14 years, this agency was created under the name of the National Agency of Evaluation and Quality Assurance of Higher Education and Scientific Research, under the law 80-12, as a public institution with financial autonomy.

The Agency missions consist of evaluating training, scientific research, cooperation programs as well as evaluating higher education institutions, both private and public.

As an institution, the Agency is managed by an Administrative Council, presided by the Prime Minister or he chooses a governmental authority. This administrative Council is constituted of full members, members designated and one elected member. Furthermore, the Council created within its organs three specialized commissions, besides ad-hoc commissions.

Regarding the Agency management, it is managed by a director, nominated according to article 92 of the Constitution.

In order to allow the Agency to fulfill its missions, it has to have employees: Some employees have been hired as the Agency's employees and staff seconded to the Agency from government departments. The Agency can also hire consultants as well as experts, either Moroccans or foreigners.

For more transparency, visibility and information, the Agency publishes annual reports to the Head of Government, to the minister as well as designated institutions.

Bibliographie

- Dahir n° 1-11-91 du 27 chaabane 1432(29 juillet 2011) portant promulgation du texte de la constitution. B.O. n° 5964 bis, du 30 juillet 2011, p : 1902.
- Royaume du Maroc : Commission spéciale d'éducation et de formation (COSEF) : La charte nationale d'éducation et de formation, 1999.
- Dahir n° 1-00-199 du 15 safar 1421 (19 mai 2000) portant promulgation de la loi n° 01-00 portant organisation de l'enseignement supérieur. B.O. n° 4800 du 1er juin 2000, p : 393.
- Dahir n° 1-03-195 du 16 ramadan 1424 (11/11/2003) portant promulgation de la loi n° 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes. BO. N° 5170 du 18/12/2003.
- Dahir n° 1-03-195 du 16 ramadan 1424 (11/11/2003) portant promulgation de la loi n° 69-00 relative au contrôle financier de l'Etat sur les entreprises publiques et autres organismes. BO. N° 5170 du 18/12/2003.
- Dahir n° 1-14-130 du 3 chaoual 1435 (31 juillet 2014) portant promulgation de la loi 80-12. (B.O. n° 6284 du 21 Aout 2014) p : 3888.
- Le décret n° 2-15-813 du 28 décembre 2015 BO n° 6431 du 18 janvier 2015, édition générale (en langue arabe) p : 342.
- <http://www.maroc.ma/fr/discours-royaux/texte-int%C3%A9gral-du-discours-adress%C3%A9-par-sm-le-roi-%C3%A0-la-nation-%C3%A0-l'occasion-du-59%C3%A8me>
- Lee Harvey : Les initiatives canadiennes d'assurance de la qualité vues dans le contexte international. Colloque sur l'assurance qualité du CMEC, Québec 27et 28 mai 2008. www.univac.ca/wp-content/
- Institut International de planification de l'éducation (UNESCO) : Assurance qualité externe : option pour les gestionnaires de l'enseignement supérieur. Module 4 : Créer et organiser une structure d'assurance qualité. 2011.
- Organisation internationale de la francophonie : Guide 5 : Conception et réalisation d'un référentiel d'évaluation p: 22
- ENQA : références et lignes directrices pour le management de la qualité dans l'espace européen de l'enseignement supérieur. Helsinki 2005, p 29.

- 3

TECHNIQUES DE MESURES D'ÉVALUATION

L'ÉVALUATION FORMATIVE EN FRANÇAIS : ANALYSE MÉTROLOGIQUE^a

Pierre Valois¹, Erick Falardeau¹, Frédéric Guay¹,
Belkacem Abdous^{1,2}, Suzanne Ngo Boumkouo¹

Résumé

Le développement de la compétence « Lire et apprécier des textes variés » est important dans la progression des apprentissages des élèves du secondaire. Pour développer cette compétence, l'élève doit construire du sens, porter un jugement critique, mettre à profit et acquérir des connaissances sur la langue, les textes et la culture, réfléchir à sa pratique de lecteur. Cela implique également que les textes ainsi que les tâches qui leur sont associées soient variés et complexes du point de vue de l'évaluation des apprentissages. En d'autres mots, pour qu'un élève consolide sa compétence à « lire et apprécier » et maîtrise tous les processus qu'elle implique, il doit disposer d'outils formatifs qui proposent un découpage clair des différentes composantes de la compétence. Dans le présent article, nous montrons les avantages d'utiliser les théories de la mesure et le logiciel convivial d'analyse d'item pour valider un outil d'évaluation formative en lecture au secondaire.

Introduction

Le développement de la compétence « Lire et apprécier des textes variés » est important dans la progression des apprentissages des élèves. Pour développer cette compétence, l'élève doit construire du sens, porter un jugement critique, mettre à profit et acquérir des connaissances sur la langue, les textes et la culture, réfléchir à sa pratique de lecteur. Cela implique également que les textes ainsi que les tâches qui leur sont associées soient variés et complexes du point de vue de l'évaluation des apprentissages. Autrement dit, pour qu'un élève consolide sa compétence à « lire et apprécier » et maîtrise tous les processus qu'elle implique, il doit disposer d'outils qui proposent un découpage clair des différentes composantes de la compétence (Falardeau, Guay, & Valois, 2014). Malheureusement, il n'existe pas, à notre connaissance, d'outil mis à la disposition des enseignants québécois leur permettant de réaliser une évaluation formative des élèves en lecture : « compréhension des éléments significatifs d'un ou de plusieurs textes; interprétation fondée d'un ou de plusieurs textes; justification pertinente des réactions à un ou plusieurs textes; jugement critique et fondé sur divers textes ou œuvres à partir de critères; recours à des stratégies de lecture appropriées » (Ministère de l'éducation, 2003, p. 101)

L'objectif général de cet article vise à montrer qu'il est possible et souhaitable d'évaluer les qualités métrologiques d'un outil d'évaluation formative de la compétence à lire et apprécier des textes variés en français au secondaire. Selon certains auteurs tel que De Ketele et Gerard (2005), c'est de faire fausse route que de prendre appui sur les théories psychométriques pour construire et valider les outils d'évaluation des apprentissages, notamment d'ordre formatif. Un tel commentaire ou une telle prise de position ne peut qu'étonner puisque finalement les principes et techniques mises au point par les spécialistes de la psychométrie ne tendent qu'à rencontrer les critères qui font qu'une mesure en est une. Si les théories de la

a- Cette étude a été réalisée grâce à une subvention du Fonds de recherche société et culture du Québec accordée aux trois premiers auteurs (chercheur principal : E. Falardeau). Une partie de l'article provient du mémoire de maîtrise de S. Ngo Boumkouo qui a été réalisé sous la supervision de P. Valois (directeur) et E. Falardeau (codirecteur).

1- Université Laval, Québec, Canada

2- Direction de la Statistique, Haut Commissariat au Plan, Rabat, Maroc

mesure sont ce qu'elles sont, cela tient à la nature même des construits en jeu ainsi qu'à la complexité de leur évaluation.

La mesure n'est pas une fin en soi, mais un moyen d'objectiver des jugements et d'assurer une plus grande précision entre les spécialistes. Le médecin qui inscrit au dossier d'un patient que sa température corporelle est de 41°C fournit à un utilisateur éventuel du dossier une information beaucoup plus objective et précise que de préciser qu'il semble faire de la fièvre. De la même façon, le professeur de français qui informe un parent que son enfant a obtenu un résultat de « x » ou « y » à une épreuve validée ou à l'une de ses composantes fournit une information beaucoup plus objective et précise que d'affirmer qu'il semble ou non éprouver des problèmes en français.

I. Méthodologie

1. Participants

L'échantillon est composé de 602 élèves qui proviennent du premier cycle du niveau secondaire (n = 279 pour le premier secondaire et n = 323 pour le deuxième secondaire) de quatre régions du Québec. Leur moyenne d'âge est de 13.1 ans (minimum = 12; maximum = 16).

2. L'outil d'évaluation formative

Il sera question, dans une première section, du choix du texte que les élèves devaient lire et apprécier. Une deuxième section sera consacrée aux différentes composantes de la compétence « Lire et apprécier des textes variés ». Par la suite, deux items ou manifestations d'une des composantes de cette compétence seront présentés (c'est pour des raisons éditoriales ayant trait à la longueur du texte que nous nous sommes limités à l'analyse de seulement deux des 12 items qui composent l'épreuve). Enfin, la méthode d'analyse d'items utilisée pour vérifier les qualités métrologiques de l'épreuve sera décrite.

2.1. Le choix du texte

Les élèves provenant du premier cycle du secondaire, il importait de choisir un texte qui répondait à certains critères propres à ce niveau académique. Premièrement, il devait leur prendre entre 60 et 75 minutes (la durée normale d'un cours) pour le lire et répondre aux questions ou items s'y rapportant. Deuxièmement, le niveau de difficulté du texte devait correspondre au niveau moyen de compétence en lecture des élèves du premier cycle du secondaire. Enfin, il devait être intéressant et instructif pour les élèves. Sur la base de ces critères, le texte de Gazaille (2002) a été retenu. C'est un texte courant qui présente des faits qu'il est possible de vérifier et il est écrit sur un ton neutre, l'opinion de l'auteur étant absent.

2.2. Les composantes de la compétence « Lire et apprécier des textes variés »

L'identification des composantes « Lire et apprécier des textes variés » s'est effectuée lors de la première année du projet. Avec la participation de deux assistantes de recherche, les chercheurs ont travaillé à l'aide du programme formation de l'école québécoise du français du 1er cycle du secondaire. Ils ont d'abord mis à plat toutes les stratégies de lecture identifiées dans le programme de même que les composantes et les sous-composantes de la compétence. Au total quatre composantes (C), comprenant au total 12 items (voir l'épreuve en annexe), ont été identifiées pour l'élaboration de l'outil d'évaluation de la compétence « Lire et apprécier des textes variés ». Les quatre composantes sont :

1. Je construis du sens (C1)
2. Je porte un jugement (C2)
3. Je réfléchis à ma pratique de lecteur (C3)
4. Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture (C4)

2.2.1. Je construis du sens (C1)

Pour maîtriser cette composante de la compétence, le lecteur doit réaliser quatre étapes correspondant aux sous-composantes planification, compréhension, interprétation et réaction à un texte.

2.2.1.1. Première sous-composante de C1 : Je planifie ma lecture

Pour planifier la lecture, le lecteur doit s'assurer qu'il a porté une attention particulière aux éléments paratextuels (première et quatrième de couverture, titre, sous-titres, illustrations, photographies, graphiques, schémas).

2.2.1.2. Deuxième sous-composante de C1 : Je comprends un texte

Pour cette composante, l'élève doit reconnaître les idées principales ou les aspects du sujet décrit ainsi qu'utiliser la stratégie liée à la compréhension des mots difficiles (en analysant la composition du mot, le contexte de la phrase, en trouvant un terme équivalent, en utilisant le dictionnaire, etc.).

2.2.1.3. Troisième sous-composante de C1 : J'interprète un texte

La sous-composante « J'interprète un texte » permet au lecteur de s'interroger sur l'intention de l'auteur. Le lecteur doit déterminer si le point de vue de l'auteur était objectif ou subjectif en s'appuyant sur des indices présents dans le texte. Pour interpréter, il doit porter une attention à la langue.

2.2.1.4. Quatrième sous-composante de C1 : Je réagis à un texte

Dans cette sous-composante, l'élève est amené à expliquer ses réactions (sentiments, réflexions, prise de conscience...) et à être attentif aux effets que le texte produisait sur lui en établissant des ressemblances ou des différences entre la vision du monde proposée dans le texte et sa vision personnelle du monde.

2.2.2. Deuxième composante (C2) : Je porte un jugement critique

La composante « je porte un jugement critique » invite le lecteur à prendre position par rapport au texte lu et en se basant sur des critères tels que la qualité de l'écriture, les caractéristiques des personnages, l'intérêt de l'histoire, la qualité de l'information, la présentation visuelle et le message transmis.

2.2.3. Troisième composante (C3) : Je réfléchis à ma pratique de lecteur

Dans cette situation, l'élève doit s'interroger sur les stratégies qu'il maîtrise pour planifier sa lecture ainsi que sur ce que le texte lui a appris.

2.2.4. Quatrième composante (C4) : Je mets à profit et j'acquiers des connaissances sur la langue, les textes et la culture

Pour l'évaluation de cette composante, le lecteur est appelé à se poser deux questions essentielles :

- Quelles connaissances sur la langue m'ont été nécessaires pour lire ce texte (p. ex. : les temps verbaux, les organisateurs textuels, les marqueurs de relation, la reprise de l'information, la ponctuation, les subordonnées, etc.) ?
- Quelles connaissances culturelles ai-je acquises en lisant ce texte (littéraires, artistiques, historiques, géographiques, scientifiques, etc.) ?

2.3. Les deux items analysés

Comme indiqué précédemment, les résultats de l'analyse métrologique ne seront présentés que pour deux des 12 items de l'épreuve. La première partie de l'item 1 permet d'évaluer si l'élève planifie sa lecture tandis que l'item 11 permet d'évaluer la qualité de la réaction d'un élève à un texte.

2.3.1. Item 1

Tu as une minute pour observer le texte. D'après toi, quelle en est l'idée générale ?

C'est une question à développement qui évalue la sous-composante « Je planifie ma lecture » et dont la correction prévoit cinq échelons où, par exemple, l'échelon 5 vaut cinq points et correspond à l'évaluation « maîtrise cette composante nettement au-delà des attentes ». L'ensemble des échelons est présenté ci-dessous :

1. Maîtrise peu ou pas cette composante selon les attentes;
2. Maîtrise cette composante en-deçà des attentes;

3. Maîtrise cette composante minimalement selon les attentes;
4. Maîtrise cette composante au-delà des attentes;
5. Maîtrise cette composante nettement au-delà des attentes.

Les critères suivants ont été utilisés pour procéder à l'évaluation des réponses des élèves :

- 5 : L'idée générale est très bien décrite, c'est-à-dire que l'élève résume les trois thèmes importants du texte (la publicité, l'adolescence et la surconsommation [consommation excessive]) et qu'il fait le lien entre eux en évoquant l'idée que la publicité exerce une influence négative sur les jeunes. Exemple de réponse : Il sera question de l'influence négative de la publicité sur les adolescents, qui sont amenés à consommer de façon exagérée.
- 4 : L'idée générale est bien décrite, c'est-à-dire que l'élève résume les trois thèmes importants du texte et qu'il fait le lien entre eux, mais il n'évoque pas l'idée de l'influence négative OU l'élève relie deux thèmes principaux par l'idée d'une influence négative. Exemple de réponse : Il sera question de l'influence de la publicité sur la consommation des jeunes, ou Les adolescents se font avoir par la publicité.
- 3 : L'idée générale est décrite à l'aide de mots-clés seulement, c'est-à-dire que l'élève identifie les trois thèmes importants, mais qu'il ne fait pas de lien entre eux. Exemple de réponse : Il sera question de l'adolescence, de la publicité et de la surconsommation.
- 2 : L'idée générale du texte est partiellement décrite, c'est-à-dire que l'élève n'évoque que deux thèmes importants et qu'il n'établit pas de lien entre eux. Exemple de réponse : La publicité et l'adolescence.
- 1 : L'idée générale est décrite minimalement, c'est-à-dire que l'élève n'évoque qu'un thème important. Exemple de réponse : L'adolescence.

2.3.2. Item 11

Selon toi, les compagnies manipulent-elles à ce point les adolescents par la publicité ? Justifie ta réponse en t'appuyant à la fois sur des éléments d'information du texte et sur ton expérience personnelle.

C'est une question à développement qui évalue la sous-composante « Je réagis à un texte » et dont la correction prévoit cinq échelons. La grille de correction est présentée ci-dessous :

- Pour justifier son opinion, l'élève s'appuie clairement sur un élément du texte et sur un élément de son expérience personnelle, et le lien entre les deux est approprié. Exemple : Oui, je crois que les adolescents sont manipulés. Les publicitaires font exprès de rendre les nouveaux produits attrayants en faisant toutes sortes de promesse, comme les cellulaires. Il est vrai qu'on s'en lasse rapidement. À mon école, tout le monde en a un parce que c'est la tendance. Pourtant, plusieurs élèves ne s'en servent même pas. (5 points)
- L'élève justifie son opinion, mais un des deux éléments de réponse ou le lien tracé entre eux est plus ou moins pertinent. Exemple : Oui, je crois que les adolescents sont manipulés par les compagnies, car elles leur font croire que leurs produits sont essentiels. Je connais quelqu'un qui est tombé dans le panneau. Il s'est acheté un iPod Touch. (4 points)
- Pour justifier son opinion, l'élève s'appuie de façon plus ou moins pertinente sur un élément du texte, mais pas sur son expérience personnelle. Exemple : Oui, je crois que les adolescents sont manipulés par les compagnies, car elles font de la publicité partout, dans les magazines, pendant les émissions de télévision, et même dans les écoles. (3 points)
- L'élève ne s'appuie pas sur un élément du texte pour justifier son opinion, même si l'élément en lien avec son expérience personnelle est pertinent. Exemple : Non, je ne crois pas que nous sommes manipulés. Les jeunes de ma classe n'ont pas tous le dernier iPod. (2 points)
- L'élève justifie son opinion à l'aide d'éléments textuels ou personnels non pertinents. Exemple : Oui, je crois que les adolescents sont manipulés par les compagnies, car les vendeuses sont énervantes avec leur vente à pression. (1 point)

2.4. Méthode d'analyse d'item

L'objectif de l'analyse d'item est de qualifier les items par rapport à un certain nombre de caractéristiques et de déterminer quels sont ceux qui méritent d'être immédiatement conservés dans l'outil d'évaluation formatif,

ceux qui exigeraient une certaine bonification et enfin ceux qui devraient être écartés complètement.

Essentiellement, cette analyse réfère à un ensemble d'évaluations quantitatives qui permettent de déterminer la valeur de chaque item au regard de certains critères. Elle implique généralement la production des résultats suivants :

- (a) les fréquences d'item pour identifier le nombre de données manquantes. C'est une information qui peut révéler la présence potentielle d'ambiguïtés, d'un niveau de difficulté élevé, mais qui peut également dépendre plus simplement de la position de l'item dans l'épreuve; les items en fin d'épreuve étant souvent moins répondus. Dans le cas des items à choix multiples, les fréquences informent également de l'attrait des leurres et permettent éventuellement de juger si ces derniers jouent effectivement leur rôle. Pour les questions ouvertes ou à développement, elles permettent d'examiner la répartition des scores partiels à l'item.
- (b) l'indice de facilité des items (p) : selon la théorie classique des tests. Notons au passage que le calcul de la variance des items aide à nuancer l'interprétation de cet indice de facilité. Il est difficile de déterminer des valeurs acceptables de niveau de facilité des items. Dans un contexte d'évaluation formative, lorsque les items ont été testés sur des échantillons représentatifs d'élèves, il semble difficile de retenir des items dont le niveau de facilité est inférieur à .30 (i.e. quand seulement 30 % des élèves réussissent une question). De la sorte et de façon relativement arbitraire, nous jugeons acceptables des niveaux de réussite compris entre .30 et 1.00.

On peut se référer à la légende suivante pour qualifier le niveau de facilité des items :

- extrêmement difficile : valeur « p » inférieure à 0.30;
- très difficile : valeur « p » comprise entre 0.30 et 0.40;
- plutôt difficile : valeur « p » comprise entre 0.41 et 0.55;
- plutôt facile : valeur « p » comprise entre 0.56 et 0.70;
- très facile : valeur de « p » comprise entre 0.71 et 0.85;
- extrêmement facile : valeur « p » supérieure à 0.85.

- (c) l'indice du degré de discrimination des items : selon la théorie classique des tests et la théorie des réponses aux items.

S'il est difficile et relativement arbitraire de fixer des critères pour juger de la difficulté d'un item, l'opération l'est encore davantage dans le cas de la discrimination pour diverses raisons dont le fait que ces deux caractéristiques (difficulté et discrimination) ne sont pas indépendantes entre elles. En théorie classique, par exemple, on estime la discrimination par la relation qu'entretient le rendement à l'item avec celui à l'ensemble de « l'épreuve ». Or l'indice de relation (la corrélation) est influencée en partie par la variance à l'item (donc par son niveau de difficulté).

Mais, au-delà de tout, un item doit de façon générale discriminer. En effet, si un item ne permet pas de différencier les élèves qui ont fait les apprentissages de ceux qui ne les ont pas réalisés, il est généralement mauvais. C'est pourquoi nous amorçons l'évaluation des items par l'examen de son pouvoir discriminant. Si un item ne présente pas un pouvoir discriminant modéré ou acceptable (0.65 selon la grille de Baker, 2001), il mérite donc d'être rejeté. Mais ce n'est pas parce qu'un item satisfait à ce critère qu'il se doit pour autant d'être conservé dans la banque. Encore faut-il que l'item affiche un niveau de difficulté à son tour acceptable selon les circonstances.

L'auteur Baker (2001, p. 34) suggère la table suivante pour interpréter ou juger des valeurs de la discrimination calculées selon la théorie des réponses aux items :

- 0.34 et moins : très faible;
- 0.35 - 0.64 : faible;
- 0.65 - 1.34 : modérée;
- 1.35 - 1.69 : bonne;
- 1.70 et + : très bonne.

Pour l'indice de discrimination calculé selon la théorie classique des tests, la table suivante est suggérée par Hopkins et Stanley (1981), les valeurs théoriques possibles allant de -1 à +1 :

- .10 et moins : très faible;
- .10 - .19 : faible;
- .20 - .29 : modérée;
- .30 - .39 : bonne;
- .40 et + : très bonne.

(d) l'analyse des courbes caractéristiques des items pour comparer les items entre eux pour ce qui est de leurs indices de difficulté et de discrimination.

(e) l'examen des courbes caractéristiques des options de réponses afin d'identifier les bons et mauvais leurres pour les questions à choix multiples (un complément formel aux indices mentionnés jusqu'ici qui permet d'améliorer le jugement global au sujet de l'item sans pour autant déboucher sur un indice quantitatif).

Afin de respecter la longueur maximum du texte permise, nous nous limiterons à présenter les résultats de deux items (sur les 12 que contient l'épreuve). Nous examinerons ces deux items à travers certaines caractéristiques de statistiques descriptives telles que l'indice de facilité (i.e. le score moyen et son équivalent rapporté sur une échelle sur 1 afin de faciliter la comparaison du degré de facilité des items entre eux) et l'indice de discrimination selon la théorie classique et selon la théorie de réponses aux items. Ces indices peuvent aider à identifier les items ambigus de même que ceux, bien que non ambigus, dont le contenu ne semble pas mesurer la compétence « lire et apprécier des textes variés ». Nous présentons également les courbes caractéristiques d'item ainsi que les courbes caractéristiques d'options de chacun des items de l'outil d'évaluation. Ces courbes, de par leur caractère visuel, facilitent l'interprétation des indices de facilité et de discrimination des items.

Les analyses statistiques ont été réalisées avec le logiciel libre eirt version 1.3.0 à l'aide de l'estimateur par maximum de vraisemblance marginalisée. Ce logiciel métrologique consiste en un module Excel qui permet d'obtenir l'estimation des paramètres d'items selon la théorie classique des tests (TCT) et de la théorie des réponses aux items (TRI) de même que les courbes caractéristiques d'item et les courbes d'options de réponses pour chaque item (Valois, Houssemand, Germain, & Abdous, 2012).

II. Résultats

Dans cette section, nous présentons et interprétons les résultats de l'analyse métrologique réalisée sur les deux items décrits précédemment. Pour chacune des questions, nous rappelons d'abord le libellé, suivi du tableau des fréquences, des indices de facilité et de discrimination, ainsi que des courbes caractéristiques d'item et celles d'options de réponses.

Item 1 : Tu as une minute pour observer le texte. D'après toi, quelle en est l'idée générale ?

Un examen du tableau des fréquences (voir le Tableau 1), nous montre que le taux de réussite est très bas. Seulement 3.7 % des répondants ont obtenu le score maximum (5), c'est-à-dire qu'ils ont été en mesure de répondre parfaitement à cet item. De plus, 68.4 % n'ont pu obtenir plus de deux points sur le score maximum de cinq. Par conséquent, seulement 17.2 % des élèves ont obtenu un score de 4 ou 5 points à cet item. La valeur de l'indice de facilité obtenue selon la théorie classique des tests (i.e. la cote moyenne des élèves à cet item) corrobore ce résultat (voir le Tableau 2). En effet, en ramenant à une échelle de 100, les résultats obtenus à cet item présentent une cote moyenne de 43 %. Si nous nous référons aux critères présentés précédemment pour qualifier le niveau de facilité des items, ce résultat indique que cet item présente un niveau de facilité qualifié de « Plutôt difficile ».

Tableau 1. Fréquences observées à l'item 1

Score	Fréquence	Pourcentage
1	221	36.7
2	191	31.7
3	87	14.5
4	81	13.5
5	22	3.7
Total	602	100.0

Tableau 2. Indices métrologiques de l'item 1

Item	Min ^b	Max ^c	Théorie classique			TRI ^a
			Moy ^d	<i>p</i> ^e	Disc ^f	Disc
1	1	5	2.16	0.43	0.30	0.82

^a TRI signifie « théorie des réponses aux items »

^b Score minimum à l'item

^c Score maximum à l'item

^d la valeur « Moy » représente la moyenne à cette item sur le score maximum

^e la valeur « *p* » représente le degré de facilité exprimé sur 1

^f la valeur « *Disc* » représente le degré de discrimination de l'item

En dépit du caractère « Plutôt difficile » de cet item, il présente néanmoins un pouvoir discriminant qualifié de « Bon » selon la théorie classique des tests (TCT), avec un indice de discrimination de 0.30, et qualifié de « Modéré » selon la théorie des réponses aux items (TRI), avec un indice de discrimination de 0.82.

L'observation de la courbe caractéristique d'item confirme ce constat (voir la figure 1). En effet, la courbe caractéristique de l'item respecte le profil recherché (fonction logistique), ce qui indique que les répondants ayant fait le plus d'apprentissage ont les scores les plus élevés à cet item. De plus, l'examen des courbes caractéristiques des options de réponse (voir la figure 2) indiquent que le score obtenu à l'item augmente en fonction du degré d'apprentissage en lecture, bien que les résultats montrent la difficulté pour l'évaluateur de faire la différence entre les degrés de compétence 2 (maîtrise cette compétence en-deçà des attentes) et 3 (maîtrise cette composante minimalement selon les attentes) et entre les niveaux de compétence 3 et 4 (maîtrise cette composante au-delà des attentes). Au regard de tous ces éléments, il apparaît que cet item présente un pouvoir discriminant tout au moins modéré et un niveau de facilité acceptable. Pour ces raisons, il mérite d'être conservé dans l'outil. Toutefois, il y aurait lieu de modifier les critères 2, 3 ou 4 de façon à mieux différencier les niveaux de compétence recherchés ou à fusionner ces niveaux, ce qui ferait passer l'échelle d'évaluation de 5 à 4 critères.

Rappelons que les critères 2, 3 et 4 suivants ont été utilisés pour procéder à l'évaluation des réponses des élèves à cet item.

4 : L'idée générale est bien décrite, c'est-à-dire que l'élève résume les trois thèmes importants du texte et qu'il fait le lien entre eux, mais il n'évoque pas l'idée de l'influence négative OU l'élève relie deux thèmes principaux par l'idée d'une influence négative. Exemple de réponse : Il sera question de l'influence de la publicité sur la consommation des jeunes, ou Les adolescents se font avoir par la publicité.

3 : L'idée générale est décrite à l'aide de mots-clés seulement, c'est-à-dire que l'élève identifie les trois thèmes importants, mais qu'il ne fait pas de lien entre eux. Exemple de réponse : Il sera question de l'adolescence, de la publicité et de la surconsommation.

2 : L'idée générale du texte est partiellement décrite, c'est-à-dire que l'élève n'évoque que deux thèmes importants et qu'il n'établit pas de lien entre eux. Exemple de réponse : La publicité et l'adolescence.

A la lecture de ces critères, nous serions enclins à fusionner les critères 2 et 3 puisqu'il ne semble pas que les deux compétences décrites soient suffisamment distinctes.

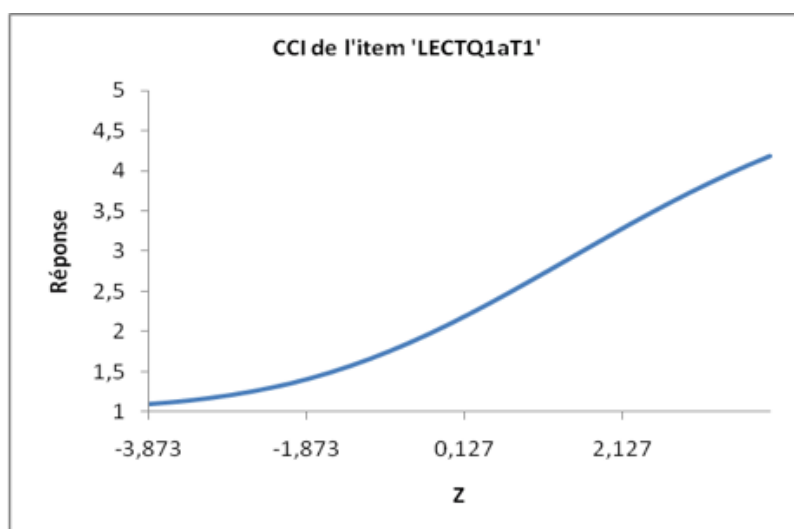


Figure 1. Courbe caractéristique de l'item 1

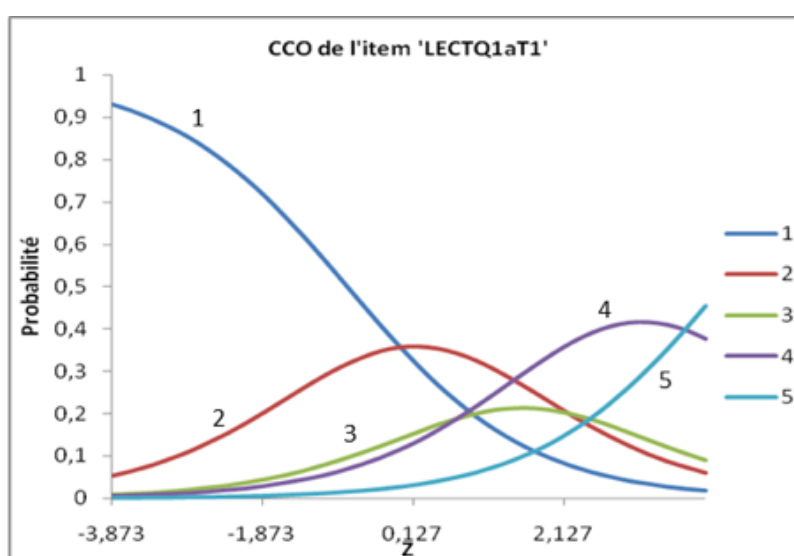


Figure 2. Courbe caractéristiques des options de réponse de l'item 1

Item 11 : Selon toi, les compagnies manipulent-elles à ce point les adolescents par la publicité ? Justifie ta réponse en t'appuyant à la fois sur des éléments d'information du texte et sur ton expérience personnelle.

Le taux de réussite à cet item est de 15,4 % (voir le tableau 3). Par ailleurs 45,8 % des sujets ont pu obtenir les scores intermédiaires de 3 et 4 points ce qui nous amène à conclure qu'il s'agit d'une question « Plutôt facile » (moyenne de 0,59; voir le tableau 4).

Tableau 3. Fréquences observées à l'item 11

Score	Fréquence	Pourcentage
1	102	16,9
2	132	21,9
3	159	26,4
4	117	19,4
5	92	15,4
Total	602	100,0

Tableau 4. Indices métrologiques de l'item 11

Item	Min ^b	Max ^c	Théorie classique			TRI ^a
			Moy ^d	p ^e	Disc ^f	Disc
1	1	5	2.94	0.59	0.38	1.16

^a TRI signifie « théorie des réponses aux items »

^b Score minimum à l'item

^c Score maximum à l'item

^d la valeur « Moy » représente la moyenne à cette item sur le score maximum

^e la valeur « p » représente le degré de facilité exprimé sur 1

^f la valeur « Disc » représente le degré de discrimination de l'item

L'observation de la courbe caractéristique (voir la figure 3) révèle que cette dernière respecte le profil recherché, ce qui indique que les élèves ayant fait le plus d'apprentissage ont les scores les plus élevés à cet item. De plus l'examen de la courbe caractéristique d'options (voir la figure 4) démontre que la grille d'évaluation est adéquate. L'indice de discrimination de cet item est qualifié de « Modéré » selon la TRI (1.16) et de « Bon » selon la théorie classique des tests (0.38) (voir le Tableau 2). Au regard de tous ces constats, nous proposons de conserver cet item dans l'outil d'évaluation.

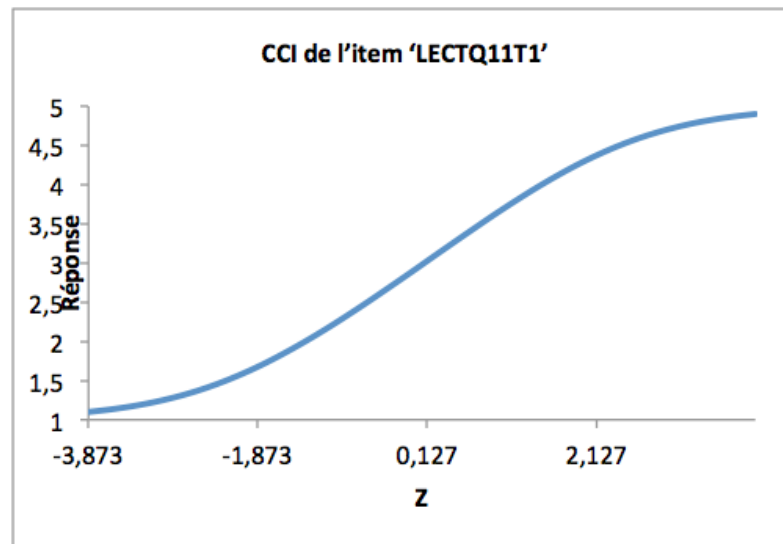


Figure 3 Courbe caractéristique de l'item 11

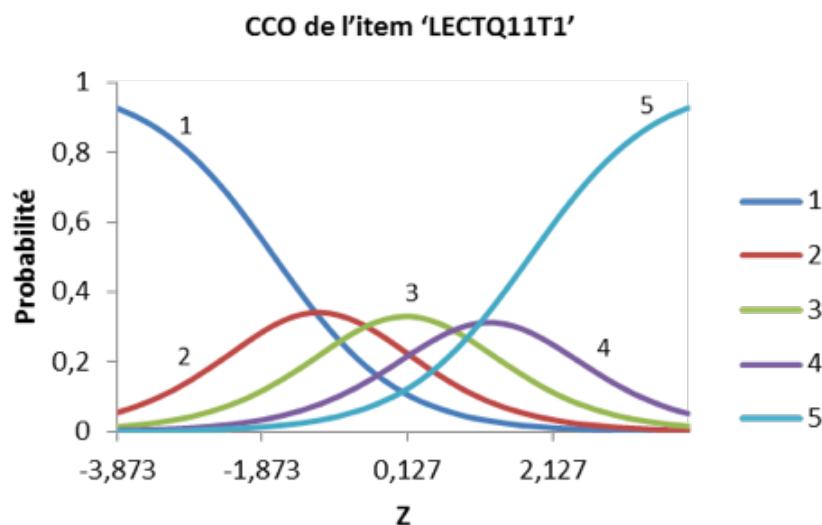


Figure 4. Courbe caractéristique d'options de l'item 11

Conclusion

L'objectif général de cette étude visait à montrer qu'il est possible et même souhaitable d'évaluer les qualités métrologiques d'un outil d'évaluation formative de la compétence à lire et apprécier des textes variés en français au secondaire. Il ressort de ces analyses que la théorie classique des tests et la théorie de réponses aux items sont très utiles pour détecter les forces et les faiblesses des questions d'une épreuve d'évaluation formative en français.

Les résultats de telles analyses psychométriques donnent aux concepteurs de tests des informations qui peuvent le guider dans la reformulation des items et des choix de réponses, pour les rendre plus clairs. En bref, il serait dommage de se priver des théories de la mesure et d'un logiciel d'analyse d'items convivial tel que eirt pour mettre à l'épreuve un outil d'évaluation - qu'il soit d'ordre normatif ou formatif - avant de l'administrer aux élèves. Il s'agit d'une question d'équité pour ces derniers.

Traduction du Résumé

Abstract Translation

Pierre Valois, Erick Falardeau, Frédéric Guay,
Belkacem Abdous, Suzanne Ngo Boumkouo

FORMATIVE EVALUATION IN FRENCH : A METROLOGICAL ANALYSIS

Surveys dedicated to assess educational achievement are important given the role they play in the evaluation of the output and quality of education systems. Instruments used in this context include tests and questionnaires. The first ones aim to measure students' performance in a particular subject when the second ones provide information on individual, family and school characteristics which can be used to explain the differences in school performances between students. Results obtained from such studies will help provide guidance to decision makers to improve the quality of education. However, the relevance of the actions taken depends on the capacity of instruments to reflect reality. This refers to important concepts, reliability and validity, that must be taken into consideration when constructing a measuring instrument. In the case of students' achievement, the validation of tests is a crucial step. It goes through item analysis using two approaches : classical approach and item response theory. The aim of this paper is to present the theoretical underpinnings of these approaches and their use in the context of National Assessment Program of Student Achievement (PNEA).

Bibliographie et Références

Baker, F. B. (2001). The basics of item response theory. College Park, MD: ERIC Clearinghouse on assessment and evaluation. Retrieved from <http://echo.edres.org:8080/irt/baker/>

De Ketele, J.-M., & Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Revue Mesure et Evaluation en Education*, 28, 1-26.

Falardeau, É., Guay, F., & Valois, P. (2014). L'impact d'un enseignement explicite des stratégies de lecture et de l'autorégulation sur les capacités d'Élèves du secondaire. Dans B. Daunay & J.-L. Dufays (Eds.). *Didactique du français : du côté des élèves* (p. 139-156). Bruxelles : De Boeck.

Gazaille, M.-P. (2002). Adolescents, publicité et surconsommation. *Parents Ados*, 3(3), 5-6.

Hopkins, K. D., & Stanley, J. C. (1981). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.

Ministère de l'éducation. (2003). *Programme de formation de l'école québécoise, enseignement secondaire, premier cycle*. Québec : Gouvernement du Québec.

Valois, P., Houssemand, C., Germain, S. and Abdous, B. (2011), An open source tool to verify the psychometric properties of an evaluation instrument, *Procedia-Social and Behavioral Sciences*, 15, 552-56.

Annexe

Questionnaire

Nom : _____

Date : _____

Adolescents, publicité et surconsommation

1- a) Tu as une minute pour observer le texte.

D'après toi, quelle en est l'idée générale?

b) Explique comment tu t'y es pris pour prédire l'idée générale.

2- Si tu avais à choisir une seule phrase du texte qui représenterait bien le message de l'auteure, laquelle choisirais-tu? Encerle la bonne réponse.

a) « Besoin d'appartenance à un groupe d'amis, identification à ses vedettes favorites, dissociation des parents par divers moyens, tout un lot d'actions posées dans le but de se découvrir et de se diriger lentement mais sûrement vers l'âge adulte et de se découvrir soi-même, de se définir en tant qu'individu à part entière. »

b) « En effet, à l'instant précis où [les adolescents] sont à la recherche de leurs valeurs et où leurs goûts changent, la publicité leur présente des produits et des services branchés qui, selon le message véhiculé, leur sont non seulement essentiels, mais les rendront également heureux. »

c) « Sachant que les enfants, jusqu'à l'âge de 5 ans, ne savent pas distinguer le contenu d'une émission des pauses publicitaires, il n'y a rien d'étonnant à constater qu'une fois adolescents, les jeunes sont déjà habitués à apprécier la pub sous toutes ses formes. »

d) « Ainsi programmés pour consommer depuis le berceau, les enfants seront déjà acclimatés à cette publicité constante et seront donc un excellent public dans quelques années, lorsqu'ils seront devenus des consommateurs indépendants... avec un pouvoir d'achat individuel entier! »

3- Que signifie l'adjectif « adulées » à la fin du premier paragraphe? Tu dois encercler une seule réponse.

a) connues

b) admirées

c) boudées

d) appréciées

4- Que signifie la phrase suivante? Encerle la bonne réponse.

« Les publicitaires ont depuis longtemps saisi que les adolescents constituent, en raison de cette recherche d'identité individuelle et de leur besoin d'appartenance et d'acceptation, un groupe plus vulnérable et, par le fait même, une cible de choix. »

a) Les publicitaires ont compris qu'il est facile d'atteindre les adolescents par la publicité, puisqu'ils cherchent qui ils sont et qu'ils ont besoin d'appartenir à un groupe.

b) Les publicitaires savent que les adolescents, qui cherchent constamment à changer d'identité et à combler leurs besoins, peuvent être attaqués facilement.

c) Les adolescents sont saisis par les publicitaires qui veulent les faire changer d'identité, et ils acceptent d'être des cibles, car ils constituent un groupe impuissant.

d) Les publicitaires savent qu'il est plus facile de viser les adolescents, qui sont un groupe fragile.

5- Dans le chapeau du texte, à quelle information réfère le groupe du nom « ces dictats »?

6- Par quel mot ou quelle expression pourrais-tu remplacer le marqueur de relation « vu » dans le dernier paragraphe?

7- Dans cette phrase du deuxième paragraphe, pourquoi l'auteure utilise-t-elle le point d'exclamation?

« Des études effectuées auprès de jeunes enfants confirment que les enfants peuvent développer des images mentales en lien avec les logos ou mascottes de certaines lignes de produits dès l'âge de... 6 mois! »

8- Selon toi, qu'a voulu démontrer l'auteure en insérant un discours rapporté dans le deuxième paragraphe?

9- Le point de vue de l'auteure est-il objectif?

a) Oui, car l'auteure apporte des statistiques pour appuyer son idée.

b) Oui, car l'auteure a écrit un texte informatif, publié dans une revue.

c) Non, car l'auteure ne présente qu'un côté de la publicité.

d) Non, car l'auteure utilise des points d'exclamation.

10-À la lecture de ce texte, qu'as-tu appris concernant la publicité et son influence sur les jeunes?

11- Selon toi, les compagnies manipulent-elles à ce point les adolescents par la publicité? Justifie ta réponse en t'appuyant à la fois sur des éléments d'information du texte et sur ton expérience personnelle.

- 12-** Si un ami désirait s'informer sur le phénomène de la consommation, lui proposerais-tu de lire ce texte? Quelle que soit ta réponse, tu dois la justifier à l'aide de deux critères. Donne un exemple pour chacun de tes critères.

Résumé

Les enquêtes portant sur l'évaluation des acquis scolaires revêtent une importance particulière eu égard aux informations qu'elles fournissent concernant le rendement et la qualité des systèmes éducatifs. Les instruments de mesure utilisés dans ce cadre consistent en des tests évaluant les élèves dans différentes matières ainsi que des questionnaires informant sur leurs caractéristiques individuelles, familiales et scolaires. Afin de mesurer, le plus fidèlement possible, le niveau des acquis des élèves ainsi que de refléter le contexte réel dans lequel ils apprennent, ces instruments doivent être fiables. Partant, s'assurer de la validité des tests à travers l'analyse des items s'avère une étape cruciale dans le processus de validation de ces instruments. Pour mener cette analyse, deux approches sont généralement adoptées. Il s'agit de l'approche classique et de l'approche se basant sur la théorie des réponses aux items. Cette communication présente les sous-jacents théoriques de ces approches et leur utilisation dans le cadre du Programme National d'Évaluation des Acquis (PNEA)¹.

I. Pourquoi une analyse des items ?

L'analyse des items est motivée par des préoccupations ayant trait à la fidélité des instruments de mesure. Dans le domaine d'évaluation des acquis scolaires, un test doit refléter, de la manière la plus fidèle possible, le niveau réel des acquis. En d'autres termes, les items constituant le test doivent tous contribuer à la mesure de la compétence ou l'habileté visée. Les items doivent également être informatifs et permettre de discriminer entre les individus selon leur niveau d'habileté.

Dans ce cadre, l'analyse des items constitue une étape primordiale dans le processus de validation de l'instrument de mesure. Cette analyse consiste à examiner les propriétés métrologiques du test et des items qui le composent en vue d'en améliorer la qualité. Elle permet, en effet, de détecter les items qui présentent des lacunes, c'est à dire ceux qui n'affichent pas le même comportement que les autres items ou qui ne leurs sont pas associés, mais également ceux qui leurs sont négativement corrélés. Une fois identifiés, ces items doivent être reformulés voire remplacés. Bien évidemment, cette opération doit être réalisée avant l'étude principale, et ce en menant au préalable une enquête expérimentale. Pour ce faire, on administre l'instrument à un échantillon d'individus dont les caractéristiques sont similaires à celles de la population ciblée. Les résultats obtenus sont utilisés pour calculer les différents indicateurs et indices psychométriques du test et des items.

L'analyse des items s'appuie sur des modèles mathématiques découlant de différentes théories de mesure. Ces modèles servent également à estimer l'habileté des individus en leur attribuant des scores permettant de les classer sur l'échelle de cette habileté.

¹ - Le Programme National d'Évaluation des Acquis (PNEA) est un programme ayant pour objectif de mesurer le niveau des acquis des élèves marocains par rapport au programme prescrit. La première étude réalisée dans ce cadre a été conduite en 2008. Elle a porté sur la 4^{ème} et la 6^{ème} années du primaire ainsi que la 2^{ème} et la 3^{ème} années du collège. Les élèves ont été évalués en mathématiques, en sciences, en arabe et en français. Une deuxième étude ayant pour cible les élèves du tronc commun du secondaire qualifiant sera menée en Mai 2016. Les matières concernées par cette évaluation sont: l'arabe, le français, les mathématiques, les sciences et l'histoire-géographie.

* Instance Nationale d'Évaluation - Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique.

II. Quels sont les modèles utilisés ?

Généralement, deux approches peuvent être adoptées pour conduire une analyse des items. Il s'agit de l'approche dite classique et l'approche reposant sur la théorie des réponses aux items. Dans ce qui suit, nous présentons les fondements théoriques des deux approches, leurs avantages et leurs limites.

1. L'approche classique

L'un des modèles reposant sur cette approche est le modèle classique (Richard Bertrand et Jean-Guy Blais, 2004). Selon ce modèle, le score obtenu par un individu dans un test, noté X , est la résultante de deux composantes non observées, à savoir le score vrai, V , et une erreur de mesure, E . Il se traduit par l'équation suivante:

$$X = V + E \quad (I)$$

Dans cette équation, le score vrai est une constante alors que l'erreur de mesure est une variable aléatoire. Il en découle que le score observé est aussi une variable aléatoire.

Le modèle classique se base sur les postulats suivants:

- L'erreur de mesure se distribue normalement avec une moyenne nulle : $\mu_E=0$. Il s'en suit que le score observé se distribue normalement autour du score vrai avec une espérance nulle : $\xi_X=V$;
- La corrélation entre l'erreur de mesure et le score vrai est nulle : $\rho_{EV}=0$. Ainsi, la variance des scores observés est égale à la somme de la variance des scores vrais et de la variance des erreurs de mesure : $\sigma_X^2=\sigma_V^2+\sigma_E^2$;
- La corrélation entre les erreurs de deux tests différents, $E1$ et $E2$, est nulle : $\rho_{E1E2}=0$;
- Deux tests sont parallèles si leurs scores vrais, $V1$ et $V2$, sont égaux et les variances de leurs erreurs de mesure sont égales : $V1=V2$ et $\sigma_{E1}^2 = \sigma_{E2}^2$. Dans ce cas, les scores observés de ces deux tests ont la même moyenne et la même variance;
- La corrélation entre les erreurs d'un test ($E1$) et le score vrai d'un autre test ($V2$) est nulle : $\rho_{E1V2}=0$;
- Deux tests sont dits τ -équivalents si leurs scores vrais diffèrent par une constante : $V1=V2+k$.

L'analyse d'items selon l'approche classique permet d'examiner la qualité de l'instrument de mesure au niveau global et au niveau local (Pierre Vriegnaud, 2006). Au niveau global, est évaluée la fidélité du test, alors qu'au niveau local, ce sont les caractéristiques métrologiques des items, en l'occurrence la difficulté et la discrimination, qui sont soumises à l'examen.

Fidélité

La fidélité renvoie à la précision avec laquelle un instrument mesure l'aptitude ou l'habileté, que l'on cherche à estimer. Dans le cadre de la théorie classique, elle est définie comme étant la corrélation entre le score observé et le score vrai. Cette définition, traduit une relation étroite entre l'erreur de mesure et la notion de fidélité. En effet, plus forte est cette corrélation, plus réduites sont les erreurs de mesure et meilleure sera la fidélité. Cette dernière, peut aussi être vue comme étant la proportion de la variance des scores observés qui est attribuable aux scores vrais: $\rho_{XV} = \sigma_V^2 / \sigma_X^2$.

Pour estimer la fidélité, trois grandes méthodes sont utilisées. Il s'agit des méthodes basées sur le test-retest, les méthodes mixtes, appelées formes parallèles et les méthodes de consistance interne.

Fidélité test-retest : comme son nom l'indique, la fidélité test-retest permet d'évaluer la stabilité du construit dans le temps. Elle consiste à administrer l'instrument en deux temps et à calculer la corrélation entre les deux mesures obtenues. Cependant, cette méthode reste relativement coûteuse car elle nécessite deux administrations.

Fidélité entre formes parallèles ou équivalentes : indique à quel point deux ensembles différents d'items peuvent mesurer le même concept. Pour ce faire, il faut construire deux formes différentes, en termes d'items qui les composent, mais qui soient équivalentes. Ensuite on administre ces deux formes au même échantillon d'individus et on calcule le coefficient de corrélation entre les scores des deux tests. Cette méthode est dite mixte car elle peut combiner les deux méthodes précédentes. En effet, si les deux formes sont administrées à deux moments différents, on mesure à la fois la consistance interne et la stabilité. Néanmoins, la construction de deux tests parallèles est une opération complexe et coûteuse. De plus, le coefficient de corrélation entre les deux tests peut être biaisé dans la mesure où les conditions de passations

de ces derniers peuvent différer.

Consistance interne : permet d'évaluer jusqu'à quel point chacun des items mesure la même dimension. En ce sens, elle reflète le degré d'homogénéité des items constituant un instrument de mesure. Contrairement aux deux méthodes précédentes, celle-ci ne nécessite qu'un seul test, administré une seule fois. Partant, elle est moins onéreuse et plus simple à mettre en œuvre, permettant ainsi d'obvier aux contraintes posées par les deux premières. Pour ces raisons, elle reste la méthode la plus utilisée.

En pratique, la consistance interne est mesurée par l'un des trois indicateurs suivants : la corrélation des items entre eux, la corrélation entre chaque item et le total des scores ou bien la corrélation entre les scores de deux groupes d'items contenant le même nombre d'éléments.

Cette dernière méthode est appelée méthode de bissection (*split-half* en anglais). Elle consiste à découper au hasard le test en deux moitiés équivalentes qui seront administrées à un même échantillon. Pour chaque individu, on obtiendra donc deux scores qui serviront au calcul du coefficient de corrélation.

Une alternative à cette approche, consiste à diviser le test, non en deux parties, mais plutôt en autant de parties que d'items². La moyenne des corrélations entre toutes les paires d'items constitue ce que l'on appelle l'Alpha de Cronbach. C'est cet indice que l'on utilise le plus souvent pour évaluer la consistance interne d'un instrument. Compris entre 0 et 1, plus il est élevé, plus l'instrument de mesure est fidèle au sens de la consistance interne. Toutefois, la valeur de l'Alpha de Cronbach est étroitement liée au nombre d'items composant l'instrument dans la mesure où sa valeur augmente avec l'augmentation de ce nombre.

Indice de difficulté

En matière d'évaluation des acquis scolaires, l'indice de difficulté d'un item est le pourcentage des élèves qui réussissent l'item. Selon cette définition, plus cet indice augmente, plus l'item est considéré facile. Un item présentant un taux de réussite très bas (item très difficile) ou très élevé (item très facile) est peu informatif et doit donc être amélioré ou supprimé. En pratique, l'indice de difficulté doit être compris entre 10% et 85%.

Indice de discrimination

L'indice de discrimination d'un item, appelé également "r-bis point", est la corrélation entre les réponses à cet item et le score total. Dans le cas des instruments mesurant le niveau d'habileté d'un individu, dont notamment les tests d'évaluation des acquis scolaires, un item n'est gardé que s'il a un pouvoir discriminant élevé, c'est à dire lorsqu'il a une grande capacité à discriminer les élèves selon leurs niveaux d'habileté. Ainsi, les items présentant de bonnes propriétés doivent avoir une corrélation modérément positif, indiquant que les élèves les plus doués réussissent l'item et que ceux l'ayant moins échouent l'item.

Le modèle classique, quoique simple, suppose que l'erreur de mesure n'est pas différenciable. Or, les erreurs peuvent avoir plusieurs sources. Elles peuvent être dues aux correcteurs, aux individus ou bien aux conditions de passation. Pour tenir de cet aspect, la théorie de la généralisabilité, qui est une extension de la théorie classique, permet de différencier les sources d'erreur. Ainsi, selon cette théorie, le score observé est la somme entre le score vrai et les différentes erreurs de mesure.

La théorie classique repose sur des conditions d'application relativement souples (Frenette et al., 2007). Elle nécessite des techniques mathématiques simples et des échantillons de tailles modérées (Hambleton et Jones, 1993). Il n'en demeure pas moins qu'elle présente certaines limites dont on cite principalement celle appelée dépendance circulaire (Frenette et al., 2007). Cette dépendance fait que les paramètres d'items dépendent de l'échantillon des répondants et que le paramètre d'habileté de ces derniers dépend de l'échantillon d'items. Autrement dit, lorsqu'on change l'échantillon des individus, la difficulté et la discrimination des items prennent d'autres valeurs. De même, lorsqu'on change l'échantillon des items, les résultats obtenus par les individus changent.

L'indépendance entre les échantillons et les paramètres estimés est une propriété importante, dite d'invariance, qui n'est pas vérifiée dans le cas des modèles de la théorie classique. Les modèles de la TRI ont l'avantage de posséder une telle propriété. En effet, ils permettent d'obtenir des paramètres d'items qui sont indépendants de l'échantillon des individus et une estimation de l'habileté qui est indépendante de l'échantillon d'items (Frenette et al., 2007).

2 - Chaque item constitue une partie

2. Théorie des réponses aux items (TRI)

La TRI, comme toute théorie de mesure, a principalement deux objectifs:

- Construire un instrument de mesure qui permet d'appréhender une habileté donnée et d'estimer ses propriétés telles que la précision et la validité;
- Attribuer aux sujets auxquels on a administré l'instrument des scores qui reflètent au mieux l'habileté visée et les classer sur l'échelle de cette habileté.

Les modèles reposant sur cette théorie sont des modèles probabilistes dans lesquels la probabilité qu'un individu réponde correctement à un item est une fonction de ses caractéristiques et celles de l'item. Cette fonction est appelée fonction caractéristique et est traduite par l'équation ci-après:

$$P(j=k|\theta, \xi) = F(\theta, \xi, k) \quad (2)$$

P : est la probabilité qu'un individu donne la réponse k à l'item j ;

θ : est une caractéristique de l'individu;

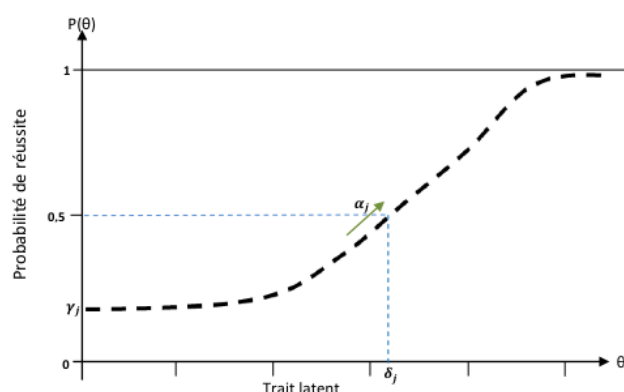
ξ : représente les caractéristiques de l'item.

Le paramètre θ est appelé trait latent. Il représente une habileté, une compétence ou une aptitude non observable et que l'on cherche à estimer à travers les modèles de la TRI. Dans le domaine d'évaluation des acquis, θ est interprété comme étant le niveau de compétence des élèves dans une matière donnée. N'étant pas directement observable, ce niveau est estimé à partir des scores obtenus dans un test.

Les caractéristiques de l'item, notés ξ , sont à estimer, simultanément, avec θ . En général, ils sont au nombre de trois: la difficulté, la discrimination et la pseudo-chance.

La fonction caractéristique est représentée graphiquement par une courbe dite courbe caractéristique de l'item (Figure 1).

Figure 1: Courbe caractéristique de l'item j



Cette courbe est croissante et monotone, montrant ainsi que plus l'habileté de l'individu, présentée par le trait latent θ , est élevée, plus sa probabilité de répondre à l'item augmente.

La fonction caractéristique donne lieu à trois modèles, selon qu'on considère un, deux ou trois paramètres.

Modèle à un paramètre

Ce modèle est présenté par la fonction ci-dessous:

$$P_j(\theta) = \frac{1}{1 + e^{-D(\theta - \delta_j)}}$$

$P_j(\theta)$: probabilité qu'un individu ayant la caractéristique θ réponde correctement à l'item j ;

δ_j : paramètre de difficulté de l'item j ;

D : une constante dont on attribue, dans la pratique, la valeur 1.7 pour que la courbe caractéristique ait une allure très proche de l'ogive normale. Lorsque cette constante prend la valeur 1, le modèle est appelé modèle de Rash.

Le paramètre de difficulté δ_j représente, comme c'est illustré dans la figure 1, la valeur de θ qui correspond à une probabilité de réussite égale à 0.5. Contrairement aux indice de discrimination et de pseudo-chance, la difficulté est présente dans tous les modèles de la TRI.

Modèle à deux paramètres

Un modèle qui est plus complexe que son prédécesseur, est le modèle à deux paramètres. En plus de la difficulté, ce modèle fait intervenir le paramètre de discrimination α_j . La fonction de probabilité s'écrit ainsi:

$$P_j(\theta) = \frac{1}{1 + e^{-D\alpha_j(\theta - \delta_j)}}$$

α_j paramètre mesure, comme dans le cas de la théorie classique, le pouvoir discriminant de l'item, c'est-à-dire sa capacité à distinguer les individus selon leur habileté. Dans le contexte de la TRI, la discrimination est exprimée par la pente maximale de la courbe caractéristique (voir figure 1).

Modèle à trois paramètres

Dans le cas des items à choix multiples, l'individu a une probabilité non nulle de choisir la bonne réponse, non pas parce qu'il la connaît, mais par le simple fait du hasard. Dans la TRI, ce phénomène est appelé la pseudo chance et les modèles qui en tiennent compte sont dits les modèles à trois paramètres:

$$P_j(\theta) = \frac{1 - \gamma_j}{1 + e^{-D\alpha_j(\theta - \delta_j)}}$$

Avec γ_j est le paramètre de pseudo chance qui représente la probabilité de réussite pour les faibles niveaux de compétence. Dans les deux premiers modèles, ce paramètre est égale à 0.

Hypothèses

L'utilisation des modèles de la TRI est soumise à certaines conditions qui sont indispensables à leur application. Il s'agit de trois hypothèses fondamentales qui doivent être vérifiées empiriquement:

- **Monotonocité**

Selon cette hypothèse, la courbe caractéristique d'item doit être une fonction monotone croissante. Cela revient à supposer que la probabilité de trouver la bonne réponse augmente avec le niveau de l'habileté de l'individu.

- **Unidimensionnalité**

L'unidimensionnalité est une condition essentielle des modèles qui se basent sur la TRI. Elle suppose que tout les items composant un instrument contribuent à mesurer chez les individus un seul et unique trait latent ou "dimension". Dans la pratique, il est difficile de trouver un cas où l'on a vraiment une dimension unique d'où l'utilisation d'une définition plus large de l'unidimensionnalité et qui se base sur la notion de dimension dominante.

- **Indépendance locale**

Cette hypothèse stipule que la réponse à un item n'est pas influencée par les réponses aux autres items. En d'autres termes, la corrélation entre les résultats des individus à deux items quelconques doit être statistiquement nulle ou proche de zéro. Il est à souligner qu'il y a une relation entre l'indépendance locale et l'unidimensionnalité dans la mesure où la vérification de l'hypothèse d'unidimensionnalité implique nécessairement celle de l'indépendance locale.

Les modèles reposant sur la TRI sont, comme en témoigne cette brève présentation, plus complexes comparativement à ceux relevant de la théorie classique. De surcroît, ils sont plus nécessaires en termes de taille d'échantillon, surtout lorsqu'il s'agit des modèles à plusieurs paramètres. Néanmoins, ils ont l'avantage de séparer ce qui relève des élèves de ce qui relève des items (Rocher, 2015). Qui plus est, ils possèdent la propriété d'invariance évoquée plus haut. Grâce à cette dernière, lorsqu'un modèle est bien ajusté à un échantillon donné, les paramètres d'items qui en découlent restent fixes. Partant, ils peuvent être appliqués à d'autres échantillons permettant ainsi de déduire les paramètres relatifs aux individus, en l'occurrence leur niveau d'habileté ou de compétence.

III. Analyse des items dans le cadre du PNEA

La TRI, malgré sa complexité, gagne de plus en plus du terrain. Elle a une portée plus étendue surtout au niveau de la modélisation des scores et de la création des échelles et des indices. Elle est également d'une grande utilité lorsqu'il s'agit d'ajustement des métriques ou de parallélisation des épreuves¹.

Plus facile à mettre en œuvre et reposant sur des hypothèses plus simples, l'approche classique constitue une étape primordiale dans l'analyse des items. Elle est d'autant plus utile dans la phase d'expérimentation de l'instrument. En effet, la taille des échantillons servant à tester ce dernier est souvent petite, ce qui fait obstacle à l'utilisation des modèles de la TRI à plusieurs paramètres.

Dans les enquêtes, à caractère national ou international², portant sur l'évaluation des acquis scolaires les deux approches sont utilisées de manière complémentaire. Les indices classiques des items³ sont toujours rapportés comme indicateurs descriptifs permettant de juger de la fidélité du test et de la qualité psychométrique des items qui le composent. Au niveau de la modélisation des scores et de la création des échelles, ce sont les modèles de la TRI qui sont adoptés à ce stade d'analyse.

Dans le cadre du programme national d'évaluation des acquis (PNEA 2016), visant l'évaluation des élèves marocains du tronc commun du secondaire qualifiant dans six matières⁴, une expérimentation des tests a été réalisée en vue d'en étudier les propriétés psychométriques. A cette étape, l'analyse des items a consisté en l'examen des éléments ci-dessous :

- **Fidélité du test** : cet aspect a été appréhendé à travers la consistance interne et plus précisément en utilisant l'Alpha de Cronbach. Nous avons pris comme seuil minimum 0.80, à partir duquel nous jugeons que le test est fidèle. Nous avons également évalué la contribution de chaque item à la fidélité globale du test et ce, en calculant l'indice Alpha(i) qui n'est que l'Alpha de Cronbach lorsque l'on supprime l'item (i) du test.
- **Difficulté des items** : c'est l'indice de difficulté tel qu'il est présenté dans la théorie classique et qui est le pourcentage des élèves réussissant l'item. Ainsi, nous considérons qu'un item est trop facile lorsque cet indice est supérieur à 0.85 et trop difficile lorsqu'il est inférieur à 0.10. Pour les questions à choix multiples avec quatre options, l'indice de difficulté doit être supérieur ou égal à 0.25. Pour celles ayant cinq options, l'indice de difficulté doit être supérieur ou égal à 0.20.
- **Discrimination des items** : c'est l'indice de discrimination utilisé dans le cadre de la théorie classique mais avec une correction qui consiste à utiliser, au lieu du score total, le score total privé du score de l'item en question. Un item est jugé discriminant lorsque cet indicateur est supérieur à 0.20.
- **Analyse des leurres** : cette analyse concerne les questions à choix multiples dont les options de réponses sont appelées leurres ou distracteurs. Elle consiste en l'examen de la répartition des élèves selon leurs réponses. Le pourcentage des élèves ayant choisi la bonne réponse, correspond à l'indice de difficulté. Un bon leurre doit, quant à lui, être choisi par au moins 10% des élèves. Le pourcentage des réponses invalides et des non réponses est également indiqué. Pour chaque leurre, le coefficient "r-bis point" corrigé est estimé. Il correspond à la corrélation entre l'indicatrice du choix du leurre et le score total des autres items. Un bon item doit avoir une corrélation modérément positive pour l'option correcte et une corrélation négative pour les leurres.
- **Analyse des questions ouvertes** : ces questions sont notées 0, 1, 2 ou plus selon que la réponse est fausse, incomplète ou incorrecte. Le pourcentage des élèves ayant obtenu chaque niveau de score, le pourcentage des réponses invalides et le pourcentage des non réponses sont calculés. Aussi, pour chaque niveau de score, le "r-bis point" corrigé est estimé. Il représente la corrélation entre ce niveau de score et le score total des autres items. Les "r-bis points" pour ce type de questions doivent augmenter et passer, de façon monotone, d'une valeur négative à une valeur positive en passant du niveau du score le plus bas à celui le plus élevé.

1 - Cette procédure, dite "equating" en anglais, consiste à positionner sur la même échelle de compétence les individus de différentes cohortes, à partir de leurs résultats observés à des évaluations différentes.

2 - Parmi les enquêtes nationales, on cite à titre d'exemple, les enquêtes françaises CEDRE (Cycle des Évaluations Disciplinaires Réalisées sur Échantillons) et les enquêtes américaines NAEP (National Assessment of Educational Progress). Les principales enquêtes internationales concernant l'évaluation des acquis scolaires sont TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study) et PISA (Programme for International Student Assessment).

3 - Il s'agit de l'Alpha de Cronbach, l'indice de difficulté et l'indice de discrimination.

4 - Il s'agit des langues arabe et française, des mathématiques, des sciences de la vie et de la terre, de la physique-chimie et de l'histoire-géographie.

Résultats de l'analyse d'items : Exemple du test de l'Histoire-Géographie

Le test de l'Histoire-Géographie est commun à toutes les filières du tronc commun du secondaire qualifiant. Deux versions équivalentes, A et B, sont construites, contenant chacune 40 questions. Parmi ces dernières, on distingue trois types : les questions à choix multiples, les questions de liaison et les questions ouvertes.

L'analyse d'items a porté sur les deux versions. La fidélité de ce test a atteint le seuil minimum requis qui est de l'ordre de 0.80, et ce pour les deux versions du test.

Les résultats concernant les indices de difficulté et de discrimination ainsi que le r-bis sont synthétisés dans des tableaux dont nous présentons l'exemple ci-dessous.

Pour aider les équipes chargées de réviser les tests, des signes indiquant, pour chaque item, la présence d'un ou plusieurs problèmes, sont rapportés dans la dernière colonne de ces tableaux. Les signes utilisés sont les suivants:

o DC : la discrimination de l'item est inférieure à 0.20 ;

o DF : la difficulté de l'item

- excède 0.85 ;
- ne dépasse pas 0.10 ;
- inférieure à 0.20 pour les items à choix multiples ayant cinq options
- inférieure à 0.25 pour ceux ayant quatre options ;

o PB : le r-bis d'au moins un leurre dans un item à choix multiples est positif ou les r-bis des scores dans une question ouverte ne sont pas ordonnés.

o L : le pourcentage des élèves ayant choisi une option dans un item à choix multiples ou ayant obtenu un niveau de score dans une question ouverte est inférieur à 10% ;

o A-i : l'Alpha de Cronbach d'un item est supérieur à celui du test.

Exemple de questions fermées

Question	Difficulté	Discrimination	Alpha-i	Pourcentages					R-bis					Observation				
				P_1	P_2	P_3	P_4	P_5	Réponse invalide	Non réponse	PB_1	PB_2	PB_3		PB_4	PB_5	Réponse invalide	Non réponse
Q1	0,14	0,01	0,80	26,6%	20,3%	14,5%	35,0%		0,6%	3,0%	0,07	-0,01	0,01	-0,03		-0,01	-0,09	DF;DC;PB
Q3	0,82	0,28	0,79	4,5%	4,0%	82,1%	7,5%		0,6%	1,3%	-0,11	-0,18	0,28	-0,14		-0,06	-0,08	L
Q8	0,34	0,15	0,80	34,4%	9,1%	34,4%	19,6%		0,6%	1,9%	0,15	-0,12	-0,03	-0,03		-0,004	-0,08	DC;L
Q11	0,67	0,26	0,79	13,4%	67,5%	9,9%	7,7%		0,5%	1,0%	-0,14	0,26	-0,12	-0,11		0,01	-0,09	L;PB
Q12	0,10	0,07	0,80	71,9%	10,5%	9,9%	5,8%		0,3%	1,6%	0,12	-0,14	0,07	-0,09		-0,03	-0,08	DF;DC;L;PB
Q21A	0,28	0,26	0,80	28,4%	28,4%	13,6%	15,2%	6,2%	2,1%	6,2%	0,26	-0,17	-0,08	0,05	-0,05	0,02	-0,10	L;PB
Q21B	0,42	0,19	0,80	10,9%	6,4%	42,2%	21,7%	11,0%	2,4%	5,5%	-0,17	-0,06	0,19	-0,003	-0,01	-0,01	-0,09	DC;L
Q21C	0,19	0,01	0,80	28,1%	20,2%	12,1%	19,0%	12,8%	3,0%	4,9%	-0,03	0,14	-0,04	0,01	-0,04	-0,01	-0,10	DF;DC;PB
Q21D	0,29	0,13	0,80	14,4%	28,9%	11,7%	21,3%	15,5%	2,0%	6,2%	-0,03	0,13	-0,04	0,004	-0,01	-0,02	-0,12	DC;PB
Q30	0,45	0,28	0,79	16,9%	18,1%	44,8%	14,7%		1,0%	4,4%	-0,06	-0,24	0,28	0,04		-0,05	-0,16	PB

Exemple de questions ouvertes

Question	Difficulté	Discrimination	Alpha-i	Pourcentages				R-bis				Observation
				P_0	P_1	Réponse invalide	Non réponse	PB_0	PB_1	Réponse invalide	Non réponse	
Q15	0,16	0,40	0,79	40,7%	16,1%		43,3%	0,07	0,40		-0,36	DF
Q25	0,07	0,24	0,80	42,2%	7,4%	0,1%	50,3%	0,18	0,24	0,01	-0,31	DF;L;PB
Q35	0,36	0,52	0,79	15,5%	35,7%	0,1%	48,8%	-0,04	0,52	0,01	-0,47	PB

Les résultats de cette analyse ont aidé les équipes chargées de l'élaboration des tests dans la validation de ces instruments et ce, en reformulant ou en changeant les items ne répondant pas aux critères requis. Les modèles de la TRI peuvent, quant à eux, servir à l'estimation des scores des élèves dans l'enquête principale¹. Il sera également intéressant d'y recourir pour exploiter les questionnaires du contexte². Ceci consiste à créer, à partir de plusieurs questions, des indicateurs composites afin de tester leur impact sur le niveau des acquis.

L'élaboration des dits questionnaires est fondée sur un cadre théorique qui définit les dimensions constituant chaque questionnaire et les questions y afférentes. Tester l'adéquation du modèle adopté³, au contexte marocain est aussi une opération inévitable. L'expérimentation des questionnaires auprès d'un échantillon d'écoles, d'enseignants et d'élèves permet de mener une analyse factorielle pour tester la validité du construit.

Conclusion

L'analyse des items est nécessaire pour l'amélioration de la qualité des tests psychologiques ou pédagogiques, dont notamment ceux centrés sur l'évaluation des acquis scolaires. Dans le cadre du PNEA, l'Instance Nationale d'Évaluation a inscrit cette étape dans le processus de validation des tests mesurant le niveau des acquis des élèves. La méthode suivie consiste à expérimenter les tests auprès d'un échantillon d'élèves et d'en déduire les propriétés psychométriques des items qui les constituent. Ces propriétés permettent de juger de la qualité de ces items et de leur contribution dans la mesure des compétences ou habiletés que l'on cherche à évaluer. Par la suite, les modifications et les ajustements nécessaires sont apportées. Les modèles adoptés reposent généralement sur la théorie classique de mesure et/ou la théorie des réponses aux items. Chacune d'elles présente des avantages et des limites faisant de la combinaison des deux un choix plus judicieux. En effet, les indices découlant de la théorie classique sont toujours utilisés dans les études d'évaluation des acquis au niveau national et international. Les modèles de la TRI sont, quant à eux, très utiles dans la modélisation des scores et la création des échelles.

Traduction du Résumé Abstract Translation

Amina Benbiga - Larbi Karani

ASSESSMENT OF STUDENTS' ACHIEVEMENTS : PNEA MODEL

Surveys dedicated to assess educational achievement are especially important given the role they play in the evaluation of the output and quality of education systems. Instruments used in this context include tests and questionnaires. The first ones aim to measure student performance in a particular subject when the second ones provide information on individual, family and school characteristics which can be used to explain the differences in school performances between students. Results obtained from such studies will help provide guidance to decision makers to improve the quality of education. However, the relevance of the actions taken depends on the capacity of instruments to reflect reality. This refers to important concepts, reliability and validity, that must be taken into consideration when constructing a measuring instrument. In the case of student achievement, the validation of tests is a crucial step. It goes through item analysis using two approaches: classical approach and item response theory. The aim of this paper is to present the theoretical underpinnings of these approaches and their use in the context of National Assessment Program of Student Achievement (PNEA).

1 - L'enquête principale est programmée pour la fin de l'année scolaire 2015-2016.

2 - En plus des tests évaluant le niveau des acquis scolaires des élèves du tronc commun, trois questionnaires sont élaborés dans le but de collecter des informations sur les caractéristiques individuelles, familiales et scolaires des élèves ainsi que sur l'environnement éducatif. Il s'agit du questionnaire élève, le questionnaire enseignant et le questionnaire directeur.

3 - L'élaboration des questionnaires du contexte est inspirée du travail de Janosz (1998) portant sur "L'environnement éducatif à l'école secondaire".

Bibliographie et Références

Richard Bertrand et Jean-Guy Blais, "Modèles de mesure : l'apport de la théorie des réponses aux items", Presses de l'Université du Québec, 2004.

Pierre Vrignaud, "La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question ?", Revue française de pédagogie, 2006.

Thierry Rocher, "Mesure des compétences : méthodes psychométriques utilisées dans le cadre des évaluations des élèves", Education et formations, N° 86-87, 2015.

Dany Laveault et Jacques Grégoire, "Introduction aux théories des tests en psychologie et en sciences de l'éducation", Editions de Boeck Université, 2e édition, 2005.

Eric Frenette, Richard Bertrand, Pierre Valois, Marc Dussault et Marie-Hélène Hébert, "Comparaison empirique des paramètres d'items/sujets de la théorie des réponses aux items et de la théorie classique des tests", Journal of Educational Measurement and Applied Cognitive Sciences, volume 1, No 1, 2007.

EVALUATION DES APPRENTISSAGES ET AMÉLIORATION DE L'ÉDUCATION¹

L'EXPÉRIENCE ARGENTINE: ONE ET IMESA

Marisa Diaz *

Résumé

Les politiques d'information et d'évaluation sont en vigueur depuis plus de 20 ans en Argentine. Durant cette période ont été consolidés les processus institutionnels et méthodologiques qui nous permettent de disposer d'un système national à caractère fédéral de grande valeur politique et pédagogique. En particulier, les politiques d'évaluation étatique sont étroitement liées aux politiques d'enseignement et orientées à la prise de décisions pédagogiques, sur une base informée, qui évaluent particulièrement la distance qui existe entre la réussite proposée et la situation de départ. C'est-à-dire quelles sont les politiques qui ne s'épuisent pas dans la production d'information diagnostique ou descriptive sur la réalité d'un système éducatif.

Dans ce sens, les ONE - opératifs nationaux d'Évaluation- qui consistent à appliquer chaque trois années de épreuves standardisées pour mesurer les connaissances des élèves, à caractère échantillonal ou général selon le besoin et pour dans des années déterminées de l'éducation obligatoire argentine - sont une source précieuse d'information pour les améliorations de la gestion institutionnelle, les pratiques d'enseignement et les apprentissages des élèves en Langue, Mathématiques, Sciences Sociales et Sciences Naturelles. Étant donné que l'Argentine est un pays fédéral et la responsabilité directe sur les classes et les écoles est de chacune de 24 provinces dans lesquelles le système éducatif est organisé, les épreuves sont construites en prenant comme référence quelques contenus minimaux réfléchis només Noyaux d'Apprentissage Prioritaire.

De la même façon, à partir du 2013 et en fonction des progressions des politiques de renforcement de l'éducation secondaire obligatoire, l'Argentine incorpore comme instrument de suivi pédagogique l'IMESA - un indice d'Amélioration de l'Education Secondaire Argentine-, destiné à orienter le travail des institutions éducatives du niveau aux processus d'améliorations des trajectoires scolaires de ses élèves.

Cet indice combine les résultats des épreuves nationales (ONE) en Langue et Mathématiques avec des indicateurs relatifs à la trajectoire des élèves et sont d'une application universelle pour les écoles secondaires, facilitant un regard intégral sur chaque institution éducative. C'est un outil pour appuyer l'autoévaluation scolaire et le conception et le développement de stratégies éducatives.

ONE et l'IMESA sont exemples de comment soutenir des processus d'évaluation externe et d'autoévaluation institutionnelle dans les institutions éducatives, qui intègrent un programme de travail qui influe d'une manière effective sur les améliorations de la qualité de l'enseignement et des apprentissages scolaires, sans négliger les processus d'inclusion de tous les élèves, le respect des cultures scolaires et le renforcement du travail des équipes d'enseignants.

Système éducatif argentin

Le système éducatif argentin est un système fédéral constitué par 24 juridictions. En 2015 le nombre total des élèves était de 10.964.618 inscrits dans 52.072 institutions éducatives². Cette population scolaire correspond aux niveaux: initial, primaire, secondaire, adultes et éducation technique.

L'Etat fédéral fixe les lignes de la politique éducative pour le système dans son ensemble et régle son application sur la base de processus fédéraux de consultation et de concertation à travers le Conseil Fédéral de l'Education (CFE). Il appuie techniquement et financièrement les provinces et il est responsable de l'enseignement universitaire.

¹ - Traduction française du papier EVALUACION DE APRENDIZAJES Y MEJORA DE LA EDUCACION - La experiencia de Argentina: ONE e IMESA. Auteur Marisa Diaz, Directrice Adjointe de Planification Educative, Ministère de l'Education Nationale, Argentine.

² - Données officielles du Relevé Annuel Statistique 2014 élaboré par le Ministère de l'Education de la Nation à travers son Service de Gestion de l'Information sur la base des données fournies par les Provinces.

* Directrice Adjointe de Planification Éducative, Ministère de l'Education Nationale, Argentine

Les provinces sont responsables de la gestion des systèmes éducatifs au sein de leurs juridictions ainsi que de l'établissement du programme provincial qui doit être en harmonie avec les standards nationaux tels que définis dans les Noyaux des Apprentissages Prioritaires (NAP) pour l'éducation obligatoire.

L'éducation publique fait l'objet de l'article 14 de la Constitution Nationale et elle est réglementée par la Loi de l'Education Nationale 26.206 de l'année 2006 qui établit: la structure du système par niveaux et par modalités, l'étendue de l'éducation obligatoire et les responsabilités de l'Etat national, des Etats provinciaux et de la ville autonome de Buenos Aires, capitale du pays.

En 2015 la loi 27045 vient modifier la loi 20.206 dans son article 16 et établit comme obligatoire l'accès à l'école des enfants de 4 ans.

Ministère de l'Education Présidence de la Nation

Système éducatif en Argentine

14 ans d'éducation obligatoire

L'éducation est garantie par la Constitution

Outil principal de réglementation: la LEN approuvée en 2006

Niveau éducatif	Âge	Zone géographique	Statut
Education pré- primaire (ISCED 0)	3-5 ans	Tout le pays	L'éducation obligatoire commence à 4 ans
Primaire (ISCED 1)	6-11 ans	12 provinces	Obligatoire
Secondaire (ISCED 2 et 3)	12-17 ans	12 provinces	Obligatoire
	13-17 ans	12 provinces	
	ETP dure une année de plus		

Sous-Secrétariat de Planification Educative

Stratégie Nationale d'Evaluation

L'Etat est tenu de procéder obligatoirement à un suivi intégral des politiques éducatives et plus particulièrement d'établir l'évaluation de leurs effets sur les systèmes, les institutions et les contenus de l'enseignement.

Par conséquent, l'évaluation est une tâche de grande valeur politique qui permet de doter de visibilité les responsabilités et les actions des différents intervenants pour l'obtention des objectifs escomptés et opère en révisant les modalités moyennant lesquelles l'action de l'Etat se déploie pour les améliorer, en installer d'autres ou, le cas échéant, les interrompre. Pour le Ministère de l'Education Nationale, l'évaluation est un processus de formation et non pas de punition.

C'est pour cela que l'Argentine développe, depuis plus de vingt ans, les Mécanismes Nationaux d'Evaluation (ONE) et participe à des études évaluatives régionales et globales. Elle effectue également des études spécifiques et des évaluations des programmes de gouvernement.

- "Mécanisme National d'Evaluation" (ONE)

L'Argentine développe, depuis 1993, une stratégie nationale d'évaluation qui combine l'application d'épreuves nationales et internationales. Dans ce cadre, le Mécanisme National d'Évaluation (ONE) mesure les résultats de l'apprentissage en langue, mathématiques, sciences sociales et sciences naturelles au sein de l'éducation fondamentale dans tout le pays. Les premières années de sa mise en application, ce processus se faisait une fois par an; depuis 2010 il est appliqué une fois tous les trois ans.

Le Mécanisme National d'Évaluation (ONE) fournit des informations et des connaissances concernant la performance des apprenants dans les écoles publiques et privées dans les 4 principales zones du programme officiel. Il s'agit de critères de mesure standardisés des acquis éducatifs cumulés en 3ème, 6ème et 9ème (échantillon) et en 12ème (totalité).

Les épreuves nationales sont élaborées par le Ministère de l'Éducation Nationale sur la base des programmes en vigueur dans les provinces, les NAP (Noyaux d'Apprentissage Prioritaires) qui établissent des standards d'apprentissage fédéraux pour toutes les matières, les manuels utilisés dans les écoles et les résultats des ONE des années précédentes. Il s'agit d'épreuves en support papier qui intègrent des QCM avec des épreuves de production pour les élèves. Le traitement des résultats se fait moyennant application du modèle Rasch.

Les résultats des élèves sont organisés par niveau de réussite:

- **ELEVÉ:** lorsque les élèves obtiennent la qualification "Destacado desempeño" ("Performance remarquable") dans tous les contenus et compétences évaluées en fonction de leur niveau de scolarité et des documents du programme pris comme référence.
- **MOYEN:** lorsque les élèves obtiennent la qualification "Satisfactorio desempeño" ("Performance satisfaisante") dans tous les contenus et compétences évaluées en fonction de leur niveau de scolarité et des documents du programme pris comme référence.
- **BAS:** lorsque les élèves obtiennent la qualification "Elemental desempeño" ("Performance élémentaire") dans tous les contenus et compétences évaluées en fonction de leur niveau de scolarité et des documents du programme pris comme référence.

Ministère de l'Éducation
Présidence de la Nation

Ecoles et élèves

ONE 2013	Ecoles	Elèves
Année 12 (ISCED 3) - total	8.889	280.000
Année 9 (CITE 2) - échantillon	1.740	34.820
Année 3 et 6 (CITE 1) - échantillon	5.400	70.000

Sous-secrétariat de Planification Educative

Résultats de l'ONE: Diffusion et implications sur le travail scolaire

Actuellement, les résultats sont publiés au niveau national et par régions géographiques³ regroupant les provinces argentines qui présentent le pourcentage des élèves inclus dans chaque rubrique ainsi qu'un rapport qualitatif qui analyse les facteurs associés aux apprentissages et aux recommandations méthodologiques adressées aux écoles et aux enseignants par niveau évalué. L'objectif recherché c'est que ces derniers puissent mener des actions correctives en partant des erreurs les plus fréquentes détectées lors du traitement des données.

Chacune des autorités éducatives provinciales a accès aux résultats de sa juridiction ainsi qu'à un élément de comparaison par rapport à sa région géographique. En ce qui concerne le recensement de la 12^{ème} année, les résultats sont, en plus, organisés par école via site web à accès restreint.

Cette formule est en vigueur depuis l'année 2006 et trouve son appui légal dans la Loi Nationale sur l'Éducation de la même année⁴. Cette réglementation fait l'objet de controverses et de débats de grande intensité promus par certains secteurs de la société et certains spécialistes qui remettent en question le fait que les résultats soient publiés avec des restrictions. Ils avancent pour cela qu'une stratégie de diffusion plus large permettrait un plus grand contrôle institutionnel et social de la qualité des établissements et du travail des enseignants.

Depuis le Ministère de l'Éducation Nationale, on soutient que les résultats de l'ONE doivent être transmis aux responsables des administrations éducatives régionales et aux écoles. Il ne s'agit pas de données isolées mais d'une partie d'un processus d'amélioration des pratiques scolaires.

Par ailleurs, ce débat remet à jour la révision des points de vue concernant l'évaluation et le degré de réussite des apprentissages. Pour le Ministère de l'Éducation Nationale, évaluer c'est comprendre, informer, prédire et interpréter. C'est un regard systémique et systématique qui renforce et favorise une lecture intégrante de l'offre scolaire dans ses processus et ses résultats.

C'est pourquoi, en évaluant les apprentissages nous évaluons l'enseignement ce dernier étant un processus complexe et englobant qui ne peut pas être mesuré/qualifié exclusivement par un seul indicateur des progrès scolaires comme c'est le cas pour les épreuves nationales. La publication massive des résultats par école comporte également le risque de la quantité de classements qui distorsionnent le fond du problème et portent atteinte aux processus effectifs d'amélioration. Dans ce sens, les modalités de communication des résultats doivent être, selon le Ministère, en parfait accord avec l'intention qui a présidé le travail sur ces

3 - La manière de publier les résultats a changé depuis la première édition, partant des décisions prises conjointement par le Ministère de l'Éducation Nationale et le Conseil Fédéral des Ministres.

4 - L'article 97 de la Loi sur l'Éducation Nationale 26.026 établit la diffusion des résultats des évaluations en préservant l'identité des élèves, des enseignants et des institutions éducatives pour éviter tout risque de stigmatisation.

résultats, laquelle doit se matérialiser en une proposition émanant de l'institution, et qui ne stigmatise pas, qui soit proactive et mesurable en termes de progrès scolaires.

En synthèse, les ONE permettent de définir des scénarios et de revoir, au sein de chaque école, l'organisation institutionnelle, la continuité des pratiques éducatives, les parcours scolaires et l'intégration des écoles dans leurs contextes.

Les ONE mettent en évidence le fait que la convergence entre les politiques éducatives et le travail dans les salles de cours à partir d'une culture de l'évaluation où tous les acteurs du système éducatif participent tous ensemble de manière responsable à un processus formatif et non pas coercitif.

Une information importante: le dernier ONE 2013 a connu la participation de 8.889 écoles et 280.000 élèves de la 12ème année⁵.

• Indice d'amélioration de l'Education Secondaire Argentine (IMESA)

La stratégie nationale d'évaluation accompagne les politiques d'amélioration et appuie le développement de l'éducation obligatoire. Elle a, depuis toujours, permis la réorientation des décisions de politique éducative au niveau national et/ou régional en se concentrant sur des problématiques directement et prioritairement en rapport avec l'enseignement.

Dans ce sens, les maigres résultats obtenus lors des dernières éditions de l'ONE de la 12ème année sont venus attirer l'attention sur le besoin d'intensifier les efforts en vue d'améliorer l'offre scolaire dans un contexte de:

- Accès récent et massif au secondaire d'élèves provenant de secteurs désavantagés (en 2006 la scolarisation au niveau secondaire est devenue obligatoire dans tout le pays)
- Des taux élevés d'abandon scolaire
- Stagnation des résultats de l'apprentissage.

A partir de ce qui a été dit depuis 2013 et dans le cadre du développement du Plan National de l'Education Obligatoire et la Formation des Enseignants 2012-2016 approuvé par le Conseil Fédéral de l'Education⁶, graduellement, les écoles secondaires argentines disposent d'un Indice d'Amélioration institutionnelle dénommé IMESA (Índice de Mejora de la Educación Secundaria Argentina) élaboré par le Ministère de l'Education Nationale et approuvé par la majorité dans sa formulation technique et son application par le Conseil Fédéral des Ministres.

L'objectif de l'IMESA consiste à offrir aux directeurs et aux enseignants des écoles secondaires de l'Argentine un outil de gestion orientée vers l'amélioration des niveaux de compréhension et d'intervention dans les problématiques de l'enseignement fondamental et son corrélat dans les apprentissages et le progrès scolaire des élèves.

Antécédents⁷ :

La proposition de l'IMESA prend comme antécédent proche l'indice de Développement de l'Education Fondamentale IDEB⁸.

L'IDEB contient deux composantes: l'une associée aux résultats de ses épreuves d'évaluation nationales, l'autre en rapport avec le parcours des élèves.

Quant aux évaluations, on prend la moyenne des épreuves de Langue et des Mathématiques du dernier niveau du cycle objet d'observation, après avoir effectué un changement d'échelle pour que les résultats prennent des valeurs entre 0 et 1.

Pour le parcours, l'indice prend la somme des inverses des taux de promotion des niveaux qui composent le cycle, qui, en supposant que l'abandon est 0, serait égal à la durée moyenne du cycle pour les élèves qui obtiennent leur diplôme.

5 - Fourni par la Direction Nationale d'Information et d'Evaluation de la Qualité Educative. Ministère de l'Education Nationale.

6 - Résolution du CFE n°188/12.

7 - Ce texte est la synthèse d'une circulaire interne élaboré par une équipe technique d'IMESA dépendante de la Direction Nationale d'Information et d'Evaluation de la Qualité Educative.

8 - L'Indice de Développement de l'Education Fondamentale a été créé en 2005 par l'Institut National de l'Evaluation et des Recherches Educatives du Brésil (INEP).

C'est ainsi que, dans sa définition, l'IDEB prend en considération les rendements en épreuves standardisées et le parcours des élèves. Pour mesurer cette dernière dimension on prend en compte le temps moyen pour conclure le niveau, mais on emploie dans le calcul la somme des inverses des taux de promotion⁹.

En ce qui concerne l'Argentine, on pense que l'hypothèse de l'absence d'abandon est difficile à démontrer. Ce problème est pris en compte à l'IMESA et pour le résoudre on considère le temps moyen pour compléter le niveau, assumant cette définition comme étant la plus appropriée.

Cependant, cet indicateur mesure seulement le temps pour les élèves qui arrivent à obtenir leur diplôme sans prendre en compte s'ils constituent 10 ou 90% de ceux qui s'inscrivent initialement dans ce niveau.

Dans ce sens, et s'agissant d'un indice lié à la qualité du processus scolaire, la prémisse est que ce succès n'est pas seulement la conquête de quelques uns; c'est pourquoi l'indice doit inclure un critère pour mesurer la capacité des écoles à retenir leurs élèves et faire en sorte qu'ils puissent compléter leurs parcours scolaires.

Cette problématique n'est pas associée seulement à la vulnérabilité des élèves; il s'agit également, par exemple, dans les écoles privées (et dans quelques unes publiques) d'une pratique très étendue et qui est celle de l'expulsion des élèves qui doivent redoubler.

Partant de cela il a été donc envisagé d'inclure le taux de réussite finale (fin de cycle et/ou obtention du diplôme) dans la construction de l'indice proposé. C'est à dire que, si un établissement obtient un taux élevé de rendement et un parcours de ses élèves proche de l'idéal comme résultat de l'expulsion de ceux parmi ses élèves qui ont le plus de problèmes, la correction par diplomation débouchera sur une diminution de l'indice qui permettra une lecture institutionnelle et pédagogique plus proche de la réalité.

La proposition de l'Argentine:

La proposition de calcul de l'IMESA se définit donc comme:

$$IMESA = \frac{R * TE}{TM} * f$$

où:

R. est une mesure relative au rendement des élèves, obtenue à partir des épreuves de Mathématiques et de Langue par l'ONE et qui varie de 0 à 1.

TE. est le taux de réussite à la fin du cycle, c'est à dire la proportion d'élèves qui s'inscrivent en première année et qu'on estime qu'ils arriveront à remplir toutes les conditions pour finir le cycle.

TM. est le temps moyen dont les élèves ont besoin pour compléter le cycle. Si les élèves complètent le cycle en temps idéal, c'est à dire sans redoubler aucune année, cet indicateur vaut 1 et augmentera au fur et à mesure de l'augmentation du redoublement.

La source de ces deux derniers indicateurs est le Relevé Annuel. Il varie de 0 à 100.

f. est un facteur d'expansion pour que l'indice varie entre 100 et 500, 100 étant la pire situation possible et 500 la meilleure, l'idéale.

Pour atteindre le niveau maximum de l'indice, tous les élèves de l'école doivent atteindre avec succès la fin du cycle, sans jamais redoubler et en ayant les meilleurs notes dans les épreuves d'évaluation de qualité. C'est ainsi que cette valeur maximale doit être comprise comme une valeur idéale difficilement réalisable dans la pratique.

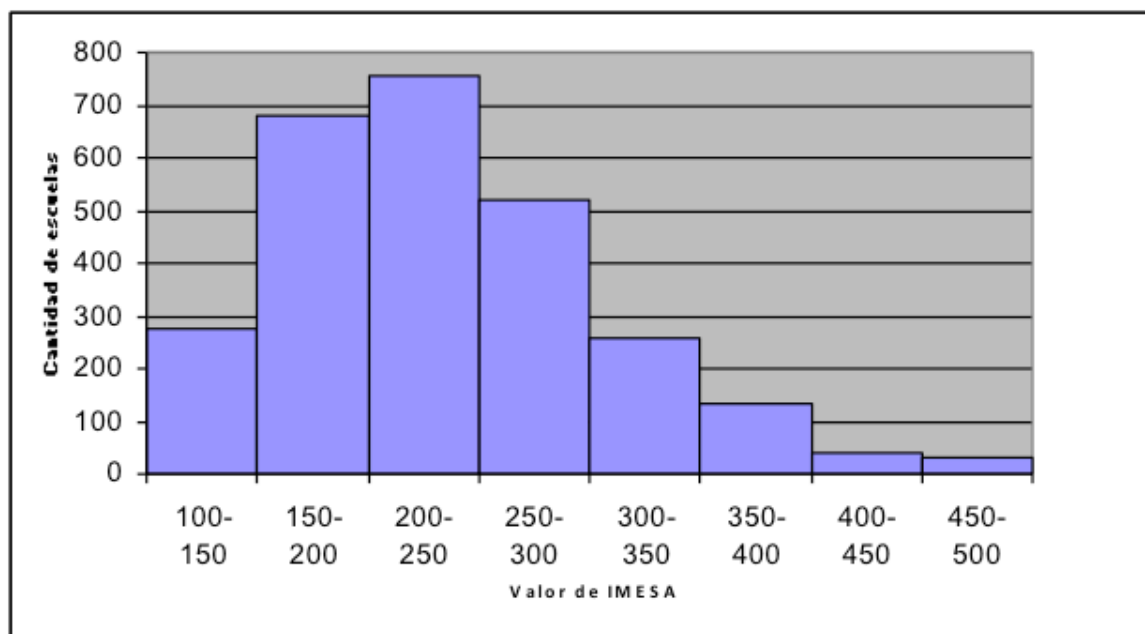
Exemple:

Ci-dessous, la distribution des fréquences de l'indice obtenu pour les écoles participantes à la première étape de l'IMESA et qui ont été sélectionnées puisqu'elles disposaient de Plan d'Amélioration¹⁰ Institutionnel en 2010.

⁹ - Note technique consultable en ligne:

http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/Nota_Tecnica_n1_concepcaoIDEB.pdf.

¹⁰ - Les Plans d'Amélioration Institutionnelle sont élaborés par les institutions éducatives avec l'appui technique et financier du Ministère de l'Éducation Nationale et le suivi des équipes techniques des régions. Ils abordent les problématiques institutionnelles de l'enseignement, les apprentissages et les indicateurs de rendement. Mis en pratique en 2010 pour le tiers des écoles secondaires publiques, actuellement ils ont été élargis pour couvrir la totalité.



Nombre des écoles selon la valeur de IMESA. Source: DINIECE.

Utilisation de l'IMESA

- Au niveau des juridictions éducatives, l'IMESA, de par ses caractéristiques, entre en relation avec les Objectifs d'Efficiéce Interne des systèmes éducatifs provinciaux approuvés par les conventions bilatérales¹¹, et se constitue donc en point d'arrivée des indicatifs. On rappellera ici les objectifs déterminés au niveau fédéral pour l'ensemble du pays pour l'année 2016¹² :

Le Secondaire Cycle Fondamental: réduire le taux d'abandon à 4% dans chaque province et réduire d'un tiers le redoublement.

Le Secondaire Cycle Orienté: réduire de 50% le taux d'abandon et de 25% celui du redoublement dans chaque province.

Les Niveaux de performance définis par l'ONE: améliorer de deux points le pourcentage des niveaux moyen et élevé dans toutes les étapes de base évaluées, en 9ème et en dernière année du secondaire.

- Au niveau des écoles, l'IMESA peut diriger les projets d'amélioration institutionnelle en permettant aux directeurs et aux superviseurs de "complexifier les lectures concernant les parcours de leurs énéves et le travail de leurs enseignants". Il s'agit également d'un complément instrumental de grande valeur pédagogique pour le maintien d'une alerte permanente sur la nécessité de disposer de visibilité concernant les processus d'inclusion/exclusion qui ont lieu dans la vie quotidienne scolaire et les conséquences qu'ils peuvent avoir sur l'aspiration à une bonne éducation pour tous.

Conclusions

- Une stratégie d'évaluation nationale construit une lecture convergente et collective du système éducatif. Ainsi, la condition d'intégralité devient l'axe central de la conception et du développement des actions étant donné qu'elle permet d'analyser et d'interpréter les processus en marche, réviser les stratégies pédagogiques, contribuer à l'amélioration de l'enseignement, identifier les faiblesses et les points forts et promouvoir le feedback nécessaire.

¹¹ - Les conventions bilatérales ont lieu entre la Nation et les provinces pour l'utilisation des financements disponibles pour le secteur éducatif. Elles sont organisées par objectifs quinquennaux et à mi-parcours.

¹² - Conformément aux Conventions Bilatérales, le Conseil Fédéral de l'Éducation a approuvé par la Résolution n°188/12 le Plan National de l'Éducation Obligatoire et la Formation des Enseignants 2012-2016 qui reprend et traduit les objectifs d'efficiéce interne en lignes d'action spécifiques en délimitant les responsabilités institutionnelles des destinataires et les délais pour leur réalisation.

- L'évaluation du système éducatif, les institutions et les classes comme une stratégie politique soutenue dans le temps, devient un outil qui permet de réaffirmer ou de revoir les pratiques; mettre en évidence les besoins en appui et/ou en consolidation et encadrer les institutions éducatives dans un programme d'amélioration et de formation continue. De ce point de vue, les instruments utilisés et l'information obtenue à travers eux deviennent des intrants pour un processus individuel et collectif de grande valeur formative qui suppose la connaissance de la réalité scolaire pour la transformer et l'enrichir.
- Même si la performance mise en évidence par les élèves dans les épreuves standardisées sont un indicateur de la qualité des apprentissages promue et/ou rendue possible par un système éducatif, un regard intégral sur la qualité requiert l'implémentation d'une approche qui valorise d'autres dimensions constitutives de la qualité éducative: les niveaux d'inclusion ou d'exclusion opérants dans le système, la manière dont l'Etat s'occupe de garantir l'égalité des chances éducatives pour tous les enfants, adolescents et jeunes, ainsi que le degré d'accompagnement lors des parcours scolaires réels des élèves.
- En synthèse, dans chaque niveau éducatif, l'école peut et doit être évaluée; et au delà des spécificités et des caractéristiques particulières que l'évaluation assume dans chacun des niveaux éducatifs, il existe un noyau de dimensions communes institutionnelles sur lesquelles il est nécessaire de faire des recherches et générer du savoir pédagogique:
 - Pratiques institutionnelles et pédagogiques inclusives,
 - Dynamique productive du temps scolaire,
 - Travail collectif par l'enseignement et les apprentissages,
 - Insertion des écoles dans leurs communautés.

L'expérience de ces dernières années en Argentine démontre que conformément aux avancées méthodologiques en rapport avec la pertinence des instruments conçus et appliqués, il y a un travail d'ordre institutionnel rigoureux et constant qu'il est nécessaire de faire sur les résultats des évaluations en tant qu'information pédagogique à complexifier dans la révision des pratiques pédagogiques.

Si l'engagement et la participation des enseignants, des équipes de pilotage et des familles fait défaut, les résultats seront des données qui circulent, mais qui ne pourront ni interpeler ni émouvoir.

Les bonnes écoles sont une construction collective de responsabilités partagées; en elles et pour elles, l'évaluation de la proposition scolaire dans toutes ses dimensions et toutes ses conséquences s'impose comme un défi politique et technique inévitable pour le système éducatif argentin.

Traduction du Résumé

Abstract Translation

Marisa Diaz

THE ARGENTINE EXPERIENCE IN THE USE OF THE EVALUATIONS OF LEARNING

Information and evaluation policies have been implemented in Argentina for more than 20 years. During this period, institutional and methodological processes have been strengthened in order to build a strong national education system, both political and pedagogical. State evaluation policies are closely associated to teaching policies and are pedagogical decisions oriented which evaluate the distance between the success suggested and the original situation. Which means: what are the policies which do not exhaust in the information production on the real situation of an education system.

In this context, NEO (National Evaluation Operators), whose mission consists of implementing, every three years, standard tests in order to measure students' skills, either on a sampling basis or general one, depending on needs, and in order to improve institutional management, teaching practices as well as students' achievements in languages, mathematics, social and natural sciences. Given the fact that Argentina is a federal country, and the direct responsibility on schools and classes is on each of the 24 provinces where the education system is organized, exams are elaborated taking as reference some minimum contents called Core Priority Learnings.

In the same way, starting from 2013 and according to education policies' evolution in high school, Argentina applies, as a follow up pedagogical means, an indicator for improving secondary education in Argentina, dedicated to direct education institutions' work.

This indicator combines national tests' results in languages and mathematics with other indicators related to students paths, and are applied in secondary schools, enabling a complete look at every education institution. It is a means to enhance auto evaluation in schools, as well as conception and development of education strategies.

The National Tests as well as the Indicator mentioned above are examples of how to support external evaluation and institutional evaluation processes in education institutions, which integrate a work program that impacts effectively quality improvement in teaching and learning, without neglecting inclusion processes of all students, school culture respect and team work reinforcement.

Bibliographie et Références

- Ley Nacional de Educación (2006) N°26.206. [portal.educación.gov.ar/leyes nacionales](http://portal.educación.gov.ar/leyes_nacionales)
- Consejo Federal de Educación (2009) Resolución N°84/09: "Lineamientos políticos y estratégicos de la educación secundaria obligatoria".
www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación (2009) Resolución N°88/09: "institucionalidad y Fortalecimiento de la Educación Secundaria Obligatoria, planes jurisdiccionales y Planes de Mejora".
www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación Resolución (2010) N°116/10: "Lineamientos para la evaluación de componentes de la calidad educativa".
www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación Resolución (2012) N° 188/12: "Plan Nacional de Educación Obligatoria 2012-2016". www.gov.ar/consejo/resoluciones
- IMESA (2014): Documento marco "IMESA-Plan de Mejora Institucional",
www.mejoraeducativa.com.ar/imesa/documentos/IMESA/IMESA_documento_PDF. 2014
- DINIECE (2014): Anuario Estadístico. [diniece. me.gov.ar](http://diniece.me.gov.ar)

-4

ÉVALUATION DE L'ENSEIGNEMENT SUPÉRIEUR



SAFEGUARDING QUALITY OF HIGHER EDUCATION: THE CASE OF THE NETHERLANDS

Joop Hartog *

Abstract

I will outline the institutional structure of monitoring and improving the quality of higher education in The Netherlands. After critical assessment of recent educational policies a Parliamentary Committee of Inquiry proclaimed the rule of Evidence Based Interventions Only. From discussion and follow-up on that report I will draw the conclusion that more is needed than just a list of recommendations.

Paper prepared for the symposium The Assessment in Education and Training: Approaches, Issues and Challenges, Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique, Rabat, 22-23 October 2015. Comments by Liesbeth van Welie are gratefully acknowledged.

Introduction

Quality of education is an elusive concept, the more so the more abstract the goal of education is taken to be. It is easy to set the goal of a course in typewriting: bring the student up to a level of typing a given amount of text within a specified time limit with the number of typing errors below a specified threshold. The lower the time limit and the lower the threshold, the higher the quality of the graduates. To measure the quality of education, we may measure the typing performance of students at entry and after graduation, to highlight the increase in the student's performance. But even in this simple case, we face the trade-off in two dimensions of quality (speed versus accuracy) and the problem of separating the contribution of the training method from the ability of the student (autonomous self-training might be more effective than a class based instruction program).

Higher education usually is a mixture of training specific professional skills and developing general analytical ability and a critical, rational attitude. Standards for professional training can be derived from required or common skills in the profession's practice, although this may be more difficult for a philosopher than for a dentist. Analytical ability can essentially only be tested against indirect and subjective standards, and experienced professionals have to set required thresholds. In the end this means that quality can only be assessed against a distribution of performance within a profession, on scales that often remain implicit. One virtue of quality assessment may be that it at least compels to seek explicit specification of these scales.

Quality standards may be upheld by professional pride and dedication, and may be safeguarded and formalised by professional organisations. In doing so, these organisations may at the same time protect their members' market position, which immediately leads to an interesting policy question: should the government support such organisations (and give them influence in setting educational standards) or should it prevent monopolistic power to be vested in these organisations? Maintaining professional quality may be in the interest of consumers, but higher prices ensuing from monopolistic power is certainly not.

Protecting quality may be a welfare improving government activity in case of experience goods. There is no need for a government rule to forbid the sale of rotten fruit: the lack of quality is obvious. But education is an experience good in two ways. First, a potential student can only properly assess the quality of an education after participating, and often even only (long) after graduation. Second, buyers of the graduate's output are often also unable to assess output quality before entering a contract: how good is a medical doctor's diagnosis and prescription, an engineer's calculation of the required strength of a steel bridge, an economist's policy advice, a school manager's educational concept? In case of experience goods, in particular with high risk from substandard quality (mortality, health, wealth, wellbeing) the government may provide a valuable service by watching over quality standards. But the existence of a potential problem is not a sufficient condition for government intervention. The market may very well generate its own solutions. Reputations may develop, quality certificates may be provided by commercial agents or interest groups (e.g. student associations, newspapers and magazines, professional organisations) and the government

* Professeur Émérite, Amsterdam School of Economics, Université d'Amsterdam, Pays-Bas

may simply fail to deliver after granting itself exclusive rights: government monopolies usually face no countervailing power.

Regulation may have negative effects as it may undermine professional motivation. Monitoring by key statistics (eg number of graduates, or number of enrolments) will lead management to focus on these numbers and ignore, or at least undervalue, unmeasured (and perhaps unmeasurable) quality dimensions in favour of measured characteristics. As professionals are usually well aware of the unmeasured characteristics, and as these characteristics may be very important for their sense of quality, such management by measurable characteristics may lead to dissatisfaction among professionals, diminish their motivation and in the end be harmful to true quality. Thus, there is an important choice between management by observable characteristics, or even by prescribing specific protocols and work methods, and investing in professional quality and quality awareness, and leave actual work methods to the discretion of the professionals themselves: trust versus bureaucracy.

Clearly, interest in quality of (higher) education raises interesting questions. Should the government intervene? If so, how? Will policies be effective? If the government itself is an important agent in the design of education and schools, is it proper to let the government also monitor and assess quality? Quis custodiet ipsos custodes, who will guard the guards themselves? In this paper I will set out the institutional and legal framework for quality control of higher education in The Netherlands, summarise recent discussions and experiences with evidence based interventions and in the end propose an independent watchdog to put policy makers on a leash.

I. The institutional and legal framework

Higher education in The Netherlands is essentially publicly provided. Even universities that once, a century or so ago, have been founded privately are now fully publicly funded. Of course, there are many conditions for funding and higher education is heavily regulated. But here, we will focus exclusively on policies and institutions aiming at safeguarding the quality of education.

1. The Law on Higher Education (1992)

The legal framework is specified in the Law on Higher Education and Academic Research ("Wet op het Hoger Onderwijs en Wetenschappelijk Onderzoek"). The Dutch government funds institutes of higher education: universities and institutes of higher professional education (professional schools¹). A key condition for funding is accreditation. A student who successfully completes a program at a funded university or professional school obtains an educational degree ("graad"). To receive government funding² and to be allowed to award recognised degrees an institute should meet requirements set for 8 topics, covering quality control, governance, curricula and student entry requirements. The Law imposes the job of accreditation to a separate institute and specifies its rules of operation. Independent accreditation was agreed by the European ministers of education when they set up the uniform European Bachelor-Master structure of higher education in 1999 (the Bologna Agreement).

The framework for assessment of an education for accreditation needs approval of the Minister of Education and should explicitly spell out the criteria for evaluation of an educational program. It should at least pay attention to:

- Intended final qualification level of the graduate, judged against what is internationally common and desirable
- Content and structure of the educational program
- Realised final qualification of the graduate, judged against what is internationally common and desirable
- Adequacy of evaluation, testing and examining students
- Quality and quantity of teaching staff and personnel policies that affect quality of the education
- General and program specific arrangements that affect the quality of the program, including student counselling and provisions that improve accessibility and feasibility of the program for students with a handicap

1- The labeling of these institutes is in a state of flux. In Dutch they are called institutes of Higher Vocational Education. Increasingly, they call themselves, in English, University of Applied Sciences. Their legal admission standard and their academic level of instruction are lower than for university.

2- Government funding is a complicated and not very transparent combination of lump sums and amounts per registered student and per graduate.

- Structure and organisation of internal quality management aimed at systematic improvement of the education

Requests for accreditation are assessed by a committee of experts. As anticipated in the introduction, education quality is not measured directly. Standards for the graduates are essentially set as standards in the field and assessed by the experts. In addition, the protocol refers to quality of the input and to procedures for quality management.

An accreditation is valid for 6 years. Accreditation can also be given at the level of the institute, rather than for an educational program. The institute for accreditation is NVAO, the Nederlands Vlaamse Accreditatie Organisatie (Accreditation Institute of the Netherlands and Flanders), hence, also covering the Dutch speaking part of Belgium.

The assessments of NVAO certainly bite, as Table 1 shows. In 2014, most existing education programs were rated "sufficient", only a few were considered "excellent", but this is as intended by the grading instructions. Sufficient means the program is up to standard. 10% of the existing university programs and 7% of the professional programs either were required to improve or clarify certain aspects, or were withdrawn altogether (which occurred for just a few programs). New programs were less easily accepted: only just under half (university) or one third (professional) of the new programs were accepted right away. Assessment at the institutional level also led to low rates of direct approval. It is clear that universities and professional schools have to take NVAO assessment seriously, as it is certainly no rubber stamp exercise.

Table 1. Accreditations processed by NVAO, Netherlands, 2014

	Existing programs					New programs			Institutes		
	total	suff	good	excell	adjust	total	Pos	adj	total	pos	Adj
University	389	312	37	1	39	24	11	13	5	2	3
Professional	257	209	24	6	15	36	13	23	10	4	6

Source: *NVAO Jaarverslag 2014*. Pos is positive assesment, adj is adjustment required or withdrawn.

It is hardly possible to evaluate this system of quality assessment on its effects. Indeed, from an international survey and discussion of quality assurance systems, Kis (2005, p 33) concludes: "However, the question of how effective quality assurance systems should be designed and implemented is subject to wide debate. There is, for instance, a lack of clarity about what the purpose of quality assurance should be, about the adequateness of diverse methods and instruments used by quality assurance mechanisms, or concerning the consequences of quality monitoring results. Identifying the features of effective quality assurance systems is rendered more complicated by the difficulties in measuring the effectiveness of such systems. It is also difficult to know how the quality of education would have changed without the implementation of quality assurance processes. Moreover, it is not easy to measure the outcomes of quality in higher education. Hence, numerous analysts seem to agree that the impact of quality assurance systems on teaching and learning is difficult to assess and is thus in need of further research."

I have participated in the system on both sides (as an evaluator in NVAO assessments and as responsible for presenting my own faculty of economics in assessment reports and site visit), which enable me to give some personal impressions and observations. In my view, parts of the evaluation protocol have developed into a highly detailed scheme of targets and (claimed) realisations that can easily be used by the universities and schools under assessment to paint a rosy picture. In particular the detailed protocol for stating educational targets and indicating where and how in the curriculum these targets are met is to a large extent a void bureaucratic exercise. True effectiveness cannot be assessed from a cross-table of targets and curriculum components and the challenge for an assessment committee is to see and sense behind the screen of smooth reports and neatly aligned presentations. Universities even train their students and staff in preparation for site visits of an accreditation committee.

Yet, I do not consider the entire exercise as useless. An important virtue of the system is indeed the obligation to reflect on goals, targets and implementation of the educational program. In that sense, the preparation for an assessment can be a useful cleansing activity, generating awareness of weak and strong points, and generating all kinds of improvements. But the whole process can be made much leaner. As the evaluation is done by professionals from the disciplines under investigation, I would suggest to trust

their judgement without imposing too much detail on how they should evaluate. Simple and verifiable indicators should certainly be used. Key verifiable variables may be structure and content of the curriculum, reading lists, textbooks and other course material used. Exams and particularly, the grading of exams and term papers are important inputs for expert assessment. Professional standing of teaching staff can be observed from their research performance. Standard administrative statistics on student flows should also be used: numbers entered, numbers graduated, details of drop-out, time to graduation. But in the end, as truly objective measures on educational quality are not available, and it is the professional judgment that counts, this expert assessment should simply be given most weight.

Actually, in The Netherlands there has been a recent move towards reducing the bureaucratic and documentation burden of the protocol. Flanders has moved towards assessment based only on the institute's procedures of quality monitoring and some experiments in this direction have also been done in The Netherlands. I do not consider that an improvement, as it would simply ignore any assessment of the actual quality that is delivered. I prefer the Dutch way: more trust, less bureaucracy.

2. The Ministry of Education

The Inspectorate for Education at the Ministry of Education has a separate section for higher education. Its task is to assess and improve the quality of higher education. The following elements of this task are explicitly listed:

- assess and improve the quality of the system of higher education
- assess and improve the quality of the Dutch system of accreditation in higher education
- assess and improve the quality of higher education
- assess and improve the quality of financial control, efficiency and continuity of funded institutes of higher education
- ad hoc investigations of institutes in case of serious problems

Since the start of accreditation through NVAO, the key task of the Inspectorate is to assess compliance with the law. The Inspectorate reports annually on the quality of higher education and its financial situation. Interestingly, the Chief Inspector reports directly to Parliament, not to the Minister of Education. Assessing the system of accreditation does not entail evaluating NVAO but assessing the effectiveness of the system, which, in light of the conclusion above, can hardly be done convincingly. One might say that the Inspectorate has a clear mission, as expressed in the first bullet point, and a very flexible program to implement this. In cases of public outcry (such as in recent years on substandard graduations and financial mismanagement), the Minister of Education orders an investigation by the Inspectorate.

The work of the Inspectorate can be exemplified by the most recent Annual Report, 2013/14. The report notes that generally, social and political pressure on higher education continues: it should deliver better education, reduce drop-out and shorten the time it takes to graduate. Growth in number of students tapers off, unemployment among graduates remains high, the Bachelor graduation rates have decreased in professional schools and increased in universities. Quality of education is evaluated by referring to outcomes of NVAO accreditation activities and surveys among students. The low quality of teacher colleges (training primary school teachers) has been a hot item in recent public debates and the Inspectorate reports on higher drop-out and lower graduation rates after reforms in the colleges have been implemented. The Inspectorate also investigates selected problems. Recent reports are on selective admission to higher education, on exams, and on demands put on students in relation to time to graduate. The reports focus on formal aspects ("have the plans been implemented as agreed?"), not on effectiveness as understood in academic research. There is little direct reference to academic literature: most references are to policy reports, to reports from advisory committees and to commissioned research reports.

Judged by its visible output, the Inspectorate serves to signal developments and to detect problems and bottlenecks, somewhat like a neighbourhood watchman. There is at best an indirect link to academic research. The impact of the Inspectorate on quality of education is, by nature, hard to assess.

3. The Council of Education

The Council of Education ("Onderwijsraad") advises the government on main features of education policies and legislation, from an independent stance. The Council, founded in 1919, advises the Minister of Education, on request and at its own initiative, but Parliament can also ask for advice. However, the annual work plan

is set by the Minister of Education. The Council covers issues at all levels of education. The Work Plan 2015 lists 4 main themes: Education and the Knowledge Economy, Core functions of Education, Governance and Organisations, The Role of Teachers. There are frequent interactions and discussions with parliamentary committees.

The Council has 10 expert members, mostly academics, and is assisted by a staff of 10 professionals. The expert members convene in regular meetings, and bring their academic expertise as input to the reports. Reports are not based on original research but on informed opinion and the existing academic literature.

4. The universities

Careers in the (Dutch) universities are built on research performance, research performance meaning publications. There are noticeable differences among disciplines, but over the recent decades quantitative indicators have gained much weight across the entire academic spectre (number of publications, publications weighted by journal prestige, citations, Hirsch indexes etc). Complaints about drawbacks of using simple quantitative measures have increased, and rightly so (maximising number of publications per research result, substituting quantity for quality, deliberate reciprocal citations by authors, citation requirements by journal editors up to outright data manipulation and fraud), but certainly in economics (which I know best) the emphasis on visible research output has led to a marked increase in high level publications³. Observed research performance of staff is informative on professional standing and in that sense on teaching potential but of course not on actual teaching performance.

Whatever the drawbacks of the focus on visible research performance, such quality indicators are not available for quality of instruction. Performance is not easily observable in the public domain. Intense monitoring of teaching performance is too costly, the productivity of teaching is variable across students and hard to measure. Hence, by necessity indirect measures will be used in formalised systems: quality of teaching material and teaching methods, teaching skills, student evaluations. This is indeed how the assessment for accreditation takes place.

Until recently, universities themselves had modest monitoring of teaching quality at best. There were student evaluations, there were external surveys by students and by a leading Dutch weekly magazine: influential but not proper measures of relevant dimensions of educational quality. Most quality assessment, if any, was by peers: informal, implicit, intuitive. Visible indicators are reading lists, student enrollments, exam grades, invisible are reputations emerging in the corridors and at the coffee tables.

The system of accreditation has given a strong impetus to explicit quality management at the universities. Most now have formalised rules and policies to monitor and improve quality of teaching. For example, the University of Amsterdam has a handbook for quality management for application in all its faculties (Kader Kwaliteitszorg, Framework Quality Management). The essence is to create awareness of goals and targets, to monitor performance permanently and to turn the evaluations into improvements when necessary and feasible.

Dutch universities have agreed to require a minimum qualification for teaching. On top of their professional academic qualification, teaching staff in all disciplines should have the Basic Teacher Qualification. As this has only recently been introduced, experienced staff can get this certificate from proven proficiency. For new staff, there are courses on the didactic methodology of teaching.

It seems obvious that quality of education should be monitored, but it is an interesting question how much outside involvement is required. Certainly at the university level, but possibly at all levels of education, teaching is done by professionals who have a strong attachment to their profession. In many cases they will be strongly motivated to remain up-to-date in their discipline and to convince their students of the value of their specialisation. There exists a literature that indicates that teacher quality matters for student learning, but much of it is measured as a teacher fixed effect, and it is hard to pin down what makes up this quality. Most evidence is on primary and secondary education and to my knowledge the effect of quality monitoring of university teaching has not been measured. And as is well known, the danger of quantitative performance measurement is driving out internal motivation. It's very well conceivable that the best policy is to invest in internalising quality awareness and quality pride in the professionals who have to do the teaching. At least it's an interesting hypothesis.

3- NWO, the Dutch organisation that distributes government funds for research, has recently changed its criteria to reduce the impact of such measures.

II. Evidence based interventions: the Dijsselbloem experience

In the old tradition of neoclassical welfare economics, the government is the wise, detached and distant authority that promotes the “public interest”, it seeks to maximize a social welfare function, by setting rules for the private sector and correcting market failures. But alas, as public policy analysis and direct evidence have shown, the government is not a neutral maximiser of general social welfare: it is subject to power games, political manipulations and the private interests of politicians and civil servants. Parliament holds the ultimate power on the design of the system of education and claims a substantial authority over the content of the educational process. This gives it a strong influence on the quality of education. If we cannot simply trust government and parliament to operate in solemn wisdom, the question arises on safeguarding quality at the very foundation of the system: are there strong enough incentives to choose the right methods for given goals? Or, as noted in the introduction: *Quis custodiet ipsos custodies?* In fact, it appears that lawmakers have subjected virtually all private sector provision of goods and services to tight regulation on quality and accountability but kept their hands free to do as they please in public provisions, including education. They simply proclaim ultimate authority.

A new drug can only be offered for sale in the Dutch market after receiving permission from CBG, the Dutch Council for Assessment of Drugs or from EMEA, the European Medicines Evaluation Agency ⁴. Criteria for admission are pharmaceutical quality, effectiveness and safety of the drug. Evidence is collected in three stages: tests on healthy volunteers for health risks, tests on patients (controlled, double blind experiments) and analysis of complaints among patients who use the drug. About 70% of the drugs passes stage 1, 50% passes stage 3. The Dutch admission authority CBG employs 300 professionals.⁵ It is worth noting however, that while drugs require testing, this is not the case for treatment methods and protocols.

Firms that offer raw animal food products (meat, fish, dairy, eggs) need a permit from the Netherlands Food Authority (NVWA). This requires compliance with rules on hygiene, used materials and administrative registration of the production process ⁶. An academic institute at Wageningen University undertakes research to assess food quality and safety; the institute is academically independent but the workplan is determined jointly by the government (formerly the Ministry of Agriculture, now part of the Ministry of Economic Affairs). The organisation and authority structure is rather complex but there is a clear link between policies and academic research⁷. The Netherlands Food Authority employs 2471 workers (in full-time equivalents)⁸.

Product safety is subject to regulation at European level. The approach here is to make the producer liable for damage caused by defaults of the product. The ruling does not apply to services⁹.

The government intervenes massively in the social domain and in education. But there are no rules on verifiable, externally checked quality standards for the public domain. In the end, government is only held responsible by voters, once every four years, on a massive amount of decisions simultaneously. In The Netherlands, I know of only one exception: the government can, by explicit legal provision, be held responsible for road unsafety. The only other basis for action against the government would be based on the principle of “decent governance”.

To illustrate the issues and the dangers, consider the recent cases of massive policy intervention in education, the widespread discontent it caused and the response to this discontent. In 2007-2008, a parliamentary commission of inquiry assessed national government policies on education (“Commissie Dijsselbloem”, the Dijsselbloem Committee, after its chairman Jeroen Dijsselbloem). The focus was on secondary education, but the analysis has general relevance. The inquiry was set up to deal with widespread dissatisfaction among all parties with a stake in education: politicians, students, teachers, school boards, parents, everyone complained. During the 1990’s, parliament had imposed three major reforms, affecting tracking in the lower stage of secondary education, teaching methods in the upper stage and reform of lower vocational education. On lower secondary education, there had been a long standing debate on postponing tracking (traditionally, the student population was split up at age 12). The reform imposed a common curriculum at different levels of depth intensity. In upper secondary education, government led a switch towards more independent student learning, a new structure of the curriculum and larger school conglomerates offering several school types in the same institution. Lower vocational education was merged with the lower level of general education and more general education was injected in the lower vocational curriculum. This simple enumeration already suggests an incredibly ambitious overhaul of secondary education.

4- Information taken from <http://www.pdsb.nl/voor-patienten/onderzoek/hoe-komt-een-geneesmiddel-op-de-nederlandse-markt.aspx>

5- <http://www.cbg-meb.nl/over-cbg/inhoud/werken-bij-het-cbg>

6- <https://www.nvwa.nl>

7- <file:///C:/Users/Beheerder/Downloads/evaluatie-wettelijke-onderzoekstaken-voedselveiligheid.pdf>

8- Private mail communication NVWA September 5 2015

9- http://www.professorvanboom.eu/pdf_files/VanBoom_VanDoorn_HandboekConsumentenrechtH13.pdf

The first and essential point to realise is that a national, centrally determined structure of education inevitably will have to be a compromise. Education has an enormous impact on individuals' life, on the social fabric of society, on economic performance and on the operation of the labour market. It's a key determinant of economic efficiency and of the distribution of wealth and welfare. The details of the production of human capital and the consequences of alternative structures, methods and policies are far from perfectly known. There is a host of opinions, an abundance of views, but only a modest amount of solid empirical knowledge. This calls for very careful decision making and a high quality and integrity of the process of selecting from the alternatives. Transparency, honesty and clear motivation are essential ingredients. The report of the Dijsselbloem Committee shows that the national government can put the quality of education at risk and that careful procedures are needed. The Dijsselbloem Committee is very critical about the decision making process, the content and the implementation of the three reforms.

Prior to curbing early selection and differentiation among lower secondary students, there were no pilots or experiments to test effectiveness in reaching intended goals. Earlier experiments in the 1970's and 1980's have been completely neglected. The time taken for implementing the reform was much too short. The Scientific Council on Government Policies advised 10 years, the few years that were actually allowed were in hindsight judged a mistake. The reforms in upper secondary education, to allow for more independent learning by students and less teacher dominated class instruction ("studiehuis", residence of learning), were inspired by the free and voluntary experience of a single school, and then promoted for all schools in one stroke¹⁰. A critical report, on pilots by 13 schools on some components of the reform and more general concerns, by the Inspectorate for Education, was delivered to Parliament only after the vote on the proposal. A very negative commissioned expert report was never presented to Parliament ("not in conformity with the agreed research question"). Publication of the Annual Report 1998 of the Social Cultural Planning Bureau (a government research agency) was postponed by two weeks, as the Minister of Education did not agree with a statement on the risk of the reforms for children from weak social-economic backgrounds. The reforms on lower vocational education have been accompanied by pilot projects with bi-annual evaluations; the Dijsselbloem Report does not mention policy effects of the evaluations.

The Dijsselbloem Committee concludes quite clearly that the scientific foundation of the reforms was insufficient. Reliable available evidence has not been sufficiently employed. Pilots and experiments were mostly absent, results from pilots that have been held were ignored, experiments that were done did not match sound research principles. Complexity of the implementation of the reforms called for much more time, too many reforms were imposed simultaneously.

As this report by members of parliament themselves makes crystal clear: beware of legislators. Sure, they list a number of lessons to be drawn from their investigations. The Inspectorate of Education should play a much stronger role, as the prime responsible agent for the quality of education. Its activities should be based on a politically sanctioned protocol. It should focus on the resulting output quality of education and leave didactic methods to schools and teachers. Research on education and development of a knowledge infrastructure is vital for innovation. The link between schools and research should be strengthened, in both directions: academic research should analyse practice, practice should have access to evidence. The government can play a stimulating role in implementing evidence based innovations in schools. The Committee even spells out a detailed protocol for such evidence based policy interventions:

Conditions for a careful policy process

- The problem analysis is clear, supported by academic research and generally recognised by actors in the field
- The need for government intervention has been clearly demonstrated
- An evaluation of earlier policy is available
- A report on assessment of policy alternatives is available
- Spillovers and relationships with other policies have been considered and assessed
- The chosen policy option has been validated by academic research. If not, the new policy will be tried first in small scale pilots, with control groups, under supervision of academic researchers
- The results of the pilots are adequately evaluated and will be identifiably integrated in the policy
- Conditions for adequate implementation (budget, expertise, time) have been met

10- There is scope for an interesting digression here. The Studiehuis was not compulsory but ex post it was generally taken to have been compulsory. Possibly, at the time of implementation, (some) schools were under the impression that it was compulsory, possibly schools did not want to be conservative and out of step with modern times.

- The operational organisation is well defined
- The agents who have to implement the new policy in practice have been actively engaged in defining this policy and were well able to foresee the consequences for their own work
- There is sufficient support among all concerned, particularly among the professionals who have to implement the new policy
- Evaluations have been planned ahead. There will be no rush to adjustments before necessity and benefits have been assessed.

I would not call this a formulation as the golden standard for new policy interventions (it's a literal translation from the Dijsselbloem report); it has duplications and some conditions are clearly specific for the education reforms and their flaws that gave rise to the investigation. But clearly, it would be an incredible improvement if all major policy interventions would be introduced according to the core of the recommendation: demonstrate, with reliable empirical evidence, that the policy intervention will realise the stated goal, and if this evidence is not available, define binding, independent evaluation simultaneous with the policy intervention.

Alas, the wise recommendations from the Dijsselbloem Committee have simply gone to waste. In a reflection six years after publication of the Dijsselbloem report, the Education Council (Onderwijsraad 2014) noted that essentially, education policies were still made by Parliament in isolation from the workforce, without sufficient interaction on the needs felt in the schools, without recognising the urgency and relevance of new policies as perceived in the field and with persisting problems of implementing the new policies. The Council concludes to negligible effects of the Dijsselbloem recommendations and continuing lack of trust between policy makers and school practice. In particular, the protocol for new policies, calling for sound evidence based interventions, has largely been ignored. In a sequel to this report the Council promotes less centralisation and uniformity in quality management in higher education, more autonomy for the universities and professional schools and emphasis on a culture of quality awareness and improvement (Onderwijsraad 2015).

III. What can be done?

Within universities, investing in internalizing quality awareness and quality ambition among teaching staff is a better option than bureaucratic formal quality control.

The government should allow truly independent assessment of its own policies, just as it requires from the private sector. Such assessment should be entrusted to a truly independent research institute. A Center for Assessment of Education could be mandated to apply the Dijsselbloem protocol to all major policy interventions in education, it should be allowed to set its own research agenda, could advise government and parliament on request and should open a counter for interaction between research and practice. The institute should do research on topics in the domain of education and maintain academic quality by the requirement to maintain a reputation in the international academic community and to publish research outcomes in leading international journals. Crudely imposing proportionality with the drug admission authority (in terms of research cost relative to medical education expenditures) would justify an education authority of 150 professionals. Crude approaches, wide margins, but substantially above the 10 part-time professionals on the Council of Education.

Traduction du Résumé**Abstract Translation**

Joop Hartog

DE LA QUALITÉ DE L'ENSEIGNEMENT SUPÉRIEUR : LE CAS NÉERLANDAIS

Dans ce papier, je mettrai en exergue la structure institutionnelle de contrôle de la qualité et de son amélioration dans l'enseignement supérieur dans les Pays Bas. Après avoir procédé à une évaluation critique des politiques éducatives, un Comité parlementaire chargé des enquêtes a proclamé la règle d'interventions fondée sur les faits uniquement. Partant des discussions et du suivi de ce rapport, je présenterai les conclusions plutôt qu'une liste de recommandations.

Ce papier a été préparé pour le colloque international sur « l'Évaluation en Education et en Formation : Approches, enjeux et défis », qui s'est tenu à Rabat les 22 et 23 octobre 2015. Les commentaires des Liesbeth van Welie sont sincèrement remerciés.

Bibliography and References

Website NVAO: www.nvao.net

http://wetten.overheid.nl/BWBR0005682/geldigheidsdatum_13-08-2015

<http://www.onderwijsinspectie.nl/onderwijssectoren/hoger-onderwijs/hoe-werkt-ons-toezicht>

<https://www.onderwijsraad.nl/>

Kis, V. (2005), Quality Assurance in Tertiary Education: Current Practices in OECD Countries and a Literature Review on Potential Effects,

<http://www.oecd.org/education/skills-beyond-school/38006910.pdf>

Klaauw, B. van der (2014), From micro-data to causality: forty years of empirical labour economics, *Labour Economics*, 30, 88-97 (Special Issue)

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink (2007), the effect of extra funding for disadvantaged students on achievement, *Review of Economics and Statistics*, 89(4) 721-736.

Leuven, E., H. Oosterbeek and B. van der Klaauw (2010), The effect of financial rewards on students achievement: Evidence from a randomized experiment. *Journal of the European Economic Association* 8(6), 1243–1265.

Onderwijsraad (2014), *Onderwijspolitiek na de commissie-Dijsselbloem*, Advies uitgebracht aan de Voorzitter van de Tweede Kamer der Staten-Generaal, Den Haag.

Onderwijsraad (2015), *Kwaliteit in het hoger onderwijs*, Advies uitgebracht aan de Minister van Onderwijs, Cultuur en Wetenschap, Den Haag

Commissie Parlementair Onderzoek Onderwijsvernieuwingen (2008). *Tijd voor Onderwijs. Eindrapport*. Kamerstukken II, 2007-2008, 31 007, nr. 6

Welie, L. van (2013), *They will get there!*, PhD thesis, TIER, Universiteit van Amsterdam

L'ÉVALUATION DES INSTITUTIONS DE L'ENSEIGNEMENT SUPÉRIEUR : UNE DESCRIPTION ET UN RETOUR D'EXPÉRIENCES DE TROIS PROJETS PILOTES (HEC¹, QHE², AQRM³)

Abdelali Kaaouachi *

Résumé

Cet article propose de dresser une description et un retour d'expériences de trois projets pilotes concernant l'évaluation des institutions de l'enseignement supérieur. Le premier projet a concerné la classification des institutions de l'enseignement supérieur de la région MENA (HEC), le deuxième projet s'est intéressé à l'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes (QHE), et le troisième projet a proposé un mécanisme africain d'évaluation de la qualité (AQRM) des institutions et des programmes. La méthodologie de ce travail repose sur une analyse documentaire et sur l'expérience de l'implication de l'auteur dans ces projets comme étant un expert et un évaluateur externe. Les résultats de cette recherche montrent que les trois projets partagent un certain nombre d'éléments comme l'usage des référentiels d'évaluation et l'implication des acteurs des institutions dans les processus de collecte des informations. En revanche, ils se distinguent par d'autres éléments comme les objectifs, la catégorie d'évaluation, le type d'organe pilotant l'évaluation, et l'usage des résultats des évaluations. Enfin, plusieurs enseignements positifs et négatifs ont été tirés des trois projets.

Mots clés : Assurance qualité, Classement, Classification, Enseignement supérieur, Evaluation, Institution, Qualité.

Introduction

Depuis très longtemps, l'évaluation de la qualité des institutions de l'enseignement supérieur a été une préoccupation des pays anglo-saxons. Dans ce sens, il suffit de rappeler que la naissance des systèmes d'évaluation externe en Grande Bretagne fut en 1832, et l'accréditation des institutions aux Etats Unis d'Amérique a débuté vers la fin du 19ème siècle à travers la création d'agences régionales d'accréditation (El-Khawas, 2001). Mais ce n'est qu'à partir des années 1990, qu'il y eut émergence et prolifération des systèmes d'évaluation de la qualité des institutions, et ce en raison de plusieurs facteurs qui sont essentiellement dus aux changements et transformations majeurs qu'a subit l'enseignement supérieur pendant les trois dernières décennies. Depuis cette période, de nombreux pays ont porté un intérêt grandissant à la qualité des institutions de l'enseignement supérieur et dès lors, à l'évaluation de cette qualité.

L'évaluation de la qualité des institutions concerne l'ensemble des activités liées à l'offre de formation, à la recherche, à la gouvernance et aux services rendus à l'environnement socio-économique et à la société en

1 - Higher Education Classification in MENA.

2 - Quality in Higher Education in Arab Countries.

3 - African Quality Rating Mechanism.

* Professeur, Université Mohammed I^{er}, Oujda, Maroc

général. Plusieurs systèmes d'évaluation de la qualité des institutions de l'enseignement supérieur ont été développés : les classements des institutions, les classifications des institutions, les activités d'assurance qualité (évaluations internes/évaluations externes), les projets d'experts visant l'évaluation de la qualité.

A l'échelle de la région MENA et celle de l'Afrique, trois projets pilotes d'évaluation des institutions de l'enseignement supérieur ont été menés lors des dernières années. Le premier projet a concerné la classification des institutions de l'enseignement supérieur de la région MENA (HEC), le deuxième projet s'est intéressé à l'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes (QHE), le troisième projet a proposé un mécanisme africain d'évaluation de la qualité (AQRM) des institutions et des programmes.

Cet article propose de dresser une description et un retour d'expériences de ces trois projets. La méthodologie de la recherche repose sur une analyse documentaire et sur l'expérience de l'implication de l'auteur dans ces projets comme étant un expert et un évaluateur externe.

L'article est constitué de cinq sections. La première section présente certains aspects de l'évaluation des institutions de l'enseignement supérieur, à savoir ses objets, son aperçu historique et le contexte de son émergence. La deuxième section propose un essai de catégorisation des systèmes d'évaluation des institutions selon quatre classes : le classement, la classification, les activités d'assurance qualité, et le projet d'experts. La troisième section décrit les trois projets en question en termes d'objectifs, de méthodologies et des principaux résultats. La quatrième section dresse un retour d'expériences relatif à ces trois projets. Le chapitre se termine par une conclusion qui discute les résultats de l'étude et présente les principaux défis associés à la question d'évaluation des institutions de l'enseignement supérieur, notamment dans la région MENA et celle d'Afrique.

I. Evaluation des institutions de l'Enseignement Supérieur : ses objets, son aperçu historique et le contexte de son émergence

L'évaluation institutionnelle porte sur l'institution d'une manière globale et vise à formuler des appréciations générales sur la réalisation des missions et la conduite des différentes activités relatives à la formation, à la recherche, à la gouvernance ainsi qu'aux services rendus à la société. Les objets de l'évaluation sont assez divers et concernent les dimensions suivantes : la vision et la mission, les orientations et les objectifs stratégiques, les processus de management, le financement, les activités de formation, les activités de recherche, les étudiants, le personnel (enseignants et staff administratif et technique), les espaces et le matériel, les équipements pédagogiques et scientifiques, le partenariat et la coopération, l'engagement sociétal, etc.

Un survol historique montre que l'exercice d'évaluation des institutions de l'enseignement supérieur a une origine anglo-saxonne qui date de plus d'une dizaine de décennies. Ainsi, la naissance des systèmes d'évaluation externe a commencé en Grande Bretagne en 1832 lorsque l'Université de Durham a demandé aux examinateurs externes de l'Université d'Oxford d'approuver ses diplômes (Lewis, 2005). En outre, l'accréditation des institutions aux Etats Unis d'Amérique a débuté vers la fin du 19ème siècle à travers la création d'agences régionales d'accréditation¹ (El-Khawas, 2001). La période 1932-1940 était caractérisée par le lancement des opérations d'évaluation des programmes et des institutions aux USA par Ralph Tyler qui a réussi à formuler une distinction nette entre les deux concepts mesure et évaluation. Vers la fin des années 1950, il y a eu une relance des activités évaluatives aux USA suite à des problèmes de plusieurs ordres : politique (les Soviétiques ont placé le premier satellite artificiel Spoutnik sur orbite), économique (crise et diminution des investissements) et d'ordre social (pauvreté, grèves des étudiants, racisme...) (Kaaouachi, 2007). Depuis les années 1990, émergent des systèmes d'évaluation, d'accréditation et d'assurance qualité dans l'enseignement supérieur.

Les activités d'évaluation des institutions de l'enseignement supérieur ont émergé dans un contexte international caractérisé par plusieurs constats :

- la croissance des systèmes de l'enseignement supérieur marquée par une massification des effectifs d'étudiants. En effet, l'Institut de Statistique de l'UNESCO a estimé, pendant l'année 1970, le nombre d'étudiants inscrits dans l'enseignement supérieur dans le monde entier à environ 32,5 millions. Cette estimation est passée à près de 178 millions d'étudiants en 2010 ;

¹ - Aux USA, six agences d'accréditation régionales se chargent de l'accréditation : Middle States Association of Colleges and Schools, New England Association of Schools and Colleges, North Central Association of Colleges and Schools, Northwest Association of Schools, Colleges and Universities Southern Association of Colleges and Schools, Western Association of Schools and Colleges. Le « Council for Higher Education Accreditation (CHEA) » coordonne l'activité d'accréditation.

- l'expansion massive du secteur privé qui représente environ 30% des inscriptions faites dans l'enseignement supérieur (Gürüz, 2008) ;
- le développement de l'enseignement supérieur transfrontières (cross-border) et de l'enseignement à distance ;
- la diversification des profils des institutions, de l'offre de formation et des programmes de recherche ;
- le besoin de la transparence et des systèmes qui lient la responsabilité à la reddition des comptes (accountability) ;
- le besoin de la mesure de la qualité et de son amélioration ;
- la réduction des contrôles directs des gouvernements et l'octroi de l'autonomie ;
- la diminution des ressources de financement confrontée par une augmentation des pressions sur les coûts et par une adoption de nouveaux modes de financement ; à cela s'ajoute la tendance de partage des budgets selon les critères de rendement ;
- l'effet de la globalisation et de l'internationalisation de l'enseignement supérieur : la mobilité et l'attraction des étudiants et des enseignants, la reconnaissance mutuelle des qualifications et des diplômes...
- les exigences et les défis de la nouvelle gestion publique² pour les institutions de l'enseignement supérieur : transparence, concurrence, évaluation de la performance (Welpé, Wollersheim, Ringelhan & Osterloh, 2015) ;
- l'évolution mondiale vers l'économie de la connaissance, ayant comme conséquence l'apparition de nouveaux acteurs et de nouveaux prestataires de l'enseignement supérieur...

II. Essai de catégorisation des pratiques d'évaluation des institutions de l'enseignement supérieur

Un examen des pratiques d'évaluation des institutions de l'enseignement supérieur propose de les catégoriser dans quatre classes (Kaaouachi, 2015) : les classements des institutions (Rankings) ; les classifications des institutions ; les activités d'assurance qualité (évaluations internes/évaluations externes) ; les projets d'experts visant l'évaluation de la qualité.

1. Le classement des institutions de l'enseignement supérieur

L'activité de classement des institutions de l'enseignement supérieur a été initiée aux États-Unis au début du 20^{ème} siècle. Le but du classement est de mesurer et de comparer la performance des institutions.

Lors des deux dernières décennies, il y a eu une expansion majeure des systèmes de classement nationaux et internationaux : plus de 150 classements nationaux et une dizaine de classements internationaux. L'arrivée des classements internationaux a coïncidé avec l'intensification de la concurrence dans l'économie internationale. Les classements internationaux des institutions ont acquis une importance croissante dans le monde entier, et ce depuis 2003, année de publication des résultats du premier classement mondial des universités par l'Université Jiao Tong de Shanghai. Après ce premier exercice de classer les institutions à l'échelle mondiale, plusieurs modèles ont été proposés comme le classement de la revue britannique Times Higher Education, et le classement Webometrics.

Les méthodologies des classements se basent sur des critères, des indicateurs et des systèmes de pondération. Le choix des indicateurs et leurs poids reflète les jugements de valeur de l'organisme de classement. Le processus typique de classement est constitué des étapes suivantes : la détermination des critères et des indicateurs, l'attribution des poids, le développement des instruments de mesure comme les questionnaires, la collecte des données, l'analyse des données et la production des tableaux de classement.

Aujourd'hui, le classement des universités est devenu l'un des plus importants outils utilisés par les universités, les étudiants, les enseignants et les chercheurs, et les employeurs à travers le monde. Les

2 - En anglais, new public management. C'est un concept né dans les années 1970 qui vise à minimiser toute différence entre gestion publique et gestion privée. Ce concept propose une approche pragmatique des problèmes et un meilleur partage des rôles entre le niveau du pilotage (le pouvoir politique) et le niveau d'exécution (le pouvoir de l'administration). Ceci afin d'améliorer le rapport coût/efficacité du service grâce à une modernisation accrue et un plus grand pragmatisme de gestion au sein des administrations publiques (Source, Wikipédia).

universités utilisent les classements pour définir leur performance, leur réputation professionnelle et leur statut, tandis que les étudiants les utilisent pour choisir leur futur lieu d'étude ou de recherche. Les enseignants et les chercheurs les utilisent pour des fins de mobilité. Les employeurs les utilisent pour des fins de recrutement des compétences et de collaboration en recherche.

Les classements sont fort contestés en raison de leurs faiblesses méthodologiques. Les principales limites sont les suivantes (Ikaouachi, 2008) :

- des indicateurs davantage liés à la recherche ou à la réputation qu'à la formation (une pondération élevée) [Shanghai, Times] ;
- les indicateurs relatifs aux prix Nobel et médailles Fields ont le reflet du passé de l'institution, davantage que l'analyse de sa situation actuelle [Shanghai] ;
- les revues « Nature et Science » sont à dominance scientifique, technologique et de santé, et ont l'anglais comme langue de rédaction [Shanghai] ;
- les indicateurs sont corrélés entre eux [Shanghai] ;
- la limite des méthodes bibliométriques dans l'estimation du volume des publications et celui des citations et l'appréciation de leur qualité [Shanghai, Times, Webometrics] ;
- l'assignation arbitraire des poids et la sensibilité des classements à des petites variations des poids utilisés [Shanghai, Times, Webometrics] ;
- le problème de subjectivité des appréciations des pairs et des recruteurs [Times] ;
- les indicateurs ne reflètent pas les progrès académiques et scientifiques des universités [Webometrics] ;
- des limites d'ordre technique, par exemple, la non citation du rattachement académique des auteurs dans leurs publications, la pluralité des noms des établissements, la dispersion et la fusion des établissements, le classement des établissements avec l'université mère [Shanghai, Times, Webometrics] ;
- les classements changent fréquemment leurs méthodologies, ce qui rend impossible de comparer les résultats pour des années différentes [Shanghai, Times, Webometrics] ;
- la variabilité et la diversité des contextes, des valeurs culturelles, des missions et des objectifs des universités [Shanghai, Times, Webometrics] ;
- les classements ne comprennent pas des mesures sur certains aspects comme l'enseignement et l'apprentissage ; la recherche en sciences humaines et sociales ; l'engagement régional ou civil ; la vie et les expériences des étudiants [Shanghai, Times, Webometrics], etc.

2. La classification des institutions de l'enseignement supérieur

La classification est un outil qui permet de grouper des entités dans des catégories selon des similarités. Elle s'applique pour connaître et décrire la diversité entre ces entités. Elle n'offre aucun jugement de valeur sur la qualité, ni un tableau de classement pour ces entités.

Deux méthodologies s'appliquent pour la classification des entités : les techniques idiographiques et les techniques nomothétiques (Bailey, 1994). Les techniques idiographiques utilisent des critères a priori pour l'attribution des entités à des catégories, alors que les méthodes nomothétiques prennent une approche inductive, en utilisant des techniques statistiques telles que l'analyse de classification pour grouper des entités dans des classes (Brint, Riddle, et Hanneman [2006]).

Les fonctions de la classification des institutions sont de plusieurs ordres :

- un outil de transparence pour les différentes parties prenantes ;
- un instrument de stratégies institutionnelles (élaboration des missions et des profils) ;
- une base pour les politiques gouvernementales ;
- un outil de recherche pour divers acteurs (étudiants, enseignants, employeurs) ;
- un instrument pour élaborer des systèmes de classement plus pertinents.

Plusieurs systèmes de classification des institutions de l'enseignement supérieur ont été proposés, à savoir

la classification Carnegie, la classification européenne U-Map, la classification de la région MENA. Ces systèmes utilisent des techniques idiographiques basées sur des critères visant à grouper les institutions dans des catégories.

Les systèmes de classification ont plusieurs utilités. Ils proposent des dimensions et des indicateurs représentant un spectre beaucoup plus large des aspects des institutions que celui établi par les systèmes de classement. Ils sont indispensables dans la compréhension et la description de la diversité d'un système de l'enseignement supérieur, et représentent une première étape essentielle pour démêler la complexité de ce système. Ils permettent aux divers utilisateurs d'identifier les forces et les faiblesses des institutions sur les aspects qui les intéressent le plus. Ils peuvent être utilisés pour créer des classements multi-dimensionnel afin de comparer la performance des institutions, comme le classement européen U-Multirank.

Cependant, deux défis concernent la classification des institutions. Le premier concerne l'alimentation des bases de données par des indicateurs et statistiques sur les institutions de l'enseignement supérieur. Le second s'attache à l'approfondissement de la réflexion sur deux éléments permettant la création des catégories de classification, à savoir le choix des critères et des indicateurs, d'une part, et la détermination des limites des catégories, d'autre part.

La sous-section 3.1. décrit le projet de classification des institutions de l'enseignement supérieur de la région MENA, qui entre dans cette catégorie d'évaluation des institutions.

3. Les activités d'assurance qualité

Depuis les années 1990, l'assurance qualité est utilisée dans l'enseignement supérieur, suite à une importation de la notion de « Total Quality Management » du secteur industriel. Au cours des dernières décennies, les processus d'assurance qualité dans l'enseignement supérieur sont devenus de plus en plus courants dans plusieurs pays du monde. Au niveau européen par exemple, c'est à Berlin, en 2003 que la conférence des Ministres a donné une forte impulsion aux mécanismes d'assurance qualité par le développement de critères et méthodologies et par la création d'agences d'assurance qualité.

L'assurance qualité vise l'amélioration de la qualité. En outre, elle permet aux diverses personnes d'avoir confiance en la qualité de l'enseignement supérieur.

Dans les activités d'assurance qualité, le mécanisme appliqué est celui de l'évaluation et/ou de l'accréditation des institutions. La procédure repose sur l'auto-évaluation qui est préparée par l'institution elle-même et sur l'évaluation externe des experts après des visites sur place. Deux approches de l'évaluation externe sont généralement utilisées : une première approche visant l'adéquation aux objectifs (fitness of purpose) ; cette évaluation se pratique à une fin d'amélioration continue de la qualité. Une deuxième approche concerne l'atteinte d'un minimum de standards ; cette évaluation se pratique à une fin d'accréditation.

Les opérations d'évaluation sont basées sur des cadres d'évaluation qui sont constitués par des référentiels d'évaluation précisant des dimensions, des standards, des critères, des indicateurs et des pondérations, et par des guides d'évaluation détaillant la démarche d'évaluation.

Les initiatives d'assurance qualité varient d'un pays à l'autre. Il y a eu des tentatives pour harmoniser certains aspects de l'assurance qualité, par exemple au niveau européen, les normes et les lignes directrices européennes³ ont été élaborées en 2005 et révisées en 2012. Parmi les convergences qui caractérisent les activités d'assurance qualité, on peut noter :

- des mêmes principes méthodologies : l'auto-évaluation, l'évaluation externe par les pairs, l'utilisation d'un référentiel pour l'évaluation externe, les processus de sélection et de formation des évaluateurs, les principes de collégialité et de responsabilité de l'action des comités d'évaluation, une évaluation combinant le travail sur documents et la visite de l'entité évaluée, les principes de dialogue avec les entités évaluées, la production des résultats de l'évaluation et leur publication, les mécanismes de retour d'expérience, l'évolution permanente des pratiques d'évaluation ;
- une indépendance par rapports aux pouvoirs publics à travers la création d'agences d'évaluation.

Par contre, plusieurs divergences caractérisent les activités d'assurance qualité :

- la finalité de l'évaluation : amélioration de la qualité contre le contrôle et la reddition des comptes ;
- le rôle des experts : appuis et recommandations contre inspection ;

3 - Elles ont été adoptées par les Ministres chargés de l'enseignement supérieur, suite à une proposition de l'Association Européenne pour la Garantie de la Qualité (ENQA) en collaboration avec l'Union des Etudiants d'Europe (ESU), l'Association Européenne des Etablissements d'Enseignement Supérieur (EURASHE) et l'Association Européenne d'Université (AEU).

- les destinataires du rapport final : acteurs internes contre acteurs externes ;
- le classement ou non ;
- l'impact ou non sur les subventions octroyées ;
- le type d'agence d'évaluation : étatique contre privé contre une forme intermédiaire.

L'impact des activités d'assurance qualité dans l'enseignement supérieur est difficile à évaluer pour plusieurs raisons : la difficulté de mesurer les changements dans la qualité (Barrow, 1999), la difficulté d'isoler l'impact de l'assurance qualité des autres facteurs qui influent sur l'enseignement supérieur (Shah, 1997), l'intérêt pour la création d'une image réussie du management de la qualité (Zbaracki, 1998). Cependant, les effets des mécanismes d'assurance qualité peuvent être observés à trois niveaux différents : sur l'enseignement et l'apprentissage (Brennan et Shah, 2000), sur les questions d'organisation et de gestion au sein des institutions (Stensaker, 2003) et enfin sur la perception des enseignants sur l'assurance qualité (Dill, 2000).

En somme, l'utilité des activités d'assurance qualité est bien marquée. Cependant, certaines limites existent dans certains contextes : le faible engagement des acteurs, la non appropriation des démarches et des outils d'assurance qualité, la faible fiabilité des bases de données, le peu d'attention accordée aux processus éducatifs et de recherche, le coût élevé des activités d'assurance qualité, le mauvais usage des résultats des évaluations, l'absence des activités de métaévaluation, etc. (Kaaouachi, 2014).

La sous-section 3.3 décrit le projet AQRM qui entre dans la catégorie d'évaluation des institutions.

4. Les projets d'experts pour l'évaluation de la qualité

De nombreux projets d'experts ont été lancés pour évaluer la qualité des institutions de l'enseignement supérieur dans plusieurs régions du monde. Il s'agit de projets élaborés à la base des initiatives des organisations savantes qui impliquent des groupes d'experts pour conduire l'exercice d'évaluation de la qualité. Généralement, les méthodologies ainsi que les dispositifs des études sont proposés par ces mêmes experts, alors que les institutions représentent le terrain de l'enquête pour ces études.

La sous-section 3.2. décrit le projet d'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes, qui entre dans cette catégorie d'évaluation des institutions.

III. Description des trois projets d'évaluation des institutions de l'enseignement supérieur

Cette section décrit trois projets pilotes concernant l'évaluation des institutions de l'enseignement supérieur. Le premier, relatif à la classification des institutions de l'enseignement supérieur de la région MENA, le deuxième à l'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes, le troisième propose un mécanisme africain d'évaluation de la qualité (AQRM) des institutions et des programmes. La section se termine par une analyse comparative de ces trois projets.

1. Projet de classification des institutions de l'enseignement supérieur de la région MENA

La classification des institutions de l'enseignement supérieur de la région MENA est un projet pilote de classification dans lequel ont participé 300 institutions de l'enseignement supérieur de 8 pays (Bhandari et El Amine, 2012). Il a été conduit par l'Association Libanaise des Sciences de l'Education (LAES) et l'Institut International de l'Education (IIE).

Plusieurs objectifs ont été assignés à ce projet régional :

- développer un modèle de classification des institutions de l'enseignement supérieur de la région MENA ;
- comprendre et décrire la diversité de ces institutions ;
- regrouper ces institutions dans des catégories selon les similarités ;
- proposer un système d'information simple qui soit utile aux étudiants et aux chercheurs, à des fins d'orientation ou de mobilité.

La classification est basée sur onze dimensions : le profil institutionnel, le curriculum, le profil d'enseignement, le profil étudiant, le profil enseignant, le profil financier, l'implication dans la recherche, l'orientation culturelle, l'orientation religieuse, l'engagement régional, l'engagement international. Pour chacune de ces dimensions, un ensemble d'indicateurs est retenu. Par exemple, l'ensemble des indicateurs de la dimension «curriculum» est composé par : l'organisation du cursus, les types d'exigences (premier cycle), les exigences d'obtention du diplôme.

Pour l'ensemble des indicateurs retenus, plusieurs catégories ont été constituées. Une analyse de ces catégories a produit les principaux résultats suivants :

- le manque de données au niveau institutionnel sur certains aspects tels que la participation à la recherche ; le profil de l'enseignement et de l'apprentissage ; le profil du corps professoral ; le profil financier ;
- le Qatar et les Emirats Arabes Unis attirent plus d'étudiants étrangers que les autres pays arabes ;
- la plupart des institutions se sont alignées sur un modèle étranger : le modèle français et le modèle américain sont répandus contre 6% pour un modèle culturel mixte ;
- la participation des institutions arabes est relativement faible au niveau international;
- la mobilité des étudiants entre les pays arabes est également faible ;
- les institutions du Qatar et des Emirats Arabes Unis avaient des proportions importantes de professeurs et d'étudiants internationaux (institutions étrangères délocalisées) ;
- le faible engagement et l'absence d'investissement institutionnel dans la recherche. Peu de moyens alloués à la recherche ;
- une offre limitée aux livres spécialisés, aux livres électroniques, aux revues papiers, aux revues électroniques et aux bases de données en ligne ;
- l'enseignement a plus de poids que la recherche et très peu d'enseignants sont actifs dans la recherche.

2. Projet d'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes

Il s'agit d'une étude pilote menée par l'Association Libanaise des Sciences de l'Education (LAES) en vue d'évaluer la qualité des institutions de l'enseignement supérieur des pays arabes (El-Amine, Eds. 2014). Six pays et une vingtaine d'institutions ont participé à cette étude.

La méthodologie, basée sur un cadre référentiel préparé par le groupe d'experts, présente les dix dimensions suivantes : la mission, la gouvernance, la recherche institutionnelle et la planification stratégique, les ressources financières, les ressources matérielles, l'environnement de l'enseignement, la recherche scientifique, le corps enseignant, les étudiants, l'assurance qualité. Aussi, pour chaque dimension, un ensemble d'indicateurs est retenu. Par exemple, l'ensemble des indicateurs de la dimension de la recherche scientifique est composé par le financement de la recherche, les bourses de recherche, la production scientifique, les lauréats de l'université ayant des prix internationaux, les enseignants de l'université ayant des prix internationaux, les contrats de recherche avec les secteurs public et privé, les ressources financières à travers les recherches et les consultations, la position de l'université dans les classements internationaux, la stratégie de la recherche.

Une analyse SWOT⁴ a été utilisée en vue d'établir les forces et les faiblesses de diverses dimensions de la vie académique. Le mode de collecte des données a concerné des documents et des entrevues avec les acteurs des universités participantes à ce projet (responsables, enseignants, étudiants).

A l'issue de ce projet, plusieurs résultats ont été établis :

- la mission a une position marginale dans la plupart des universités des pays arabes ;
- les institutions publiques sont similaires dans leur manque d'indépendance financière et administrative ;
- le manque d'indépendance, à certains égards académiques (dans certains pays) tels que la sélection des étudiants et des professeurs, et la conception des programmes de formation ;

4 - Cet acronyme désigne 4 mots anglais : *Strengths, Weaknesses, Opportunities, Threats*.

- la gouvernance fait défaut de la responsabilisation, de la transparence et des partenariats ;
- la recherche institutionnelle n'est pas connue alors que toutes les universités disposent de projets de développement même si ces projets n'ont pas été mis en œuvre dans un certain nombre de cas ;
- les universités sont généralement financées unilatéralement par les Etats, leurs ressources sont rares et le coût par étudiant est faible ;
- les universités disposent de bâtiments et de locaux. Parfois, il y a des problèmes en termes de sureffectif des étudiants, du faible niveau des services, de bâtiments dispersés, de manque d'entretien et d'inadéquation des équipements ;
- pas suffisamment d'informations sur la production scientifique du corps professoral ; cela reflète la marginalité de l'activité scientifique ;
- le nombre de personnel enseignant est suffisant, mais il y a des problèmes sérieux concernant leur promotion, leur développement professionnel et l'évaluation de leur performance ;
- l'assurance qualité en Egypte est une activité existante sous la responsabilité gouvernementale. Inexistence d'un système complet d'assurance qualité au Maroc et au Liban mais il y a quelques expériences ;
- certaines universités privées au Maroc et au Liban sont engagées dans le processus d'assurance de la qualité via des organismes internationaux.

3. Projet de Mécanisme africain d'évaluation de la qualité (AQRM)

Il s'agit d'un projet d'évaluation de la qualité des institutions et des programmes de l'enseignement supérieur des pays africains. Il est conduit par la Commission de l'Union Africaine et s'inscrit dans l'axe de revitalisation du domaine de l'enseignement supérieur, 9ème domaine prioritaire de la deuxième décennie du plan d'action de l'éducation en Afrique pour la période 2006-2015. Plusieurs objectifs ont été assignés à ce projet :

- revitaliser et renforcer les institutions africaines de manière à être compétitives au niveau mondial et à être pertinentes au niveau local ;
- utiliser l'outil pour l'analyse comparative de la qualité de l'enseignement supérieur en Afrique ;
- encourager les institutions à entreprendre l'auto-évaluation et développer une culture de la qualité.

La démarche du projet AQRM est basée sur une autoévaluation faite par l'institution elle-même et sur une évaluation externe menée par les experts après visite sur place.

Le questionnaire AQRM est articulé autour de deux parties : une première partie qui concerne les informations générales dans laquelle on demandait à l'institution de fournir des statistiques sur les étudiants, les enseignants, les équipements ainsi que des informations sur la gouvernance et la gestion, ainsi que le lien avec l'industrie. Et une deuxième partie, dite d'autoévaluation, formulée en critères, standards, notation et appréciation.

Les six critères retenus sont la gouvernance et le management, les infrastructures, la finance, l'enseignement et la formation, la recherche, la publication et l'innovation et l'engagement sociétal. Pour chaque critère, il y a un ensemble de standards. Au total, 49 standards ont été considérés.

La notation adoptée est l'échelle d'appréciation suivante : 0 = médiocre, 1 = insuffisant, 2 = satisfaisant, 3 = bien, 4 = excellent. Le score agrégé est calculé par une moyenne de ces notes. Cela a permis de générer cinq niveaux de qualité : médiocre (note moyenne inférieure à 1), insuffisante (note moyenne entre 1 et 1.99), satisfaisante (note moyenne entre 2 et 2.79), qualité bonne (note moyenne entre 2.8 et 3.5) et excellente (note moyenne supérieure à 3.5). Concernant l'appréciation, pour chaque critère, l'institution doit formuler des points forts, des domaines nécessitant une amélioration et des recommandations.

A travers le projet AQRM, les institutions ont élaboré des rapports d'autoévaluation. A leur tour, les divers comités d'experts ont produit des rapports d'évaluation externe qui ont décrit la situation des institutions et ont formulé les points forts et les points à améliorer.

4. Analyse comparative des trois projets

Le tableau 1 dresse une analyse comparative entre les trois projets et ce relativement aux aspects suivants : l'objectif, la catégorie d'évaluation, l'organe évaluateur, la méthodologie, les résultats, le plan d'action.

Tableau 1 : Analyse comparative des trois projets HEC, QHE, AQRM

	HEC	QHE	AQRM
Objectif	Description de la diversité	Appréciation de la qualité de façon unilatérale	Appréciation de la qualité de façon collective
Catégorie	Classification	Projet d'experts	Assurance qualité
Organe	Service statistique de l'institution et chercheurs associés au projet	Personnes de contact de l'institution et chercheurs associés au projet	Comité d'autoévaluation et comité d'experts
Méthodologie	Remplissage d'un questionnaire spécifique au projet sur la base d'un référentiel d'évaluation	Recherche des données (documents et entrevues) sur la base d'un référentiel d'évaluation	Remplissage d'un questionnaire (information et appréciation) sur la base d'un référentiel d'évaluation
Résultats	Catégoriser les indicateurs	Liste des points forts et points faibles	Liste des points forts et points nécessitant des améliorations
Plan d'action	Aucun	Aucun	Oui

Tableau 1 : Analyse comparative des trois projets.

Il ressort de ce tableau que les trois projets partagent un certain nombre d'éléments, comme l'usage des référentiels d'évaluation et l'implication des acteurs des institutions dans les processus de collecte des informations. Cependant, ils se distinguent par d'autres éléments, comme les objectifs, la catégorie d'évaluation, le type d'organe pilotant l'évaluation, l'usage des résultats des évaluations.

IV. Retour d'expériences

Suite à l'implication de l'auteur dans les trois projets d'évaluation des institutions de l'enseignement supérieur dans la région MENA et celle de l'Afrique, il y a lieu de dresser la liste des constats suivants :

- la culture qualité est peu développée dans les institutions de l'enseignement supérieur [HEC, QHE, AQRM] ;
- l'engagement des acteurs est faible dans l'exercice d'évaluation de la qualité [QHE] ;
- le manque de documents et de preuves qui renseignent le niveau de la qualité atteint [QHE, AQRM] ;
- la transparence n'est pas assurée. Manque de données sur certains profils tels que les aspects financiers et les aspects de la recherche scientifique [HEC] ;
- la création des catégories est faite en recherchant des quartiles sur les z-scores. Problème des observations proches des quartiles (points de coupure) [HEC] ;
- la non implication des institutions dans le processus d'élaboration des méthodologies et des dispositifs des projets [QHE] ;

- l'activité d'autoévaluation était purement mécanique (remplissage du questionnaire). Elle n'est ni descriptive ni évaluative [AQRM] ;
- la grande part des standards est de type qualitatif. Par exemple, « Avoir des procédures établies pour attirer des financements de l'industrie ». L'utilisation des données institutionnelles ou certains indicateurs quantitatifs aurait pu aider à apprécier la qualité [AQRM] ;
- les réponses sont parfois arbitraires (les institutions ont répondu selon leur compréhension des standards) [AQRM] ;
- aucune indication si l'approche de consultation a été utilisée pour remplir le questionnaire ; ce qui pourrait aider à l'amélioration [AQRM] ;
- la difficulté d'apprécier la qualité de toutes les activités d'une université lors d'une visite en 3 jours par les experts [AQRM] ;
- aucune information sur la mise en place des plans d'actions et sur leurs suivis [AQRM].

Conclusion

Au cours des dernières décennies, plusieurs demandes ont fait appel à la transparence et à la reddition de comptes sur les moyens utilisés dans l'enseignement supérieur. En outre, plusieurs demandes ont appelé à l'amélioration de la qualité de diverses prestations liées à la formation, à la recherche et aux services rendus à la société. Cela a permis de développer, entre autres, plusieurs systèmes d'évaluation des institutions tels que les classements, les classifications, les activités d'assurance qualité, et les projets d'experts visant l'évaluation de la qualité.

Cette recherche a dressé une description et un retour d'expériences de trois projets pilotes concernant l'évaluation des institutions de l'enseignement supérieur : HEC, relatif à la classification des institutions de l'enseignement supérieur de la région MENA, QHE, qui s'est intéressé à l'évaluation de la qualité des institutions de l'enseignement supérieur des pays arabes, et AQRM qui a proposé un mécanisme africain d'évaluation de la qualité des institutions et des programmes.

Ce sont des expériences intéressantes qui ont permis d'apprécier la qualité des institutions selon une variété de méthodologies. Le retour d'expériences relatif à ces trois projets montre que des acquis importants ont été enregistrés à travers l'exercice d'évaluation, cependant plusieurs points méritent d'être pris en compte pour améliorer les dispositifs et les approches des futures opérations d'évaluation.

Concernant l'effet sur la qualité, le projet de classification HEC offre la transparence sur les systèmes d'enseignement supérieur : il contribue à l'amélioration de la qualité au niveau des systèmes, mais seulement d'une manière indirecte à travers la

stimulation de la concurrence. Il n'est pas conçu comme un instrument

d'assurance qualité interne au sein des institutions. Par ailleurs, les deux projets QHE et AQRM ont une portée beaucoup plus importante et peuvent bien être des dispositifs intéressants pouvant améliorer la qualité au niveau des institutions.

Le débat autour de l'évaluation de la qualité de l'enseignement supérieur demeure vivace et même chaque fois renouvelé, en raison des transformations et changements qui affectent en continu ce secteur. Certes, les classements des institutions permettent d'alimenter les systèmes d'information et servent à plusieurs usages. Mais, compte tenu de leurs limites, les institutions ont besoin de développer un des mécanismes d'assurance qualité qui permettra d'améliorer leur qualité. La piste d'assurance qualité est alors nécessaire à entreprendre pour les institutions de l'enseignement supérieur, notamment dans la région MENA et celle d'Afrique.

Toutefois, quelques défis majeurs caractérisent la question d'évaluation de la qualité de ces institutions :

- développer la promotion d'une culture qualité auprès des acteurs de l'enseignement supérieur. Cette culture de la qualité peut contribuer à mener des autoévaluations qui soient pertinentes et efficaces. Elle peut aussi contribuer à favoriser une ambiance de coopération entre les évalués et les évaluateurs notamment lors de la visite de site. La sensibilisation et la formation en assurance qualité sont deux éléments clés dans la mise en place d'une culture qualité ;
- accroître la volonté et l'engagement des acteurs dans l'exercice d'évaluation de la qualité. L'implication des acteurs internes est nécessaire pour mener de bonnes pratiques d'évaluation, en raison de la nature sensible du travail d'évaluation de la qualité qui est un enjeu des relations

complexes entre les différentes parties prenantes. Il est primordial alors de lancer un dialogue approfondi et continu entre les acteurs concernés sur les procédures d'évaluation et sur l'usage des résultats des évaluations ;

- améliorer la transparence en développant un système d'information complet sur les divers indicateurs de la vie des institutions et en mettant disponible tous les documents et toutes les preuves qui renseignent le niveau de la qualité. Les instruments de transparence ont pour objet d'offrir une information pertinente, fiable et valide des intervenants sur les niveaux de la diversité dans les systèmes d'enseignement supérieur.

Il reste un long chemin à parcourir, notamment pour les institutions de l'enseignement supérieur de la région MENA et celle de l'Afrique, pour l'évaluation de la qualité. De plus, les aspects de culture, de volonté, d'engagement et de transparence sont des préalables indispensables avant l'entrée dans le processus de mesure de la qualité et de son amélioration.

Traduction du Résumé Abstract Translation

Abdelali Kaaouachi

ASSESSMENT OF HIGHER EDUCATION INSTITUTIONS : A RETURN ON EXPERIENCE ON 3 PILOT PROJECTS (HEC, QHE, AQRM)

This article proposes to draw up a description and a return of experience of three pilot projects concerning evaluation of higher education institutions. The first project involved the classification of institutions of higher education in the MENA region (HEC); the second project focused on evaluating the quality of institutions of higher education in the Arab countries (QHE); the third project described the Mechanism of evaluation of institutions and programs in Africa (AQRM). The methodology of this work is based on a literature review and on the experience of the involvement of the author in these projects as an expert and an external evaluator. The results of this research show that the three projects share a number of elements such as the use of evaluation standards and the involvement of institutional actors in the information-gathering process. However, they are distinguished by other elements such as the objectives, the evaluation category, the type of organ driving the evaluation and the use of evaluation results. Finally, several positive and negative lessons have been learned from these three projects.

Bibliographie et Références

- Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: Sage Publications.
- Barrow, M. (1999). *Quality-management Systems and Dramaturgical Compliance, Quality in Higher Education*, Vol. 5, No.1.
- Bhandari, R., El Amine, A. (2012). *Higher Education classification in the Middle East and North Africa*. New York: Institute of International Education.
- Brennan, J., Shah, T. (2000). *Quality assessment and institutional change: Experiences from 14 countries*, Higher Education, Vol. 40.
- Brint, S., Riddle, M., Hanneman, R. (2006). Reference sets, identities, and aspirations in a complex organizational field: The case of American four-year colleges and universities. *Sociology of Education*, 79, 229-252. doi: 0.1177/003804070607900303.
- Dill, D.D. (2000). *Designing Academic Audit: lessons learned in Europe and Asia*, Quality in Higher Education, Vol. 6, No. 3
- El-Amine, A. (Eds.) (2014). *Quality Issues in Higher Education Institutions in Arab Countries*. Beirut, the Lebanese Association for Educational Studies.
- El-Khawas, E. (2001). *Accreditation in the USA: origins, developments and future prospects*. International Institute for Educational Planning.
- Guruz, K. (2008). *Higher Education and International Student mobility in the Global Knowledge Economy*. Albany, State University of New York Press.
- Kaaouachi, A. (2007). An Evaluation Model of Higher Education Institutions in the Arab World. *Journal of the Association of Arab Universities (AAU)*, 501-526.
- Kaaouachi, A. (2008). Les classements internationaux des universités : quels enjeux pour l'université marocaine ? *Prospectives Universitaires*, 159-173.
- Kaaouachi, A. (2014). *Analyse des systèmes d'évaluation des institutions de l'enseignement supérieur*. Colloque EVALUP 2014, Oujda 2014.
- Kaaouachi, A. (2015). Les systèmes d'évaluation des institutions de l'enseignement supérieur : une analyse de pratiques. In A. Kaaouachi (dir.), *L'évaluation de l'enseignement supérieur et de la recherche : description, analyse et étude de cas*. Imprimerie Jossour, pp. 309-340.
- Lewis, R. (2005). *External Examiner System in the United Kingdom: Fresh Challenges to an Old System*. Public Policy for Academy Quality Research Program.
- Martin, M., Stella, A. (2007). *External quality assurance in higher education: making choices*. Paris: UNESCO: International Institute for Educational Planning.
- Shah, T. (1997). *Quality Management, Quality Assessment and the Decision-Making Process*. The IMHE Project on Institutional Impact, in Brennan, J. de Vries, P. and Williams, R. (Eds.) *Standards and Quality in Higher Education*, Higher Education Policy Series 37, Jessica Kingsley.
- Stensaker, B. (2003). *Trance, Transparency and Transformation: the impact of external quality monitoring on higher education*, Quality in Higher Education, Vol. 9, No. 2
- Welppe, I. M., Wollersheim, J., Ringelhan, S., Osterloh, M. (Eds.) (2015). *Incentives and performance: Governance of research organizations*. Cham et al.: Springer International Publishing.
- Zbaracki, M.J. (1998). The rhetoric and reality of total quality management, *Administrative Science Quarterly*, Vol. 43.

L'ÉVALUATION DE L'INSERTION DES DIPLÔMÉS DU SUPÉRIEUR AU MAROC : LE RENDEMENT DE L'UNIVERSITÉ

Saad-Ellah Berhili, Fatima Berahou, Sanaa Cheddadi, Nawal Zaaï *

Résumé

L'insertion professionnelle, composante clé de la relation formation-emploi, constitue un indicateur pertinent du rendement externe du système d'éducation et de formation. Les enquêtes d'insertion réalisées au Maroc renvoient principalement aux enquêtes emploi du HCP qui restent un outil de mesure de la situation macro-économique de l'emploi, mais à des niveaux d'agrégation très généraux.

Cette intervention présente le dispositif d'évaluation de l'insertion des diplômés de l'université, élaboré par l'Instance Nationale d'Evaluation et expérimenté pour la première fois sur 3 universités marocaines couvrant l'ensemble de leurs établissements (accès ouvert et accès régulé). Il permet de mesurer de manière très fine le degré d'insertion des diplômés de l'université et de comprendre la dynamique de la relation Formation-Emploi au sein de cette population de sortants du système éducatif. La méthodologie et la rigueur scientifique utilisées confèrent à ce dispositif une valeur très importante et une comparabilité avec les dispositifs les plus avancés dans le monde en matière d'évaluation du cheminement professionnel des diplômés de l'université.

L'implémentation de ce dispositif a été faite sous forme d'une enquête pilote auprès d'un échantillon représentatif de diplômés des 3 universités citées ci-haut. Cette enquête d'évaluation de l'insertion et du cheminement des diplômés de l'université, basée sur un calendrier longitudinal rétrospectif sur une durée de 3 années, traite des problématiques importantes et émergentes pour lesquelles on dispose de peu d'informations aujourd'hui pour pouvoir les cerner et les analyser. Il s'agit principalement de recueillir de nouvelles données sur l'insertion professionnelle des jeunes lauréats et d'analyser les phénomènes de mobilité, de précarité des emplois, de déclassement par rapport au niveau des études réussies et l'épanouissement personnel des diplômés vis à vis de leurs situations professionnelles.

L'institutionnalisation et la généralisation de ce dispositif au niveau de toutes les universités et établissements d'enseignement supérieur à une plus grande échelle (formation des cadres, secteur privé, formation professionnelle post-baccalauréat) présente plusieurs avantages par rapport aux études sectorielles et ponctuelles qui existent jusqu'à présent. Elle permettra d'assurer une meilleure compréhension des articulations Formation-Emploi à un niveau national et qui constitue un des grands défis aujourd'hui du secteur de l'Enseignement Supérieur au Maroc, ainsi que d'identifier le rôle des facteurs macro-économiques et sociaux dans la détermination de l'emploi des diplômés.

Introduction

Dans un contexte national connaissant une forte dynamique économique et sociale et des changements profonds de la structure du marché de travail aux niveaux national et régional, l'université marocaine se trouve devant le défi de renforcer son attractivité ainsi que la qualité de son enseignement, de sa recherche et de ses modes de gestion et de gouvernance mais également devant l'enjeu majeur d'améliorer son rendement externe.

D'autre part, et pour faire face à l'augmentation de la demande sociale, particulièrement après l'instauration du système LMD en 2003, l'université a eu recours à la massification du nombre d'étudiants, ce qui a entraîné l'accroissement du nombre de diplômés sur le marché du travail. Cette augmentation a eu des effets sur le taux de chômage des lauréats de l'enseignement supérieur créant ainsi une crise de confiance vis-à-vis des établissements universitaires nationaux.

* Instance Nationale d'Evaluation, CSEFRS, Octobre 2015

D'où la nécessité de mener une réflexion sur la mise en place d'une stratégie d'amélioration de l'employabilité des lauréats universitaires à laquelle doivent adhérer tous les acteurs de l'enseignement supérieur et du secteur professionnel tant à l'échelle nationale que territoriale. Toutefois, la pertinence de toute stratégie est liée à l'efficacité du dispositif de suivi à mettre en place pour appréhender les différentes dimensions liées au rendement externe de l'enseignement supérieur.

Les enquêtes d'évaluation de l'insertion réalisées au Maroc renvoient principalement aux enquêtes emploi du Haut-Commissariat au Plan (HCP) qui restent un outil de mesure de la situation macro-économique de l'emploi à des niveaux d'agrégation très sommaires (le niveau supérieur par exemple renvoie à tous les diplômes Baccalauréat et plus). Aussi, en analysant les résultats de ces enquêtes, il est très difficile d'approcher de manière précise la relation formation-emploi puisqu'on n'a aucune information sur les trajectoires professionnelles de ces diplômés depuis leur sortie de l'université jusqu'au moment de l'enquête.

Le présent papier décrit les étapes qui ont conduit à la mise en place, pour la première fois au Maroc, d'un dispositif d'évaluation longitudinal de l'insertion des diplômés de l'université, dont la conceptualisation, le raffinement et le suivi sont menés par l'Instance Nationale d'Evaluation au sein du Conseil Supérieur de l'Education de la Formation et de la Recherche Scientifique.

I. Les fondements théoriques de l'évaluation de l'insertion

1. Vers une définition consensuelle de l'insertion

L'insertion professionnelle constitue de plus en plus une clé cruciale des politiques publiques mais aussi une question pertinente de la recherche scientifique. Sa définition a toujours suscité l'intérêt des spécialistes. Elle a de tout temps fait l'objet de controverses entre les sociologues et les économistes. Les premiers refusent de l'aborder indépendamment d'un projet sociétal. En effet, toute précarité touchant la vie professionnelle implique une altération de la vie de famille, des aspirations personnelles et des conditions de vie. Une vie professionnelle épanouie implique aussi avoir une place dans la société caractérisée par des échanges satisfaisants avec son environnement. L'insertion professionnelle est alors, indissociable de l'insertion sociale. La difficulté d'accéder à l'emploi engendre aussi d'importants coûts au niveau économique puisque l'on assiste à un manque à gagner en termes de perte de revenus et une diminution de consommation et d'épargne. Cette situation entrave également la capacité des pays à innover et par conséquent diminue les possibilités de leur compétitivité à l'échelle internationale.

L'insertion professionnelle constitue une étape potentielle d'améliorer les chances d'échapper à la pauvreté. Elle implique aussi « l'acquisition de compétences, la possibilité d'expérimentation et la recherche de promotion sociale » (Gauthier, 1997). Les économistes définissent l'insertion professionnelle comme un processus marqué par l'entrée dans la vie active suivie d'un ensemble d'actes rationnels pour construire « un projet de vie ». L'insertion selon cette optique traduit « la période qui suit la sortie du système de formation et qui correspond au moment où l'individu va chercher à négocier les savoirs acquis, sanctionnés ou non par un diplôme, au sein de ce système pour accéder à un emploi », (Tanguy, 1986).

La vision selon laquelle l'insertion professionnelle est une transition vers la vie active est très simpliste. La limite qui marque le début et la fin d'une période d'insertion n'est pas toujours facile à définir avec précision. Ce processus peut prendre la forme d'un emploi à temps partiel ou d'études prolongées, de fréquents changements d'emplois, de retours à la formation, de périodes répétées de chômage, stages, etc. (Laflamme, 1996; Tremblay, 1994; Trottier et al., 1995). Ainsi, l'insertion professionnelle ne s'effectue plus en un seul mouvement dans le temps ; elle est plutôt une succession de statuts qui se transforment pour chaque individu de manière irrégulière. Vincens (1981) insiste sur la nécessité d'analyser l'insertion professionnelle d'un individu depuis sa sortie de l'établissement de formation jusqu'à l'accomplissement de son projet de vie. L'insertion professionnelle est alors, définie comme un processus au cours duquel se succèdent la formation, la recherche d'emploi et de nombreuses entrées et sorties du marché du travail constituant ainsi un cheminement professionnel.

Ce processus fait appel à trois mécanismes : la formation professionnelle, la transition professionnelle et l'intégration professionnelle (Laflamme et Baby, 1993). Le premier mécanisme représente les acquis et les compétences académiques cumulées durant le cursus scolaire. Le deuxième mécanisme englobe toutes les étapes franchies pour décrocher un emploi. Ces étapes dépendent du chercheur d'emploi mais dépendent aussi de la conjoncture économique et de la structure du marché du travail qui régissent son fonctionnement. Le dernier mécanisme désigne l'aboutissement de la recherche d'emploi et représente principalement les conditions socioprofessionnelles de l'emploi occupé. Plusieurs indicateurs sont utilisés

pour décrire l'intégration professionnelle des diplômés : durée d'accès au premier emploi, statut d'emploi (contrat à durée déterminée ou indéterminée), catégorie socioprofessionnelle de l'emploi, durée du chômage, relation formation-emploi, type du contrat, durabilité de l'emploi, qualité de l'emploi, sans nier pour autant l'autonomie financière ainsi que la possibilité de maintenir l'employabilité. Tous ces critères doivent être pris en compte dans la définition de l'insertion socioprofessionnelle (Goyette, 2006).

L'ampleur des définitions proposées dans la littérature démontre la complexité d'asseoir une définition précise et définitive du processus de l'insertion professionnelle. En effet, les frontières entre les différentes situations occupées sur le marché du travail sont souvent étroites. La recherche d'emploi n'est pas toujours régulière et continue dans le temps. Cette recherche n'est pas toujours institutionnalisée et cadrée, ce qui rend difficile d'appréhender le processus de l'insertion professionnelle. Néanmoins, les chercheurs montrent un consensus tacite sur le caractère longitudinal du processus de l'insertion professionnelle. Ils se sont rendu compte de la nécessité de dépasser l'analyse ponctuelle et statique, et ont adopté une approche longitudinale et dynamique pour expliquer le chômage des diplômés (Kalachek, 1980). L'insertion est alors analysée à partir des études longitudinales à travers des flux d'entrées sur le marché du travail (Giret, 2000). Ces études concernent généralement des groupes d'individus qui ont en commun le même événement initial (même niveau d'instruction, même type de diplôme, etc.). Elles synthétisent et situent dans le temps les positions et les différents événements qui ont marqué le parcours professionnel des lauréats. Il s'agit des situations occupées sur le marché du travail et concentrées sur quelques années : poursuite d'études, stages, inactivité, service militaire, reconversion professionnelle, etc. Cette séquence de situations professionnelles forme la trajectoire professionnelle¹ de chaque individu observée en général sur quelques années. C'est en comparant dans le temps les trajectoires professionnelles des différentes catégories de population que nous pouvons saisir le mieux ce qui fait leur singularité (Maillard, 2005). Ces trajectoires permettent de comparer les débuts professionnels de la population d'intérêt, de mieux cerner l'entrée dans la vie active et la dynamique des parcours sur le marché du travail (Mansuy, 2001). Cependant, il reste très difficile de cerner ce que veut dire un individu inséré ou définir les critères statistiques qui permettent de définir un état d'insertion définitive. Une même situation occupée sur le marché du travail ne peut être perçue de la même manière pour des groupes différents de diplômés. Cependant, la réponse à cette question peut s'articuler autour de deux pôles, un pôle de l'extériorité et l'autre de l'intériorité (Vincens, 1997). Le premier définit un état, et considère qu'une fois le diplômé est dans une situation d'emploi, il est inséré. Cette définition reste exogène vu que nous ne nous intéressons pas à la perception du diplômé vis-à-vis de l'insertion. D'autres auteurs voient l'insertion d'un autre angle, celui de l'intériorité. Seul l'individu est capable de déterminer s'il est inséré ou non. Dans ce cas, la notion de l'insertion sera centrée sur le diplômé (Vincens, 1997).

2. Le contexte de l'insertion des jeunes au Maroc

Engagé dans des politiques d'ajustement structurel puis dans la préparation de son économie à l'entrée dans la zone de libre-échange de l'Union européenne, le Maroc a enregistré d'importants dysfonctionnements au niveau de son marché du travail. Cette situation a engendré une croissance du nombre des diplômés en chômage et une précarisation de l'emploi et l'extension de l'emploi informel (Charmes, 2004).

Les efforts conjugués des pouvoirs publics et des partenaires sociaux au Maroc n'ont pas réussi à freiner un chômage qui sévit notamment parmi les jeunes diplômés de l'enseignement supérieur. En 2013, le taux de chômage des diplômés de niveau supérieur² a été évalué à 18,8% contre seulement 4,5% pour les non diplômés (Direction de la Statistique, 2013). Par ailleurs, les plus touchés par le chômage en 2013 sont les individus âgés entre 15 et 24 ans, et entre 25 et 34 ans avec respectivement 19,1% et 13,2% contre 5,6% et 2,3% respectivement pour les 35-44 ans et les 45 ans et plus (Direction de la statistique, 2013). Par ailleurs, le Maroc connaît un chômage structurel, où dominant le chômage de longue durée³ et le chômage de primo-insertion⁴. En effet, plus de la moitié des chômeurs (50,4%) sont des primo-demandeurs d'emploi (Direction de la statistique, 2013), d'une part ; et de l'autre part le chômage de longue durée représente 64,2% du volume global du chômage. Cette proportion atteint 79% parmi les diplômés de niveau supérieur (Direction de la statistique, 2013).

1 - En sciences sociales, on appelle trajectoire professionnelle la succession de passages d'un statut à un autre sur le marché du travail. L'analyse de ces passages appelés transitions professionnelles s'avère primordiale pour comprendre la dynamique du marché du travail.

2 - Niveau supérieur : Baccalauréat, diplômes de cadres moyens et diplômes d'enseignement supérieur (facultés, grandes écoles et instituts).

3 - Selon le Haut-Commissariat au Plan, le chômage de longue durée est défini comme celui dont la durée est supérieure ou égale à un an.

4 - Le chômage de primo insertion renvoie au chômage des individus qui ont effectué pour la première fois une recherche active de l'emploi

Le chômage des diplômés au Maroc trouve son explication, entre autres, dans le déficit global constaté au niveau de l'emploi, dans l'inadéquation des filières de formation aux exigences effectives de l'économie marocaine (Conseil Economique et Social, 2011) et dans la portée limitée des programmes publics de promotion de l'emploi (Idmaj, Taehil et Moukawalati) (DEPF, 2014).

Le Maroc a réagi en prenant une série de mesures en faveur de l'emploi et de l'insertion professionnelle. En 2000, l'Agence Nationale pour la Promotion de l'Emploi et des Compétences (ANAPEC) a été créée pour aider les jeunes dans leur accès au marché du travail en encourageant l'initiative. Dans ce sens, l'ANAPEC a conçu et implanté trois programmes d'envergure : (i) Moukawalati (mon entreprise), pour appuyer les projets de création d'entreprises par les jeunes ; (ii) Idmaj (intégration), avec l'objectif d'améliorer l'employabilité des jeunes diplômés par l'acquisition de nouvelles compétences professionnelles, notamment à travers une première expérience en entreprise ; et (iii) Taehil (qualification), qui vise à atténuer l'inadéquation entre le système éducatif et le marché du travail en permettant aux chercheurs d'emploi d'acquérir des compétences nécessaires pour occuper des postes d'emploi potentiels.

La performance du marché du travail a été traditionnellement évaluée à travers le taux de chômage et le taux d'emploi. Or, la diminution du taux de chômage n'est pas toujours synonyme d'une augmentation du taux d'emploi. Par exemple, la situation où les individus au chômage arrêtent de chercher un emploi se traduit par une baisse du taux de chômage mais peut aussi se traduire par une augmentation de l'inactivité dans le cas des chômeurs découragés par exemple, sans pour autant affecter le niveau de l'emploi. L'imperfection de l'information apportée par ces deux mesures statiques a engendré un intérêt accru envers l'analyse des flux sur le marché du travail à travers plusieurs états sur le marché du travail.

Les données disponibles actuellement issues des enquêtes statistiques (particulièrement celles du Département de la Formation Professionnelle) permettent une analyse plus fine de l'insertion professionnelle. Toutefois, il faut différencier entre les enquêtes à court terme d'entrée dans la vie active dites "enquêtes d'insertion professionnelle" et les enquêtes à moyen et long terme dites "enquêtes de cheminement professionnel". Les premières produisent une image instantanée de la situation professionnelle occupée sur le marché du travail quelques mois après la sortie de formation et permettent de mieux rendre compte du premier emploi et des éventuelles poursuites d'études, tandis que les enquêtes de cheminement professionnel retracent le parcours des sortants de l'enseignement supérieur sur une période plus longue (généralement plus de 18 mois). Les enquêtes de cheminement professionnel permettent d'informer sur l'évolution dans le temps des situations occupées sur le marché du travail.

3. Les dispositifs d'évaluation de l'insertion des autres pays

Dans un contexte de plus en plus concurrentiel, les institutions de l'enseignement supérieur ont besoin d'un dispositif statistique de mesure de l'insertion professionnelle. Ce dispositif d'information leur permet de communiquer régulièrement sur le devenir professionnel de leurs diplômés en fournissant des informations sur le processus de l'insertion professionnelle.

Les enquêtes d'insertion professionnelle permettent de comparer, du point de vue de l'efficacité externe, les spécialités et les voies de formation et d'examiner la contribution de l'enseignement supérieur au développement des compétences nécessaires pour répondre aux exigences du marché du travail (Reflex, HEGESCO, PROFLEX). Elles peuvent également étudier la dynamique de l'insertion professionnelle en reconstituant les parcours des diplômés à l'aide d'un calendrier et de distinguer, d'une génération à l'autre, les tendances de l'insertion professionnelle (CEREQ).

La caractérisation des emplois occupés sur le marché du travail par les diplômés représentent un élément fondamental d'appréciation de la qualité de l'insertion professionnelle (BibEr). Les enquêtes d'insertion professionnelle permettent aussi une auto-évaluation des compétences acquises et des modes d'enseignement. Ainsi, les diplômés peuvent évaluer la pertinence de la formation suivie en formulant des réponses subjectives aux questions posées (REFLEX). Les enquêtes d'insertion professionnelle servent également à favoriser la reddition de comptes des universités (Almalauréa), ce qui permet de reconnaître l'excellence et l'apport de l'université à la société.

Cette analyse comparative révèle plusieurs dimensions constitutives de l'insertion professionnelle, à savoir, les performances éducatives, la durabilité de l'emploi, la qualité des emplois occupés, le degré de satisfaction de la formation, etc. Une dimension à elle seule ne suffit pas pour appréhender l'insertion professionnelle. D'où la nécessité d'une approche multidimensionnelle à même d'embrasser d'une même vision l'ensemble des dimensions susceptibles d'influencer l'insertion professionnelle.

II. Les objectifs et avantages de l'enquête sur l'insertion

Le dispositif d'évaluation de l'insertion mis en place par l'Instance Nationale d'Evaluation, et dédié à l'ensemble des universités marocaines, fixe des objectifs ambitieux s'articulant autour de cinq axes principaux :

- Contribution à la construction d'un **système national d'information** dédié à l'évaluation de l'insertion des diplômés de l'enseignement supérieur
- L'assurance d'une **meilleure visibilité des articulations formation – emploi** qui constitue un des grands défis du secteur de l'enseignement supérieur national
- Proposer des pistes d'orientation en faveur d'une politique d'enseignement supérieur **efficace et efficiente**
- Contribution au montage d'un **observatoire national des entrées dans la vie active**
- Une évaluation exhaustive de la **relation à l'emploi** des lauréats des universités conditionnellement à leurs caractéristiques individuelles, sociodémographiques et leurs filières et spécialités de formation.

Ce dispositif a été expérimenté dans le cadre d'une enquête de cheminement pilote en 2012, auprès d'un échantillon de diplômés de trois universités marocaines. Il a permis de mesurer quantitativement et qualitativement l'insertion professionnelle en dynamique des lauréats des universités concernées ainsi que d'éclairer la relation formation-emploi avec toute la rigueur qu'exige une approche pertinente de ce type d'analyse.

La généralisation de cette expérience au plan national présente plusieurs avantages par rapport aux études sectorielles et ponctuelles qui existent jusqu'à présent. En éliminant tous les biais liés aux particularités de chaque établissement universitaire, l'enquête nationale permettra de renseigner sur la relation formation-emploi, d'identifier le rôle des facteurs macro-économiques et sociaux dans la détermination de l'emploi des diplômés.

III. L'approche multidimensionnelle de l'insertion professionnelle

La mesure de l'insertion professionnelle est plus complexe qu'elle n'en a l'air et elle ne peut jamais être appréhendée par une mesure unidimensionnelle. En effet, le benchmark international présenté auparavant révèle plusieurs dimensions constitutives de l'insertion professionnelle. L'insertion est ainsi analysée comme un processus dynamique décrivant le passage par plusieurs situations professionnelles dans le temps (emploi à temps plein, emploi à temps partiel, stage, chômage, inactivité,...). Ces situations interfèrent, à leur tour, avec plusieurs composantes, comme les performances éducatives, la durabilité de l'emploi, la qualité des emplois occupés, le degré de satisfaction de la formation, etc.

D'où la nécessité d'une approche multidimensionnelle à la fois pour décrire le processus d'insertion dans son ensemble ou pour décrire les diverses dimensions susceptibles d'influer sur l'insertion professionnelle et qui font l'objet d'une analyse détaillée de la relation formation-emploi.

Les différentes théories explicatives de la relation formation-emploi et de la transition entre le système éducatif et le marché du travail, trouvent leur fondement dans la théorie néoclassique. On distingue deux grandes théories, la première est celle du « capital humain » de Becker (1964), selon laquelle le revenu s'explique par la productivité elle-même résultant de l'investissement en capital humain. Une relecture de cette théorie faite par Eicher J.C (1979) recommande l'ajout d'autres variables (âge, sexe, territoire, durée du travail, ...) pour que la formation explique le niveau de revenu.

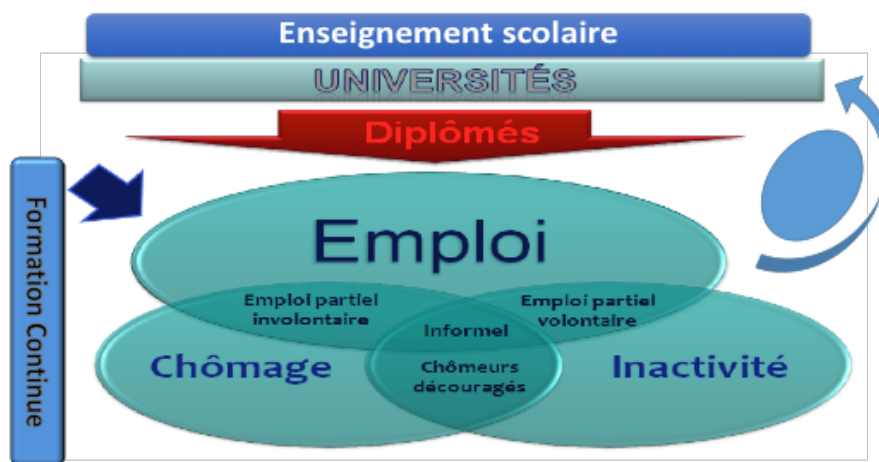
La deuxième théorie est celle du « filtre » (Spence, 1973) qui postule que le niveau de formation joue le rôle d'un filtre permettant d'identifier les individus les plus productifs. La productivité marginale étant liée à des caractéristiques innées, elle est révélée par un signal donné par le niveau de formation. Les travaux de Lemistre P. (2000), à propos de cette théorie, montrent que tous les individus ne sont pas en mesure d'atteindre les mêmes niveaux d'éducation. Le niveau d'éducation (à travers le diplôme) est un outil utilisé comme filtre permettant aux employeurs d'identifier les personnes les plus aptes pour les emplois disponibles.

Il ressort de ces deux théories toute la flexibilité qui existe pour expliquer de différentes manières la relation entre la formation et le niveau d'insertion que ça soit au niveau du choix des variables ou du type de la relation existante, d'où l'importance d'opter pour une approche multidimensionnelle de l'insertion professionnelle. Un seul indicateur, comme le taux d'emploi ou le niveau de salaire, ne suffit pas à rendre compte de l'ensemble

des composantes de l'insertion professionnelle. D'un côté, le taux d'emploi peut masquer une insertion de « mauvaise qualité », sur des postes peu qualifiés, avec des salaires plus bas ou sur des contrats à durée déterminée. D'un autre côté, vu la corrélation existante entre niveau de formation, filières de formation, taux de chômage, vitesse d'accès à un emploi à durée indéterminée et niveau de rémunération, une approche unidimensionnelle ne peut en aucun cas répondre aux différents questionnements : Quelle vitesse d'accès à l'emploi ? Quel risque de connaître des difficultés durables d'insertion ? Quels types de postes possibles, avec quelle rémunération et quelles conditions d'emploi ?

Schéma 1
L'évaluation multidimensionnelle de l'insertion

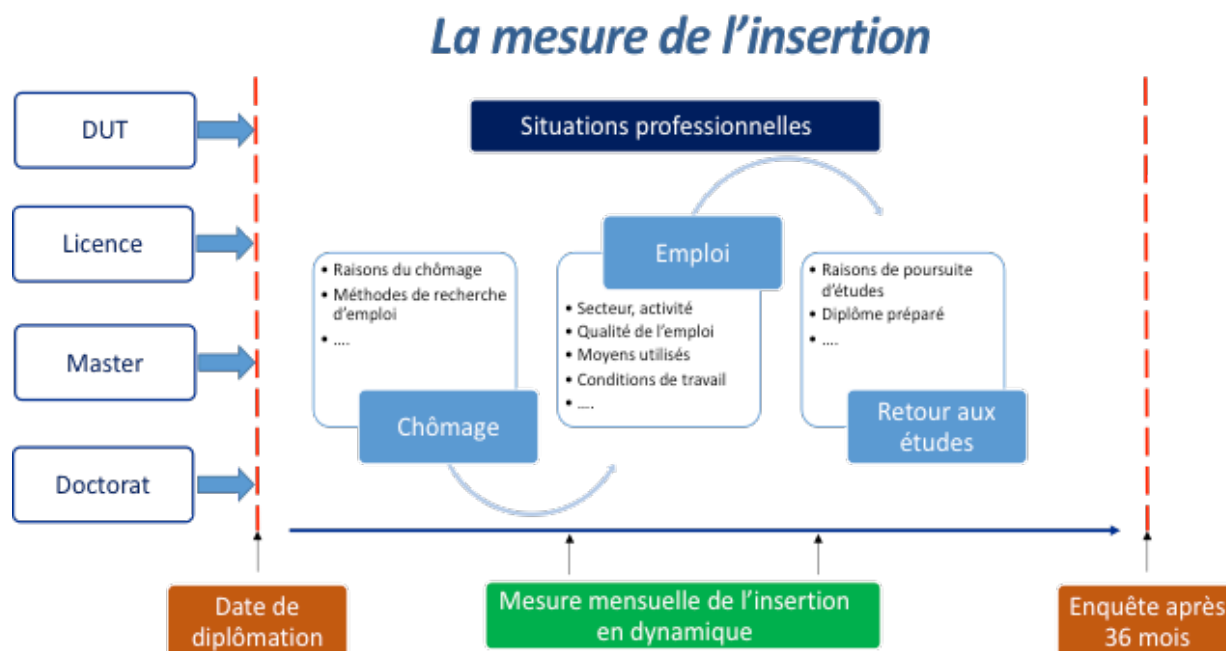
Une évaluation multidimensionnelle de l'insertion



IV. L'approche méthodologique de l'évaluation de l'insertion

La mesure de l'insertion des diplômés du supérieur repose sur un dispositif d'enquête de type longitudinal qui retrace de façon rétrospective, moyennant un calendrier mensuel sur une durée de 36 mois, les différents parcours d'insertion des lauréats des universités. Tous les diplômes de l'université ont été concernés par l'enquête.

Schéma 2
La mesure dynamique de l'insertion



1. L'instrument de mesure : Les questionnaires

Le questionnaire principal de l'enquête a été réalisé par l'équipe de l'Instance Nationale d'Évaluation. Sa conceptualisation a tenu compte des questions nouvellement mis au point par la littérature sur l'insertion telles que les phénomènes de déclassement, de précarité, d'emploi ainsi que la mobilité régionale.

Le dispositif construit par l'INE s'est inspiré de plusieurs expériences internationales dans le domaine de l'évaluation de l'insertion professionnelle des diplômés du supérieur, telles l'enquête génération CEREQ, France, l'enquête Reflex, ou encore l'enquête d'insertion de la Tunisie,...

Une démarche participative a été adoptée en associant les universités et les étudiants dans la conception des questions par l'organisation d'ateliers. Les questionnaires ont été ensuite validés afin d'être utilisés comme outil principal d'évaluation.

Le questionnaire, rédigé en français, contient 220 questions, il se compose de six grandes parties qui se subdivisent en sous-questions selon les thématiques suivantes :

- **Parcours général** : fournit des informations sur le parcours scolaire du lauréat depuis le collège jusqu'à l'obtention du diplôme ;
- **Calendrier professionnel** : définit l'itinéraire professionnel en reconstituant de façon séquentielle dans le temps les différentes situations rencontrées depuis la sortie du système de formation : emploi, chômage, poursuite d'études, inactivité ;
- **Questionnaire emploi** : analyse la situation de travail et comment le lauréat, s'est trouvé du travail ainsi que le lien entre sa formation et son emploi ;
- **Questionnaire chômage** : cerne la réalité du chômage dans sa durée et dans le motif à son origine ;
- **Questionnaire études** : repère les formations susceptibles d'être valorisées sur le marché du travail, que ce soit les diplômes obtenus en formation initiale ou continue ;
- **Fiche signalétique** : contient les informations personnelles du lauréat, la catégorie socioprofessionnelle et le niveau d'instruction de ses parents.

Les dimensions principales abordées par le questionnaire principal tel qu'il a été élaboré par l'INE sont les suivantes :

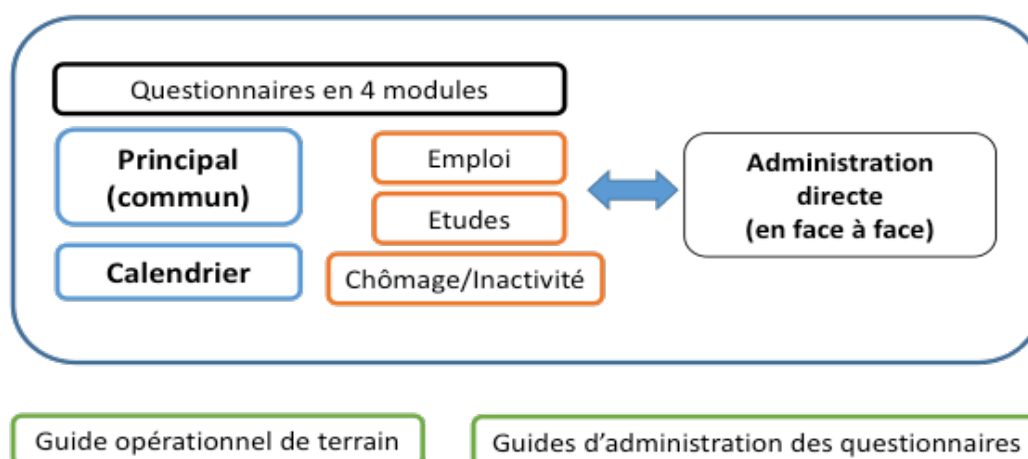
- Caractéristiques personnelles et socioéconomiques
- Parcours scolaire et universitaire
- Transition études/emploi
- Situation à court terme et moyen terme des diplômés
- Caractéristiques de l'emploi et les compétences mises en œuvre
- Formation continue
- Perception de la formation universitaire et de la trajectoire professionnelle

Il faut aussi souligner que l'enquête comporte un calendrier sur 34 mois autour des quatre principales situations que nous pouvons identifier lors d'un processus d'insertion : retour aux études, chômage, emploi et inactivité.

La méthodologie utilisée pour l'élaboration du questionnaire s'appuie sur une vision temporelle hiérarchique des séquences renseignées. Ainsi, on distingue trois temporalités : avant la formation universitaire, pendant la période de formation et après l'obtention du diplôme. Chaque séquence produit les indicateurs nécessaires pour l'analyse à posteriori.

Schéma 3
Le dispositif d'évaluation proposé

Le dispositif : les questionnaires et les guides



Une fois rempli sur le terrain par la méthode du face à face, le questionnaire fait l'objet d'une opération de saisie des questions sur une plateforme web, développée par l'équipe informatique de l'INE, et qui permet de sauvegarder les données en temps réel en les acheminant vers le serveur de l'INE avec un niveau de sécurité et de confidentialité élevés.

2. Les données sur les diplômés de l'enseignement supérieur

Aujourd'hui dans l'enseignement supérieur, le seul système qui permet d'avoir un suivi individuel des étudiants est APOGEE (Application pour l'Organisation et la Gestion des Enseignements et des Etudiants).

Cette application d'origine française et adaptée pour le système marocain, est utilisée depuis 2003 pour accompagner la réforme pédagogique. C'est une application multi-domaines qui gère actuellement plusieurs modules allant de la première inscription de l'étudiant dans l'établissement, jusqu'à l'édition du diplôme obtenu en passant par tout le parcours académique (caractéristiques sociodémographiques, notes, validation de modules, etc.).

Toutefois, le système APOGEE n'est pas encore généralisé à l'ensemble des universités marocaines et

jusqu'à ce moment, il n'existe pas de système d'information national intégré qui permet de renseigner sur la situation des diplômés de toutes les universités marocaines. Aussi, les informations de contact des diplômés sont parfois inexistantes ou mal renseignées, d'où la nécessité de systématiser la collecte des informations de contact des étudiants par l'université avec une mise à jour régulière, de renseigner leurs informations individuelles depuis l'accès à l'université jusqu'à la sortie et de les harmoniser pour l'ensemble des établissements d'enseignement supérieur.

Dans le cas où les établissements ne disposent pas des adresses et des coordonnées actualisées des diplômés, d'autres méthodes alternatives permettent de retracer ces lauréats (ANAPEC, réseaux sociaux), ou tout autre moyen susceptible de fournir leurs coordonnées.

3. La méthodologie d'échantillonnage

3.1. Choix de la population

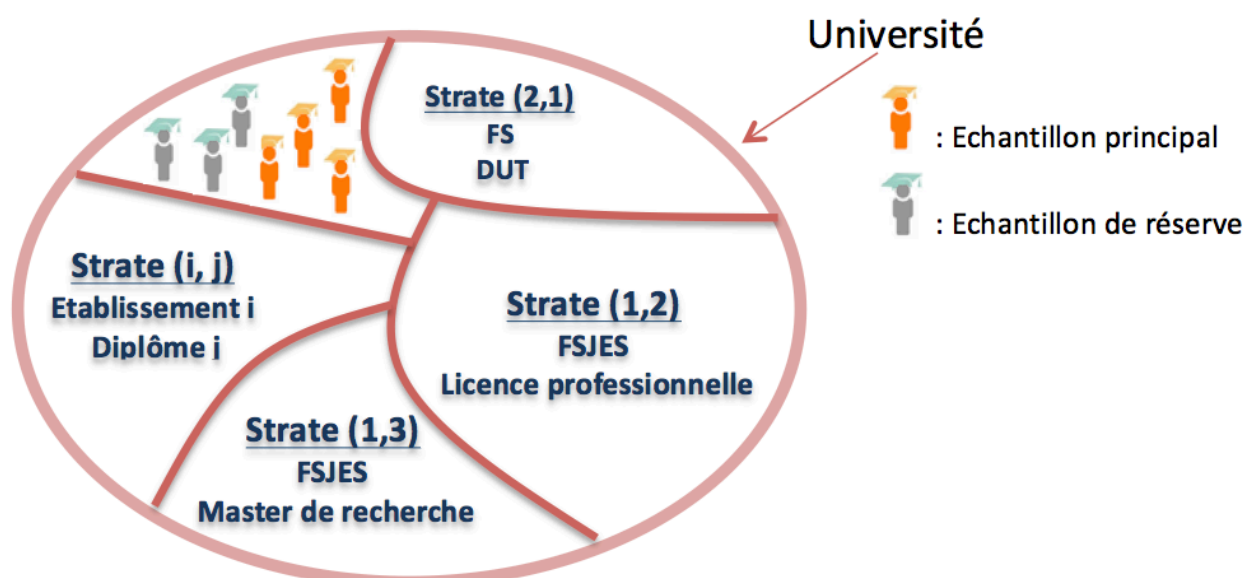
Le choix de la population cible porte sur les diplômés d'une année universitaire donnée qui ont achevé un cycle complet d'études universitaires, sanctionné par un diplôme de 1er, de 2ème ou de 3ème cycle, qu'ils aient choisi d'arrêter leurs études ou non. Tous les établissements de l'université sont concernés, qu'ils soient à accès ouvert ou à accès régulé. Par ailleurs, l'ensemble des diplômes présents dans le système universitaire national sont retenus, à savoir les diplômes universitaires techniques (DUT), les diplômes de licence fondamentale (LF) et professionnelle (LP), les diplômes de master de recherche (MR) et de master spécialisé (MS), les diplômes de l'ENCG, etc.

3.2. Méthode d'échantillonnage

La connaissance préalable de la population cible permet d'opter pour une méthode d'échantillonnage stratifié à allocations proportionnelles aux tailles des strates de manière à être représentatif par établissement et par type de diplôme. La taille de l'échantillon est calculée selon un seuil de signification et une marge d'erreur fixés au départ selon les standards internationaux et ce, dépendamment du niveau d'inférence souhaité (établissement, filière, etc.). En effet, l'échantillon est stratifié par établissement et par type de diplôme. Ensuite, un tirage systématique est utilisé au sein de chaque strate afin d'obtenir un échantillon représentatif de la population mère (voir schéma ci-dessous).

Un échantillon de réserve (composé de jumeaux statistiques) est prévu afin de remédier au problème de non réponses éventuelles.

Schéma 4
Illustration de la technique d'échantillonnage stratifié à allocations proportionnelles



4. Collecte, saisie et préparation de la base de données

Les données sont collectées principalement par la méthode du face à face. L'opération terrain est menée par des enquêteurs qui sont contrôlés par des superviseurs. Cette méthode est celle qui permet de réduire le plus le biais de non réponse. Une équipe d'enquêteurs se déplace dans toutes les régions du Maroc afin d'administrer le questionnaire aux lauréats sélectionnés dans l'échantillon principal. La durée de l'entretien est estimée à environ 45 minutes.

Les questionnaires sont ensuite sauvegardés et acheminés vers la salle de saisie pour entamer l'étape de saisie des données. La saisie des questionnaires est faite via une plateforme web développée par les informaticiens de l'INE. Les données saisies sont sauvegardées dans un serveur spécial dédié à cet effet et qui se trouve dans les locaux de l'INE. Toutes les données personnelles saisies sont garanties du droit à la confidentialité totale et au respect des codes déontologiques en vigueur.

La préparation d'une base de données finale prête à l'exploitation constitue l'une des principales étapes de l'enquête. Un nettoyage à posteriori est effectué pour corriger les éventuelles erreurs de saisie et normaliser à la fois la nomenclature des catégories renseignées et des variables finales. Cette étape est assurée par l'équipe de travail lors de la vérification de la validité des réponses de chaque répondant. Certains diplômés peuvent être interrogés une deuxième fois par voie téléphonique pour confirmer quelques-unes de leurs réponses et pour renseigner des champs restés vides lors du premier entretien.

V. Les résultats de l'enquête

Les résultats obtenus à travers une première expérimentation de ce dispositif offrent quelques indications sur le rendement externe des universités. Cette enquête a concerné 385 diplômés de l'Université Hassan Ier de Settat, 736 diplômés de l'Université Hassan II de Casablanca et 504 diplômés de l'Université Mohammed V - Agdal de Rabat, tous de la promotion 2009, soit 1625 diplômés enquêtés au total. La population de référence de l'enquête est « tous les lauréats des trois universités diplômés à une même date (Juin 2009) et qui sont suivis de façon longitudinale pendant une période de 34 mois ».

Même si l'enquête n'a concerné qu'un échantillon d'universités de l'axe Rabat-Casablanca-Settat, ses résultats ont permis quand même, à travers l'analyse du parcours longitudinal des jeunes sur un calendrier mensuel qui retrace leurs différentes situations professionnelles (emploi, chômage, études ou inactivité), d'avoir une idée très sommaire sur la dynamique de l'insertion professionnelle des diplômés de l'université dans cette région du Maroc.

Les résultats montrés ci-après ne représentent qu'une partie des indicateurs ressortis de cette enquête très riche en résultats. Les résultats détaillés sont disponibles à l'Instance Nationale d'Évaluation et feront l'objet d'autres publications futures.

1. L'évolution de l'insertion dans le temps

Un taux d'insertion entre la date de sortie et la date de l'enquête a été calculé mensuellement. Il rapporte le nombre de lauréats en situation d'emploi au nombre de lauréats actifs à chaque mois. Les lauréats en situation d'études ou d'inactivité ne sont pas pris en compte dans le calcul de ce taux. Pour l'ensemble de la population des lauréats enquêtés, le taux d'insertion enregistré 1 mois après la sortie de l'université est de 38,6%. Ce taux s'améliore progressivement pour atteindre 81% après 34 mois de la sortie.

Par ailleurs, le taux d'insertion varie selon l'établissement de formation. En effet, les diplômés de l'ENCG, de la Faculté de Médecine Dentaire et de l'EMI enregistrent un taux d'insertion de 100%, 34 mois après leur sortie du système universitaire, tandis que les lauréats de la Faculté Polydisciplinaire de Khouribga sont les moins bien insérés avec un taux de 63%. Les lauréats de la FSJES s'insèrent mieux (79%) que ceux de la Faculté des Sciences (77%), de la FLSH (74%) et de la FST (70%).

Les lauréats ayant un diplôme d'ingénieur, un diplôme de l'ENCG ou encore un doctorat s'insèrent tous au bout de 34 mois, suivis des titulaires d'un Master de recherche (93%) et d'un Diplôme Universitaire de Technologie (DUT) (93%). Les détenteurs d'un DESA ou d'un Master spécialisé affichent des taux d'insertion de 92% et 90% respectivement. Pour les licenciés, les taux d'insertion sont de 78% pour la licence professionnelle et de 69% pour la licence fondamentale.

2. Les séquences temporelles d'accès à l'emploi

Avant d'obtenir leur premier emploi, les jeunes lauréats des trois universités qui ont trouvé un emploi attendent en moyenne 9 mois environ dont 4 mois en chômage, 4 mois dans les études et 1 mois environ dans l'inactivité. Cependant, les lauréats qui n'ont jamais travaillé tout au long des 34 mois qui suivent la sortie (32% des lauréats), passent en moyenne 10 mois dans le chômage, 21 mois environ dans les études et le reste de la période dans l'inactivité.

L'analyse par établissement montre que les lauréats de l'ENCG sont ceux qui attendent le moins dans le chômage (moins de 2 mois). Les diplômés des facultés à accès ouvert attendent plus de 4 mois en chômage pour obtenir leur premier emploi. Lorsqu'on considère toutes les situations possibles d'attente avant le premier emploi (chômage, études et inactivité), le délai moyen d'attente de ces derniers s'allonge à plus de 9 mois.

3. La qualité des emplois occupés

Type d'emploi : Prédominance du secteur privé

Le secteur privé est le premier employeur des lauréats universitaires. Ainsi, 62% des lauréats des trois universités ayant obtenu un premier emploi, débutent leur carrière dans des entreprises privées. Le secteur public n'embauche que 32% des diplômés qui se sont insérés dans leur premier emploi. Le reste des employés travaille dans le secteur semi-public et seulement 1% s'installent à leur propre compte.

Le type de contrat : des emplois majoritairement stables

Les lauréats des trois universités sont embauchés pour 55% selon un CDI dans leur premier emploi. 26% sont recrutés en tant que fonctionnaires et seulement 11% signent un contrat CDD. En revanche, la part des lauréats embauchés selon un CDD diminue de 5 points entre le premier et le dernier emploi au profit du contrat CDI dont la part augmente pour atteindre 61%.

Le niveau de salaire

L'examen des salaires des primo-demandeurs d'emploi montre que le salaire net moyen se situe autour de 6400 DH/mois avec des disparités entre les diplômés et les établissements. Entre le premier et le dernier emploi, le salaire moyen perçu augmente de 900 DH en moyenne.

Conclusion

L'enquête sur l'insertion professionnelle des diplômés de l'enseignement supérieur de l'INE est la première enquête du genre à être testée sur 3 universités marocaines. Les résultats confirment la rigueur scientifique avec laquelle a été menée l'approche méthodologique de construction du dispositif. L'INE travaille dès maintenant à la généralisation de ce dispositif à l'ensemble des universités et établissements d'enseignement supérieur du Maroc.

En effet, l'institutionnalisation et la généralisation de ce dispositif au niveau de tous les établissements présente plusieurs avantages par rapport aux études sectorielles et ponctuelles qui existent jusqu'à présent. Elle permettra de comprendre la dynamique de la relation formation-emploi au niveau national, ce qui constitue un des grands défis aujourd'hui du secteur de l'Enseignement Supérieur au Maroc, ainsi que d'identifier le rôle des facteurs macro-économiques et sociaux dans la détermination de l'emploi des diplômés.

Par ailleurs, l'analyse des données relatives à l'insertion des diplômés de l'université, quoique ne concernant qu'un échantillon réduit des universités de l'axe Rabat-Casablanca-Settat, fait ressortir quelques constats importants :

1. Les diplômés de l'université de cette région sont plus actifs dans la durée et leur taux d'activité dépasse les 75% après 34 mois de leur diplomation. Les taux d'activité sont en général plus élevés pour les diplômés des établissements à accès régulé ou ceux qui ont fait des études dans des filières à caractère professionnel.
2. Les taux d'insertion des diplômés des établissements universitaires à accès régulé sont de loin meilleurs que ceux des diplômés des établissements à accès ouvert.
3. La durée moyenne de chômage avant de trouver un premier emploi est de 4 mois, avec toutefois

des disparités entre les diplômés et les établissements. Aussi, les difficultés d'insertion ont tendance à s'estomper au fur et à mesure que le niveau d'éducation s'élève.

Finalement, et afin d'avoir une idée sur les attentes des offreurs de travail par rapport aux diplômés du supérieur et avoir une meilleure visibilité des articulations Formation-Emploi du point de vue à la fois de la demande et de l'offre, il est nécessaire de compléter ce dispositif par des enquêtes auprès des autres intervenants dans l'insertion des étudiants (les entreprises, les universités, les enseignants, ...).

Traduction du Résumé **Abstract Translation**

Saad-Ellah Berhili, Fatima Berahou, Sanaa Cheddadi, Nawal Zaaï

INTEGRATION SURVEYS IN MOROCCO : THE PERFORMANCE OF UNIVERSITY

The professional integration into the workplace is a key component in understanding the relationship between training and employment. It constitutes a relevant indicator of the external efficiency of the whole educational and training system. The surveys on professional integration in Morocco refer mainly to the employment surveys made by the Haut-Commissariat au Plan. These constitute an evaluation tool of the macro-economic situation of the employment, at very aggregated levels.

This paper presents the monitoring system that the National Authority of Evaluation carried out to evaluate the integration of university graduates in the labor force. This system was implemented for the first time on 3 Moroccan universities covering all their faculties and schools (open and regulated access). It can measure with high precision the degree of integration of these university graduates as well as understand the dynamics of the relationship training-employment among this population who left the educational system with a diploma. The methodology and the scientific tools used to construct this monitoring system give it an important value and a good comparability with the monitoring systems of the most developed countries who are interested in the evaluation of the professional integration of their university graduates.

This monitoring system took the form of a pilot survey addressed to a representative sample of graduates from these 3 universities. This survey is based on a 3-year monthly retrospective calendar and deals with very important and emerging subjects that are not easy to assess today. It's all about collecting new data about occupational integration of young graduates, studying specific issues like spatial and temporal mobility, analyzing the problem of job insecurity, or over-education and the self-fulfillment of graduates in their professional career.

The institutionalization and the generalization of this evaluation tool to all the universities and components of higher education at a larger scale (engineer schools, private sector, post-graduate vocational training) has several advantages compared to the limited sectorial studies realized until now. This will actually guarantee a good understanding of the relationship training-employment at a national level and constitutes today one of the big challenges of the higher education sector in Morocco, as well as identify the role of macro-economic and social factors on determining graduates' employment.

Bibliographie et Références

- Biggeri, L., Bini, M. and Grilli, L. (2001). The transition from university to work: a multilevel approach to the analysis of time to obtain the first job, *Journal of the Royal Statistical Society (Series A)* 164:293–305.
- Charmes, J. (2004). Secteur informel et emploi informel au Maghreb: un état des lieux et des connaissances par rapport à d'autres expériences dans le monde. Communication au colloque international «La question de l'emploi en Afrique du nord, tendances récentes et perspectives 2020.
- Coupié, T., & Mansuy, M. (2004). L'insertion professionnelle des débutants en Europe : des situations contrastées. *Economie et Statistique* , 378-379.
- DEPF. (2014). Direction des Etudes et des Prévisions Financières. Rapport Economique et Financier: Défis et opportunités de l'économie marocaine dans un contexte international changeant.
- Gauthier, M. (1997). Les jeunes et le travail : un terrain mouvant Les jeunes et le travail, dans J-P Dupuis et A. Kuzminski (dir.). (G. Morin, Éd.)
- Giret, J. F. (2000). Pour une économie de l'insertion professionnelle des jeunes. CNRS éditions.
- Goyette, M. (2006). Préparation à la vie autonome et insertion socioprofessionnelle des jeunes pris en charge par l'État : quelles interventions ? *Sociétés et jeunesses en difficulté* (1).
- Hamel, J. (2003). Pour une vue longitudinale sur les jeunes et le travail. *Cahiers internationaux de sociologie* , CXV, 255-268.
- Kalachek, E. (1980). les enquêtes longitudinales et le marché du travail des jeunes. Dans *Le chômage des jeunes:causes et conséquences* (p. 172). Paris: OCDE.
- Laflamme, C. (1996). Inflation des diplômes et insertion professionnelle des jeunes : situation des diplômés du secondaire professionnel et du cégep technique sur le marché de l'emploi. *Revue des sciences de l'éducation* , XXII (1), 47-72.
- Laflamme, C., & Baby, A. (1993). Integration of youth into the labor market : Some theoretical considerations on a structural/conjunctural issue. *Alberta Journal of Educational Research* , XXXIXd, 97-109.
- Maillard, F. (2005). L'ambivalence de la politique éducative : le CAP entre déclin et relance. *Formation Emploi* (89).
- Mansuy, M. (2001). Génération 92 : un regard renouvelé sur les parcours de débutants ? Introduction au dossier. *Formation Emploi* (73), 23-29.
- Tansel, A., & Tasçi, H. (2004). Determinants of unemployment duration for men and women in Turkey. IZA discussion Paper (1258).
- Tanguy, L. (1986). L'introuvable relation formation et emploi. *La Documentation Française* .
- Tremblay, D.-G. (1994). Chômage, flexibilité et précarité d'emploi : aspects sociaux. Dans S. L. F. Dumont, *Traité des problèmes sociaux* (pp. p. 623-652). Québec: Institut Québécois de Recherche sur la Culture.
- Trottier, C., Perron, M., & Diambomba, M. (1995). Les cheminements scolaires et l'insertion professionnelle des étudiants de l'université. *Perspectives théoriques et méthodologiques* , 1-14.
- Vincens, J. (1981). Problématique générale de l'insertion dans la vie active. Conférence prononcée au Colloque sur l'insertion professionnelle à la sortie des études postsecondaires. Louvain: Institut des sciences du travail, Université catholique de Louvain.
- Vincens, J. (1997). L'insertion professionnelle des jeunes. A la recherche d'une définition conventionnelle. *Formation Emploi* (60), 21-36.

ENJEUX ET EXPÉRIENCES DE L'ÉVALUATION DANS LE SUPÉRIEUR

Mustapha Bennouna *

Résumé

L'évaluation universitaire en Afrique suscite de l'intérêt ainsi qu'un grand débat.

Des formations à l'autoévaluation et à la rédaction des plans stratégiques d'universités ou de leurs composantes connaissent de l'attrait, et préfigurent la mise en place de contrats pluriannuels.

Certains pays ont promulgué des lois incitant les responsables universitaires à la pratique de l'autoévaluation, et annonçant la mise en place d'agences nationales ou régionales d'évaluation.

Dans cette communication, l'objectif est de livrer un premier retour d'expériences sur des études de cas d'universités africaines impliquées récemment dans ce processus.

I. Missions des Universités

Les universités commencent à diversifier leurs missions qui traditionnellement étaient limitées à la formation et à la recherche et qui s'élargissent de plus en plus à : l'insertion professionnelle, l'innovation, l'internationalisation, la mobilité (entrante et sortante), ...

Dans le cadre de ces missions, les universités sont amenées à adapter et à adopter des standards internationaux pour l'ensemble de leurs activités afin d'être « lisibles et visibles » à l'international et de pouvoir affirmer leur spécificité : la diversité n'étant pas un obstacle mais une richesse. L'assurance qualité est un des outils majeurs pour atteindre de tels objectifs.

II. Vision de l'Assurance Qualité

Le changement dans l'enseignement supérieur et la mise en place d'un processus d'assurance qualité national et international sont motivés par :

- la fréquentation croissante de l'enseignement supérieur;
- le souci d'améliorer les normes pour l'emploi au niveau local et national et la reconnaissance internationale;
- la nécessité de s'adapter à un environnement concurrentiel toujours plus marqué;
- la mise en œuvre du processus de Bologne ;
- les dispositions de la législation nationale.

Tous ces éléments contribuent à la réforme des programmes et à l'assurance qualité, qui sont intimement liées.

Les étudiants ont besoin de standards pour leur mobilité et leur évolution professionnelle.

Les employeurs ont besoin d'informations et de standards internationaux pour leurs recrutements.

Cette approche permet de lutter contre les « usines à diplômes » et d'instaurer une culture d'amélioration continue dans les formations

* Conseiller du Président de l'Université Euro-Méditerranéenne de Fès.

Bases de l'assurance qualité :

L'assurance qualité nécessite la mise en place d' « agences » indépendantes et l'existence de « parties prenantes »

Les outils de l'assurance qualité sont :

- L'autoévaluation : Processus continu d'audit interne, avec une boucle pour l'amélioration continue
- L'évaluation externe : Audit par une équipe d'experts qui établit un rapport non contraignant
- L'accréditation : Jugement de conformité avec un guide externe de références (ISO, CTI, ABET...) et qui aboutit au décernement d'un label de qualité
- L'habilitation : Décision gouvernementale basée (ou non) sur les résultats de l'évaluation/accréditation

L'assurance qualité a connu un développement remarquable dans l'espace européen de l'enseignement supérieur dont le modèle LMD a été adopté par un grand nombre de pays africains. Ainsi ont été créées des agences nationales (ou régionales) d'évaluation ou d'accréditation qui ont défini des standards et des procédures d'évaluation pour les universités (et écoles). De même un registre des agences conformes/habilitées à opérer a été développé (EQAR)

III. Cas du Maroc : Loi 01-00 (Mai 2000)

Article 77 : Le Système de l'Enseignement Supérieur est soumis, dans sa globalité, à une évaluation régulière, portant sur sa rentabilité interne et externe (auto évaluation et sondage des partenaires).

Article 78 : Les Etablissements d'Enseignement Supérieur publics et privés mettent en place un système d'auto-évaluation.

Article 79 : Pour la réalisation des audits et de l'évaluation, il sera procédé à la création d'instances spécialisées de régulation notamment une instance nationale d'évaluation et un observatoire pour l'adéquation des enseignements supérieurs à l'environnement économique et professionnel.

La mise en œuvre de ces articles reste cependant partielle et incomplète

1. Maroc : Projets de coopération en assurance qualité

Le Maroc a connu l'organisation d'un grand nombre de séminaires et de formations afin de découvrir les différentes expériences internationales : USA, Canada, France, Belgique, Espagne, Pays bas, Finlande,...

Des projets de coopération avec des pays tels que la Canada (PRICAM en 1996) et européens (TEMPUS à partir de 2003) , ont permis aux universités de se constituer en réseau (national et international) pour participer et gérer plusieurs actions qui ont abouti à la formation de ressources humaines, la mise en place de cellules qualité, l'organisation de clubs qualité, l'enseignement de modules qualité, le développement de filières (licence et Master) en assurance qualité,....

En particulier, le programme Tempus de l'Union européenne qui vient de s'achever, a œuvré en faveur de la mise en place de réseaux de coopération entre les universités européennes et leurs homologues marocaines, et a été parfaitement positionné pour les aider améliorer la qualité de leurs services et à mettre en place des mécanismes d'assurance qualité. Le Maroc a ainsi acquis un bon savoir-faire notamment en matière d'évaluation comme outil d'assurance qualité,

Ci joints une liste non exhaustive de quelques projets de coopération en assurance qualité auxquelles des universités marocaines ont contribué :

PRICAM : Assurance qualité ISP 9000. Coopération Maroc-Canada : Formation d'un grand nombre de qualitiens prêts à passer les examens d'auditeur qualité et aptes à accompagner des entreprises dans leur démarche qualité ou de certification ISO

FOREVALE : Formation à l'Evaluation externe (2006), initiée par le Ministère de l'Enseignement Supérieur (projet européen TEMPUS) :

- Formation de Ressources Humaines dans l'évaluation externe.
- Différentes universités marocaines ont participé à ce projet.

TEMPUS : EVAL- Evaluation de l'Université Med V (2005-2007)

QAREM : Integration of QA System into the New University Reform in Morocco (2005-2008).

- Evaluation institutionnelle pilote
- Deux phases, interne et externe, couvrant ESR et la bonne gouvernance;
- Développer des mécanismes internes d'assurance qualité (mise en œuvre du LMD)
- Vers l'élaboration du système AQ

TEMPUS : « Management de l'Université par la Qualité » Université Cadi Ayad de Marrakech UCAM (2008-2010)

- Mise en place d'un Système de Management de la Qualité (SMQ)
- Certification ISO 9001-Version 2000 des services administratifs
- Sensibilisation et diffusion du SMQ
- Formation d'un « Responsable Qualité » par Université.

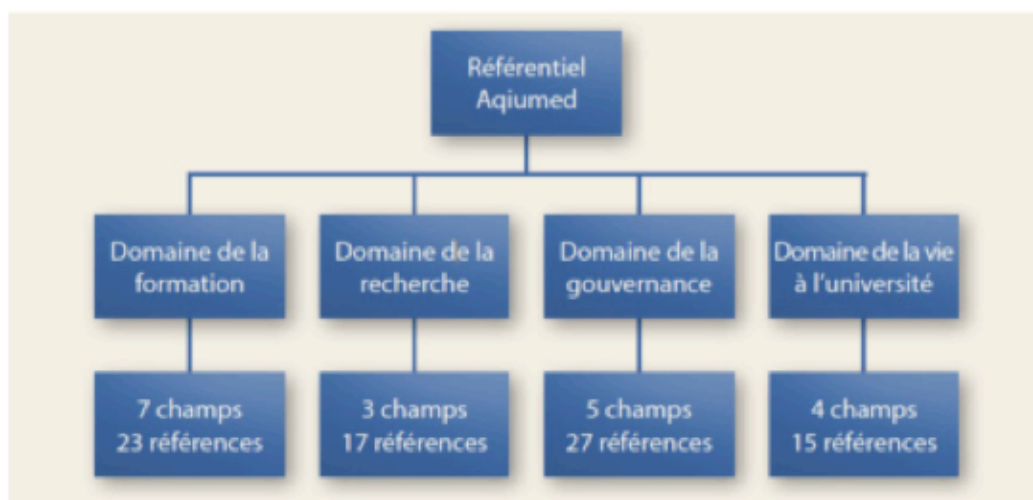
AQI-UMED

- TEMPUS « mesure structurelle » (2010-2013)
- Thématique : assurance qualité des universités
- Axe central : évaluation interne des universités
- Partenaires sud : Universités d'Algérie, Maroc et Tunisie
- Partenaires nord : agences d'évaluation européennes (Belgique, Espagne et France)
- Autres partenaires : AUF, CIEP
- Des experts indépendants

Le résultat atteint : Publication du Référentiel Aqi-Umed adopté actuellement par un grand nombre d'universités du Maghreb et par l'Agence Universitaire de la Francophonie (AUF)



Référentiel AqiUmed



2. Expériences d'évaluation menées par des universités marocaines

D'un autre côté, quelques universités maghrébines ont mené leur propre autoévaluation (volontairement ou dans le cadre de l'appel d'offres lancé par le bureau AUF Maghreb) et ont invité des experts pour effectuer l'évaluation externe.

2.1. Cas de l'Université Abdelmalek Essaadi (UAE 2010)

Objectif :

- Dresser le bilan des deux mandats du Président
- Etre en phase avec la contractualisation (plan d'urgence 2009 – 2012)
- Instaurer la culture et les pratiques d'évaluation
- Evaluer les modalités d'élaboration et de mise en œuvre de la stratégie de Gouvernance

L'autoévaluation a été menée selon le référentiel de l'Agence de l'Évaluation de la Recherche et de l'Enseignement Supérieur (AERES France) et l'évaluation externe a été menée par des experts de cette agence qui connaissent le contexte marocain pour avoir participé au projet de coopération France Maroc : Fond de Solidarité Prioritaire d'Appui à la Réforme de l'Enseignement Supérieur au Maroc (FSP ARESM)

Les domaines de gouvernance à évaluer sont :

- Organigrammes
- Pilotage des missions fondamentales de l'UAE
- Structures et outils de gestion au service de la communauté universitaire
- Vie étudiante
- Relations extérieures et internationales
- Identité de l'UAE et communication externe
- Impact de la réforme de l'enseignement supérieur (loi 01-00)

2.2. Cas de l'INSA Euro-Méditerranée -Université Euro-Méditerranéenne de Fès(2015)

- Organisme d'accréditation : CTI (Commission des Titres de l'Ingénieur – France)
- Résultat atteint : habilitation à délivrer le Diplôme d'Ingénieur Français par un établissement au Maroc

Standards CTI :

- Mission et Organisation
- Ouvertures et Partenariats
- Recrutement et Formation des élèves ingénieurs
- Recherche
- Employabilité des lauréats
- Démarche qualité et Amélioration continue

2.3. Cas de l'Université Al Akhawayn (2015)

Organisme d'accréditation : NEASC (USA)

Standards for Accreditation :

- Mission and Purposes
- Planning and Evaluation
- Organization and Governance
- The Academic Program

- Faculty & Students
- Library and Other Information Resources
- Physical and Technological Resources
- Financial Resources & Public Disclosure
- Integrity

IV. Organismes d'assurance qualité en Afrique

- Maroc :

- Instance Nationale de l'Évaluation au sein du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique (2006)
- Agence Nationale de l'évaluation et de l'assurance qualité de l'enseignement supérieur et de la recherche (2014)

- Tunisie : l'Instance Nationale de l'Évaluation, de l'Assurance Qualité et de l'Accréditation (IEAQA)

- **Sénégal** : L'Agence Nationale d'Assurance Qualité dans le Supérieur (ANAQ-Sup) (2012) a élaboré 2 référentiels d'évaluation des institutions et des programmes. Chaque université publique a signé avec l'État un contrat de performance avec un volet assurance qualité, dont l'un des objectifs est de faire accréditer, chaque année, un certain nombre de programmes par l'ANAQ-Sup

- **Côte d'Ivoire** : Programme de coopération intitulé : Appui à la Modernisation et à la Réforme des Universités et Grandes Écoles de Côte d'Ivoire (AMRUGE-CI)

- **Mauritanie** : Projet de coopération : Fond de Solidarité Prioritaire (FSP) Renforcement de la Gouvernance Universitaire et Rédaction des Plans Stratégiques

- **Mali** : Projet d'Appui au Développement de l'Enseignement Supérieur (PADES) dont une partie concerne l'évaluation des universités en vue de leur contractualisation

Coté associatif citons les associations suivantes qui disposent de programmes et activités en assurance qualité :

- > AUA : L'Association des Universités Africaines dispose d'un programme riche en assurance qualité
- > AROQA: Arab Organization for Quality Assurance in Education
- > Bureaux AUF - l'Institut de la Francophonie pour la Gouvernance Universitaire (IFGU) à Yaoundé
- > CAMES : Conseil Africain et Malgache pour l'ES qui a mis en place l'agence d'Assurance Qualité (déclaration de Bangui, 2008) pour les objectifs qui suivent :
 - Harmonisation des enseignements (LMD)
 - Crédibilité des Diplômes
 - Mise en place d'une démarche d'AQ
 - L'Organisation d'ateliers d'AQ
 - L'Elaboration de référentiels d'évaluation
 - L'Elaboration de guides méthodologiques pour les accompagner

V. Analyse SWOT des expériences menées (Plans Stratégiques et Autoévaluation)

Points forts

- Examen critique, effort sincère et constant de discernement
- L'exercice porte sur toutes les étapes et modes de fonctionnement de l'institution
- Autoévaluation : Constitue une expérience de liberté, non pas une imposition subie
- Choix conscient et responsable sans obligation.
- Exercice dynamique : tourné vers le présent, créatif, prospectif, proactif.

Points faibles et Problèmes fréquemment rencontrés

- Absence de Projet stratégique (vision, valeurs,...)
- Absence d'Analyse SWOT
- Absence d'Identification des «bons» objectifs ? Comment la pertinence des objectifs peut-elle être mesurée?
- Confusion : évaluation vs jugement !
- Faible Mobilisation : environnement interne et externe
- Difficile Collecte et traitement de données (Indicateurs)
- Non Couverture de la totalité des domaines du référentiel choisi
- Faible Couverture de tous les établissements de l'université
- Les rapports soumis sont souvent un portrait de l'université mais ne reflètent pas une vision de leadership ou un plan stratégique d'établissement.
- Difficulté de la gestion du temps
- Commission Autoévaluation pas toujours représentative : faible participation du staff administratif, corps enseignant, étudiants
- Les projets sont souvent des vœux
- Absence de leadership et de communication
- Confusion entre commission de rédaction et cellule responsable de l'assurance qualité.
- Démarche axée moyens et non résultats

VI. Recommandations et Conclusions:

La dynamique AQ est lancée dans un grand nombre d'universités africaines et en particulier du Maghreb. Un certain nombre d'éléments demandent à être affinés :

- Analyse du Contexte
- Viabilité
- Indicateurs pertinents propres (Impact, efficacité, efficience,...)
- Cohérence
- Valeur ajoutée

L'AQ permet de se poser les bonnes questions :

- Que cherche à faire l'université?
- Comment l'université cherche-t-elle à atteindre ses buts ?
- Comment l'université se sait-elle engagée dans la bonne voie ?
- Comment l'institution peut-elle changer en vue de progresser ?

Malgré ces premiers acquis, l'évolution de l'assurance qualité semble être toujours à ses débuts dans un nombre de pays africains et les établissements n'ont pas tous acquis une véritable culture de la qualité. Le concept d'amélioration de la qualité et l'implication des principaux intervenants (étudiants, enseignants chercheurs, administratifs partenaires externes) sont plutôt limités et dans le cas des employeurs, relativement rares. Dans certains pays, un fossé existerait entre les exigences juridiques officielles et la mise en œuvre effective des mesures d'assurance qualité. Les processus d'assurance qualité et leurs résultats ne sont pas suffisamment communiqués publiquement. Ces constatations s'appliquent également au processus d'accréditation. Tandis que certains établissements d'enseignement supérieur ont une bonne connaissance de ces processus et obtiennent un retour sur les résultats, la publication de ces informations pour les étudiants, les employeurs et le grand public semble être moins généralisée. Les pays africains doivent promouvoir le développement de cadres nationaux ou régionaux des certifications comme partie

intégrante du processus d'assurance qualité. Dans bon nombre de pays, il est nécessaire de mettre en place un système d'accréditation totalement indépendant et des agences d'assurance qualité facilement accessibles et d'informer le public des procédures et des codes de conduite existants. Dans un souci de transparence, les indicateurs de performance nationaux et institutionnels devraient également être publiés. Notons toutefois que bien que des modèles européens ou anglo-saxons pour l'assurance qualité puissent être utilisés à des fins de conseil et d'orientation, ils ne doivent pas être simplement transposés dans un contexte africain. Les modèles et les mécanismes d'assurance qualité sont et doivent rester ancrés dans les traditions et la culture universitaires propres à un pays ou à un établissement. L'aide doit porter sur le développement de capacités à concevoir, sans se limiter à la copie et à la mise en œuvre des mécanismes d'assurance et d'amélioration de la qualité.

Traduction du Résumé

Abstract Translation

Mustapha Bennouna

EVALUATION IN HIGHER EDUCATION : ISSUES AND EXPERIENCES

University assessment in Africa generates interest as well as many debates. Training sessions to auto evaluation as well as to university strategic planning or their components show great attractiveness. Some countries have enact legislations to encourage university deans and professors to implement auto evaluation, and set national or regional evaluation agencies.

This paper aims at giving a first experience on study case of African universities recently involved in this process.

Bibliographie et Références

- J.Reilly et A.Jongsma : Améliorer la qualité dans l'enseignement supérieur (Rapport sur les projets AQ du programme Tempus) 2008-
- Rapport d'évaluation interne de l'Université Mohammed V – Agdal (2006)
- Référentiels du Haut Conseil de l'Evaluation de la Recherche et de l'Enseignement Supérieur (HCERES-France anciennement AERES) : <http://www.hceres.fr>
- Actes du 25ème Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU)-Montpellier 2008
- Standards and Guidelines for QA in the European Higher Education area by European Association for QA in Higher Education (ENQA): <http://www.enqa.eu/>
- Appui au programme d'AQ de l'Enseignement Supérieur en Afrique : Association des universités africaines : www.aau.org
- Rapport de l'UNESCO (2014): « Vers l'assurance qualité de l'enseignement supérieur dans les pays d'Afrique francophones de l'Ouest »
- Peter Materu : Rapport de la Banque Mondiale « Amélioration de l'enseignement supérieur en Afrique subsaharienne. Etat de la question, défis, opportunités et pratiques positives » Juin 2007
- Heldenbergh Anne dir. Les démarches qualité dans l'enseignement Supérieur en Europe. Paris. L'Harmattan 2007
- Actes du Séminaire International : Autonomie de l'Université (expériences nationales et internationales) Rabat - 2010
- Rapport : Le Projet TEMPUS AqiUmed (www.aqiumed.org) (Renforcement de l'AQ interne dans les universités de la méditerranée 2010-2013) édité par la commission européenne
- Rapports internes de l'UAE : autoévaluation et évaluation externes (2010)
- Standards CTI version 2016 (<http://www.cti-commission.fr/>)

Les documents mis en ligne comportent trois livres :

- **le Livre 1** présente les principaux cadres de formations d'ingénieur, le profil général des compétences de l'ingénieur et les critères majeurs d'évaluation en vue de l'habilitation,
- **le Livre 2** est un outil d'auto-évaluation des établissements et vient illustrer les critères majeurs ;
- **le Livre 3** décrit l'ensemble des procédures d'accréditation en vue de l'habilitation en France des formations d'ingénieurs selon les différentes voies, en vue de l'admission par l'état des cursus dispensés à l'étranger et en vue de l'attribution du label européen EUR-ACE Master.

-5

ÉVALUATIONS ET POLITIQUES PUBLIQUES



LES ÉVALUATIONS EN ÉDUCATION. LEURS EFFETS SUR LES POLITIQUES ÉDUCATIVES

Georges Felouzis *

Résumé

À quelles conditions les évaluations internationales peuvent-elles guider les politiques éducatives? Pour répondre à cette question ce texte examine d'abord ce que mesurent ces évaluations, les choix méthodologiques dont elles procèdent, les résultats qu'elles produisent. Ensuite, il interroge les conditions d'appropriation par les États des résultats produits. Enfin, il aborde une réflexion plus vaste sur la nécessité de développer au plan national des chercheurs et experts en sciences de l'éducation pour garantir, au sein de chaque pays, une indépendance intellectuelle et comprendre au mieux comment mettre en œuvre des politiques efficaces et efficaces au plan scolaire.

Introduction

Les comparaisons internationales en éducation basées sur des tests standardisés sont devenues aujourd'hui une référence en matière d'évaluation des politiques éducatives. En considérant les grandes enquêtes pionnières de l'IEA (IEA, 2012a ; 2012b)), PISA pour les pays de l'OCDE et au-delà (OCDE, 2013), ainsi que PASEC (2015) et SACMEQ dans nombre de pays d'Afrique, nous assistons à un développement sans précédent des enquêtes comparatives en éducation. Il est désormais possible de comparer des pays très différents par leur développement économique, la nature de leur politique éducative ou l'ampleur de leur scolarisation, au plan des acquis de leurs élèves, de la nature de ces acquis et de leur répartition au sein de la population. Au-delà des « palmarès » que suscitent parfois la publication des résultats de ces comparaisons, il est déterminant de comprendre comment, et à quelles conditions, les états peuvent s'en saisir pour améliorer leur école à partir d'un diagnostic global, rationnellement construit sur des études empiriques à large échelle (Felouzis & Charmillot, 2012). En effet, les comparaisons internationales en éducation ne concernent pas seulement la recherche, dont le but est de produire de la connaissance sur les processus éducatifs. Elles concernent aussi, et peut-être surtout, l'action publique en éducation dans une logique d'obligation de résultats, d'amélioration des systèmes, de production de capital humain, indissociable d'une stratégie de développement à long terme (Pons, 2011).

Ce développement des évaluations internationales en éducation et leur poids grandissant dans l'orientation de l'action publique, pose bien entendu un ensemble de questions auxquelles nous tenterons d'apporter quelques réponses dans ce texte. La première concerne la nature de la connaissance produite par les évaluations internationales : que mesurent-elles réellement ? La deuxième traite de la façon dont chaque nation s'approprie les résultats produits. Enfin, nous apporterons des éléments de réflexion sur leurs liens entre les résultats produits par ces enquêtes et les politiques éducatives (Felouzis & Hanhart, 2011).

I. Trois questions sur les évaluations internationales en éducation.

Une première question est de comprendre ce que mesurent ces enquêtes internationales et sur quelles hypothèses de recherche elles s'appuient. On a coutume de distinguer, en matière d'éducation comparée, les approches universalistes des approches particularistes. Les premières s'attachent à montrer en quoi l'éducation conduit à transmettre un bien universel que l'on peut qualifier de « savoir », de « compétences » ou simplement d'acquis, fruit d'un processus qui permet aux individus d'acquérir les connaissances accumulées par les générations antérieures. En ce sens, quel que soit le pays ou le continent concerné, l'éducation a le

* Professeur, Université de Genève, Suisse

même but partout et peut ainsi être comparée en fonction de ses résultats mesurés par ce que savent et savent faire les élèves. Pour les approches particularistes, la question posée n'est pas tant celle de l'ampleur des connaissances acquises que leur nature, en lien avec l'histoire, les traditions, la culture propre à chaque nation. L'anthropologie de l'éducation est probablement la discipline la plus emblématique de cette option (Woods, 1990). Car cette discipline cherche, dans les processus éducatifs, les modalités de transmission d'une culture, d'une tradition, d'un habitus social propre à chaque société, en lien avec son histoire propre. Il est clair que le développement des évaluations internationales, leur succès auprès des politiques comme des opinions, relève de l'approche « universaliste » (Felouzis & Charmillot, 2012). Dès lors que le but est de comparer, cette hypothèse est indispensable, ne serait-ce que pour fonder théoriquement la comparaison et la rendre possible. Cela signifie-t-il que tout soit dit par ces enquêtes, que ce qu'elles mesurent est le tout de l'éducation dans chaque pays ? Assurément pas ! Il est donc nécessaire de garder à l'esprit que les enquêtes telles que PISA, PIRLS ou PASEC ne donnent qu'une vision possible des conséquences de l'éducation en gommant les spécificités éducatives de chaque pays.

La deuxième question que l'on peut poser à propos des évaluations internationales est celle de leurs modalités d'appropriation par les États (van Zanten, 2011). S'interroger sur cette appropriation, c'est tenter de construire un lien empirique entre le politique en matière d'éducation et l'évaluation des acquis des élèves à l'école. Ce lien, on le sait, n'est jamais simple ni direct. Il n'est même pas systématique. Il passe par des médiations, des transformations, voire par des processus de traduction. Traduire les résultats des enquêtes PISA en action publique est le fruit d'un long processus qui implique une volonté politique d'améliorer le système, une capacité à mobiliser les acteurs pertinents des politiques éducatives, de penser tout le processus de mise en place des politiques publiques, de leur conception à leur mise en œuvre et in fine à l'évaluation de leurs résultats. Or, on sait par expérience que les réactions des États aux enquêtes PISA notamment, sont d'une grande variété. Ces réactions dépendent de plusieurs facteurs : la nature des résultats, la capacité d'intégrer le diagnostic produit dans le débat public, la nature et le niveau de développement de la statistique publique entre autres. Un rapide coup d'œil sur les réactions aux résultats PISA dans plusieurs pays européens donne à voir leur grande variété. Dans le cas de la France, dont les résultats des élèves sont en baisse notable depuis 2000 et les inégalités d'acquis en hausse, PISA n'est qu'une source parmi d'autres pour éclairer le fonctionnement et les dysfonctionnements du système éducatif. En conséquence, les réactions à PISA sont faibles dans ce pays, tout juste cite-t-on l'enquête de l'OCDE pour argumenter dans le débat public sur les insuffisances de tel ou tel apprentissage. En Allemagne, les résultats de l'enquête PISA 2000 produisirent un véritable choc. Les compétences globales très moyennes des élèves de 15 ans de ce pays et l'ampleur des inégalités sociales d'acquis étaient telles que des réformes furent entreprises pour limiter la ségrégation scolaire dans les filières du secondaire obligatoire et réformer les plans d'étude comme les pratiques d'enseignement. Dans ce cas, la volonté politique s'est articulée, voire construite, sur des résultats nationaux en rupture avec l'idée que ce pays se faisait de son école. Enfin, d'autres pays, comme la Pologne (Le donné, 2014), le Portugal ou encore la Turquie (Establet et al, 2016), se sont emparés des résultats PISA et des recommandations pour améliorer leurs politiques éducatives, ce qui a produit un ensemble de réformes dont les résultats s'observent dans l'évolution des résultats des élèves entre 2000 et 2012. Ce trop rapide panorama des modes d'appropriation des résultats PISA par certains pays montre la diversité des options et les liens complexes qui relient l'action publique et l'évaluation des systèmes éducatifs par les acquis des élèves.

Enfin, il faut se questionner sur les liens entre les résultats des évaluations internationales et les politiques d'éducation. Pour cela, il est utile de reconnaître que les simples constats en termes d'efficacité et d'équité des systèmes ne suffisent pas à susciter des réformes scolaires. Il faut, en plus des constats avancés par ces évaluations, identifier les mécanismes qui produisent trop peu de compétences chez les élèves en fin de scolarité obligatoire ou en fin de primaire (Felouzis, 2014). Par « mécanismes », nous entendons les causes efficaces qui permettent d'expliquer les faits observés dans tel ou tel pays. Si les inégalités d'acquis sont plus marquées ici que là, encore faut-il identifier les sources, les mécanismes, les causes de ces inégalités pour avoir une chance d'y remédier. De même si un système est trop peu efficace – c'est-à-dire que son coût est trop élevé compte tenu des résultats qu'il obtient en termes d'apprentissage – il ne suffit pas de se limiter au constat. Encore faut-il identifier les sources qui produisent cette faible efficacité. Ce n'est qu'à cette condition que les politiques éducatives peuvent – à des degrés divers, on l'a vu – s'emparer de la question et la traiter comme un problème pertinent à résoudre dans leur champ de compétence. De ce point de vue, un des apports des évaluations internationales de l'OCDE ou de l'IEA est de raisonner non pas seulement en termes de pédagogie, mais aussi en termes de système. En d'autres termes, elles ont le mérite de mesurer la part respective des mécanismes « macro-scolaires » d'un côté et des modalités d'enseignement de l'autre, et ainsi inciter les acteurs des politiques scolaires à agir dans ces deux domaines simultanément de façon à améliorer leur système éducatif.

II. L'apport scientifique et politique des évaluations internationales

Dans ce contexte il est pertinent de rendre compte des apports scientifiques et politiques des évaluations internationales. Nous mettons ici « scientifiques » et « politiques » dans le même sac, non pas pour signifier qu'il s'agirait d'une seule et même chose, mais pour insister sur les liens nécessaires entre ces deux entités pour produire des améliorations des systèmes éducatifs. Si l'on se centre sur les enquêtes PISA – mais le même type de réflexion pourrait être conduite dans d'autres comparaisons internationales des systèmes éducatifs – il ressort des résultats très parlant au plan de l'amélioration des systèmes éducatifs. Notamment les pays qui, entre 2000 et 2012, ont augmenté leur taux de scolarisation (la Turquie, le Brésil et bien d'autres) sont aussi des pays qui progressent en termes de compétences moyennes des élèves (OCDE, 2013). Ce résultat est contre intuitif car on aurait pu s'attendre à ce que les élèves nouvellement scolarisés soient moins compétents au plan scolaire dans ces pays. Or, il n'en est rien car les systèmes éducatifs qui augmentent leur « couverture scolaire » sont aussi des systèmes qui s'améliorent, deviennent plus efficaces et plus équitables. L'effort éducatif est donc ici récompensé par un meilleur rendement du temps scolaire, de la scolarisation, de l'enseignement.

Un autre exemple peut être pris dans le domaine de l'organisation scolaire. On sait par exemple depuis longtemps qu'une séparation trop nette entre élèves en fonction de leur niveau d'apprentissage dans des filières différenciées et hiérarchisées dans le secondaire obligatoire est un puissant facteur d'inégalité (Gamoran & Mare, 1989). Les enquêtes PISA donnent une vérification empirique à ce phénomène en montrant que les pays les plus équitables – c'est-à-dire dans lesquels la répartition des compétences est la moins inégalitaire entre les élèves de différents groupes définis par le niveau socio-économique de leur famille, leur parcours migratoire éventuel, leur sexe, etc. – sont aussi ceux qui proposent des systèmes intégrés jusqu'à la fin de l'enseignement obligatoire : tous les élèves sont soumis aux mêmes programmes, dans des classes et des établissements hétérogènes au plan socio-économique comme migratoire. Inversement, les systèmes plus segmentés renforcent les inégalités entre élèves en leur proposant des formes et des contenus d'enseignement eux-mêmes inégaux. Un des corollaires de ce résultat est que les systèmes éducatifs les plus ségrégués – en fonction de filières ou des établissements – sont aussi les moins équitables (Merle, 2012).

Enfin, un des résultats est aussi que les voies de l'efficacité sont multiples. Il n'y a pas une « recette » qui serait applicable partout, quelles que soient les circonstances. Les analyses produites par Nathalie Mons (2007) sont assez parlantes sur ce point. Comparant plusieurs modèles d'organisation scolaire, elle relève que les modèles de « l'intégration individualisée » (Finlande notamment) et de « l'intégration à la carte » (Canada et Royaume-Uni par exemples) sont bien plus efficaces que les modèles « segmentés » (Allemagne et Autriche) et de « l'intégration uniforme » (France et Italie). Cela montre que la façon dont on conçoit la structure des études, ainsi que la façon dont on prend en compte la diversité du niveau et des besoins pédagogiques des élèves, sont des facteurs déterminants de l'efficacité des apprentissages.

III. La nécessité de relais scientifiques au plan national

Il faut pourtant se rendre à l'évidence : ces apports scientifiques et politiques des comparaisons internationales restent à un niveau élevé de généralité et d'abstraction. Cela est dû au fait que comparer des pays aussi différents que le Japon, la Turquie, les Etats-Unis et le Vietnam demande de se situer à un niveau d'abstraction assez élevé pour que la comparaison soit possible. Or, pour qu'une politique éducative soit efficace, il est impossible de se satisfaire de généralités. Celles-ci sont indispensables et utiles, mais non suffisantes au plan concret. Une des conséquences de cette situation est que les évaluations internationales ne disent pas tout sur l'éducation, loin de là ! Et si elles ne disent pas tout, cela signifie qu'au niveau national, il est nécessaire qu'un corps de chercheurs, scientifiques, universitaires, soit formé et mobilisé pour comprendre dans le détail et de façon contextualisée, les mécanismes de l'efficacité et de l'équité. Les pays les plus avancés sur ce point ne s'y trompent pas. Les Etats-Unis, la Grande-Bretagne, la France et bien d'autres encore ont un appareillage statistique et scientifique sur l'éducation qui leur permet d'aller bien au-delà des constats et analyses de l'OCDE ou de l'IEA. Dans le contexte des pays du Sud, cette question est plus pertinente encore puisque développer la recherche en sciences de l'éducation, sociologie, économie, pédagogie, etc. est une façon de garantir une certaine indépendance intellectuelle. C'est aussi une nécessité pour comprendre au mieux comment mettre en œuvre des politiques efficaces et efficientes au plan scolaire. Pour cela, il est nécessaire de développer simultanément des formations de haut niveau en sciences de l'éducation pour former le vivier des scientifiques qui auront demain pour tâche d'identifier les mécanismes sur lesquels agir pour améliorer l'école et évaluer les résultats des politiques mises en œuvre.

IV. Perspectives

En définitive, l'ampleur de la diffusion ainsi que le succès scientifique et public des comparaisons internationales en éducation montrent la force des processus de globalisation de l'éducation. Dans le concert des nations, la dimension éducative est devenue un enjeu majeur, tant au plan économique – production d'un capital humain lui-même producteur de richesses – qu'au plan proprement politique – la démocratie ne va pas sans une éducation généralisée offrant à tous les mêmes chances et les mêmes opportunités. Ce contexte international rend plus pertinentes encore les formes de gouvernances basées sur les résultats des systèmes, sur ce qu'ils produisent réellement en termes de compétences, d'équité, d'efficacité. Et ce défi ne peut être relevé sans développer un couplage fort entre la production de connaissances sur l'éducation et les politiques éducatives mises en œuvre.

Traduction du Résumé

Abstract Translation

Georges Felouzis

INTERNATIONAL EVALUATIONS IN EDUCATION AND THEIR ROLE IN EDUCATION POLICIES

At what conditions international evaluations can guide education policies? In order to answer this question, this paper examines what these evaluations measure, methodological choices they make as well as the results they produce. Then, this paper shows also how countries take own of the results. Finally, it initiates a larger reflection on the necessity of developing, at the national level, expert researchers in educational sciences in order to guarantee, in every country, an intellectual independence and understand better how to implement efficient school policies.

Bibliographie et Références

- Establet R, Felouzis G et Charmillot S, (à paraître), La Turquie dans le concert mondial de l'éducation. Une analyse des enquêtes PISA 2003 et 2012, Turcica.
- Felouzis G., (2014), Les inégalités scolaires, Paris, Puf.
- Felouzis G & Charmillot S., (2012), Les enquêtes Pisa, Paris, Puf.
- Felouzis G., Hanhart S., (Éds) (2011), Gouverner l'éducation par les nombres ? Usages, débats et controverses, Bruxelles, De Boeck « Raisons éducatives ».
- Gamoran, A., & Mare, R. D. (1989). Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?. *American Journal of Sociology*, 94(5), 1146–1183.
- IEA (International Association for the Evaluation of Educational Achievement) (2012a), Timss 2011 International Results in Mathematics (Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Alka Arora), Boston.
- IEA (International Association for the Evaluation of Educational Achievement) (2012b), PIRLS 2011 International Results in Reading (Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Kathleen T. Drucker), Boston.
- Le Donné N., (2014), La réforme de 1999 du système éducatif polonais. Effets sur les inégalités sociales de compétences scolaires, *Revue française de sociologie*, Vol 55(1).
- Merle P., (2012), La ségrégation scolaire, Paris, La Découverte.
- Mons N., (2007), Les nouvelles politiques éducatives, Paris, Puf.
- OCDE (2013), Résultats du PISA 2012 : Savoirs et savoir-faire des élèves : Performance des élèves en mathématiques, en compréhension de l'écrit et en sciences, PISA, Éditions OCDE.
- PASEC (2015), PASEC2014, Performances des systèmes éducatifs en Afrique subsaharienne francophone : compétences et facteurs de réussite au primaire, PASEC, COFEMEN, Dakar.
- Pons X., (2011), L'évaluation des politiques éducatives, Paris, Puf.
- Woods P., (1990), L'ethnographie de l'école, Paris, A. Colin.
- Van Zanten A., (2011), Les politiques d'éducation, Paris, Puf.

EVALUATION D'IMPACT DES POLITIQUES DE RÉFORME DU SYSTÈME ÉDUCATIF MAROCAIN : PROBLÉMATIQUES DE CONCEPTION ET DE MISE EN ŒUVRE

Abdelhaq EL HAYANI *

Résumé

Le présent article s'inscrit dans un contexte marqué par des défis majeurs en matière de performance du système éducatif Marocain. Ce contexte est actuellement orienté par la mise en place de la nouvelle vision stratégique du Conseil supérieur de l'éducation, de la formation et de la recherche scientifique. Cette vision qui s'appuie sur quatre grandes finalités dont celle à portée transversale relative à la gouvernance et au leadership. Parmi les dimensions qui y sont abordées, il y a la question de l'évaluation. L'objectif principal du présent article est à double perspectives. D'abord analyser le dispositif actuel de l'évaluation du système éducatif Marocain à travers une revue de diagnostic documentaire tout en se référant aux cadres théoriques les plus préconisés en matière de l'évaluation dans l'éducation. Ensuite tenter de préconiser quelques pistes d'amélioration et de correction de certaines limites constatées de la situation actuelle.

Les approches conceptuelles telles que rapprochées dans la revue des cadres théoriques ont permis de guider et de structurer notre analyse, un ensemble de limites ont été mis en évidence sur le dispositif d'évaluation dans sa globalité, mais également de manière spécifique aux évaluations respectives des élèves, des enseignants et des établissements scolaires. Un modèle de conceptualisation de l'évaluation pour le cas du système éducatif Marocain, a été préconisé. Il est structuré autour de trois composantes saillantes. Une première composante liée à la conception de politique globale d'évaluation du système éducatif. Une deuxième composante portant sur la mise en place d'une plateforme d'alignement technique traitant les préoccupations posées pour un bon alignement entre les normes, curricula et processus d'évaluation. Enfin, une dernière composante relative à l'implémentation d'une plateforme d'alignement social clarifiant de manière explicite les conditions de réussite de l'implémentation de ces processus d'évaluation dans les établissements scolaires en relation avec l'adhésion, la motivation, l'appropriation et les capacités des acteurs locaux.

Mots Clés : système éducatif, politique de réforme, évaluation, alignement technique, alignement social.

Introduction

Éléments du contexte

L'éducation se trouve aujourd'hui interpellée quant à ses réels objectifs à poursuivre et défis à relever dans un contexte mondial marqué par l'accélération des changements techniques, économiques, sociaux et culturels. L'ampleur des incertitudes engendrées font que les acteurs et décideurs sont dans l'incapacité à connaître l'évolution du monde dans une portée temporelle de cinq ans, quelles sont alors les chances d'avoir une connaissance fiable des choix pertinents à poursuivre en termes de compétences et de profils dans l'éducation d'un enfant pour une période d'enseignement de 14 ans en moyenne ?

Les efforts de l'État et sa volonté d'appuyer ce secteur vital sont largement mis en avant par l'intérêt accordé

* Directeur de la stratégie, des statistiques et de la planification / MENFP- Chercheur Doctorant - Faculté des Sciences de l'Éducation - Université Mohammed V de Rabat

à ce secteur stratégique par la plus haute autorité du Pays et sont en outre matérialisés par la tendance haussière et continue des ressources financières consacrées à l'éducation depuis le milieu des années 2000.

Il est important de souligner sur la base de ces constats que finalement les progrès réalisés actuellement, certes sont utiles, mais ne sont pas en alignement avec les vrais enjeux de l'éducation et de la formation pour l'avenir du Pays à long terme sur les plans politique, économique et social. Comme souligné clairement dans le Discours Royal du 20 août 2012, la performance quantitative du système éducatif, ne constitue pas seule la priorité nationale et ne correspond pas aux attentes urgentes de la population et des acteurs politiques et économiques. Les vrais enjeux sont aujourd'hui situés au niveau des aspects qualitatifs, la qualité des apprentissages et son adéquation avec les évolutions et les attentes du marché d'emploi, constituent la principale priorité du Pays. « [...] **Il est donc impératif de se pencher avec sérieux et résolution sur ce système que Nous plaçons, d'ailleurs, en tête de nos priorités nationales** »¹.

Au regard de cette situation, le Conseil supérieur de l'éducation, de la formation et de la recherche scientifique (CSEFRS) et le Ministère de l'éducation nationale et de la formation professionnelle (MENFP) ont engagé de manière coordonnée un processus de diagnostic du système éducatif et de concertation avec l'ensemble de ses acteurs et parties prenantes sur les quatre niveaux central, régional, provincial et local au sein des établissements scolaires. S'appuyant sur les orientations royales², sur les fondements de la nouvelle constitution de 2011³ et sur les résultats de ce processus de diagnostic et consultatif, une nouvelle vision stratégique pour la période 2015-2030⁴ a été élaborée et approuvée.

Questions de recherche

Dans le contexte historique des différentes réformes mises en place, trois questions méritent d'être posées :

- (1) D'abord, dans quelle mesure nous pouvons constater avec un niveau de fiabilité acceptable que les politiques de réforme du système éducatif Marocain sont pertinentes, appropriées et ont les chances d'avoir les impacts positifs attendus sur sa performance et sur la qualité de ses outcomes ?
- (2) Ensuite, il est aussi important de s'interroger sur l'efficacité du processus de mise en œuvre de ces politiques de réforme ?
- (3) Le système éducatif est-il doté, d'un dispositif d'évaluation d'impact, cohérent, fiable et pérenne, de ces politiques de réforme ?

I. Méthode adoptée

Le présent article se propose de répondre à ces trois questions par le biais d'une méthodologie basée sur une approche qualitative et est structurée autour de trois étapes :

- (i) la première consiste à effectuer une première investigation documentaire en vue d'analyser le cadre théorique d'évaluation dans l'éducation. Le but étant de formaliser un modèle de stratégie d'évaluation des systèmes éducatifs.
- (ii) la deuxième étape est consacrée à une seconde investigation documentaire permettant de réaliser un diagnostic du dispositif d'évaluation actuellement en vigueur dans le système éducatif Marocain. Le but est de mettre en évidence ses forces et ses limites en rapport avec les références et modèles adoptées dans le cadre théorique.
- (iii) la troisième étape est dédiée à la formulation d'un ensemble de pistes de recommandations visant à préconiser une amélioration et un réajustement du dispositif d'évaluation dans le système éducatif qui soit adapté au contexte national Marocain.

1 - Extrait du Discours de SM le Roi Mohammed VI à l'occasion du 59ème anniversaire de la Révolution du Roi et du peuple (20 Août 2012).

2 - Notamment (i) le Discours de SM le Roi Mohammed VI à l'occasion du 59ème anniversaire de la Révolution du Roi et du peuple (20 Août 2012) et (ii) le Discours de SM le Roi Mohammed VI à l'occasion du 60ème anniversaire de la Révolution du Roi et du peuple (20 Août 2013).

3 - En particulier (i) l'article 31 stipulant que « l'Etat, les établissements publics et les collectivités territoriales œuvrent à la mobilisation de tous les moyens disponibles pour faciliter l'égal accès des citoyennes et des citoyens aux conditions leur permettant de jouir du droit [...] à une éducation moderne, accessible et de qualité [...] » et (ii) l'article 32 stipulant que « L'enseignement fondamental est un droit de l'enfant et une obligation de la famille et de l'Etat ».

4 - La vision stratégique, selon les termes du rapport du CSEFRS, « a pour finalité la mise en place d'une Ecole nouvelles dont les fondements sont : (i) l'équité et l'égalité des chances ; (ii) la qualité pour tous ; (iii) la promotion de l'individu et de la société ». En plus d'un fondement transverse optant pour « un leadership efficient et une nouvelle conduite de changement ».

II. Cadre théorique de l'évaluation dans le système éducatif

1. Cadre de l'évaluation d'impact spécifique aux politiques publiques de l'éducation

De prime abord, il semble pertinent de poser deux définitions clés selon la réflexion de Marguerite Clarke de la Banque Mondiale respectivement pour le concept de l'évaluation (Assessment) et pour le concept de système d'évaluation (Assessment System). Selon elle ⁵,

(i) « **Assessment is the process of gathering and evaluating information and what student know, understand, and can do in order to make an informed decision about next steps in the educational process** ». (ii) « **An Assessment System is a group of policies, structures, practices, and tools for generating and using information on student learning and achievement** ».

Il semble d'emblée de par ces deux définitions, que le cadre d'évaluation d'impact dans le secteur de l'éducation est fondé d'une part, sur le process lui-même de collecte et d'évaluation des informations sur les apprentissages des élèves et d'autre part, sur le mix de mise en œuvre de ce process (politiques, structures, pratiques et outils) en relation avec l'utilisation des informations se rapportant aux apprentissages des élèves.

L'évaluation des apprentissages des élèves semble être au centre des préoccupations de l'évaluation d'impact de toute politique éducative. Toutefois, plusieurs auteurs ont mis l'accent sur l'importance de faire la différence entre d'une part, une stratégie d'évaluation à long terme avec implémentation d'un monitoring régulier des apprentissages des élèves et de leur achèvement et d'autre part, un débat ponctuel autour de résultats, une action à court terme, et une absence de visibilité suffisante qui n'est disposée que par un monitoring pérenne.

D'ailleurs, toujours selon Marguerite Clarke, différentes études de recherche ont démontré deux corrélations intéressantes vis-à-vis de la bonne qualité des apprentissages des élèves, (i) la première entre cette qualité des apprentissages et la qualité de l'évaluation formative dans les classes « **a strong link between high-quality, formative⁶ classroom assessment activities and better student learning outcomes** »⁷. La deuxième entre la qualité des apprentissages des élèves avec l'utilisation des données issues des systèmes d'évaluation pour la redevabilité⁸ des écoles et des acteurs⁹.

Le succès de l'implémentation du mécanisme « Test-Based-Accountability » dans un système d'évaluation dépend notamment de la manière dont les informations issues de l'évaluation sont effectivement utilisées en particulier dans l'accompagnement des écoles et des enseignants qui sont en situation de déficit de performance. « **Key determinants of whether the effects of test-based-accountability exercises are more positive than negative include the technical quality of the tests themselves, the alignment between the design and the way test results are used, and the extent to which supports are in place to help schools and teachers identified as underperforming** »¹⁰.

P. Santiago a développé également cette question cruciale de comment les politiques d'évaluation pourraient fonctionner ensemble de manière cohérente et interdépendante pour l'amélioration des apprentissages des élèves. Cette préoccupation de ne pas se contenter à des actions ponctuelles et disparates d'évaluation est confirmée. Il s'agit de construire une politique globale et intégrée des évaluations à mettre en place avec comme seul but fédérateur est celui d'améliorer en continue les apprentissages des élèves. A cette préoccupation, P. Santiago propose une architecture du système d'évaluation dans l'éducation, basée sur cinq composantes.

5 - Ravela et al., 2009, cité par (Marguerita Clarke, « What Matters Most for Student Assessment Systems : A framework paper », Banque Mondiale, 2012, P. 6).

6 - Ce concept d'évaluation formative est défini couramment comme étant une évaluation effectuée en classe et permettant d'établir un diagnostic des forces et des faiblesses pour une prise en compte dans la mise en œuvre des actions de remédiation.

7 - Marguerita Clarke, « What Matters Most for Student Assessment Systems : A framework paper », Banque Mondiale, 2012, P. 8.

8 - Concept de Test-Based-Accountability.

9 - Carnoy and Loeb, 2002, cité par (Marguerita Clarke, « What Matters Most for Student Assessment Systems : A framework paper », Banque Mondiale, 2012, P. 8).

10 - Ravela, 2005, cité par (Marguerita Clarke, « What Matters Most for Student Assessment Systems : A framework paper », Banque Mondiale, 2012, P. 9).

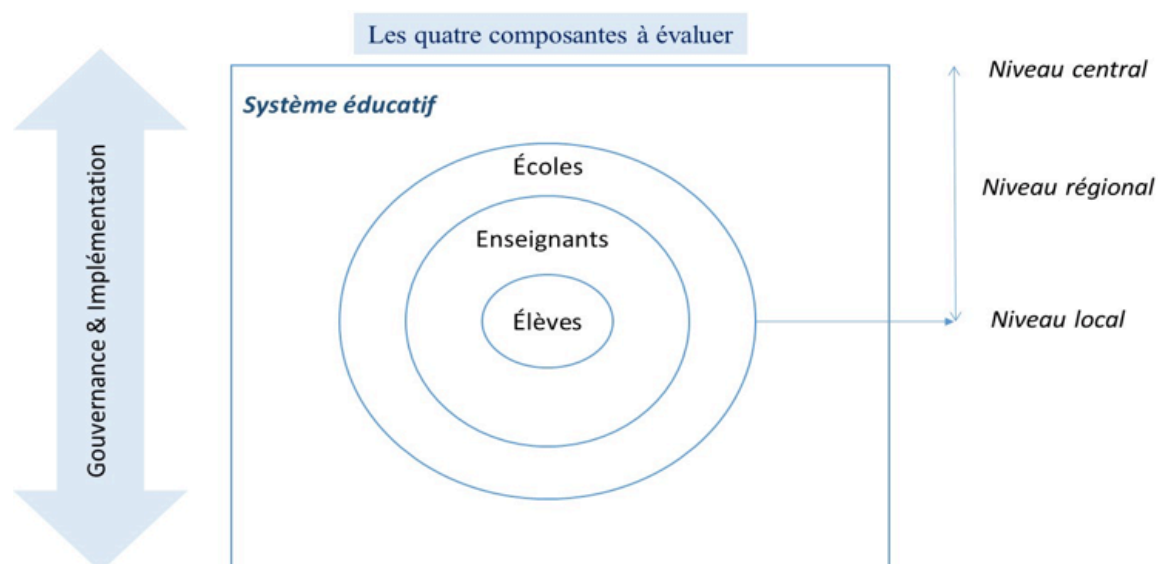


Figure 01 : Schéma de l'architecture des composantes du système d'évaluation dans l'éducation¹¹

Il y a d'une part, une composante transversale traitant de l'approche et des modalités de gouvernance et de mise en œuvre sur terrain de cette architecture d'évaluation, et il y a d'autre part, quatre composantes qui constituent les éléments clés de cette architecture nécessitant une évaluation régulière, pérenne et de manière interdépendante et cohérente, à savoir : (ii) élèves, (iii) enseignants, (iv) écoles et (v) systèmes éducatifs.

S'agissant de la première composante de gouvernance et d'implémentation, l'auteur (P. Santiago)¹² a mis le focus sur quatre priorités majeures. D'abord, il est crucial de voir comment les quatre composantes de cette plateforme d'évaluation s'interagissent pour une meilleure complémentarité tout en évitant la redondance. La deuxième priorité porte sur la nécessité de développer les compétences des acteurs clés (enseignants, chefs d'établissements et inspecteurs) dans l'évaluation et l'utilisation des résultats constatés. Il importe en particulier de donner légitimité à ces acteurs, de les impliquer dans tout le processus d'implémentation et de s'assurer que les passerelles de feedback des résultats sont maîtrisées vis-à-vis des pratiques de la classe, des projets d'établissements et de la politique éducative de manière globale. La troisième priorité concerne l'importance de la sécurisation du lien avec les pratiques de la classe.

Cette priorité relatée a trait avec les défis de cette implémentation et qui ne sont pas faciles à dépasser. Ces défis portent essentiellement sur le déficit en expériences et expertises professionnelles des acteurs clés en matière d'évaluation. L'auteur met le focus sur l'importance des concertations avec l'ensemble des partenaires impliqués, de l'approche d'expérimentation en pilote réduit et du renforcement des capacités et de l'expertise d'évaluation des acteurs. « **The policy process needs to recognise that: reaching agreements on the design of the evaluation and assessment framework requires time for discussions and consultations with all stakeholders; developing expertise in the system, [...] is expensive and requires time; conducting evaluation processes induces additional workload for school agents; and [...] more educational resources** »¹³.

En ce qui concerne la deuxième composante elle porte sur l'évaluation des élèves. La question prioritaire à cet égard est de voir comment l'alignement fortement requis entre trois éléments clés, est implémenté et effectivement concrétisé. Il s'agit de veiller à aligner l'évaluation des apprentissages des élèves d'une part, avec les standards¹⁴ et d'autre part, avec les curricula¹⁵. « **If the assessments do not well match the curriculum and the standards, then results have little value in judging how well students are learning and in diagnosing school or student needs** »¹⁶.

En particulier, l'alignement avec les standards exige selon des réflexions de plus d'un auteur de mettre le focus sur la stratégie de mise en œuvre d'évaluation à large échelle à travers des tests standards tout en mettant l'accent sur la validité, la fiabilité et l'utilisabilité des résultats. Il est aussi suggéré de balancer entre les évaluations sommatives effectuées par les enseignants en classes et qui restent qualifiées par

11 - Schéma conçu sur la base de l'article : Paulo Santiago, « Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, P. 1.

12 - Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, P. 2-3.

13 - Op. Cit. P. 3.

14 - Définition des connaissances et compétences que doivent acquérir les élèves pour chaque niveau d'enseignement.

15 - Définition des contenus d'enseignement et des objectifs à atteindre selon les standards arrêtés.

16 - Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, P. 3.

une certaine limitation « [...] **the approach to summative assessment can, in turn, have a strong impact on the learning process itself** »¹⁷ et l'évaluation externe¹⁸ assurant une meilleure crédibilité et fiabilité aux résultats et plus d'équité entre les élèves. « **The major advantage of external assessment is its high reliability. It ensures that all students are assessed on the same tasks and that their results are measured by the same standards** »¹⁹.

Enfin, une autre préconisation venant compléter les deux précédentes, stipule la pertinence d'intégrer l'évaluation formative dans la plateforme de l'évaluation des élèves. Il s'agit de renforcer la validité et la fiabilité des tests issus de l'évaluation sommative en assurant une évaluation interactive du progrès des élèves sur la base des normes et objectifs définis dans les standards et curricula. L'avantage est de suivre de manière continue la progression des enseignements dans le sens de l'atteinte des standards en termes de connaissances que de compétences dont la mesure s'avère parfois complexe. « **Strategies to achieve such integration include a closer interface between formative assessment and summative assessment** »²⁰.

S'agissant de la troisième composante portant sur l'évaluation des enseignants, la question centrale est de voir comment assurer le bon dosage dans l'utilisation des résultats de l'évaluation des élèves vis-à-vis de l'amélioration des enseignements, du renforcement des compétences des enseignants, de la redevabilité et des mécanismes d'incitation des enseignants. Ce bon dosage préconisé interpelle d'abord la nécessité d'utiliser les résultats de l'évaluation des élèves dans l'évaluation de la performance des enseignants et l'implication de celle-ci dans la gestion et l'évolution de leurs carrières. Il y a en second lieu, la complexité de doser les mécanismes de sanctions et d'incitations en fonction des résultats de l'évaluation des élèves, une seule optique semble engendrer en général des conséquences négatives pour les apprentissages des élèves. « **Test-based accountability systems are supposed to strengthen incentives for teachers to commit themselves to helping all students to meet important centrally defined standards and fulfil goals within the national curriculum** »²¹.

Pour la quatrième composante, elle porte sur l'évaluation des écoles. Elle intègre les deux précédentes composantes, à savoir l'évaluation des élèves et des enseignants. Tout en soulignant qu'un tel dispositif d'évaluation des écoles est situé dans le contexte de renforcement de leur autonomie, la principale préoccupation est située au niveau de comment aligner l'évaluation externe des écoles avec l'évaluation interne. La finalité est de combiner entre deux perspectives qui peuvent être perçues comme paradoxales. D'une part, l'évaluation interne, appelée en général par différents chercheurs par autoévaluation (self evaluation), vise à développer les capacités, la performance et l'appropriation des écoles vis-à-vis des résultats de cette évaluation. D'autre part, ces résultats ne sont pas toujours faciles à communiquer aux acteurs et partenaires externes dont les parents des élèves et la hiérarchie relevant du système éducatif, lorsqu'ils ne sont pas satisfaisants.

Le niveau de confiance et de transparence accompagnant ce dispositif se trouve ainsi influencé négativement. D'où la nécessité d'instaurer en appui aux évaluations internes, des évaluations externes ayant la caractéristique d'être neutres, indépendantes et crédibles. Ces évaluations externes doivent servir ainsi comme, outils de redevabilité vis-à-vis de tous les acteurs de l'école. « **Self-evaluation has the merit of being immediate, responsive to the school's specific needs and circumstances and its results are 'owned' by the school. However, self-evaluation which serves the needs of accountability is subject to inevitable tensions between rigour and depth on the one hand and a natural desire not to undermine the confidence of parents and superiors on the other** »²².

La dernière composante a trait avec l'évaluation du système éducatif tout entier. L'organisation, la politique, la planification et la gouvernance mis en place sur les différents paliers du système éducatif, servent pour assurer une meilleure performance dans les apprentissages des élèves. La question qui s'impose est dans quelle mesure, les décideurs du système éducatif réussissent à assurer la mise en place d'un système d'information permettant de rendre disponible ces données de performance sur tous les niveaux du système : central, régional et local. La fiabilité, la disponibilité des données en temps réel pour tous les acteurs et partenaires des différents paliers du système et surtout l'assurance que ces données portent sur l'évaluation des apprentissages des élèves selon les standards et les objectifs des curricula, constituent un vrai défi posé pour la plupart des systèmes éducatifs nationaux. « **The challenge is to develop strategies to collect valid, reliable and broad outcome measures to monitor performance against key national educational**

17 - Op. Cit. P. 4.

18 - Examens conçus et notés hors écoles.

19 - Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, P. 4.

20 - Op. Cit. P. 5.

21 - Op. Cit. P. 6.

22 - Ibid.

goals over time, for different sub-national areas and student groups»²³.

2. Focus sur la complexité de mise en œuvre de l'évaluation dans les systèmes éducatifs

Les différentes préconisations développées ci-dessus sont certes pertinentes et cruciales pour l'atteinte des objectifs d'impact des politiques éducatives. Toutefois, il est légitime de s'interroger sur les entraves qui malheureusement ne favorisent pas fréquemment le succès du processus d'implémentation d'une plateforme d'évaluation dans les systèmes éducatifs. En général ces entraves touchent beaucoup plus aux dimensions sociales et culturelles que techniques.

Nous proposons de présenter de manière synthétique des éléments constatés à ce sujet. Selon les réflexions de Janet Looney (OCDE)²⁴, tout système d'évaluation dans les systèmes éducatifs, est fondé sur un alignement rigoureux de trois éléments clés qui sont les normes²⁵, les programmes²⁶ et l'évaluation²⁷ elle-même des élèves et des écoles. Cet alignement est requis à travers une veille sur un bon dosage d'équilibre et de cohérence entre les dimensions techniques et sociales des trois composantes citées ci-dessus. Le constat préliminaire à souligner est le coût énorme de non alignement eu égard à la complexité du système éducatif. « **Yet, no system can achieve perfect alignment. A number of commentators have pointed to the complexity of systems as barrier to tight alignment** »²⁸.

S'agissant de la dimension technique, les préoccupations peuvent être appréhendées à travers les enjeux et les difficultés mis en exergue par la mise du focus sur chacune des trois composantes concernées : standards, curricula et évaluation des élèves, des enseignants et des écoles. En ce qui concerne les standards, ils sont définis par Cizek comme étant « [...] **requiring a blend of artistic, political and cultural ingredients, to this list we might add knowledge of how students learn and progress** »²⁹, cette composante met en avant des conflits et parfois des tensions autour de la détermination de ces standards. Toutefois, il y a consensus sur la pertinence du concept de « learning-to-learn » intégrant en particulier les compétences nécessaires pour la résolution de problèmes, analyse critique et autonomie des élèves.

Pour les curricula, les principales préoccupations portent sur la consistance des thèmes pouvant être adressés sur une période limitée de temps et la structure pertinente à garantir pour chaque discipline en cohérence avec les normes et objectifs fixés au niveau des standards. Tout en soulignant les enjeux qu'ont ces deux aspects sur les apprentissages, la motivation et le niveau de pression sur les enseignants.

Par rapport à la composante d'évaluation des élèves, des enseignants et des écoles, le débat est focalisé sur les critères de mesure de la qualité des informations issues de ces évaluations en alignement avec les composantes de normes et de curricula. Trois critères semblent être prioritaires par la majorité des chercheurs à savoir la validité³⁰, la fiabilité³¹ et l'utilisabilité³². L'enjeu se situe au niveau de l'importance pour les enseignants dans les classes et les chefs d'établissements dans les écoles à disposer en continu d'informations sur les résultats d'évaluation des apprentissages des élèves qui soient fiables, valides et utilisables. « **In classrooms, teachers need more detailed and frequent information on student learning** »³³.

Pour l'alignement social, la question qui semble cruciale pour une bonne appréhension de cette dimension est la suivante : avec quel degré ces données sont utilisées pour le développement de l'enseignement ? , cela dépend de l'alignement social au niveau des écoles. « **It is at this intersection of the technical and social where practitioners align efforts where change and improvement can occur** »³⁴. En d'autres termes, la question fait référence aux capacités du système éducatif à développer un environnement de collaboration efficace et de capital social évolué entre les acteurs clés de l'école en vue de faire améliorer les apprentissages des élèves surtout lorsqu'on est face à des systèmes complexes. « **In socially aligned systems, institutions and actors**

23 - Op. Cit. P. 8.

24 - Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, P. 5.

25 - Définies par l'auteur en tant que connaissances et compétences que les élèves sont censés avoir acquis à différents stades de leur éducation.

26 - Définies comme étant couvrant les objectifs identifiés dans les normes.

27 - Définie en tant que dispositif servant essentiellement à mesurer le niveau réalisé des normes.

28 - Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, P. 5.

29 - Cizek, 2001, cité par (Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, P. 11).

30 - Définie, selon J. Looney, par le niveau d'alignement des résultats avec les standards et les curricula.

31 - Définie, selon J. Looney, par le niveau de consistance et de stabilité des résultats.

32 - Définie, selon J. Looney, par le niveau de clarté des sens qu'ont les décideurs, chefs d'établissements et enseignants par rapport à ces résultats, en vue d'y apporter l'efficacité requise des réponses et des remédiations.

33 - Black & William, 1998, cité par (Janet Looney, « Alignment in Complex Education Systems: Achieving Balance and Coherence », OCDE, 2011, P. 14).

34 - Janet Looney, « Alignment in Complex Education Systems: Achieving Balance and Coherence », OCDE, 2011, P. 16.

work together to define challenges and to consider alternative courses of action. This alignment is vital for system learning and improvement »³⁵.

Toutefois, ces éléments de qualification d'un environnement adéquat d'alignement social ne semblent pas facilement accessibles, en raison de carence de mécanismes d'incitation et de motivation pour une collaboration professionnelle entre les enseignants et pour une participation effective en faveur de développement de leur école. Cette carence reste liée au même temps, au-delà de la dynamique de l'école, au déficit de l'interaction et l'interdépendance entre les couches du système éducatif.

III. Analyse documentaire du dispositif d'évaluation du système éducatif Marocain

Après avoir effectué une revue globale du cadre théorique portant sur les cadres d'évaluation spécifiques aux politiques éducatives, nous proposons dans ce volet de présenter de manière synthétique le dispositif d'évaluation dans le système éducatif Marocain. Ce dernier est structuré en quatre composantes : (i) évaluation des élèves ; (ii) évaluation des enseignants ; (iii) évaluation des établissements scolaires et (iv) évaluation du système éducatif dans sa globalité.

1. Evaluation des élèves

Celle-ci est répartie entre évaluations internes au système et évaluations externes mises en œuvre dans le cadre d'enquêtes évaluatives nationales et internationales.

a) Evaluations internes

Pour ce qui est des évaluations internes, elles sont implémentées à travers des contrôles continus durant toute l'année scolaire, un examen de passation en fin d'année scolaire en l'occurrence le certificat du cycle primaire, le brevet de fin de cycle collégial et le Baccalauréat spécifique à la fin du cycle qualifiant. En plus de l'évaluation des prérequis lancée en 2009 pour les niveaux de 2ème année et 5ème année primaire et pour les niveaux de 1ème année et 3ème année du collégial.

En matière d'analyse de ce dispositif d'évaluation interne des élèves, il est à noter d'abord qu'il reste essentiellement sommatif focalisé beaucoup plus sur les connaissances avec une prédominance des programmes par rapport au curriculum. **«The program (what the teacher must teach) prevails on the curriculum (what the students should learn). Pedagogies are very often frontal with lectures; based on a deductive approach, and not interactive enough»³⁶.**

En outre, le rapport des consultations nationales implémentées par le ministère en 2014³⁷ avec l'ensemble des acteurs et des partenaires, soulève un ensemble de reproches vis-à-vis du système d'évaluation des élèves. Nous retenons notamment les trois constats qui suivent :

- Il s'agit d'une évaluation à portée limitée, dans la mesure où elle est unidimensionnelle. Elle s'intéresse uniquement aux performances des élèves ;
- Ces évaluations portent globalement sur les connaissances académiques et ne permettent pas d'apprécier le niveau des compétences et des habilités des élèves ;
- Faible intérêt apporté aux activités de l'évaluation diagnostique et formative en particulier dans les niveaux non certificatifs ;

D'autres analyses ont mis en exergue les limites de la formation initiale et continue des enseignants leur permettant d'être en mesure de mettre en œuvre tout le processus de diagnostic et de remédiation au niveau des pratiques en classes. **« Initial and continuing training of teachers is insufficient in terms of assessment so that teachers have very limited knowledge on the development of assessment tests, tests award, and correction of tests, diagnosis elaboration and remedial measures development »³⁸.**

b) Enquêtes nationales d'évaluation

S'agissant des enquêtes nationales d'évaluation, le conseil supérieur de l'enseignement (CSE) a lancé en

35 - Op. Cit. P. 17.

36 - Bernard Huggonier, « Evaluation : An Indispensable Tool for Improving Education Quality in Education in Morocco », Morocco Compact II, MCC, 2015, P. II.

37 - Consultations inscrites dans le cadre du processus d'élaboration de la nouvelle vision stratégique de réforme de l'éducation et de la formation 2030-2015.

38 - Bernard Huggonier, « Evaluation : An Indispensable Tool for Improving Education Quality in Education in Morocco », Morocco Compact II, MCC, 2015, P. II.

2008 le programme national d'évaluation des acquis (PNEA), qui se veut un référentiel national d'évaluation des acquis, compétences et savoirs fondamentaux des élèves. Il est basé sur des tests disciplinaires (français, arabe, mathématiques, sciences et physique-chimie) et des questionnaires contextuels (questionnaires élève, parent, enseignant et établissement). Les tests menés en 2008 ont porté sur une cohorte d'élèves des 4e et 6e années du primaire et des 2e et 3e années du collège.

Ce dispositif PNEA devrait en principe appuyer les évaluations internes par un focus sur la dimension formative en assurant les feedback de remédiation aux écoles, aux enseignants et au système éducatif et également en complétant les aspects d'évaluation des connaissances par ceux de l'évaluation des compétences. Or, trois limites majeures handicapent clairement ces finalités essentielles. La première concerne la non pérennité du dispositif, il devrait être lancé tous les trois ans. Depuis sa première version de 2008, il est prévu un lancement en 2016. La deuxième limite porte sur le fait que le dispositif est implémenté sur un échantillon ce qui engendre une difficulté à avoir des éléments d'évaluation formative exhaustifs pour l'ensemble des écoles et des élèves. « **Making assessments on samples does not provide information on all students and all schools. The diagnosis is thus incomplete and does not allow establishing relevant remedies for all cases (students and schools)** ». La troisième limite a trait avec la déficience de la diffusion et de la dissémination des résultats de ces évaluations auprès des AREF, délégations provinciales et écoles.

c) Enquêtes internationales d'évaluation

Le Maroc a participé aux sessions de 1999, 2003, 2007 et 2011 de l'étude internationale TIMSS. Cette participation vise à évaluer les performances des élèves des niveaux transitoires des cycles primaires et secondaires collégiaux dans les disciplines mathématiques et sciences. De même, il participe de façon régulière à l'étude PIRLS depuis sa première session en 2001. Cette étude vise à évaluer les performances des élèves du cycle primaire en matière de lecture et ses compétences.

En matière d'analyse de la participation Marocaine à ce dispositif international, il est important de relever deux limites fondamentales. La première est liée à la problématique de diffusion des résultats aux différents acteurs des niveaux de gouvernance du système éducatif et de leur exploitation. La seconde porte sur le décalage entre le curricula du système Marocain et le curricula généralement adopté par les pays participants.

En synthèse, cinq constats majeurs en matière d'analyse du dispositif d'évaluation des élèves se déclinent comme suit :

- Limites de l'effet de diagnostic du dispositif d'évaluation des élèves sur le feedback aux choix des politiques éducatives ;
- Insuffisance en matière de remédiation locale dans le dispositif d'évaluation des élèves ;
- Prédominance de l'évaluation des connaissances par rapport aux compétences ;
- Carence en évaluations formatives ;
- Faiblesse de l'alignement de l'évaluation des élèves avec les standards ;

2. Evaluation des enseignants

L'évaluation des enseignants est assurée exclusivement par les inspecteurs, elle intervient à plusieurs niveaux. Elle se fait sous forme d'une évaluation normative à visée purement sélective au niveau du recrutement pour une formation initiale. En ce qui concerne l'inspection ou la visite de classe, même si elle se veut souvent formative, elle est très fréquemment couronnée par une note chiffrée (et donc quantitative). Le passage de l'inspecteur dans une classe est perçu par tous, comme étant un moment crucial qui va décider d'un avancement. Ainsi, on pourrait se demander où se situe la part de la formation dans une telle opération.

En matière d'analyse de ce dispositif d'évaluation des enseignants par les inspecteurs, le Conseil supérieur de l'éducation (CSE), dans son rapport « Etat et perspective du système d'Education et Formation, volume 4 : Métier de l'Enseignant », affirme que « **En fait, les visites effectuées par les inspecteurs sont des visites sommaires, espacées et interviennent essentiellement à l'occasion de l'échéance d'une promotion. En plus, à défaut de critères et d'objectifs précis d'évaluation et d'encadrement, la qualité des conditions de réalisation de ces visites peut être compromise** »³⁹. Ce constat est confirmé avec le constat de B. Huggonier dans son rapport de mission pour le compte du MCC, « **Teachers are, for now, assessed on a non-regular basis by inspectors** »⁴⁰.

39 - CSE, « Etat et perspective du système d'Education et Formation, volume 4 : Métier de l'Enseignant », Volume 2008, 4, P. 52.

40 - Bernard Huggonier, « Evaluation: An Indispensable Tool for Improving Education Quality in Education in Morocco », Morocco Compact II, MCC, 2015, P. 10.

En guise de synthèse, les principales limites sur l'évaluation des enseignants se déclinent comme suit :

- Évaluations effectuées exclusivement par les inspecteurs ;
- Elles sont globalement orientées par la promotion administrative ;
- Insuffisance de la régularité de cette évaluation ;
- Insuffisance de l'alignement avec les normes et les curricula accentuant l'impact sur le développement professionnel ;
- Absence du rôle du chef d'établissement dans l'évaluation des enseignants ;
- La composante du self-évaluation fait défaut tout en soulignant son incompatibilité avec le contexte culturel du Pays.

3. Evaluation des établissements scolaires

L'évaluation des établissements scolaires se trouve véritablement en situation paradoxale. D'un côté, un arsenal de dispositions réglementaires est disponible intégrant un minimum de mécanismes devant en principe permettre de maintenir une évaluation périodique des établissements leur assurant un niveau acceptable de prestations sur les plans pédagogique, administratif et en matière de moyens nécessaires à leur bon fonctionnement. Toutefois, la réalité est tout-à-fait en opposition avec ce qui est prévu dans ledit arsenal. Malgré les limites des composantes d'évaluation développées ci-dessus, pour les élèves et les enseignants, en principe les résultats de ces dernières devraient alimenter les mécanismes d'évaluation des établissements.

Pour ce volet particulier d'évaluation des établissements scolaire, il importe de mettre l'accent sur une première limite liée à l'absence d'un dispositif d'évaluation externe et périodique des établissements scolaires. « **All schools, [...] should be regularly (every 5 years) assessed (external assessment). This assessment should be standardized and made public** »⁴¹. Celui-ci donnerait plus de crédibilité à ses résultats, il gagnerait à être standardisé et diffusé au public. Il permettrait de repérer les faiblesses et les forces de chaque établissement, mettre en place le plan de remédiation approprié et assurer l'allocation des ressources adéquates. « **The national and local public authorities could therefore identify the strengths and weaknesses of the institutions and consequently develop a remediation plan if required by allocating more resources to the notoriously troubled institutions/schools** »⁴².

La deuxième limite est relative à l'implémentation du projet d'établissement, le système éducatif Marocain a une chance actuellement avec la mesure en cours de mise en œuvre pour assurer cette implémentation. Sauf que cette dernière se trouve confronté à trois handicaps : d'une part, le manque de ressources efficacement allouées à l'établissement en cohérence avec la taille et la consistance du projet d'établissement, d'autre part, l'absence d'un dispositif de self-évaluation qui devrait accompagner la mise en œuvre du projet d'établissement, son monitoring interne et son pilotage opérationnel et financier. « **To assist the implementation of « projet d'établissement », a self-assessment should be carried out regularly** »⁴³. Enfin, les capacités de gestion des chefs des établissements et de leurs équipes ne sont pas en mesure d'assurer la concrétisation de gestion et d'implémentation des projets d'établissements, de leur suivi et évaluation et de la mise en œuvre des plans de remédiation.

La troisième limite est liée à la non application rigoureuse des dispositions prévues par l'arsenal réglementaire dont en priorité le fonctionnement effectif des conseils d'établissement, cela constituerait une bonne plateforme de gouvernance qui vient renforcer les projets d'établissements et le dispositif de self-évaluation.

4. Evaluation du système éducatif Marocain

L'évaluation du système éducatif Marocain constitue le réceptacle des outputs des composantes d'évaluation déjà développées ci-dessus : des élèves, des enseignants et des écoles. Elle devrait être basée sur un alignement avec ces trois évaluations avec un background constitué des standards et des curricula, une cohérence entre les paliers du système des établissements scolaires jusqu'au niveau central et un feedback régulier et pérenne avec la politique de réforme du système éducatif.

Malgré les recommandations de la Charte et les attributions confiées au ministère en matière de l'évaluation,

41 - Op. Cit. P. 9.

42 - Ibid

43 - Ibid

ainsi que son importance incontournable dans le pilotage et la régulation du système éducatif, les évaluations, menées de manière systématique et régulière, font défaut. Et si elles sont faites, elles n'obéissent pas de manière satisfaisante aux normes préconisées selon les règles de l'art de l'évaluation.

Nous résumons ci-après les principales limites de l'évaluation du système éducatif Marocain comme suit :

- Absence d'une politique intégrée de l'évaluation du système éducatif Marocain avec en plus un déficit en termes de régularité et de pérennité ;
- Prédominance de la dimension efficacité au détriment des dimensions qualité des apprentissages et efficience ;
- Alignement insuffisant avec l'évaluation des élèves, des enseignants et des écoles ;
- Cohérence non déclinée au niveau des paliers terrain du système notamment les établissements scolaires ;
- Publication et redevabilité non implémentées ;
- Absence du processus de feedback aux politiques et aux acteurs locaux ;
- Impacts sur l'insuffisance de la remédiation surtout au niveau des établissements scolaires et des classes.

IV. Pistes d'amélioration

Prenant en compte les éléments du cadre théorique développé en particulier dans le présent article et les constats émergeant de l'état des lieux, la recommandation saillante et fédératrice qui semble prioritaire est de mettre en place une politique intégrée d'évaluation du système éducatif Marocain. L'intégration préconisée vient du fait que cette politique doit accorder une attention cruciale à quatre préoccupations essentielles. La première porte sur la pérennité de l'implémentation et du fonctionnement de la politique d'évaluation, et ce en vue de garantir une continuité dans le pilotage de l'évolution de la situation du système éducatif dans sa globalité et des prestations éducatives servies aux élèves dans les écoles et les classes.

La deuxième préoccupation est liée à la nécessité d'assurer un alignement sans faille entre les composantes clés préfigurant la politique d'évaluation. Cet alignement est à double dimension, une dimension technique et une dimension sociale. Nous développerons par la suite les éléments saillants de l'approche de cet alignement. La troisième préoccupation porte quant à elle sur une délimitation précise des responsabilités des acteurs du système éducatif des quatre paliers central, régional, provincial et local en matière de conception, de mise en œuvre, de diffusion et d'exploitation des résultats de la politique d'évaluation.

La dernière préoccupation que nous pensons pertinente dans la mise en place d'une politique d'évaluation, porte sur la capacité du système éducatif Marocain à concrétiser l'impératif préalable à assurer un feedback pérenne et systématique des résultats des évaluations aux acteurs, aux parties prenantes et surtout aux processus d'actualisation des stratégies de réforme de l'éducation en particulier pour le niveau national. Les mécanismes de mise en œuvre efficace de ces feedback sont à développer, à tester et à implémenter en cohérence avec ladite politique d'évaluation.

Ceci étant souligné, nous proposons, sur la base de ce qui ressort de notre investigation documentaire et théorique, un modèle de conception et de mise en œuvre d'une politique d'évaluation spécifique au système éducatif Marocain. Ce modèle est préconisé de manière à ce qu'il soit adapté au contexte Marocain et aux spécificités du système éducatif essentiellement en termes de sa culture, de son organisation administrative et pédagogique, de son niveau de décentralisation mais également tenant compte des modalités actuelle et prospective de sa gouvernance eu égard aux choix émanant de la vision stratégique 2015-2030 du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique. Ce modèle est structuré en trois composantes saillantes : (i) composante de conception globale, (ii) composante d'alignement technique et enfin, (iii) composante d'alignement social.

1. Composante de conception globale

S'agissant de la dimension de ce modèle liée à une conception globale et cohérente, nous estimons que le dispositif d'évaluation issu de la dite politique d'évaluation du système éducatif Marocain, doit porter sur deux niveaux d'objectifs comme l'illustre schématiquement la figure O2 ci-après. Il est en effet impératif de commencer la conception par la fixation des objectifs qui constituent le référentiel par rapport auquel la politique d'évaluation va se mettre en œuvre. Le premier niveau porte sur la fixation de quatre catégories

d'objectifs : de qualité, d'efficacité, d'efficience et d'équité. Les finalités ciblées dans la fixation de ces objectifs concernent :

- la progression du niveau des apprentissages que le système est appelé à réaliser pour les objectifs de qualité ;
- l'évolution de la part des élèves qui sont en mesure de réaliser une meilleure rétention et un achèvement élevé et également l'évolution de la baisse de la part des élèves qui sont en difficultés d'apprentissage pour les objectifs d'efficacité ;
- le suivi du niveau de retour sur investissement à travers l'analyse des dépenses engagées en comparaison avec les résultats réalisés de la qualité et de l'efficacité pour les objectifs d'efficience ;
- l'évolution du niveau d'équité d'accès aux prestations de l'éducation dans le sens de limitation des effets négatifs des déterminants sociaux tels que genre, milieu et revenus des ménages pour les objectifs d'équité.

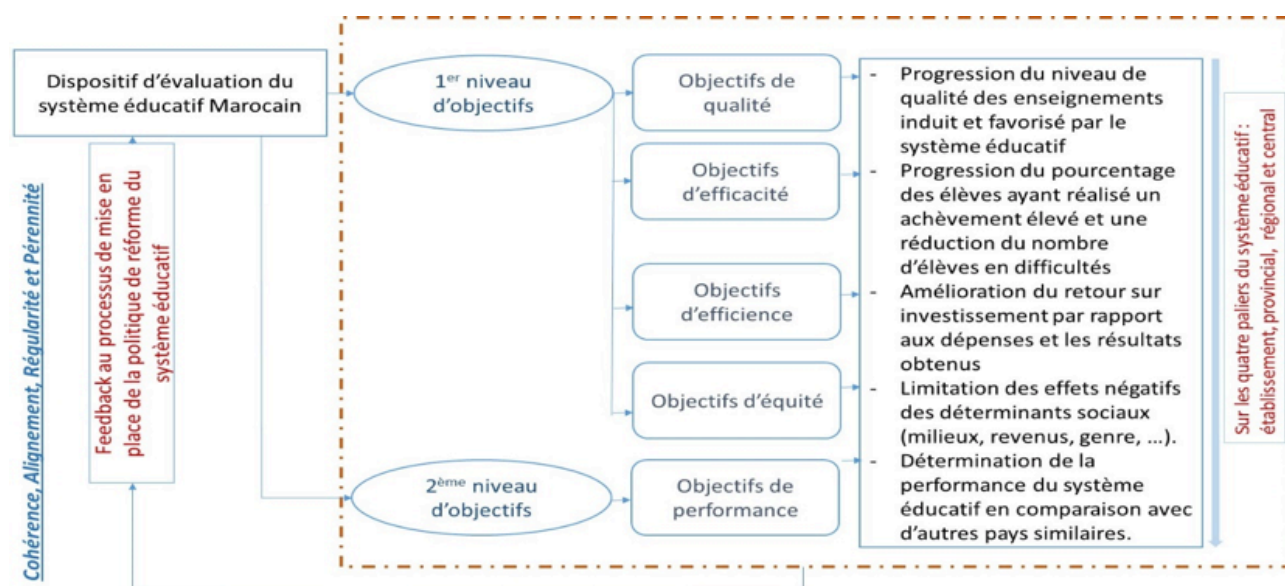


Figure O2 : Schéma de fonctionnement requis du dispositif d'évaluation du système éducatif Marocain⁴⁴

Le second niveau a trait avec la fixation des objectifs de performance en mettant l'accent sur la comparaison de la performance du système éducatif Marocain en termes des objectifs précités de qualité, d'efficacité, d'efficience et d'équité avec les performances réalisées par d'autres pays similaires. Une fois ces objectifs sont fixés, il est plus qu'impératif de décliner ceux-ci sur les différents paliers du système éducatif, régional, provincial et au niveau des établissements scolaires. Nous pensons prudent de décentraliser le dispositif d'évaluation au même titre que la décentralisation des processus de mise en œuvre de la stratégie de réforme du système éducatif. La mesure du niveau d'atteinte des objectifs précités doit partir du niveau local vu l'intérêt prioritaire accordé aux élèves qui sont au centre de toute stratégie de réforme, mais aussi se positionner aux niveaux intermédiaires de management du système pour servir les préoccupations de pilotage, régulation et en terme de finalité ultime, permettre de mettre à jour les choix et modalités de la stratégie et assurer les remédiations qui s'imposent.

La problématique qui se voit à notre sens complexe après fixation et déclinaison des objectifs à évaluer, porte sur la question de comment assurer la disponibilité des informations relatives aux résultats des évaluations effectuées sur les différents paliers de décentralisation du système éducatif. Ce qui renforce la pertinence de mettre en place un système d'information intégré, accessible par les acteurs des quatre paliers éducatifs et rendant disponible les résultats des différentes évaluations retenues, des élèves, des enseignants, des établissements scolaires et du système éducatif dans sa globalité.

2. Composantes d'alignement technique

Cette composante vient expliciter la logique globale du modèle préconisé en matière d'évaluation dans l'éducation sur une deuxième perspective après la première portant sur la conception globale de la politique

⁴⁴ - Schéma inspiré des réflexions de (Bernard Huggonier, « Evaluation: An Indispensable Tool for Improving Education Quality in Education in Morocco », Morocco Compact II, MCC, 2015, P. 2).

d'évaluation du système éducatif Marocain. Cette perspective met le focus sur la nécessité de concrétiser l'alignement technique entre trois éléments majeurs qui sont les normes des apprentissages, les curricula et l'évaluation. Ce dernier élément doit cerner bien évidemment l'évaluation des élèves, des enseignants et des établissements scolaires. La finalité qui semble prioritaire pour le contexte Marocain, est d'éviter la situation actuelle de procéder à des évaluations de tous ces éléments mais de manière isolée et disparate.

Nous voulons insister sur la pertinence voire l'urgence de tisser des passerelles de cohérence et synergie entre les résultats des évaluations réalisées. Il s'agit ainsi de mettre en place une plateforme d'alignement technique dans le sens clarifié par le schéma O3 ci-après. Le préalable qui vient donner amorce à l'implémentation de cette plateforme est de procéder à la vérification que les normes d'apprentissage sont bien explicites et pertinentes en relation avec la nouvelle stratégie de réforme qui vient d'être produite et approuvée par le CSEFRS. Le défi particulier dans la détermination de ces normes consiste à cerner, pour les différents niveaux et cycles d'enseignement en plus des connaissances de base, les compétences dites transversales « soft skills » liées à l'analyse critique, à la résolution de problèmes et de communication.

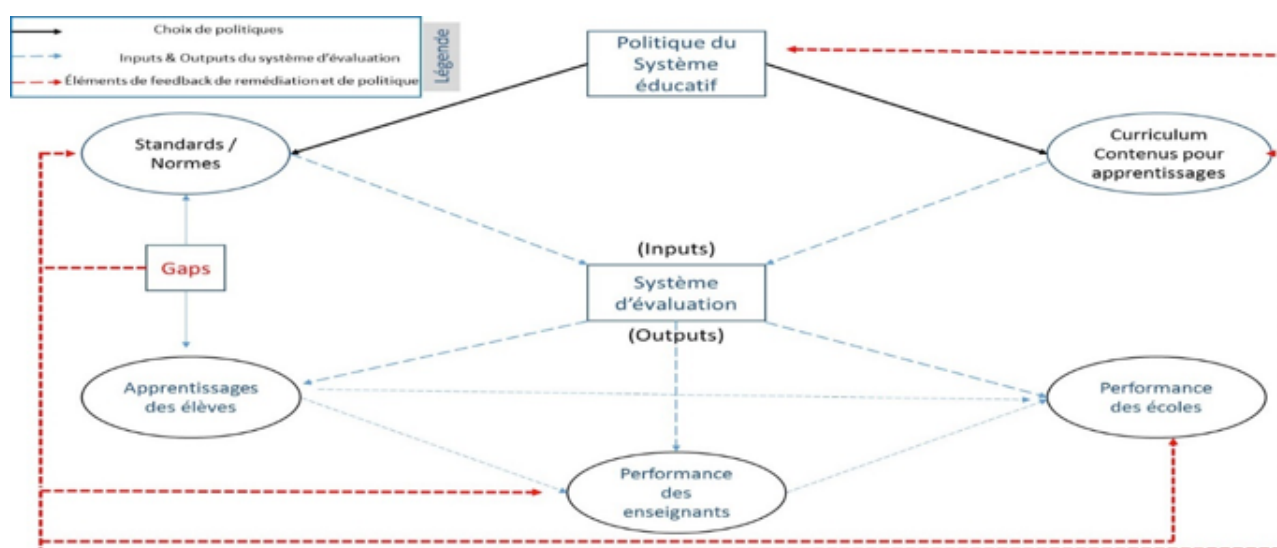


Figure O3 : Schéma de la logique d'alignement technique entre les composantes du système d'évaluation⁴⁵

Après ce préalable de normes et de curricula, la plateforme d'alignement technique doit s'articuler dans une perspective de structure générale et cohérente autour de trois éléments : évaluation des élèves, évaluation des enseignants et évaluation des établissements scolaires. De manière synthétique et centrée sur le cas du système éducatif Marocain, nous déclinons les recommandations saillantes de chaque élément parmi ces trois.

S'agissant de l'évaluation des élèves, il s'agit en priorité de rendre opérationnel et exhaustif pour tous les établissements scolaire, l'évaluation diagnostic des acquis des élèves en s'inscrivant dans la logique de l'évaluation formative. Celle-ci constitue un levier majeur qui permet d'assurer l'amélioration des pratiques pédagogiques en classes par le biais du processus de remédiation locale. Ce levier vient en complément aux examens certificatifs et du programme d'évaluation nationale des acquis (PNEA). La deuxième priorité concerne la nécessité d'aligner ces évaluations formatives aux normes et standards fixés. La difficulté de diagnostiquer les acquis des élèves en matière de compétences transversales doit être prise en compte et surmontée. Une troisième priorité doit être consacrée à la capacité du système à assurer une diffusion et une accessibilité aux informations issues des évaluations par les élèves, les parents, les enseignants, les chefs des établissements et les autres parties prenantes. Une dernière priorité est accordée à l'amélioration du dispositif du PNEA en mettant l'accent en priorité sur trois aspects : la pérennité de sa mise en œuvre, la couverture de tous les cycles et niveaux d'enseignement et enfin, la mise en place d'un cadre institutionnel assurant la prise en charge de cette évaluation dans le sens d'accorder à cette dernière une crédibilité et une neutralité suffisantes.

En ce qui concerne l'évaluation des enseignants, celle-ci, mérite d'être revue et développée à travers un mix d'interventions synergiques. D'abord, rendre les évaluations par les inspecteurs plus formatives, plus fréquentes, plus alignées aux standards et curricula et guidées par des finalités de développement

45 - Schéma inspiré de la réflexion de Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, P. 11-14.

professionnel des enseignants et pas seulement de promotion administrative. Cela constitue à notre sens un défi majeur pas facile à concrétiser sans une approche de concertation avec le corps des inspecteurs sur une base de dialogue transparent, et ouvert et de responsabilité partagée.

Néanmoins, même avec une amélioration substantielle de l'évaluation des enseignants par les inspecteurs, il est également impératif de compléter ces évaluations par deux autres catégories. D'une part, des évaluations utiles des chefs d'établissements qu'il y a lieu de les focaliser sur les aspects d'animation éducationnelle, de gestion des classes et de dynamisation des relations avec les élèves et les parents. Il s'agit également de mettre en place les self-évaluations qui peuvent certes perçues comme incompatibles avec le contexte culturel du pays mais cela peut être géré par le biais d'un bon système de standardisation, de supervision et de motivation des acteurs clés dont enseignants et chefs d'établissements scolaires.

Pour ce qui est de l'évaluation des établissements scolaires, cette évaluation se trouve en situation de déficit accentué alors que c'est à ce niveau que les informations nécessaires à une évaluation pérenne et pertinente des politiques de réforme et à une redevabilité concrète vis-à-vis des acteurs clés, sont créées et rendues disponibles et accessibles. Il nous semble qu'il est vivement préconisé de construire et d'implémenter un dispositif d'évaluation des établissements scolaires en tenant compte de trois priorités fondamentales.

La première est de rendre concret et applicables sur terrain les dispositions réglementaires en matière de fonctionnement des conseils de gestion de l'établissement scolaire. La deuxième priorité est de mettre en place une self-évaluation qui soit alignée avec le management des projets d'établissement. Il s'agit de procéder à une évaluation prise en charge et opérée par les enseignants et les chefs d'établissement intégrant essentiellement l'orientation de l'alignement tel que décrit précédemment. Alignement basé sur une cohérence entre les standards, les informations issues de l'évaluation des élèves, l'évaluation des enseignants et l'évaluation de la performance des établissements. La troisième priorité porte sur la pertinence d'équilibrer la self-évaluation des établissements scolaires par une évaluation externe, dans le sens où elle doit être gérée et prise en charge par des acteurs du système éducatif mais externe de l'établissement. La finalité étant de booster la crédibilité des résultats de l'évaluation de l'établissement, qui n'est certes pas facile à obtenir de la self-évaluation et d'avoir la possibilité logique de procéder à la diffusion de ces résultats aux différents acteurs et parties prenantes selon un cadre de transparence et de redevabilité.

3. Composante d'alignement social

Le modèle préconisé de la politique d'évaluation est complété par cette troisième composante d'alignement social. Cet alignement vient compléter la plateforme d'alignement technique. En effet, au-delà de toute l'attention à accorder à l'alignement technique entre les ingrédients du dispositif d'évaluation, malgré son importance cruciale, il est fondamental de voir de plus près les préoccupations liées à une implémentation concrète de ces différentes évaluations au sein de l'établissement scolaire eu égard au niveau requis en termes d'appui, de motivation et d'adhésion des acteurs et parties prenantes majeurs.

Comme le montre le schéma O4 ci-après, nous avons structuré cette plateforme d'alignement social autour de quatre catégories d'acteurs et parties prenantes clés. La plateforme vient tenter solutionner les principaux défis posés devant une meilleure implémentation d'un environnement approprié d'alignement social. La finalité de cet alignement se décline autour du noyau central de tout système éducatif à savoir l'élève et de manière corollaire l'environnement qui le prend en charge à savoir l'établissement scolaire. C'est ainsi que ladite finalité consiste à assurer une dualité de buts synergiques, l'un alimente et booste l'autre ; la progression des apprentissages des élèves et le développement professionnel des établissements scolaires.

S'agissant de la première catégorie des enseignants, malgré toute la visibilité qui se dégage de l'alignement technique, les défis sérieux sont au niveau de la capacité du système éducatif Marocain à assurer une implémentation efficace sur terrain dans les établissements scolaires. Les enseignants ont besoin de formation continue, d'appuis et d'encadrements permanents à leurs responsabilités dans l'évaluation diagnostic des apprentissages des élèves et de mise en œuvre des remédiations adéquates. Un feedback régulier et pérenne doit être assuré pour les enseignants si nous voulons espérer réussir dans la progression des apprentissages en relations avec les normes et standards fixés. Il s'agit là d'une dynamique prometteuse mais complexe à réussir. En plus du handicap d'encadrement et de renforcement des compétences des enseignants, le manque de ressources constitue pour le cas de notre système éducatif un second handicap qu'il importe de solutionner dans le cadre d'une approche de gouvernance qui soit rationnelle et acceptables par les parties prenantes gouvernementales.

En ce qui concerne la deuxième catégorie des chefs des établissements scolaires, celle-ci vient assurer

l'environnement adéquat à la relève des défis liés aux enseignants. Le mot clé est sans doute le leadership. Il importe de voir comment assurer un leadership à travers la sélection des chefs d'établissements, leurs formations initiale et continue. Ce leadership s'avère fondamental dans la collaboration nécessaire pour tous les processus d'évaluation interne et externe et de mise en place des projets d'établissement. Le point central dans tous ces processus est de garantir un travail collectif de mesure de la progression des apprentissages des élèves en relation avec le référentiel des normes et standards, d'exploitation des résultats et de leur prise en compte dans l'élaboration des plans d'intervention à court et à moyen termes. A ce niveau également deux handicaps mis en évidence, les capacités de leadership et les ressources nécessaires.

En plus des enseignants et des chefs d'établissements, la troisième catégorie est liée aux inspecteurs. Nous avons mis le focus sur les limites du système d'évaluation tel que pratiqué sur terrain, il s'agit en priorité de revoir les critères d'évaluation des enseignants en s'appuyant sur un alignement avec les programmes et normes. Toutefois, le défi majeur reste à notre sens lié à réussir à mettre en place une approche d'évaluation centrée sur le partage d'expériences et de feedback envers les enseignants. C'est d'ailleurs la finalité de l'encadrement de proximité que préconise la nouvelle stratégie de réforme 2015-2030 du CSEFRS. Il importe également que cette nouvelle approche puisse consacrer une meilleure collaboration et partage de bonnes pratiques entre les établissements scolaires dans le cadre des communautés de pratiques professionnelles⁴⁶ par le biais particulier des relais des inspecteurs.

La dernière catégorie a trait avec les établissements scolaires et leurs communautés. Il s'agit plus prioritairement de mettre en place un environnement de dialogue et de coopération avec la communauté de l'établissement scolaire dont très particulièrement les parents et les élèves. L'implication des parents et des élèves s'avère plus que primordiale et ce en particulier dans l'interprétation et l'exploitation des résultats, les actions de remédiation et la mise en place des projets d'établissement et leurs suivi et pilotage.

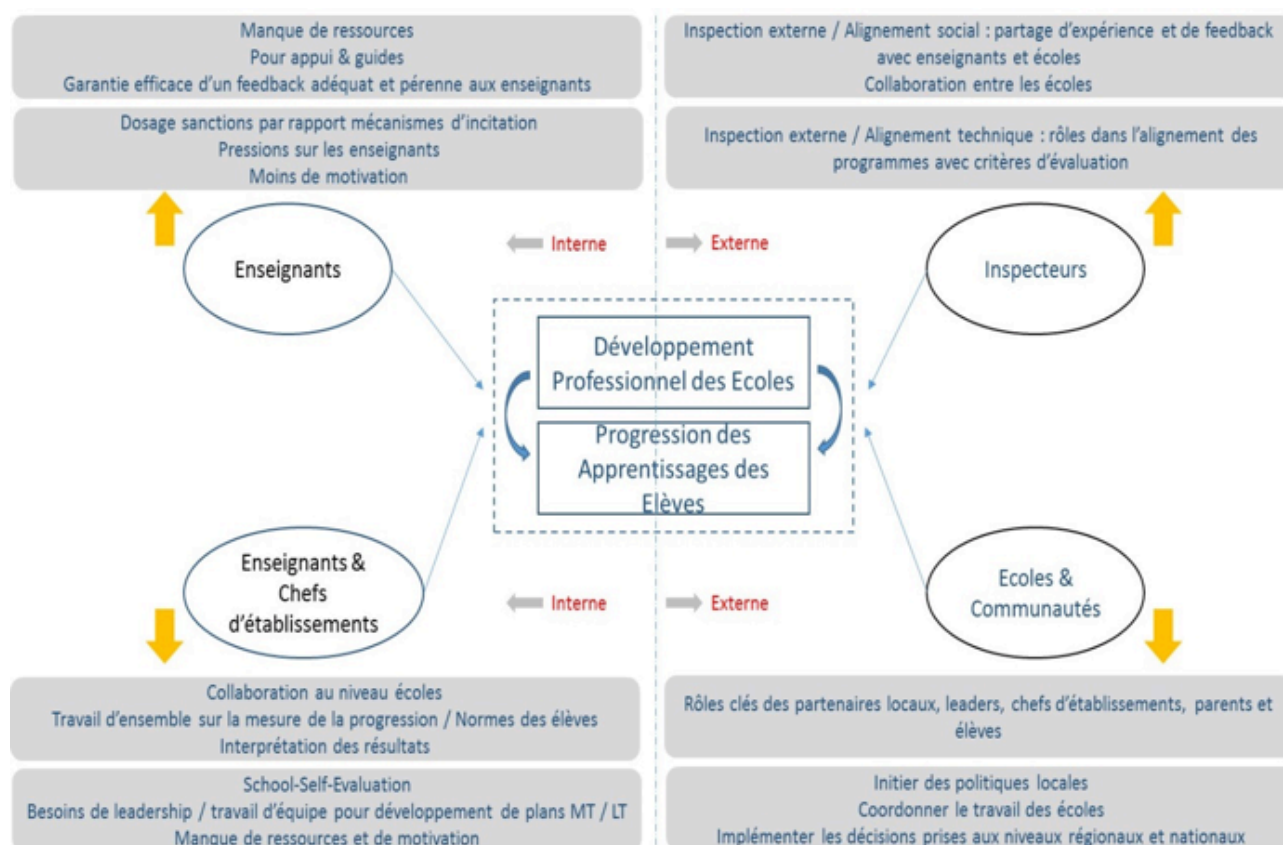


Figure O4 : Schéma des défis d'implémentation d'un environnement approprié d'alignement social / système d'évaluation dans l'éducation⁴⁷

46 - Abrégées en CPP, ce dispositif est actuellement conçu et expérimenté dans le cadre du programme de coopération avec le Canada (PAGESM : Programme d'Appui à la Gestion des Etablissements Scolaires du Maroc).

47 - Schéma inspiré des réflexions tirées de (Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, P. 18-24).

Conclusion

Le présent article s'est proposé d'aborder une problématique bien délimitée, il s'agit de voir d'abord, par le biais d'un dispositif approprié d'évaluation, si nous disposons de la possibilité et des capacités institutionnelles à constater si les politiques de réforme sont pertinentes et appropriés quant à leurs impacts sur la qualité, l'efficacité, l'équité et l'efficacité du système éducatif. Il convient aussi de voir, par le même biais, si nous pouvons nous prononcer sur l'efficacité du processus de mise en œuvre de ces politiques de réforme. La finalité ultime étant d'apprécier la pertinence de la conception de la politique et l'efficacité de sa mise en œuvre.

Les résultats obtenus méritent à notre sens d'être rappelés de manière synthétique. D'abord sur un plan macro de l'analyse effectuée, la limitation observée en matière de pérennité de nos politiques de réforme impacte la pérennité et la régularité du dispositif d'évaluation. Ce déficit de régularité est accentué par l'absence d'une politique cohérente et intégrée d'évaluation du système éducatif. Il y a certes des évaluations de différents profils et catégories qui sont réalisés mais elles sont mises en place de manière isolée et disparate. Ce qui limite l'impact positif et attendu sur la conception et la mise à jour des politiques de réforme mais également atténue substantiellement la portée du feedback aux établissements scolaires et du processus de remédiation au niveau des pratiques pédagogiques en classes.

Sur un plan micro, il est constaté le déficit d'alignement très requis et vital dans tout dispositif cohérent d'évaluation. D'une part, alignement qui reste insuffisant entre les normes et standards d'apprentissage, les curricula et l'évaluation. D'autre part, alignement également insuffisant entre les quatre niveaux retenus dans ledit dispositif d'évaluation : niveau d'évaluation du système éducatif de manière globale, niveau d'évaluation des élèves, niveau d'évaluation des enseignants et niveau d'évaluation des établissements scolaires. Des éléments de diagnostic sont développés de manière détaillée pour chacune des quatre composantes à évaluer système éducatif, élèves, enseignants et établissements scolaires. Les progrès ayant été acquis et limites sont explicités et analysés.

Sur la base de ces résultats, et en matière de pistes préconisées de correction de la présente situation, un modèle a été suggéré pour la mise en place d'une politique d'évaluation du système éducatif Marocain, ce modèle est architecturé autour de trois composantes. Une composante liée à la conception de politique globale d'évaluation du système éducatif Marocain. Une deuxième composante portant sur la mise en place d'une plateforme d'alignement technique traitant les préoccupations d'alignement entre les normes, curricula et processus d'évaluation lui-même. Enfin, une dernière composante relative à l'implémentation d'une plateforme d'alignement social, celle-ci apporte une vue précise sur l'environnement approprié et nécessaire et les facteurs clés de succès de la mise en œuvre du processus d'évaluation au niveau des établissements scolaires.

Au-delà de ces éléments de diagnostic et de recommandations, il importe de soulever une contrainte majeure à cette réflexion. Elle porte sur la démarche adoptée mettant le focus essentiellement sur la problématique de l'évaluation. Toutefois, la question de la performance de toute politique de réforme de systèmes éducatifs est complexe dans la mesure où ces systèmes sont eux-mêmes qualifiés d'un niveau profond de complexité en plus du fait que cette performance dépend de plusieurs autres facteurs autres que l'évaluation.

Traduction du Résumé**Abstract Translation**

Abdelhaq EL HAYANI

**EVALUATION OF THE IMPACT OF MOROCCAN
EDUCATION SYSTEM REFORM POLICIES : OVERALL
REVIEW OF DESIGNING AND IMPLEMENTATION ISSUES**

This article is part of a context qualified by major challenges in relation with the performance of the Moroccan education system. This context is currently guided by the implementation of the new strategic vision 2015-2030 of the Higher Education Council, training and scientific research. This vision is based on four major goals including the scope to cross on the governance and leadership. Among the dimensions that were discussed for this, there is the question of evaluation. The main objective of this article is about two perspectives. First analyze the current system of evaluation of the Moroccan education system through a review of documentary diagnosis while referring to the most recommended theoretical frameworks for evaluation in education. Then try to advocate some areas for improvement and to correct certain identified limits of the current situation.

Conceptual approaches such as declined in the review of theoretical frameworks have helped guiding and structuring our analysis, a set of limits are highlighted on the evaluation system as a whole, but also specifically to the respective assessments students, teachers and schools. A model of conceptualization of the evaluation for the case of the Moroccan educational system was recommended. It is structured around three key components. A first component related to the design of global evaluation policy of the education system. A second component on the establishment of technical alignment platform addressing the concerns raised for good alignment between standards, curricula and assessment process. A final component relating to the implementation of a social platform alignment clarifying explicitly successful implementation conditions of these assessment processes in schools according to adherence, motivation, ownership and the capacity of local actors.

Keywords: education system, reform policy, evaluation, technical alignment, social alignment.

Bibliographie et Références

1. Bernard Huggonier, « Evaluation : An Indispensable Tool for Improving Education Quality in Education in Morocco », Morocco Compact II, MCC, 2015, 18P ;
2. Charte Nationale de l'Éducation et de Formation ;
3. CSE : « Etat et Perspectives du Système d'Éducation et de Formation », volume 2 rapport analytique, 2008 ;
4. CSE : « Programme National d'Évaluation des Acquis des Elèves (PNEA) », 2009, <http://www.csefrs.ma/pageFr.aspx?id=247> ;
5. CSE : « Etat et Perspectives du Système d'Éducation et de Formation », Volume 4 : métiers de l'enseignant, 2008 ;
6. CSEFRS : Rapport analytique, « La mise en œuvre de la charte nationale d'éducation et de formation 2000-2013 acquis, déficits et défis », décembre 2014.
7. Danièle Lamarque, « L'évaluation des politiques publiques Locales », L.G.D.J, 2004, 100P. ;
8. Décret n°2.02.376 le 6 Joumada I 1423 (17 juillet 2002) portant Statut particulier des établissements d'éducation et d'enseignement publics ;
9. Décret n° 2-02-382 du 6 joumada1 1423 (17 juillet 2002) portant attributions et organigramme du ministère de l'éducation nationale ;
10. Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, 100P. ;
11. Janet Looney, « Alignment in Complex Education Systems : Achieving Balance and Coherence », OCDE, 2011, 100P. ;
12. Khalid EL MIR et Idriss KASSIMI : « Législation administrative et gestion pédagogique (en arabe) », Imarci (imprimerie des arts du 21ème siècle), 2006 ;
13. Loi n° 07.00 relative à la création des AREF (BO n°4798 du 25 mai 2000) ;
14. Marguerita Clarke, « What Matters Most for Student Assessment Systems: A framework paper », Banque Mondiale, 2012, 100P. ;
15. MEN (direction de la statistique, de la prospective et de la programmation) : « Evaluation du niveau d'acquisition des élèves de la 4^e année fondamentale », mai 1995 ;
16. MEN : Rapport résumé sur les résultats globaux de la participation du Maroc à l'étude internationale TIMSS, 2011 (format électronique) ;
17. MEN : Rapport résumé sur les résultats globaux de la participation du Maroc à l'étude internationale PIRLS, 2011 (format électronique) ;
18. MENESFCRS (DSSP) : « cadre stratégique de développement du système éducatif », avril 2005 ;
19. MENESFCRS : « Pour un nouveau souffle de la réforme, état des lieux (7 problématiques) », mai 2008 (format électronique) ;
20. MENESFCRS : Rapport National sur le Bilan Intermédiaire du Plan d'Urgence au titre des années 2009 et 2010 et le 1er semestre de 2011, rabat septembre 2011 ;
21. MENFP : Rapport national des rencontres consultatives sur l'école marocaine (en arabe, format électronique), septembre 2014 ;
22. Moha JMAD : Article de recherche « Evaluation des pratiques d'enseignants », <http://jmadoux.e-monsite.com/pages/evaluation-des-pratiques-d-enseignants.html>;
23. Paulo Santiago, « Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges, OCDE, 2009, 100P.

MODES DE GOUVERNANCE, PRATIQUES D'ÉVALUATION ET LEUR IMPACT SUR LES POLITIQUES ÉDUCATIVES PUBLIQUES

Amina DEBBAGH *

Résumé

La communication vise à contextualiser la pratique d'évaluation dans le secteur de l'éducation au regard des modes de gouvernance de ce secteur. En passant en revue de façon synthétique, les différentes pratiques, en relatant leur diversité par rapport à leur objet (les apprentissages, les pratiques enseignantes, les établissements et autres structures éducatives, les programmes et projets éducatifs...), à leurs méthodologies, à leurs objectifs et aux structures qui en ont la charge, nous tenterons de montrer que leur efficacité et leur pertinence sont largement déterminées par les modes de gouvernance éducative adoptés. L'objectif de la communication est aussi de s'interroger sur l'utilité de la pratique évaluative, son degré d'objectivité et de fiabilité ainsi que son impact sur les politiques éducatives publiques, en référence à quelques expériences internationales et au cas marocain.

Des systèmes éducatifs dans un monde en transformation

De par le monde, les systèmes éducatifs ont évolué et se sont considérablement transformés durant les trente dernières années. Cette évolution et ces transformations ont touché divers aspects de ce secteur sous la double pression de l'expansion des effectifs et de la diversité de leurs besoins et attentes d'une part et des mutations économiques, technologiques et sociales qui ont accompagné le processus de mondialisation, d'autre part. Ces mutations ont profondément modifié les attentes de la société en termes de capacités, de qualifications et de compétences que l'éducation doit assurer aux jeunes générations et leur adéquation aux nouvelles exigences de la mondialisation.

Selon leur niveau de développement, pratiquement tous les pays se sont engagés dans des processus de réforme de leurs systèmes éducatifs visant à améliorer la qualité des enseignements et apprentissages, à mettre à niveau les curricula, les programmes et les méthodes pédagogiques, à réduire les inégalités sociales et territoriales face à l'éducation et à assurer l'ouverture des établissements d'enseignement et de formation sur leur environnement local, national et international.

De ce fait, la plupart des systèmes éducatifs ont connu de profonds changements dans leur organisation interne et leurs modes de gouvernance, avec une redéfinition des rôles des acteurs et des partenaires et un renouvellement des outils de leur pilotage (et de leur contrôle) avec un recours massif à l'évaluation.

* Economiste de l'éducation, Professeur à la Faculté des Sciences de l'Éducation, Université Mohammed V de Rabat

Dans ce contexte, une littérature abondante s'est développée ces dernières années, avec un lexique tout à fait « nouveau » dans le domaine éducatif et qui est largement emprunté aux champs du management privé et du nouveau management public (New Public Management) : gouvernance, pilotage, régulation, responsabilisation, performances, accountability/redevabilité, etc.

La communication vise à contextualiser la pratique d'évaluation dans le secteur de l'éducation au regard des modes de gouvernance de ce secteur. En passant en revue de façon synthétique, les différentes pratiques, en relatant leur diversité par rapport à leur objet (les apprentissages, les pratiques enseignantes, les établissements et autres structures éducatives, les programmes et projets éducatifs ...), à leurs méthodologies, à leurs objectifs et aux structures qui en ont la charge, nous tenterons de montrer que leur efficacité et leur pertinence sont largement déterminées par les modes de gouvernance éducative adoptés. L'objectif de la communication est aussi de s'interroger sur l'utilité de la pratique évaluative, son degré d'objectivité et de fiabilité ainsi que son impact sur les politiques éducatives publiques, en référence à quelques expériences internationales et au cas marocain.

I. Modes de gouvernance et pratiques évaluatives en éducation

De la gouvernance

Il est aujourd'hui reconnu que « la gouvernance » est un concept polysémique, dont le contenu et l'objet sémantique ne sont pas toujours clairement identifiés et peuvent parfois différer en fonction du domaine concerné (gouvernance politique, gouvernance économique, gouvernance d'entreprise, gouvernance territoriale ... et gouvernance éducative). Toute une littérature s'est développée autour de cette thématique au cours des vingt dernières années. « Si la gouvernance n'existait pas, il faudrait l'inventer », disait Alain Bouvier (2007)¹ pour montrer son importance dans le contexte actuel de complexification du système éducatif, alors que pour d'autres, il ne s'agit que « d'un terme à la mode (qui) confère parfois un aspect moderne à des réalités anciennes² » ou encore « une idéologie du désengagement de l'État-Providence, voire une théorie de la décomposition de l'État, à l'œuvre depuis le tournant néolibéral des années 1980 » qui signe le passage à « une société pragmatiste, particulariste et utilitariste, garante d'intérêts économiques singuliers, dans laquelle la notion de bien commun n'a plus de place véritable³ »

Au-delà de ce débat entre détracteurs et partisans de la « gouvernance », et nonobstant son intérêt, il est important de préciser que l'emprunt de la notion de gouvernance au domaine du management du secteur privé a commencé avec la réflexion sur le rôle de l'État et du secteur public dans la vie économique, dans le développement social et, bien entendu, dans le domaine de l'éducation ainsi que dans le cadre de la renégociation du partage des responsabilités publiques entre les acteurs sociaux, les structures déconcentrées et les entités territoriales (régions, départements, districts, provinces, communes, etc.)⁴.

La notion de gouvernance renvoie ainsi à l'art de gouverner et recouvre un ensemble de composantes ayant trait aux dispositifs et procédures de gestion des affaires publiques, au style de leadership, aux mécanismes de consultation, de coordination et de prise de décisions et à la nature et aux types de relations entre les structures et les acteurs pour la réalisation d'objectifs communs⁵.

Principaux « modèles » de gouvernance de l'éducation

Au regard de l'évolution des systèmes d'éducation dans le monde, on peut affirmer qu'il n'y a pas un modèle unique de gouvernance mais bien des systèmes de gouvernance qui se basent sur des approches théoriques diverses. A la lumière des expériences de la plupart des pays, et sur la base d'un ensemble de critères relatifs tant aux aspects organisationnels et ceux de gestion de leurs ressources, qu'aux modalités

1 - A. Bouvier, La gouvernance des systèmes éducatifs, 2007

2 - Bernard Toulemonde, De nouveaux modes de gouvernance, Cahiers français n°368, 2012, p.47

3 - J.-C. Mathias, Politique de Cassandre, Sang de terre, 2009

4 - Selon Louis CÔTÉ, Benoît LÉVESQUE et Guy MORNEAU, la notion de « gouvernance est intéressante en ce qu'elle permet d'appréhender le fait que des responsabilités autrefois dévolues à l'État sont actuellement progressivement partagées avec les acteurs sociaux et qu'émergent ainsi de nouvelles articulations entre l'État et la société civile », in « L'évolution du modèle québécois de gouvernance », Politique et Sociétés, vol. 26, no.1, 2007, p.2. cités par Isabelle Lacroix et Pier-Olivier St-Arnaud, La gouvernance : tenter une définition, Cahier de recherches en politique appliquée, n°3, Automne 2012

5 - L'acteur principal étant l'État, qualifié tantôt « d'État stratège », « d'État régulateur », « d'État évaluateur », voir « d'État négociateur », in L'évolution des modes de régulation de l'action publique en éducation, Y. Dutercq et A. Van Zanten, Revue Education & Société, n°8, 2001-02, p.7. Le fait est qu'on assiste aujourd'hui à « Une redéfinition des formes d'intervention de l'État dans la gestion de l'éducation ... dans différents contextes culturels et nationaux », Malet Régis, Les politiques d'imputabilité en direction des enseignants en Angleterre et aux États-Unis, in Education comparée n°5, 2011

de fonctionnement des structures éducatives et de leur financement, on peut globalement distinguer deux grands types de modèles :

- Le « **modèle de contrôle ou de régulation** », avec ses diverses variantes (centralisé, déconcentré et décentralisé selon la répartition des compétences pédagogiques, administratives et financières entre les entités centrales, régionales et locales) et dans lequel prédomine le mode de gestion bureaucratique puis post bureaucratique (hiérarchie, injonction réglementaire, financement et contrôle centralisés ...)
- Le « **modèle de développement professionnel ou transformationnel** » caractéristique de la plupart des pays anglo-saxons ayant un régime politique fédéral ou confédéral (USA, Canada, Allemagne, Suisse ...) et dans lequel de larges prérogatives et une autonomie effective sont octroyées au niveau local (échelons intermédiaires et établissements) et aux corps professionnels, essentiellement les enseignants.

Il faut préciser que la frontière entre le « modèle de contrôle de type décentralisé » et le « modèle de développement professionnel » est difficile à tracer dans certains cas et au regard de l'évolution récente de leur système de gouvernance, suite à leurs résultats dans les enquêtes internationales de type PISA (cas des USA et du Royaume Uni). De même qu'avec l'expansion des systèmes éducatifs et l'accroissement des attentes et des besoins, généralement dans une conjoncture de limitation des ressources, on enregistre une tendance générale vers le transfert de larges compétences aux entités locales⁶, avec l'impératif d'en évaluer les prestations et les résultats/performances.

Aujourd'hui, le paradigme de la gouvernance est centré sur trois principes fondamentaux que sont celui de la responsabilité et son corolaire, l'obligation de rendre des comptes (ce qui suppose l'existence d'une culture de la responsabilisation/accountability), celui de la gestion basée sur les résultats (culture de la performance), et celui de la nécessité du suivi permanent et de la mesure de ces résultats (culture de l'évaluation). L'opérationnalisation de ces principes dans le domaine éducatif n'est pas sans poser quelques difficultés et contraintes, compte tenu des enjeux qui sont à l'œuvre dans ce secteur.

De l'évaluation en éducation

Tous les travaux consacrés à cette problématique attestent du recours massif à l'évaluation quel que soit le mode de gouvernance adopté ou le type de politique éducative publique menée. Autre constat relevé par ces travaux est que, depuis l'élaboration et l'utilisation des premiers tests « papier-crayon », élaborés par des psychopédagogues au début du siècle dernier, la pratique évaluative a connu d'importants changements aux niveaux de :

- Sa localisation avec le déplacement des initiatives d'évaluation du local au central (ou l'inverse selon les pays) ;
- Son objet, passant de l'évaluation des élèves à celle des enseignants et des établissements ;
- Ses méthodologies et ses outils de collectes et d'analyse des données ;
- Ses séquences et sa fréquence, d'une activité ponctuelle elle est devenue un processus permanent ;
- Ses destinataires avec le passage d'une évaluation à l'intention des autorités publiques, institutionnelles et professionnelles, à une évaluation destinée aux pairs, aux usagers, aux partenaires, aux médias, etc.

Par ailleurs, le recours massif à l'évaluation et à la standardisation des normes et leur utilisation comme outils de pilotage et d'aide à la prise de décision a connu de grandes évolutions et une diversification des types de dispositifs et de mécanismes, avec des influences inégales sur les conceptions et les pratiques d'enseignement et de formation, qu'elles soient d'ordre pédagogique ou managérial.

⁶ - Cinq raisons militent en faveur du transfert des compétences aux entités locales : produire des services en adéquation avec les besoins locaux ; renforcer la capacités d'innovation des établissements et donc améliorer les résultats des élèves ; renforcer la démocratie et l'approche participative en vue d'une meilleure implication et motivation des acteurs (enseignants, communauté locale ...) ; permettre une meilleure allocation des ressources éducatives (réduction des coûts de fonctionnement et des délais de prise de décision, notamment). Voir N. Mons, Les politiques d'accountability. Du changement institutionnel aux transformations locales, in Education comparée, n°5, 2011

Typologies des dispositifs d'évaluation et d'accountability

L'établissement d'une typologie est toujours un exercice délicat, voire risqué. D'autant plus risqué qu'on assiste actuellement à un « fort processus d'hybridation » (Xavier Pons) des modèles, processus dans lequel peuvent cohabiter des dispositifs de régulation de type bureaucratique (injonctions réglementaires, procédurales et normatives) que des dispositifs d'accountability de type « school-based management » (autonomie des entités locales).

La proposition d'une synthèse des typologies ci-dessous est établie à partir de la littérature sur la question : sur la base de cinq critères de distinction principaux, on peut identifier pas moins de 12 types d'accountability fondés sur des dispositifs d'évaluation spécifiques.

Critères de distinction	Catégories	Type d'accountability
Niveau de la reddition des comptes (modèle de Carnoy et Loeb)	Niveau central	Formes anciennes
	Niveau local	Formes nouvelles
Conséquences sur les acteurs (modèle de Harris et Herrington)	Mécanismes de sanctions ou d'incitants	Accountability à forts enjeux
		Accountability à faibles enjeux
Principes normatifs (ex : légitimité) et les formes d'organisation (modèle de Maurice Kogan)	Entité ayant légitimité de demander des comptes	Contrôle public et hiérarchique
		Modèle professionnel
		Modèle consumériste
Buts de l'organisation éducative (Modèle de Leithwood et Earl)	Résultats des élèves	Modèle professionnel
	Pratiques de gestion	Modèle managérial
	Place sur le marché éducatif	Modèle du marché
Outils d'action publique (Modèle de Christian Maroy)	Outils légaux, réglementaires et budgétaires	Old accountability (sur les inputs et les processus)
	Standards et indicateurs de performance, tests ...	New accountability (sur les outputs et les résultats)

En se basant sur le niveau où se situe la reddition des comptes, Carnoy & Loeb⁷ distinguent les formes anciennes et les formes nouvelles d'accountability. Les premières se situent essentiellement au niveau central avec une focalisation sur les moyens (inputs) et les processus (les méthodes et dispositifs d'apprentissage), alors que les secondes sont davantage centrées sur les résultats (outputs) aussi bien des élèves/étudiants que des structures éducatives (classes, établissements). L'évolution récente des pratiques d'évaluation attestent de la prédominance des formes nouvelles d'accountability, avec le développement des politiques de gestion publique basée sur les résultats.

Le modèle de Harris & Herrington⁸ se fonde, lui, sur le critère des conséquences de la reddition des comptes sur les acteurs éducatifs. En fonction de la mise en place ou non d'une politique d'incitations et de sanctions afin d'amener les enseignants et les établissements à améliorer leurs prestations éducatives et, donc, les résultats des élèves, ils distinguent l'accountability à forts enjeux (high stake) et celle à faibles enjeux (low stake). L'adoption de l'une ou l'autre de ces politiques a un impact différent sur l'évolution des pratiques et des carrières des acteurs (enseignants et gestionnaires), comme l'attestent les cas de pays européens et nord-américains.

Le troisième modèle (de Maurice Kogan) se base sur le principe normatif de la légitimité (qui doit rendre compte à qui ?) et sur les formes d'organisation du dispositif de reddition des comptes. Il distingue ainsi trois types d'accountability : celui du contrôle public hiérarchique dans lequel l'autorité publique et/ou politique (gouvernementale ou élue) demande des comptes au travers de la chaîne administrative hiérarchique ; le type professionnel dans lequel ce sont les pairs (directeurs, superviseurs pédagogiques ...) qui exercent

7 - Carnoy, M ; Loeb, S. Does external accountability affect student outcomes ? A cross-state analysis, in Educational Evaluation & Policy Analysis, vol.24, n°4, pp.305-331, 2002, cité par C. Maroy et A. Voisin, Les transformations récentes des politiques d'accountability en éducation : enjeux et incidences des outils d'action publique.

8 - C. Maroy et A. Voisin, op.cit.

le contrôle et demandent des comptes et, enfin, le type consumériste dans lequel ce sont les usagers, les partenaires qui ont légitimité de demander des comptes.

Le modèle de Leithwood et Earl⁹ adopte le critère des objectifs de l'organisation éducative et distingue, de ce fait, trois types d'accountability : le modèle professionnel, le modèle managérial et le modèle du marché. Dans le modèle professionnel, l'objectif premier recherché est l'amélioration des pratiques professionnelles et des résultats des élèves. Ce sont donc les enseignants qui doivent rendre des comptes à leurs pairs (enseignants, tuteurs, superviseurs pédagogiques ...). Le modèle managérial, lui, fait référence aux pratiques de gestion, de pilotage et de management des institutions et structures éducatives, notamment les établissements scolaires. Dans ce cas, les responsables administratifs, les directeurs d'établissements en particulier, sont les premiers concernés par la reddition des comptes à leurs supérieurs hiérarchiques. Le modèle du marché, enfin, se base sur les mécanismes de la compétition sur le « marché éducatif » comme régulateur du « bien éducatif » et de la qualité des prestations des établissements scolaires. Ceux-ci ont donc obligation de rendre compte de leurs performances aux usagers, les parents essentiellement qui ont le pouvoir de choisir les établissements qui répondent le mieux à leurs exigences et attentes.

Le modèle de C. Maroy part du constat que « les évolutions des politiques seraient mieux saisies au travers des « outils » et leur opérationnalisation, qu'au travers des principes normatifs ou des philosophies politiques qui les justifient ». Que l'accountability soit de type professionnel, managérial ou basée sur les mécanismes de marché, son évolution est perçue au travers de l'évolution des « outils » de l'action publique mis en œuvre. Au côté des outils traditionnels adoptés durant les années 1970-80 qui relèvent du domaine légal, réglementaire et du contrôle budgétaire et qui concernent l'évaluation des inputs et des processus (Old accountability), de nouveaux outils font leur apparition dans les années 1990-2000 avec la prédominance des référentiels, normes et standards et des indicateurs de performance du fait qu'on s'intéresse davantage aux outputs et aux résultats du système (New accountability)¹⁰.

En fonction des divers paramètres et critères de distinction envisagés par ces différents modèles et avec des combinaisons et articulations variées, plusieurs cas de figures peuvent être dégagés au regard des modes de gouvernance des systèmes éducatifs. Dans le cas de systèmes de type hiérarchique/bureaucratique (modes de gouvernance centralisé et déconcentré), le rôle clé revient aux autorités administratives aux différents échelons aussi bien pour demander des comptes que pour en rendre, principalement sur les aspects inputs et processus. Lorsque le système est véritablement décentralisé, avec une large autonomie pédagogique, administrative et financière aux entités infra-centrales (régionale et locale), la reddition des comptes concerne, en premier lieu, les professionnels et les acteurs directs (enseignants et directeurs) et peut porter aussi bien sur les moyens, les ressources et les processus que sur les outputs et les résultats. Dans ce cas de figure, on note généralement la prédominance de la pratique de l'auto-évaluation et de la reddition des comptes aux pairs et aux « supérieurs hiérarchiques » immédiats, notamment pour l'évaluation de la pertinence et de l'efficacité des moyens mis en œuvre au regard des résultats atteints et des performances réalisées.

II. Modes de gouvernance et dispositifs d'évaluation et de redevabilité : quels enseignements ?

Au regard de l'évolution et de l'importance donnée à la pratique de l'évaluation dans le secteur de l'éducation, il est légitime de s'interroger sur son rôle dans l'amélioration du fonctionnement des structures éducatives et celle des apprentissages et des acquis des élèves.

- Quels usages fait-on des dispositifs et des résultats des évaluations ?
- Utilise-t-on ces dispositifs et résultats comme outils de diagnostic ou outils de pilotage ?
- Ont-ils un impact sur les acteurs directs et indirects ?
- Ont-ils un impact sur l'amélioration des apprentissages et des performances ?
- Impulsent-ils de nouvelles orientations des politiques éducatives ?

Ce sont-là des questions de fond qui méritent des investigations approfondies et on ne peut ici qu'apporter quelques éléments de réponse.

9 - C. Maroy et A. Voisin, op.cit.

10 - Christian Maroy et Annelise Voisin, Les transformations récentes des politiques d'accountability en éducation : enjeux et incidences des outils d'action publique

Pendant longtemps, de nombreux spécialistes de l'éducation ont défendu l'idée de la décentralisation et de l'autonomie aux échelons régionaux et locaux comme solution aux problèmes qu'affronte le secteur et pour améliorer ses performances en termes de qualité et d'équité. Aujourd'hui, et même si le principe de la décentralisation n'est pas remis en cause en lui-même, cette assertion est fortement nuancée et des doutes sont émis à ce propos, surtout lorsque le processus de transfert des compétences du niveau central aux niveaux régional et local ne se fait pas correctement, c'est-à-dire avec un dispositif d'accompagnement des acteurs locaux et une répartition claire des attributions des différentes parties prenantes, y compris les collectivités territoriales.

Les principaux points d'interrogation concernent les effets et l'impact de la décentralisation sur le fonctionnement global du système éducatif, sur les performances des établissements de formation et sur les résultats des élèves et la qualité des apprentissages. De ce point de vue, les leçons des expériences internationales montrent que même lorsque la décentralisation se traduit par une amélioration globale du fonctionnement du système et par l'octroi de l'autonomie aux établissements, elle ne débouche pas systématiquement sur l'amélioration des résultats des élèves, surtout si l'attention est plus portée aux structures et aux dispositifs de gouvernance du système qu'aux aspects pédagogiques. N. Mons affirme que « les politiques de décentralisation qui ont eu des répercussions positives sur les performances des élèves (sont celles qui ont été) davantage orientées sur l'autonomie scolaire que sur le transfert de compétences aux collectivités locales ». De même que, à la lumière des exemples de certains Etats américains comme Chicago où l'octroi de l'autonomie aux établissements à la fin des années 80 n'a pas conduit aux innovations pédagogiques souhaitées et « n'a pas amélioré les performances des élèves de façon significative. Ce n'est qu'avec la réforme de 1995, avec le renforcement du contrôle du district sur les résultats des élèves (établissements de standards pédagogiques) que l'attention a été davantage portée sur les activités pédagogiques que sur les structures du système »¹¹.

Pour ce qui est de l'utilisation des données et des résultats de l'évaluation et leur effet sur les politiques éducatives en général et sur le quotidien des établissements et les pratiques professionnelles (qu'elles soient pédagogiques ou de gestion), le constat global que l'on peut faire est que la culture de l'évaluation et de l'accountability n'est pas encore enracinée dans le monde de l'éducation. On constate généralement, d'une part, « une relative intégration du discours sur l'évaluation par les chefs d'établissement » et, d'autre part, une réticence du personnel pédagogique à l'égard du paradigme de la standardisation de la mesure des performances des enseignants et des établissements. Rares sont les expériences où les données collectées de l'évaluation se traduisent par des politiques de remédiation et de perfectionnement des prestations éducatives.

De ce fait, la tendance dominante est à l'utilisation des données et des résultats de l'évaluation plus en tant qu'outil de diagnostic que comme outil de pilotage et d'amélioration de l'existant, en l'absence d'un dispositif de reddition des comptes. Bien plus, certains auteurs considèrent que ces données sont souvent utilisées pour justifier des orientations politiques, avec un usage sélectif des données allant dans ce sens. « En quelque sorte, passant très vite à côté d'un dangereux accroissement de la reddition des comptes par rapport au passé (accountability), certains décideurs orienteraient insidieusement l'usage (des) évaluations internationales vers une forme de justification de ce qui devrait advenir et de la nécessaire mobilisation de bonnes pratiques »¹²

La réticence des enseignants à exploiter les données des évaluations des apprentissages et des acquis des élèves dans l'amélioration de leurs pratiques pédagogiques n'est pas non plus sans poser des difficultés. A propos de tests internationaux comme PISA, ils sont nombreux à penser que « la standardisation des évaluations des élèves par des tests et la mise en rapport de la qualité des enseignants et de leur formation avec ces résultats, dévoient la nature même de la profession qui consiste essentiellement à entraîner les élèves à ces tests (teaching to the test), de façon d'autant plus fébrile que la situation professionnelle des enseignants, placés sous régime probatoire permanent, en dépend »¹³. On reproche aux pratiques d'évaluation des performances, à leur méthodologie et à l'exploitation de leurs résultats d'éluder les apports de la recherche scientifique sur la complexité des phénomènes d'apprentissage des élèves et des déterminants de la réussite scolaire et de réduire ainsi la fonction enseignante à la simple capacité à préparer les élèves aux tests.

De ce fait, certains auteurs n'hésitent pas à dénoncer les dégâts que peuvent causer certaines données des évaluations nationales et/ou internationales et surtout leur interprétation, à l'exemple des Etats-Unis où les résultats du PISA 2000, avec la mise en œuvre du programme No Child Behind (2005), ont eu un impact

11 - Nathalie Mons sur la décentralisation et l'autonomie et leurs effets sur l'apprentissage

12 - Marc Demeuse, Les évaluations internationales et l'éducation comparée, Education comparée, n°5, 2011, p.7

13 - Régis Malet, Politiques d'accountability et devenir de la profession enseignante aux Etats-Unis, in Evaluer, Journal International de Recherche en Education et Formation, 1 (1), 2015, p.39

négalif sur la politique de formation/certification et de recrutement des enseignants avec une nette baisse de leur qualification et un accroissement des départs précoces¹⁴.

Autres effets pervers de la course aux performances, les proportions prises par les « cours de soutien ». En France, le marché du soutien scolaire a été estimé à 470 millions d'euros en 2008. En Grèce, 80 % des élèves bénéficient de cours de soutien. En plus de son poids financier, l'apport pédagogique du soutien scolaire dans les systèmes éducatifs nécessite sa prise en considération dans l'élaboration des indicateurs de performance et dans l'interprétation des résultats des évaluations¹⁵.

Concernant les dispositifs, les méthodologies et les outils utilisés dans l'évaluation et particulièrement les outils statistiques et les indicateurs mobilisés, de nombreuses limites et réserves sont émises à l'égard des approches quantitatives exclusives. S'il est vrai qu'il serait naïf de croire qu'on peut « gouverner l'éducation par les nombres »¹⁶ et légitime de dénoncer la « dictature des standards » (l'exemple de la bibliométrie évaluative dans le domaine de la recherche scientifique) et pointer du doigt les effets pervers et dégâts de la course aux performances, il est tout aussi important de reconnaître que « l'approche quantifiée peut préserver des prises de position plus idéologiques que scientifiques, qui sont encore légion, dès qu'on traite de l'éducation »¹⁷. Quant aux indicateurs, la fiabilité de certains d'entre eux est sérieusement remise en question, comme c'est le cas de ceux relatifs aux inégalités scolaires qu'il s'agisse des inégalités sociales ou celles des performances des élèves et qui sont adoptés par les enquêtes internationales comme PISA. De nombreuses études ont analysé la qualité de mesure qu'offrent ces indicateurs et mettent en évidence divers problèmes méthodologiques dans leur construction, parmi lesquels le fait de se baser sur les réponses des élèves qui peuvent se tromper sur les données relatives au revenu ou au niveau éducatif de leurs parents. Ce qui pose un sérieux problème d'exactitude et de fiabilité des informations recueillies et servant de base à l'élaboration des indicateurs et leurs comparaisons.

Il est, de ce fait, tout aussi nécessaire de recourir aux dispositifs statistiques et aux indicateurs clés pour l'évaluation que de tenir compte de la complexité du système éducatif et de sa spécificité, notamment par rapport aux innombrables aspects qualitatifs, difficilement mesurables par absence d'indicateurs pertinents et capables d'en rendre compte. Ce n'est, par conséquent, pas une simple clause de style que d'affirmer la nécessaire complémentarité entre les approches et les méthodes d'évaluation des performances, entre le quantitatif/mesurable et le qualitatif, entre les normes nationales et internationales et leur nécessaire contextualisation.

III. Gouvernance et évaluation de l'éducation au Maroc

Les enseignements des expériences internationales sont précieux et peuvent aider dans la mise en place au Maroc d'un dispositif d'évaluation national fiable et pertinent. Bien qu'il soit difficile de parler de l'évaluation en tant que méthodologie structurée et pratique systématisée avec des outils et des instruments fiables, on peut aussi tirer quelques leçons des exemples d'évaluation du processus de réforme entamé il y a maintenant plus de 15 ans.

Caractéristiques dominantes du SEF marocain : quelques rappels

Le système éducatif marocain fait souvent l'objet du constat suivant : malgré les efforts de réforme et les moyens financiers et humains consacrés au secteur, les résultats demeurent en-deçà des attentes et des objectifs fixés par les projets de réforme. En effet, hormis la réalisation de l'objectif de généralisation du cycle primaire, le système continue de souffrir des mêmes déficits, comme l'attestent les derniers rapports et études sur le secteur, notamment le rapport du CSEFRS de 2014. Globalement, les principaux déficits enregistrés concernent :

- Son rendement interne avec la persistance de taux de redoublement et d'abandon élevés (2.4 millions élèves abandonnent sur 6.4 M scolarisés, taux de non achèvement de 37% en 2012), la faiblesse des acquis fondamentaux qui est confirmée aussi bien par les tests nationaux et internationaux (PIRLS, TIMSS, PNEA, PISA) et la persistance des inégalités scolaires spatiales (rural, urbain et périurbain), de genre et entre catégories sociales ;
- Son rendement externe qui demeure faible en raison des difficultés d'insertion professionnelle des «

14 - R. Malet, op.cit.

15 - Gouvernance et autonomie, colloque de Poitiers, ESEN, 2008

16 - G. Felouzis & S. Hanhart, Gouverner l'éducation par les nombres ? Usages, débats et controverses. De Boeck, coll. « Raisons éducatives », 2011

17 - M. Duru-Bellat, Progrès statistiques et glissements conceptuels dans l'analyse des inégalités sociales à l'école, Revue française de pédagogie, n°161, p.12

produits » du SEF et l'inadaptation des profils formés par rapport aux besoins du marché de l'emploi, lui-même peu créateur d'emplois qualifiés ;

- La gestion et la gouvernance du secteur qui demeure « suradministré » et « sous-piloté » dans la mesure où le processus de déconcentration/décentralisation entamé depuis le début des années 2000 avec la création des Académies Régionales n'a pas encore débouché véritablement sur l'autonomie (pédagogique, administrative et financière) et le renforcement des pouvoirs des acteurs éducatifs locaux, ni sur l'amélioration de la qualité de gestion du secteur.

Les facteurs explicatifs de ce bilan sont d'ordre à la fois interne au système et externe et qui sont relatifs à son environnement. Ce dernier présente, en effet, plusieurs aspects aggravant la situation du secteur éducatif : la pression démographique qui a fait que les conditions et le rythme de la généralisation n'ont pas permis d'améliorer la qualité des prestations éducatives, surtout en milieu périurbain et rural. Rappelons, de façon synthétique, ces principaux facteurs :

- Absence d'une politique intersectorielle intégrée, spécifiquement pour le périurbain et le monde rural ;
- Conditions socio-économiques défavorables et prédominance de l'analphabétisme des ménages les plus défavorisés, ce qui se traduit par une faible implication des parents dans la scolarisation de leurs enfants ¹⁸ ;
- Fragilité du tissu économique (prédominance des PME et des entreprises familiales peu créatrices d'emploi et création d'emplois peu qualifiés ...)
- Environnement général favorisant les comportements déviants (violence, drogue, fraudes, corruption ...) ¹⁹

Pour ce qui est des raisons propres au système, les mêmes faiblesses persistent depuis la mise en œuvre de la réforme dont la faible qualité des programmes et méthodes pédagogiques ; une qualification problématique des enseignants et du personnel de gestion, malgré les efforts de formation continue fournis dans le cadre du Plan d'Urgence (2009-2012)²⁰ ; la faible qualité des infrastructures et des supports didactiques, surtout en milieu rural et péri urbain et les défaillances des dispositifs d'orientation et de supervision pédagogique.

Toutes ces difficultés et faiblesses sont aggravées par les modes de gouvernance du secteur éducatif marocain, modes devenus à la fois obsolètes par rapport à l'expansion des besoins et des attentes des acteurs et inadaptés pour vaincre les « résistances au changement » au sein même du système, ce qui pose la question de la pédagogie de la réforme et de la conduite du changement, et de la place qu'occupent l'évaluation et la redevabilité dans le système.

La place de l'évaluation dans la gouvernance du SEF marocain

Parmi tous les aspects expliquant les faibles performances du secteur éducatif marocain, la question de la gouvernance a pu bénéficier ces dernières années d'une attention particulière. Le dernier rapport du CSEFRS reconnaît que « La gouvernance demeure parmi les problématiques transversales qu'affronte l'école depuis des décennies, aussi bien au niveau de la gestion qu'au niveau de la participation, de la transparence et de la corrélation entre la responsabilité, l'évaluation et la reddition des comptes »²¹.

La réforme des modalités de gouvernance du secteur a débuté depuis la mise en place des AREF (loi 07-00, 2002), avec le transfert de certaines compétences du niveau central au niveau régional : gestion administrative, pédagogique, planification, gestion des infrastructures, des RH, budgétaire. Depuis la Charte et le rapport 2008 du CSE, le choix de la « décentralisation » a été affirmé comme « ...un choix irréversible et un chantier prioritaire de la réforme du système d'éducation et de formation » et « une nécessité impérieuse pour s'adapter aux besoins locaux et régionaux », l'objectif étant de « simplifier, rationaliser et accélérer les

18 - Parmi les facteurs expliquant les difficultés d'apprentissage des élèves, l'environnement familial joue un rôle important. A titre d'exemple, et selon l'enquête EGRA, les 3/4 des 1436 ménages enquêtés ne disposent d'aucun livre à la maison, Early Grade Reading Assessment, USAID, 2012

19 - Les données des différentes enquêtes TIMSS et PIRLS, qui se sont déroulées durant la décennie de la Charte, relatives au climat de l'école au Maroc lié aux enjeux de discipline, font état d'une dégradation entre 2001 et 2011. Les résultats de ces enquêtes révèlent que 60% des élèves marocains en quatrième année du primaire, et 49% en deuxième année collégiale sont scolarisés dans des écoles où les problèmes de discipline et de sécurité sont fortement présents (absentéisme, retard, tricherie, vandalisme, menace, etc.), voir le Rapport analytique du CSEFRS, 2014.

20 - « Entre 2009 et 2010, les masses budgétaires allouées aux AREF ont connu une augmentation de l'ordre de 300 à 400% atteignant ainsi des volumes jusqu'alors jamais constatés pour la formation continue », in Evaluation de mesures et de stratégies mises en œuvre dans le cadre du programme d'urgence pour l'éducation au Maroc, Rapport final, MENFP, Décembre 2013, p.120.

21 - Rapport du CSEFRS, 2014, p.

procédures de gestion, apporter des réponses efficaces aux problèmes de proximité et libérer les initiatives constructives à tous les échelons ». Parallèlement, des mesures ont été initiées pour propulser une nouvelle dynamique au niveau local, celui des établissements, avec la mise en place des Conseils de Gestion des Etablissements Scolaires (COGES) et le lancement des projets d'établissement.

Or, les constats relevés dans le rapport 2014 du CSEFRS sont pratiquement les mêmes que les dysfonctionnements déjà mentionnés dans le rapport 2008 du CSE. Citons en particulier :

- L'inadéquation des structures du Département de l'enseignement avec ses nouvelles attributions et la persistance du mode ancien de fonctionnement de l'administration centrale en interne et dans ses relations avec les échelons subalternes (Exemple : la gestion du programme de FC du personnel de l'EN) ;
- La confusion dans les attributions des AREF et des délégations provinciales et l'absence de structure de coordination entre les délégués et les directeurs des AREF ;
- La centralisation des ressources et des compétences par les AREF au détriment des niveaux provinciaux et locaux ;
- Une autonomie et un financement limités des établissements scolaires, ce qui se traduit par des COGES non fonctionnels avec une faible implication des acteurs (enseignants, parents d'élèves) et des partenaires et par une qualité déficiente et un impact limité des projets d'établissement sur les apprentissages des élèves et sur le fonctionnement administratif et pédagogique des établissements ;
- Un déficit persistant de la pratique du suivi et de l'évaluation à tous les échelons

Ce dernier aspect revêt une importance particulière pour atteindre l'objectif d'instauration d'un système de gouvernance basé sur les résultats et les performances. Certes, la Charte a prévu dans ses dispositions que « le Système d'Education et de Formation sera soumis, dans sa globalité, à une évaluation régulière, portant sur sa rentabilité interne et externe, et touchant tous les aspects pédagogiques et administratifs ». Elle a recommandé, en outre, « la mise en place et la généralisation d'opérations d'audit, le développement de l'auto-évaluation des établissements et les études de satisfaction des acteurs éducatifs et de leurs partenaires hors système » (article 157 de la Charte)²².

Force est de constater que, quinze ans après, l'évaluation n'a pas connu d'avancées significatives en tant que pratique maîtrisée avec des outils et des dispositifs organisés, des indicateurs et des standards normalisés et prévoyant aussi des mécanismes clairs de responsabilisation et de reddition des comptes. Mise à part l'évaluation des acquis des élèves effectuée dans le cadre du Programme National d'Evaluation des Acquis (PNEA, 2008) par l'Instance Nationale d'Evaluation sur la base des données du Centre National des Examens, le fonctionnement du système éducatif n'a pas encore bénéficié d'une évaluation globale et par composantes depuis la mise en œuvre de la réforme. A noter également l'absence de tout dispositif d'auto-évaluation et la seule structure centrale à avoir procédé à un audit de son fonctionnement est la Direction du Système d'Information²³.

Le contexte politique et socio-économique actuel incite pourtant à aller dans le sens de la concrétisation des choix fondamentaux²⁴ et l'amélioration des conditions de leur mise en œuvre, tels le choix de la « décentralisation » avec le processus entamé de la régionalisation avancée ainsi que le choix de « l'autonomie » aux établissements et aux structures éducatives de proximité pour en améliorer la qualité et les performances (GAR, démarche contractuelle, culture de l'évaluation), pas seulement au niveau du dispositif de gestion administrative et financière mais aussi et surtout au niveau des activités pédagogiques et de l'amélioration de la qualité des apprentissages et des performances des élèves.

Pour l'heure, et globalement, on constate que notre pays éprouve des difficultés à assurer la transition du paradigme bureaucratique à un nouveau paradigme plus professionnel. Les chantiers de la décentralisation et la réforme du mode de gouvernance, avec l'institutionnalisation de la pratique d'évaluation sont encore à leurs débuts. Le dispositif d'évaluation est en cours de constitution²⁵ et la réflexion sur les indicateurs de performance et sur les mécanismes de responsabilisation et de redevabilité à tous les échelons est encore balbutiante. Les avancées à ces divers niveaux sont à même de permettre non seulement d'instaurer un système d'évaluation pertinent et fiable mais aussi d'en faire un outil puissant d'aide à la décision et de régulation du système éducatif.

22 - La Constitution marocaine de 2011 met l'accent sur les principes de la bonne gouvernance, de la reddition des comptes et de l'évaluation, principes qui sont sensés être intégrés dans le fonctionnement du système éducatif.

23 - Rapport du CSEFRS, 2014, p.49

24 - Choix affirmés par la Constitution de 2011 et par la Vision Stratégique 2015-2030 du CSEFRS.

25 - Citons la création de l'INE, de l'Agence Nationale d'Evaluation de l'Enseignement Supérieur et de la Recherche Scientifique et le renforcement des systèmes d'information et de collecte des données éducatives (exemple : le système MASSAR)

Conclusion

L'évolution de la pratique de l'évaluation dans le monde et le développement de ses outils et de sa méthodologie permettent aujourd'hui de disposer d'un arsenal de données et d'informations qu'il s'agit d'exploiter en vue d'un meilleur fonctionnement du SEF. Ces outils et méthodologies dépendent fondamentalement des modes de gouvernance des secteurs éducatifs. La diversité des expériences internationales en matière d'évaluation est, de ce fait, le reflet de la diversité de leurs systèmes de gouvernance qui, eux-mêmes, sont conditionnés par l'évolution de leurs contextes politique, social et économique.

Si cette importance donnée aujourd'hui à l'institutionnalisation de l'évaluation et, son corollaire la redevabilité, comme dispositifs et pratiques d'une gestion saine et transparente du système, est largement justifiée, il est néanmoins indispensable de relativiser sa portée et son impact sur un secteur aussi multidimensionnel et complexe que celui de l'éducation. L'approche quantifiée permet, certes, « de préserver des prises de position plus idéologiques que scientifiques, qui sont encore légion, dès qu'on traite de l'éducation » (Duru Bellat), il est cependant important de tenir compte de la complexité du système éducatif et de sa spécificité, notamment par rapport aux innombrables aspects qualitatifs, difficilement mesurables par absence d'indicateurs pertinents et capables d'en rendre compte, comme c'est le cas pour la question des inégalités sociales en éducation.

Même si le Maroc a participé à des enquêtes internationales d'évaluation des acquis et a procédé à son premier PNEA, son expérience en matière d'évaluation est encore à ses débuts. Elle est appelée à se développer avec le lancement du processus de la régionalisation avancée et le nécessaire développement des prérogatives des échelons régionaux et locaux, comme cela a été réaffirmé dans la Vision 2015-2030 du CSEFRS.

Traduction du Résumé Abstract Translation

Amina Debbagh

GOVERNANCE MODES, EVALUATION PRACTICES AND THEIR IMPACT ON EDUCATION PUBLIC POLICIES : A COMPARATIVE APPROACH

This paper aims at contextualizing evaluation practices in education in views of governance approaches in this sector. While reviewing some practices, taking into consideration their diversity (learnings, teaching methods, schools, education programs...), their methodology and their objectives, we will try to show that their efficiency and their appropriateness are largely determined by governance approaches that are adopted. This paper also aims to question the evaluation practice usefulness, its objectivity and reliability extent as well as its impact on public education policies, with reference to some international experiences, and to the Moroccan one.

Bibliographie et Références

Bouvier Alain (2007), La gouvernance des systèmes éducatifs, De Boeck.

Colloque de Poitiers (2008), Gouvernance et autonomie, ESEN.

Conseil Supérieur de l'Education, de la Formation et de la Recherche Scientifique/INE (2014), La mise en œuvre de la charte nationale d'éducation et de formation, Acquis, déficits et défis, Rapport analytique

Conseil Supérieur de l'Education, de la Formation et de la Recherche Scientifique (2014), Pour une école de l'équité, de la qualité et de la promotion, Vision stratégique de la réforme 2015-2030

Demeuse Marc (2011), Les évaluations internationales et l'éducation comparée, Education Comparée, n°5.

Duru-Bellat Marie, Progrès statistiques et glissements conceptuels dans l'analyse des inégalités sociales à l'école, Revue française de pédagogie, n°161

Felouzis, G & S. Hanhart (2011), Gouverner l'éducation par les nombres ? Usages, débats et controverses. De Boeck, coll. « Raisons éducatives ».

Lacroix Isabelle et Pier-Olivier St-Arnaud (2012), La gouvernance : tenter une définition, Cahier de recherches en politique appliquée, n°3.

Malet Régis (2011), Les politiques d'imputabilité en direction des enseignants en Angleterre et aux États-Unis, in Education comparée n°5.

Malet Régis (2015), Politiques d'accountability et devenir de la profession enseignante aux Etats-Unis, in Evaluer, Journal International de Recherche en Education et Formation, 1 (1).

Maroy, C. et A. Voisin, Les transformations récentes des politiques d'accountability en éducation : enjeux et incidences des outils d'action publique.

Mons Nathalie (2011), Les politiques d'accountability. Du changement institutionnel aux transformations locales, in Education Comparée, n°5.

Toulemonde Bernard (2012), De nouveaux modes de gouvernance, Cahiers français n°368.

LA COUR DES COMPTES : MÉTHODES D'ÉVALUATION DES INSTITUTIONS ÉDUCATIVES

Mohamed HADHOUDI *

Résumé

Depuis de nombreuses années, la Cour des comptes a porté un intérêt particulier aux politiques de l'éducation, de la formation, de l'enseignement supérieur et de la recherche scientifique.

Ses travaux ont pour objectif d'apprécier tous les aspects de la gestion et de formuler des suggestions sur les moyens susceptibles d'améliorer et d'accroître l'efficacité et le rendement des services publics.

Ainsi, la Cour des comptes a inscrit dans ses programmes l'évaluation de la gestion de nombreux acteurs chargés de la mise en œuvre des politiques éducatives du Royaume. Il s'agit notamment des Académies régionales d'éducation et de formation, des universités, des écoles et instituts supérieurs et de l'Office de la formation professionnelle.

L'approche évaluative de la Cour des comptes :

- *examine si les programmes publics atteignent leurs objectifs avec un souci d'efficacité, d'efficience et d'économie et en conformité avec la législation et la réglementation en vigueur ;*
- *examine la mise en œuvre des politiques publiques et ne se prononce pas sur leur opportunité ;*
- *est contradictoire. Elle tient compte des avis et commentaires de toutes les parties prenantes ;*
- *est collégiale. Tous les rapports sont délibérés par des formations composées au moins de cinq membres ;*
- *formule des recommandations et assure le suivi de leur mise en œuvre.*

La Cour des comptes communique les résultats de ses travaux par plusieurs moyens, comme le rapport annuel, les rapports thématiques, les rapports particuliers, les référés du premier président, les lettres des présidents de chambres...etc.

Introduction

La Cour des comptes est l'institution supérieure de contrôle des finances publiques du Royaume. Elle a pour mission la protection des principes de valeurs de bonne gouvernance, de transparence et de reddition des comptes de l'Etat et des organismes publics.

Les compétences de la Cour des comptes sont prévues par la constitution et par la loi. Ainsi, La Cour des comptes est chargée d'assurer le contrôle supérieur de l'exécution des lois de finances. A ce titre, elle exerce un double contrôle : un contrôle juridictionnel et un contrôle de la gestion axé sur l'appréciation des résultats et des performances.

* Président de la troisième chambre de la Cour des comptes

Elle s'assure de la régularité des opérations de recettes et de dépenses des organismes soumis à son contrôle en vertu de la loi et en apprécie la gestion. Elle sanctionne, le cas échéant, les manquements aux règles qui régissent lesdites opérations. La Cour des Comptes contrôle et assure également le suivi des déclarations du patrimoine, audite les comptes des partis politiques et vérifie la régularité des dépenses des opérations électorales.

Elle assiste aussi le Parlement et le Gouvernement dans les domaines relevant de sa compétence en vertu de la Loi et rend compte à Sa Majesté le Roi de l'ensemble de ses activités dans le cadre de son Rapport Annuel.

Depuis 2006, la Cour des comptes a accordé une attention particulière aux institutions éducatives chargées de l'élaboration et de la mise en œuvre de la politique publique de l'éducation. Plusieurs missions d'audit et d'évaluation ont été menées par la Cour des comptes. Elles ont concerné certaines Académies régionales d'éducation et de formation et plusieurs universités.

Les résultats des travaux de l'audit et de l'évaluation sont publiés dans les rapports annuels de la Cour des comptes (www.courdescomptes.ma).

La présente présentation porte principalement sur les méthodes d'évaluation des institutions éducatives. Elle donne une synthèse des principales constatations et recommandations formulées par la Cour des comptes.

I. Aperçu historique

L'institution du contrôle supérieur des finances publiques au Maroc, s'est faite d'une manière progressive marquée par cinq étapes essentielles :

1960 : Institution de la Commission Nationale des comptes

La Commission avait pour compétence d'apurer les comptes des comptables publics, c'est-à-dire vérifier la régularité et la légalité des opérations reprises dans ces comptes.

Elle était composée d'un Président nommé par Sa Majesté le ROI, sur proposition du Ministre de la Justice, et d'Inspecteurs des Finances désignés par le Ministre des Finances, La Commission devait également produire à la Chambre des Représentants, en annexe de la Loi de règlement, un rapport sur l'exécution de la Loi de Finances.

Le contrôle de cette Commission sur les finances publiques était très limité du fait du manque de moyens humains et matériels et aussi du fait que ce contrôle revêtait le caractère d'un contrôle comptable supérieur de nature administrative plutôt qu'un contrôle de nature juridictionnelle.

1979 : Création de la Cour des Comptes, en vertu de la loi n° 12-79

L'année 1979 a connu l'institution de la Cour des Comptes en tant qu'organe juridictionnel chargé d'assurer le contrôle supérieur de l'exécution des lois de finances et ce en vertu de la loi n°12-79. La Cour s'assure de la régularité des opérations de recettes et de dépenses des organismes soumis à son contrôle et, le cas échéant, réprime les manquements aux règles qui régissent lesdites opérations. Elle contrôle et apprécie la gestion des organismes soumis à son contrôle. Elle rend compte à Sa Majesté le ROI de l'ensemble de ses activités.

1996 : Elévation de la Cour des Comptes au rang d'Institution constitutionnelle

En vue de faire jouer pleinement son rôle à la Cour en tant qu'Institution Supérieure de Contrôle, la Constitution du 13 septembre 1996 l'a érigée en Institution Constitutionnelle : Le titre X de la Constitution précise que la Cour des comptes est chargée d'assurer le contrôle supérieur de l'exécution des Lois de finances.

Par ailleurs, dans le cadre de la politique de décentralisation et de déconcentration et en vue d'améliorer la gestion des Collectivités Locales, la Constitution a également prévu la création de Cours Régionales des Comptes.

2002 : Promulgation de la Loi n°62-99 formant code des Juridictions financières

En application des dispositions constitutionnelles, la loi n° 62-99 formant Code des Juridictions financières a été promulguée le 13 juin 2002. Cette loi, composée de trois livres, a marqué une étape importante dans l'itinéraire de la Cour des comptes du fait qu'elle a explicitement fixé les attributions, l'organisation et le fonctionnement de la Cour (livre I), des Cours Régionales des Comptes (livre II) ainsi que le statut particulier des magistrats de ces juridictions (livre III).

2011 : Renforcement des attributions constitutionnelles de la Cour des Comptes

La nouvelle constitution a introduit plusieurs innovations de nature à consolider les acquis des juridictions financières, après presque une décennie d'exercice élargi du contrôle, et à renforcer le rôle de celles-ci dans les domaines importants de la gouvernance publique.

Aujourd'hui, la Cour a vocation à contrôler et conseiller le Gouvernement, à assister le Parlement et à apporter son appui à l'autorité judiciaire. La Cour des comptes sera également amenée à publier, en plus du rapport annuel, les rapports particuliers se rapportant aux missions réalisées.

II. Principales caractéristiques des approches de l'évaluation

L'évaluation exercée par la Cour des comptes porte sur tous les ministères et organismes publics chargés de l'éducation, de l'enseignement supérieur et de la formation professionnelle. Elle porte également sur toutes les Académies régionales d'éducation et de formation, sur toutes les universités et les établissements de la formation professionnelle. Elle s'étend enfin aux instituts et écoles supérieures rattachés à d'autres départements ministériels Comme l'Agriculture, l'Energie ...etc.

Les méthodes d'évaluation adoptée par la Cour des comptes sont celles prévues pour l'exercice de ses attributions en matière de contrôle de la gestion et de l'emploi des fonds publics telles que prescrites par la loi n°62-99 relative à la Cour des comptes.

Ainsi, les travaux de la Cour visent l'appréciation de la qualité de la gestion et la formulation des suggestions et recommandations susceptibles d'en améliorer les méthodes et d'en accroître le rendement et l'efficacité. Ils portent sur tous les aspects de la gestion, sur l'efficacité des contrôles internes et sur la régularité des opérations financières.

L'évaluation, quant à elle, **a pour objectif d'établir sur la base des réalisations, dans quelle mesure les objectifs assignés ont été atteints, au regard des moyens mis en œuvre.**

Dans l'exercice de ses missions, la Cour des comptes respecte et applique les principes fondamentaux garantissant des résultats indépendants et objectifs. Ces principes peuvent être résumés comme suit :

Indépendante : Elle est garantie par la Constitution comme prévue par son article 147 et se concrétise par son autonomie par rapport aux autres pouvoirs : Législatif, Exécutif et Judiciaire. Ses membres ont la qualité de magistrat et sont inamovibles.

Externe : L'évaluation exercée par la Cour des comptes est qualifiée d'externe. En effet, elle est réalisée par l'Institution supérieure de contrôle, constitutionnelle et indépendante.

Contradictoire : Tous les rapports de la Cour des comptes tiennent compte des commentaires et avis des responsables des organismes publics. Ainsi, les responsables gestionnaires sont sollicités et invités par les conseillers de la Cour à formuler leurs remarques et avis sur les conclusions et recommandations durant toutes les étapes des missions de l'évaluation.

Collégiale : Tous les rapports sont délibérés par des formations composées de plusieurs membres (au moins cinq membres). Les conseillers rapporteurs préparent les projets de rapports qui ne deviennent définitifs qu'après leur adoption par des formations collégiales.

Publique : Depuis la promulgation de la loi 62-99 en 2002, la Cour publie les résultats de ses travaux dans son rapport annuel. Actuellement, la Constitution de 2011, prévoit la publication de l'ensemble des travaux de la Cour, y compris les rapports particuliers et les décisions juridictionnelles www.courdescomptes.ma.

Porte sur la mise en œuvre : L'évaluation exercée par la Cour ne porte pas sur l'opportunité politique des politiques publiques. Elle n'est pas partisane. Néanmoins, elle examine la pertinence de ces politiques publiques.

Utilise les travaux des autres institutions et organisations internationales : En effet, la Cour des comptes recourt aux travaux et études réalisés par les autres institutions comme le Conseil supérieur de l'éducation, de la formation et de la recherche scientifique. Les travaux de diagnostics et de vision en matière d'éducation réalisés par cette Institution constituent des référentiels principaux dans les évaluations de la Cour. En plus, la Cour utilise tous les travaux des inspections générales des ministères et des audits internes des autres organismes publics.

III. Le processus de l'évaluation

Le processus de l'évaluation comprend les phases de la programmation, de l'exécution, du rapport et du suivi des recommandations.

La programmation des missions de l'évaluation

La Cour des comptes a programmé plusieurs missions d'évaluation des institutions éducatives. Ces missions ont porté principalement sur organismes publics chargés de la mise en œuvre des politiques de l'éducation, de l'enseignement supérieur et de la formation professionnelle.

Le choix des organismes publics et des thématiques est effectué par le Comité des programmes et des rapports. Ce comité, présidé par le premier président de la Cour et des présidents des chambres, joue un rôle stratégique dans la programmation des activités de la Cour.

La programmation des évaluations des institutions éducatives prend en compte les risques inhérents aux capacités des différentes parties prenantes pour la mise en œuvre des réformes des politiques publiques de l'éducation. Ainsi, la Cour avait constaté que la majorité des organismes publics, autres que les ministères, sont nouvellement créés et ont des compétences nationales ou régionales. Les premières missions de l'audit et de l'évaluation ont eu pour objectif de **s'assurer que les acteurs chargés de la mise en œuvre des réformes des politiques éducatives ont les capacités et les compétences requises garantissant l'atteinte des objectifs des réformes.**

C'est ainsi qu'il a été constaté que les Académies régionales d'éducation et de formation sont de création récente pour mettre en œuvre une réforme ambitieuse de l'éducation sur le plan régionale. Néanmoins, elles dépendent fortement du ministère de l'éducation nationale dans l'acquisition des moyens.

- Les ressources humaines sont encore allouées et gérées par le ministère de tutelle ;
- Les ressources financières dépendent à plus de 90% des subventions du budget général de l'Etat ;
- Le transfert du patrimoine aux Académies n'est pas encore effectué.

Quant aux universités, elles sont chargées de la mise en œuvre de la réforme Licence Master et Doctorat LMD, mais l'acquisition des moyens dépend principalement du ministère de tutelle et du budget général de l'Etat. Les moyens alloués pourront ainsi constituer un risque élevé dans la réalisation des objectifs assignés.

D'autres risques sont identifiés comme le cumul de plusieurs métiers. Les académies régionales et les universités cumulent la gestion de l'offre de formation, la réalisation des investissements en infrastructure, la gestion de la restauration (cantines et internats), l'hébergement ...etc. Ces missions requièrent des moyens et des compétences qui ne sont pas toujours disponibles.

La gestion de la qualité constitue également une préoccupation majeure pour la Cour des comptes. En l'absence d'une autorité indépendante pour l'évaluation de la qualité de l'éducation, la Cour a inscrit dans ses travaux l'examen des conditions de l'offre scolaire et de l'offre de formation.

En général, la programmation de l'évaluation des institutions éducatives est caractérisée par les points suivants :

- La pluriannualité : En effet, plusieurs acteurs sont chargés de la mise en œuvre des politiques éducatives : les travaux de la Cour portent chaque année sur certains organismes ou certaines thématiques ;
- Les travaux de la Cour ne se limitent pas à la seule évaluation. En effet, ayant constaté des insuffisances dans les capacités des organismes publics, la Cour procède à des missions d'audit, d'inspection et d'investigation pour s'assurer de l'efficacité des contrôles internes, de la sincérité des transactions et de la réalité des prestations, des fournitures et des travaux effectués ;

- L'utilisation des travaux des audits internes et des inspections : La Cour examine toujours la capacité des organes de l'audit interne et des inspections pour s'assurer de leur indépendance et l'allocation des moyens suffisants leur garantissant la réalisation des objectifs qui leur sont assignés. Les travaux de ces organes sont souvent utilisés dans les travaux de la Cour ;
- L'orientation actuelle est vers l'évaluation des métiers et des programmes : Ainsi, la Cour entreprend actuellement des missions d'évaluation des thématiques communes à plusieurs organismes publics comme l'évaluation de la formation initiale dispensées par les universités.

L'exécution des missions

D'une manière générale, le déroulement des missions de l'évaluation est effectué sur place. Les visites permettent l'observation des conditions de la scolarité, des infrastructures et les entretiens et les dialogues avec les gestionnaires.

Pour des motifs de proximité, la Cour des comptes recourt actuellement à plusieurs cours régionales des comptes pour réaliser des missions de contrôle de la gestion de certaines Académies régionales.

Les rapports de la Cour des comptes

Plusieurs moyens sont utilisés par la Cour des comptes pour la communication des résultats de ses travaux. Elle prépare des rapports provisoires, des rapports particuliers et le rapport annuel.

- Les rapports provisoires sont communiqués aux responsables des organismes publics pour commentaires et avis ;
- Les rapports particuliers sont adressés par le Premier président au Chef du gouvernement, au ministre chargé de l'économie et des finances et au ministre de tutelle, lesquels peuvent formuler leurs observations et exprimer leurs avis ;
- Le rapport annuel est soumis à Sa majesté le Roi. Il est également transmis au Chef du gouvernement, et aux Présidents des deux chambres du Parlement. Ce rapport est publié au Bulletin officiel.
- Un exposé des activités de la Cour est présenté par le Premier président devant le Parlement. Il est suivi d'un débat.

Le suivi des recommandations

La Cour des comptes a procédé à plusieurs missions de suivi de la mise en œuvre de ses recommandations formulées dans ses rapports annuels et entreprend actuellement le suivi des recommandations de ses rapports particuliers.

IV. Principales recommandations formulées par les évaluations des Académies régionales

D'une manière générale le contrôle de la gestion des Académies régionales d'éducation et de formation a donné lieu à des recommandations portant sur :

- La gouvernance des Académies, en l'occurrence le fonctionnement des organes de gouvernance et l'ambiguïté du rattachement de certaines structures ;
- Les insuffisances liées à l'activité scolaire en termes d'infrastructure d'accueil, des équipements des établissements scolaires et des conditions de scolarité ;
- L'encadrement des enseignants et des élèves ;
- Les défaillances constatées dans la gestion des ressources financières, humaines et patrimoniales

Gouvernance, organisation et système d'information

En matière de la gouvernance et de l'organisation des académies, la Cour avait recommandé de :

- **Clarifier les compétences des directeurs des Académies et de lever les chevauchements des attributions et la confusion hiérarchique des délégués.**

La Cour avait observé que les directeurs des Académies n'exerçaient pas toutes les compétences qui leur sont attribuées par l'article 6 du décret 2.00.1016 du 29/06/2001 pris pour l'application de la loi n°07/00, notamment en matière de nomination des délégués et des autres responsables au niveau des Académies et des délégations. Les chevauchements des attributions entre le département de tutelle, les directeurs des Académies et les délégués persistent encore.

- **Adapter l'organigramme aux missions de l'Académie.**

Les structures prévues par l'organigramme ne couvrent pas toutes les compétences et attributions de l'Académie et certaines structures sont mises en place sans qu'elles soient prévues par l'arrêté du ministre de l'Education Nationale n° 122 du octobre 2002.

- **Mettre en œuvre un système d'information fiable et performant.**

Le système d'information est la pierre angulaire dans la prise de décision et l'évaluation. Actuellement, l'information sur la performance n'a pas atteint le niveau de fiabilité et de disponibilité souhaité. La Cour a noté que le Ministère est en train d'instaurer un système d'information intégré visant à gérer l'activité administrative et pédagogique des différents intervenants (Ministère, Académies, Délégations, établissements scolaires, Inspecteurs...).

Activité Scolaire

Plusieurs des recommandations formulées lors du contrôle des Académies par la Cour se sont rapportées à l'offre scolaire, à l'équipement des établissements scolaires et aux conditions de scolarité.

- **Améliorer le système de planification et de programmation des projets de construction des établissements scolaires.**

Plusieurs contraintes ne favorisent pas l'amélioration du système de programmation des projets telles que:

- le transfert tardif, par le département de tutelle, des projets de construction des établissements scolaires pour les années 2010 et 2011 ;
- la difficulté de consolider les crédits de paiement ce qui ne permet pas la réalisation et l'achèvement des travaux de construction et d'extension des établissements scolaires.

- **Améliorer la qualité d'enseignement au niveau de l'équipement des établissements scolaires, du pilotage de la scolarité (encombrement et classes à cours multiples) et d'admission en cycles.**

A ce niveau, plusieurs mesures ont été prises visant l'extension de l'offre scolaire, la création des écoles communautaires dotées d'un internat et des logements pour les enseignants et la dotation des établissements en équipement et matériel didactique. Toutefois, Certaines Académies souffrent encore du phénomène d'encombrement et de classes à cours multiples ainsi que du problème de sous équipement de certains de leurs établissements.

Gestion des ressources

La gestion des ressources concerne les ressources humaines, les ressources financières et le patrimoine.

A ce niveau, la Cour avait recommandé la clarification de la gestion des ressources humaines entre l'Académie et l'autorité de tutelle et un meilleur redéploiement des enseignants.

Traduction du Résumé

Abstract Translation

Mohamed HADHOUDI

METHODS OF EDUCATIONAL INSTITUTIONS EVALUATION AT THE COURT OF ACCOUNTS

The Court of Accounts has had for several years a particular interest in policies, education, training policies, as well as higher education and scientific research policies.

The Court's works aim at examining management aspects, and elaborating suggestions as well as recommendations on the means to be used to improve public services' efficiency and performance.

Hence, the Court of Accounts has included in its work schedule the management evaluation of several actors in charge of implementing education policies in the Kingdom, namely the Regional Academy of Education, universities, schools, institutes of higher education and the Vocational Training Office.

The evaluation approach of the Court of Accounts:

- Examines if public programs reach their objectives in an efficient way, and in compliance with the applicable rule and legislation;
- Examines public policies implementation, but does not speak about their opportunities;
- Is contradictory: takes in consideration opinions and remarks of all stakeholders;
- Is a collegial body; All reports are discussed with committees constituted by at least 5 members;
- Makes recommendations and does follow ups for their implementation.

The Court of Accounts communicates through different ways, such as the annual report, thematic reports, specific reports, the Chamber President's Letters ...

Bibliographie et Références :

La constitution de 2011 ;

La loi n°62-99 formant codes des juridictions financières ;

Les Rapports annuels de la Cour des comptes pour les années : 2008 ; 2009 ; 2010 ; 2012 et 2013 :

www.courdescomptes.ma

LES APPROCHES ADOPTÉES PAR L'ONDH POUR ÉVALUER LES POLITIQUES PUBLIQUES EN ÉDUCATION

Ikbal Sayah *

Résumé

L'ONDH, qui s'est surtout penché, dans une première phase, sur l'évaluation de l'INDH, a progressivement étendu son champ d'investigation à l'évaluation d'autres politiques publiques, notamment celles relatives à l'éducation, l'un des 5 nœuds du futur identifiés par le RDH 50 et qui figurent parmi les politiques publiques de référence à évaluer par l'ONDH du double point de vue de son impact sur le développement humain et de l'application de ses principes structurants (participation, intégration, etc.).

Deux approches ont alors été privilégiées. La première, l'approche qualitative, a été adoptée dans le cadre d'une étude sur l'accessibilité à l'éducation de base au niveau national. Cette étude, qui s'est appuyée sur l'analyse des données qualitatives recueillies à l'issue d'investigations menées dans onze communes du Royaume, a eu pour objectif de vérifier l'hypothèse selon laquelle les objectifs que la politique d'éducation de base s'est fixée pourraient être facilement atteints dès lors qu'au niveau opérationnel, les normes d'intervention étaient déclinées en tenant compte de la diversité des milieux et de la disparité des espaces, des situations socioculturels, linguistiques et socio-économiques.

Cependant, cette première approche ne peut répondre à elle seule à un certain nombre de questionnements essentiels qui nécessitent le recours à des études quantitatives poussées afin d'enrichir la compréhension de questions éducatives cruciales, comme celles des causes de l'échec ou de l'abandon scolaire, et à éclairer les choix des décideurs et des acteurs. Pour ce faire, l'ONDH a mis en place une enquête de panel de ménages afin de mieux appréhender des questions complexes touchant à l'inéquité scolaire, d'autant plus qu'elles se situent au cœur des mécanismes de transmission de la pauvreté intergénérationnelle.

Les premiers résultats de la première vague de cette enquête, en plus de ceux issus des investigations qualitatives effectuées par l'ONDH, feront l'objet de la présente communication.

Introduction

S'il existe une politique publique pour laquelle il y a plus qu'exigence d'évaluation, mais devoir d'évaluation, c'est bien celle de l'éducation et de la formation qui, malgré l'injection de ressources budgétaires considérables, ne parvient pas à les transformer en résultats.

En effet, tous les rapports, notamment ceux du CSE, l'OCDE ou de l'ONDH, pour n'en citer que quelques-uns parmi les récents, dressent un bilan très mitigé de l'efficacité du système éducatif national :

- Faible maîtrise des compétences de base par plus des deux-tiers des élèves au terme de leur scolarité obligatoire ;
- Important taux d'échec scolaire, surtout en milieu rural ;
- Haut niveau d'inégalités scolaires d'origine sociales qui de plus s'accroissent au cours des études.

* Directeur des Etudes, Observatoire National du Développement Humain

Tous ces constats montrent bien que si l'éducation nationale a pu relever, en grande partie, le défi de la « massification », elle ne progresse guère sur le chemin de l'égalité des chances, ce qui n'est pas sans handicaper aujourd'hui, mais encore plus demain, l'insertion de notre pays dans l'économie et du savoir, dans la mesure où l'accès à celle-ci est en réalité conditionné à des connaissances et compétences transmises par l'éducation.

C'est dire les dangers portés par l'inaction, tant sur l'humain, que sur la société, l'économie et la démocratie. L'inaction conduira, de façon irréfragable, à la sclérose de notre société, faute de renouvellement de ses élites et de mobilité sociale. Mais elle conduira aussi sûrement à la pauvreté économique lorsque l'on sait que dans notre pays, les inégalités freinent la croissance deux fois plus que la pauvreté et qu'une hausse de 10% des élèves ne maîtrisant pas les fondements scolaires se traduit par une baisse de 1,5 point du PIB national.

Plus que jamais, il est important de rechercher les causes de l'inefficacité du système à conduire à la réussite tous les élèves qui lui sont confiés, c'est-à-dire de favoriser chez chacun d'entre eux, le développement de leurs capacités personnelles, afin qu'ils aient une chance de réussir leur vie personnelle, sociale et professionnelle.

Pour ce faire, l'Observatoire National du Développement Humain, qui a pour mission d'évaluer les politiques de développement humain, a adopté deux approches évaluatives complémentaires.

I. Les approches évaluatives développées par l'Observatoire

La première approche, qui est qualitative, a été développée dès 2010, en vue d'identifier les déterminants de l'accessibilité à l'éducation de base de qualité. L'Observatoire souhaitait alors conforter son hypothèse selon laquelle les objectifs que la politique d'éducation de base s'étaient fixés pourraient être facilement atteignables dès lors que sur le plan opérationnel, les normes d'intervention étaient déclinées en tenant compte de la spécificité des milieux, de la disparité des espaces, des situations socioculturelles, linguistiques et socioéconomiques.

L'Observatoire a alors entrepris dans ce cadre des études de cas dans onze communes représentatives de la diversité du territoire national, tant en termes de milieux, de situations socio-économiques ou de réalités socio-culturelles. Dans chacune de ces communes, les disparités ont été analysées en fonction d'un référentiel de normes –internationales ou nationales- sur le droit à l'éducation d'une part, et en rapport avec les spécificités territoriales en matière d'offre scolaire, d'appui scolaire, de qualité de rendement et de gouvernance, d'autre part.

Ces études ont permis de conformer l'hypothèse émise par l'ONDH. Elles ont révélé de manière éclatante les conséquences néfastes pour des millions d'enfants, de l'absence d'une vision cohérente de l'aménagement du territoire et d'une véritable coordination de l'action publique au niveau local, mais ont également montré la nécessité de miser sur les solutions locales, notamment à l'attention des enfants en situation d'échec scolaire, généralement les plus pauvres, pour améliorer les performances du système éducatif.

La seconde approche déployée par l'ONDH est cette fois-ci quantitative.

Elle repose sur deux outils. Le premier est le système d'information territorial dont la pertinence tient au fait que les éléments de contextes territoriaux sont susceptibles d'agir sur le registre de l'échec scolaire dans ce que certains ont appelé une approche écologique de l'éducation. Celle-ci considère que le niveau d'éducation est le résultat d'interactions entre l'individu et ses environnements. Elle permet donc d'associer l'éducation et une graduation de contexte sans pour autant chercher à établir des causalités.

C'est justement l'un des objectifs de l'enquête panel de ménages de l'ONDH qui présente l'avantage, en plus du suivi de l'évolution globale d'un grand nombre d'indicateurs d'intérêt en rapport avec le secteur éducatif, de prendre en considération les facteurs exogènes qui peuvent sérieusement affecter les trajectoires scolaires. Il s'agit là d'une analyse indispensable si l'on souhaite enrichir la compréhension de questions éducatives cruciales et éclairer les choix des décideurs et des acteurs.

A cet égard, l'ONDH suit régulièrement depuis 2012 un échantillon représentatif au niveau national, national urbain et national rural, composé de 8.000 mêmes ménages, soit près de 37.500 individus. Cet échantillon sera étendu à d'autres ménages dès 2017, en vue de disposer de résultats robustes à l'échelon régional.

L'un des avantages de ce panel est de couvrir un grand nombre de dimensions (pauvreté, éducation, santé, revenus, dépenses, logement, accès aux services de base, etc.), et donc de permettre une analyse plus fine des facteurs de différenciation des parcours scolaires des élèves, dès les classes préscolaires, mais aussi après que ces derniers ont quitté le système éducatif.

Aussi, ces données de panel longitudinales sont d'un grand apport pour mesurer et évaluer l'égalité des chances dans les domaines de la préscolarisation, de la scolarisation ou de la survie scolaire.

Les premiers résultats de l'enquête panel de ménages de 2012 ont fait l'objet d'une publication en 2014. Ceux de l'enquête de 2013 sont en cours d'analyse et ils ne manqueront pas d'apporter un éclairage inédit et pertinent sur des questions aussi importantes que celles touchant à l'avenir de nos enfants, du fait simplement que l'on ne soit plus soumis à la tyrannie de la moyenne.

En effet, en appréhendant les dynamiques individuelles, les décideurs, les chercheurs ou les acteurs concernés disposent désormais d'informations sur les flux entrants et les flux sortants, alors qu'auparavant, ils ne disposaient que d'un solde et donc de deux fois moins d'informations.

En l'occurrence, s'ils s'intéressent au décrochage scolaire, ils auront des informations sur :

- Les caractéristiques des décrocheurs qui quittent l'école ;
- Les causes de leur décrochage ;
- Les caractéristiques des décrocheurs qui finissent par rejoindre l'école et qui y perçoivent (peut-être) le sens et l'utilité, notamment suite à la mise en place de dispositifs scolaires ou non scolaires appropriés.

De même, si ces mêmes acteurs s'intéressent à la part prise par le secteur privé dans la scolarisation des enfants, un flux important de ces derniers du secteur privé vers celui public devrait les conduire à s'interroger sur la pertinence des stratégies éducatives adoptées par les parents, notamment ceux issus des classes moyennes. Ces classes, plus dépendantes de l'école que les autres catégories sociales pour la reproduction de leur statut social, ne seraient-elles plus en mesure de supporter les frais engendrés par la scolarisation de leur progéniture ? Il s'agit donc de nouvelles interrogations que les données longitudinales mettent désormais en exergue. Il convient alors d'y répondre compte tenu du rôle joué par les classes moyennes pour assurer la pérennité de notre projet démocratique.

II. Perspectives

De ce qui précède, il apparaît que l'ONDH s'est surtout investi sur la détermination des facteurs exogènes à l'origine des inégalités scolaires. Il s'agit d'un sujet d'importance car si ces inégalités ne sont pas corrigées, elles se transmettront aux générations suivantes sous forme de nouvelles inégalités, cette fois-ci sociales, qui malheureusement, freineront à coup sûr la concrétisation du « Maroc souhaitable ».

Les approches développées par l'Observatoire peuvent aider à répondre à ces problématiques d'inefficacité scolaire, mais ne permettent pas d'appréhender les facteurs endogènes qui sont à même de nuire à la qualité de l'éducation. Ces derniers touchent à la qualité des installations et des bâtiments, à la disponibilité des fournitures scolaires, au ratio élèves par enseignant, à la qualité des programmes, à l'utilisation de l'évaluation formative ou à l'utilisation de pédagogies basées sur les TIC.

En effet, il ne faut pas oublier que si les données permettent d'assurer un pilotage plus aisé des politiques éducatives, les enjeux sont tout d'abord pédagogiques. La réussite de chacun est en fait entre les mains de toute la communauté éducative, et surtout entre celles des enseignants, détenteur de la plus noble des autorités, celle du savoir et de sa transmission.

Il est donc essentiel d'évaluer la qualité des prestations des services éducatifs. C'est l'objet de la prochaine étude de l'ONDH qui, de la sorte, aura complété son dispositif d'évaluation des politiques éducatives.

Traduction du Résumé**Abstract Translation**

Ikbal Sayah

THE OBSERVATORY APPROACHES TO EVALUATE PUBLIC POLICIES

Human Rights National Observatory, which has focused in its first mission to evaluate the Human Development National Initiative, has progressively enlarged its field of investigation to evaluating other public policies, namely those related to education, as one of the 5 nodes of the future identified in the Human Development Report 50, and which figure among the reference public policies to evaluate by the Observatory, given the impact on human development.

Two approaches have hence been favored. The first, qualitative, has been adopted in a study on the access to basic education at the national level. This study, which was based on the analysis of quality data, collected through surveys led in 11 localities in Morocco aimed at verifying the hypothesis, which say that objectives targeted by the basic education policy might be reached if norms interventions are set, taking into consideration space diversity, sociocultural, linguistic and socio-economic situations.

Yet, this first approach cannot provide the expected feedback as it requires more in depth quantitative studies in order to enrich the understanding of educational issues, such as school drop, failure and enlighten decision makers as well as actors' choices. To this end, the Observatory has set a survey of household panel in order to better understand complex questions related to inequity in school especially that they are in the heart of intergenerational poverty transmission mechanisms.

This paper will present the first results of this survey as well as those from qualitative investigations.

CONCLUSION



RÉSUMÉ DES TRAVAUX DU COLLOQUE ET CONCLUSIONS

Albert Sasson *

Le colloque international sur «l'évaluation en éducation et en formation : Approches, enjeux et défis», organisé à Rabat, les 22 et 23 octobre 2015, par l'Instance Nationale d'Évaluation du Système d'Éducation, de Formation et de Recherche Scientifique (INE), a le mérite de se fonder sur une note de présentation, diffusée à l'ensemble des participants et précisant les objectifs et les résultats attendus du colloque. Ces derniers ont été rappelés dans les deux allocutions introductives, celles du président du CSEFRS, Omar Azziman, et de la Directrice de l'INE, Rahma Bourqia. Un colloque de deux jours, d'une rare densité intellectuelle, avec 33 communications, dont 17 présentées par des experts étrangers (certains d'entre eux ou elles représentant des organismes intergouvernementaux comme l'OCDE, l'UNICEF ou la Banque Mondiale). Un bon et sage équilibre entre l'expertise ou l'expérience nationales et celles d'origine externe.

Un colloque marqué par la participation active de l'assistance, malgré les contraintes de temps ; une parité hommes/femmes qui reste insuffisante ; quelques chevauchements pour certaines présentations ; une large place accordée à l'enseignement supérieur, mais fort peu à l'évaluation de la recherche qui devrait aller de pair ; une place relativement modeste pour l'évaluation concrète du système éducatif scolaire national.

Mais il s'agit d'une première tentative, qui est globalement réussie et qui sera suivie d'autres opérations (journées d'étude ou de réflexion, ateliers, séminaires nationaux et internationaux), de façon à impliquer encore plus étroitement tous les acteurs et les institutions, et à traiter certains sujets avec plus de détail et de discussion. En cela, il convient de faire confiance à l'INE et de l'encourager dans son œuvre de stimulation, de coordination et de promotion de la culture de l'évaluation - avec modération, mais aussi avec pertinence.

Quels enseignements retenir de ces deux jours d'écoute et de discussion, quels messages forts recueillir et emporter en vue de l'action future ?

a. En premier lieu, l'évaluation, qu'il s'agisse des acquis des élèves ou des étudiants, des établissements scolaires ou universitaires, des enseignants, ou encore des autres acteurs du système éducatif, est une tâche complexe et difficile. Cette complexité et cette difficulté, qu'il faut bien réduire ou surmonter, provient du fait qu'il s'agit de porter un jugement sur la valeur, de jauger, d'apprécier, d'estimer une série d'activités, et cela par les méthodes et moyens divers (questionnaires, enquêtes, vérifications, entretiens, inspections), de façon participative (c'est-à-dire en dialoguant avec tous les acteurs) et en disséminant très largement les résultats de l'exercice (et en poursuivant le débat que ces résultats génèrent). On peut donc, selon l'étymologie latine, procéder à l'appréciation de la valeur - satisfaisante, voire excellente, ou à l'inverse, insuffisante, voire trop faible. En langue arabe, le mot *taqīm* contient aussi le concept de jugement, d'appréciation de la valeur et de sa mesure (*qias*). C'est aussi le cas du mot *assessment* qui implique le jugement, par exemple le contrôle des connaissances, ainsi que des méthodes (*methods of assessment*).

b. En second lieu, l'évaluation donne lieu à des enquêtes, à des comparaisons ainsi qu'à des exercices internationaux, auxquels il n'est pas aisé de participer (méthodologiquement et financièrement). Leurs acronymes ou sigles sont bien connus (PISA, TIMSS, PIRLS, SABER, etc.), leurs résultats aussi, ainsi que les critiques de ces derniers. Le Maroc s'efforce d'y participer et surtout d'en appliquer les leçons à ses propres évaluations. On connaît les louanges faites aux systèmes éducatifs des pays scandinaves, où l'enfant, l'élève, est au cœur du dispositif d'apprentissage, où l'accompagnement des plus faibles est une obligation pour les enseignants, où les maîtres jouissent de la considération de leurs concitoyens et ont droit à des rémunérations tout-à-fait satisfaisantes. Le slogan *No child left behind* s'applique beaucoup plus au cas des pays scandinaves qu'à celui encore inégalitaire prévalant outre-Atlantique.

* Membre du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique, Membre du Conseil Économique, Social et Environnemental, Membre de l'Académie Hassan II des Sciences et Techniques.

Le Maroc doit absolument réussir sa réforme éducative, après plusieurs décennies marquées par de gros investissements dans l'éducation nationale, mais avec des résultats peu satisfaisants. Qu'il s'agisse de l'enseignement des langues nationales et des langues d'ouverture sur le monde contemporain, de l'histoire, de l'éducation scientifique et technique, de l'éducation à la citoyenneté, ou de la formation professionnelle, l'évaluation est un instrument précieux pour éclairer les programmes et les curriculums, les méthodes pédagogiques et, au final, l'efficacité de l'ensemble du système.

c. En troisième lieu, l'accréditation des universités et des établissements d'enseignement supérieur, ainsi que l'évaluation de la qualité des enseignements prodigués par les enseignants-chercheurs et des acquis des étudiants, constituent une part importante de l'évaluation des systèmes éducatifs, au Maroc et ailleurs. Plusieurs exemples ont été donnés durant le colloque et le message fort est qu'il faut poursuivre les efforts déjà déployés, en s'inspirant des normes ou standards internationaux, mais aussi en veillant à l'étroite relation qui doit exister entre l'évaluation des acquis des élèves et de ceux des étudiants. L'amélioration indispensable des premiers est une aide précieuse aux efforts faits pour redonner à l'enseignement supérieur une qualité qui a tendance à se dégrader.

Les tentatives faites pour évaluer le « rendement externe » de l'université ou de certaines de ses composantes, constituent une démarche intéressante s'il s'agit d'accompagner l'insertion des diplômés dans le monde du travail, à l'instar de ce que font le plus souvent les écoles d'ingénieurs ou de gestion d'entreprises (dans le cadre de la collaboration étroite entre l'entreprise et la formation technique). L'expression « rendement externe », empreinte d'un souci primordial de l'économie, aurait avantage à être remplacée par une autre mettant en relief la volonté d'accompagnement de jeunes hommes et femmes, qui ont besoin de conseils, de compassion et d'assistance.

d. En quatrième lieu, les acteurs engagés dans l'évaluation de systèmes éducatifs au sein de politiques ou de stratégies sont relativement nombreux : Cour des Comptes, INDH-ONDH, Ministère de l'Education Nationale et de la Formation Professionnelle, Ministère de l'Enseignement supérieur et de la Recherche Scientifique. La concertation devrait s'imposer, l'échange d'expériences aussi, le risque de chevauchement ou d'empilement d'institutions ou de dispositions réglementaires doit être évité. L'Instance Nationale d'Evaluation doit ici jouer un rôle important.

En conclusion, ce colloque international sur l'évaluation est en définitive une très heureuse initiative, prise dans le cadre de la Vision stratégique 2015-2030, proposée par le CSEFRS et approuvée pour une mise en œuvre rapide avec tous les acteurs concernés (après une campagne de communication, de restitution et de débats sur les réformes annoncées en 2015-2016 à l'échelle nationale, entre novembre 2015 et janvier 2016). Ce colloque a sans doute permis d'attirer l'attention sur les efforts actuellement déployés pour engager un assez large éventail d'activités d'évaluation, mais aussi sur ceux qui restent à faire, dans le cadre de la Vision stratégique 2015-2030, pour accompagner les indispensables réformes de l'enseignement scolaire, de la formation professionnelle et de l'enseignement supérieur et de la recherche scientifique et technique.

ANNEXES



EVALUACION DE APRENDIZAJES Y MEJORA DE LA EDUCACION.

LA EXPERIENCIA DE ARGENTINA: ONE E IMESA

Marisa Diaz *

Resumen

Las políticas de información y evaluación tienen más de 20 años de vigencia en Argentina. Durante este período se han consolidado procesos institucionales y metodológicos que nos permiten contar con un sistema nacional de carácter federal de gran valor político y pedagógico. En particular, las políticas de evaluación estatal están ligadas estrechamente a las políticas de enseñanza y orientadas a la toma de decisiones pedagógicas, sobre una base informada, que valora particularmente la distancia que existe entre el logro propuesto y la situación de la que se parte. Es decir que son políticas que no se agotan en la producción de información diagnóstica o descriptiva de la realidad de un sistema educativo.

En tal sentido, los ONE —Operativos Nacionales de Evaluación que consisten en la aplicación cada tres años de pruebas estandarizadas para medir cuánto saben los estudiantes, de carácter muestral o censal, según corresponda, en determinados años de la educación obligatoria argentina— son una valiosa fuente de información para la mejora de la gestión institucional, las prácticas de enseñanza y los aprendizajes de los alumnos en Lengua, Matemática, Ciencias Sociales y Ciencias Naturales. Dado que Argentina es un país federal y la responsabilidad directa sobre las aulas y las escuelas es de cada una de las 24 provincias en las que se organiza el sistema educativo, las pruebas se construyen tomando como referencia unos contenidos mínimos acordados que se denominan Núcleos de Aprendizaje Prioritario.

De igual modo, a partir del año 2013 y en función de los avances de las políticas de fortalecimiento de la educación secundaria obligatoria, la Argentina incorpora como instrumento de seguimiento pedagógico el IMESA —Índice de Mejora de la Educación Secundaria Argentina—, destinado a orientar el trabajo de las instituciones educativas del nivel en procesos de mejora de las trayectorias escolares de sus estudiantes.

Este índice combina los resultados de las pruebas nacionales (ONE) en Lengua y Matemática con indicadores relativos a la trayectoria de los alumnos es de aplicación universal para las escuelas secundarias, y posibilita una mirada integral sobre cada institución educativa. Es una herramienta para apoyar la autoevaluación escolar y el diseño y el desarrollo de estrategias educativas por escuela, pero también es un instrumento de seguimiento y de evaluación para el estado nacional y las provincias, permitiendo monitorear el impacto de políticas educativas.

Tanto el ONE como el IMESA son ejemplos de cómo sostener procesos de evaluación externa y de autoevaluación institucional en las instituciones educativas, que integren una agenda de trabajo que incida de manera efectiva en la mejora de la calidad de la enseñanza y los aprendizajes escolares, sin descuidar los procesos de inclusión de todos los estudiantes, el respeto de las culturas escolares y el fortalecimiento del trabajo de los equipos docentes.

El Sistema Educativo Argentino

El sistema educativo argentino es un sistema federal conformado por 24 jurisdicciones. En 2015 su matrícula total fue de 10.964.618 estudiantes, atendidos en 52. 072 servicios educativos¹. Esta población escolar corresponde a los niveles inicial, primario, secundario, adultos y educación técnica.

El estado nacional fija los lineamientos de política educativa para el sistema en su conjunto y regula su aplicación en base a procesos federales de consulta y concertación a través del Consejo Federal de Educación (CFE). Apoya técnica y financieramente a las provincias y es responsable de la educación universitaria.

¹ - Datos oficiales del Relevamiento Anual Estadístico 2014 elaborado por el Ministerio de Educación de la Nación a través del Área de Gestión de la Información en base a datos provistos por las jurisdicciones.

* Directrice Adjointe de Planification Éducative, Ministère de l'Éducation Nationale, Argentine

Las provincias son responsables de gestionar los sistemas educativos jurisdiccionales y de establecer el currículo provincial en consonancia con los estándares nacionales definidos en los Núcleos de Aprendizajes Prioritarios para la educación Obligatoria (NAP).

La educación pública está consagrada en el artículo 14 de la Constitución Nacional y regulada por la Ley de Educación Nacional 26.206 del año 2006 que establece: la estructura del sistema por niveles y modalidades; la extensión de la educación obligatoria y las responsabilidades del estado Nacional los estados provinciales y la ciudad autónoma de Buenos Aires, capital del país.

En 2015 la ley 27045 modifica la ley 20.206 en su artículo 16 y establece como obligatoria la sala de 4 de años.



Sistema educativo en Argentina

14 años de educación obligatoria.

La educación esta garantizada en la Constitución.

La regulación principal es la LEN aprobada en 2006.

Nivel educativo	Edad	Área geográfica	Status
Educación pre-primaria(ISCED 0)	3-5 años	Todo el país	Education obligatoria empieza a los 4 años
primaria (ISCED 1)	6-11 años	12 provincias	Obligatoria
secundaria (ISCED 2 et 3)	12-17 años	12 provincias	Obligatoria
	13-17 años	12 provincias	
	ETP dura un año mas		

Subsecretaría de Planeamiento Educativo

Estrategia Nacional de Evaluación

El Estado tiene la responsabilidad ineludible de realizar un seguimiento integral de las políticas educativas y, particularmente, evaluar sus efectos en los sistemas, las instituciones y las aulas.

Por consiguiente, la evaluación es una tarea de gran valor político que otorga visibilidad a las responsabilidades y acciones de los distintos protagonistas para el logro de los objetivos propuestos, y opera revisando los modos en que la acción estatal se despliega para mejorarlos, instalar otros o, si fuera necesario, interrumpirlos. Para el Ministerio de Educación de la Nación la evaluación es un proceso formativo y no punitivo.

Por ello desde hace más de 20 años, Argentina desarrolla los Operativos Nacionales de Evaluación (ONE) y participa de estudios evaluativos regionales y globales. De igual modo lleva adelante estudios específicos y evaluaciones de programas de gobierno.

• Operativo Nacional de Evaluación (ONE)

Desde 1993, Argentina desarrolla una estrategia nacional de evaluación que combina la aplicación de pruebas nacionales e internacionales. En ese marco, el Operativo Nacional de Evaluación (ONE) mide

resultados de aprendizaje en lengua, matemática, ciencias sociales y naturales en la educación básica en todo el país. En sus primeros años de implementación, fue un operativo anual, desde 2010 se aplica cada 3 años.

El ONE provee información y conocimiento sobre los desempeños de los estudiantes de escuelas públicas y privadas en las 4 áreas básicas del currículo oficial. Es una medición estandarizada de los logros educativos acumulados en 3° 6° 9° (prueba muestral) y 12°. (prueba censal).

Las pruebas nacionales son elaboradas por el Ministerio de Educación Nacional tomando como base los currículos vigentes en las jurisdicciones; los NAP (Núcleos de Aprendizaje Prioritarios) que establecen estándares de aprendizajes federales para todas las áreas; los libros de textos usados en las escuelas y los resultados de los ONE de años anteriores. Son pruebas en papel que integran el múltiple choice con preguntas de producción para los alumnos. Los resultados se procesan mediante la aplicación del modelo Rasch.

Asimismo, el ONE aplica cuestionarios de factores de contexto para directores docentes y estudiantes que relevan información sobre: aspectos sociales y económicos de los estudiantes, condiciones escolares y condiciones de enseñanza.

Los resultados de los estudiantes se organizan por niveles de logro inclusivos:

- **ALTO:** los estudiantes alcanzan destacado desempeño en todos los contenidos y capacidades evaluadas de acuerdo a su nivel de escolaridad y los documentos curriculares tomados como referencia.
- **MEDIO:** los estudiantes alcanzan satisfactorio desempeño en todos los contenidos y capacidades evaluadas de acuerdo a su nivel de escolaridad y los documentos curriculares tomados como referencia.
- **BAJO:** los estudiantes alcanzan elemental desempeño en todos los contenidos y capacidades evaluadas de acuerdo a su nivel de escolaridad y los documentos curriculares tomados como referencia.

Escuelas y estudiantes

ONE 2013	Escuelas	Estudiantes
Año 12 (ISCED 3) - censo	8.889	280.000
Año 9 (CITE 2) - muestra	1.740	34.820
Año 3 y 6 (CITE 1) - muestra	5.400	70.000

Subsecretaría de Planeamiento Educativo

Difusión y trabajo escolar con los resultados del ONE

En la actualidad los resultados se informan a nivel nacional y por regiones geográficas² que agrupan a las provincias argentinas presentando el porcentaje de alumnos incluidos en cada rango conjuntamente con un informe cualitativo que analiza los factores asociados a dichos aprendizajes y recomendaciones metodológicas para las escuelas y los docentes por grado evaluado a los efectos que puedan llevar a cabo acciones correctivas a partir de los errores más frecuentes detectados en la resolución de las pruebas.

Cada gobierno educativo tiene acceso a los resultados de su jurisdicción y aun comparativo con respecto a la región geográfica que integra. Para el caso de la prueba censal de 12° año, además, los resultados se entregan por escuela vía sitio web de acceso restringido.

Este modo de informar está vigente desde 2006 y tiene resguardo legal en la ley Nacional de Educación del mismo año³. Esta regulación es motivo de controversias y debates de gran intensidad promovidos por algunos sectores de la sociedad y especialistas que cuestionan que los resultados sean publicados con restricciones argumentando que una estrategia de difusión más amplia permitiría un mayor control institucional y social de la calidad de las escuelas y el trabajo de los docentes.

Desde el Ministerio de Educación Nacional se sostiene que los resultados del ONE deben informarse a los responsables de las administraciones educativas jurisdiccionales y a las escuelas como parte de un proceso de mejora de las prácticas escolares y no como un dato aislado.

Este debate a su vez actualiza la revisión de las posiciones respecto a la evaluación y los logros de aprendizaje. Para el Ministerio de Educación Nacional evaluar es comprender, informar, predecir e interpretar. Es una mirada sistémica y sistemática que fortalece y propicia una lectura integradora de la propuesta escolar en sus procesos y resultados.

2 - La manera de informar los resultados fue variando desde la primera edición hasta el presente a partir de decisiones consensuadas entre el Ministerio de Educación de la Nación y el Consejo Federal de Ministros. En la actualidad se toma como criterio las 5 regiones del país: NEA, NOA, CENTRO, CUYO y PATAGONIA.

3 - El artículo 97 de la Ley de Educación Nacional 26.026 establece la difusión de los resultados de las evaluaciones preservando la identidad de alumnos, docentes e instituciones educativas evitando cualquier forma de estigmatización.

Por lo tanto, al evaluar aprendizajes evaluamos enseñanza y esta, es un proceso complejo y abarcativo, que no puede ser medido/calificado de modo excluyente por un solo indicador de los progresos escolares como lo son las pruebas nacionales. Asimismo, la publicación masiva de resultados por escuela conlleva el riesgo del armado de rankings que distorsionan el foco del problema y atentan contra procesos de mejora efectivos. En tal sentido los modos de comunicar los resultados deben estar, a criterio del Ministerio Nacional, en profunda y clara relación con la intencionalidad de trabajo sobre los mismos. Intencionalidad que tiene que materializarse en una propuesta de autoría institucional que no estigmatice; que sea proactiva, progresiva, y mensurable en términos de progresos escolares.

En síntesis, los ONE informan y trazan escenarios. Permiten revisar en el seno de cada escuela la organización institucional la continuidad de las prácticas de enseñanza las trayectorias escolares y la integración de las escuelas a sus contextos.

Los ONE hacen evidente que la convergencia entre las políticas educativas y el trabajo en las aulas se afianza a partir de una cultura de la evaluación en donde todos los actores del sistema educativo participan corresponsablemente de un proceso formativo y no punitivo.

Un dato, en el último ONE 2013, censal participaron 8889 escuelas y 280.000 estudiantes del 12° año⁴.

• Índice de Mejora de la Educación Secundaria Argentina (IMESA)

La estrategia nacional de evaluación acompaña y potencia las políticas de mejora y fortalecimiento de la educación obligatoria. A lo largo del tiempo ha posibilitado el re direccionamiento de decisiones de política educativa a nivel nacional y/o jurisdiccional haciendo foco en problemáticas vinculadas prioritariamente a la enseñanza.

En tal sentido y en particular los magros resultados obtenidos en las últimas ediciones en el ONE censal del 12 ° año escolar vienen alertando sobre la necesidad de intensificar los esfuerzos para la mejora de la propuesta escolar en un contexto de:

- reciente acceso masivo de estudiantes de sectores desventajados a la secundaria (en 2006 se establece obligatoria la educación secundaria en el país).
- altos niveles de abandono.
- estancamiento de resultados de aprendizaje.

Por lo dicho desde 2013 y en el marco del desarrollo del Plan Nacional de Educación Obligatoria y Formación Docente 2012-2016 aprobado por el Consejo Federal de Educación⁵, de forma gradual, las escuelas secundarias argentinas cuentan con un Índice de Mejora institucional denominado IMESA (Índice de Mejora de la Educación Secundaria Argentina) Elaborado por Ministerio de Educación de la Nación y consensuado en su formulación técnica y aplicación por el Consejo Federal de Ministros.

El propósito del IMESA es ofrecer a los directores y docentes de las escuelas secundarias de Argentina, una herramienta de gestión orientada a mejorar los niveles de comprensión e intervención en las problemáticas de enseñanza de las áreas básicas y su correlato en los aprendizajes y el progreso escolar de los estudiantes.

Antecedentes⁶ :

La propuesta del IMESA toma como antecedente cercano al índice de Desarrollo de la Educación Básica IDEB⁷.

El IDEB contiene dos componentes: uno asociado a los resultados de sus pruebas de evaluación nacionales y otro relacionado con la trayectoria de los alumnos.

Respecto de las evaluaciones, se toma el promedio de las pruebas de lengua y matemática del último grado del ciclo que se desea monitorear, luego de haber realizado un cambio de escala para que los resultados tomen valores entre 0 y 1.

Respecto de la trayectoria, el índice toma la sumatoria de las inversas de las tasas de promoción de los

4 - Dato provisto por la Dirección Nacional de Información y Evaluación de la Calidad Educativa. Ministerio de Educación de la Nación

5 - Resolución del CFE N°188/12

6 - Este texto es una síntesis de un documento de circulación interna elaborado por equipo técnico de IMESA dependiente de la Dirección Nacional de Información y Evaluación de la Calidad Educativa

7 - El Índice de Desarrollo de la Educación Básica es creado en 2005 por el Instituto Nacional de Evaluación e Investigaciones Educativas de Brasil (INEP)

grados que componen el ciclo, que, si se supone que el abandono es 0, sería igual a la duración media de ciclo para los alumnos que egresan.

Es así que, en su definición, el IDEB considera los rendimientos en pruebas estandarizadas y la trayectoria de los alumnos. Como medida de esta última dimensión se considera el tiempo medio para concluir el nivel, pero se emplea en el cálculo la sumatoria de las inversas de las tasas de promoción⁸

Para el caso argentino se considera que el supuesto de ausencia de abandono es difícil de sostener. Este problema es tenido en cuenta en el IMESA y para solucionarlo se ha considerado el **tiempo medio** para completar el nivel asumiendo esta definición como la más apropiada.

Sin embargo, este indicador solo mide el tiempo para aquellos alumnos que egresan; sin importar si son el 10 o el 90% de los que se inscriben al nivel.

En tal sentido por tratarse de un índice ligado a la calidad del proceso escolar la premisa es que este logro no sea solo la conquista de algunos pocos, por ello, el índice debe incluir alguna medida de la capacidad de las escuelas de retener a sus alumnos y lograr que completen sus trayectorias escolares.

Esta problemática no solo está asociada a la vulnerabilidad de los alumnos, sino que, por ejemplo, en las escuelas privadas (y algunas estatales) es una práctica muy extendida la expulsión de los alumnos que deben repetir el año escolar.

Por lo expuesto se ha considerado entonces incluir la tasa de egreso dentro de la construcción del índice propuesto. Esto es, si un establecimiento tiene alto rendimiento y una trayectoria de sus alumnos cercana a la ideal como resultado de la expulsión de sus alumnos con más dificultades, la corrección por egreso resultará en una disminución del índice que permitirá una lectura institucional y pedagógica más cerca a la realidad.

La propuesta de Argentina:

La propuesta de cálculo del IMESA se define entonces como:

$$IMESA = \frac{R * TE}{TM} * f$$

Donde:

R es una medida relacionada con el rendimiento de los alumnos obtenida a partir de las pruebas de matemática y lengua del Operativo nacional de Evaluación que varía de 0 a 11.

TM : es el tiempo medio que les toma a los alumnos completar el nivel. Si los alumnos completan el nivel en el tiempo ideal, esto es sin repetir, este indicador vale 1 y será mayor a medida que aumente la repitencia.

TE : es la tasa de egreso, esto es, la proporción de alumnos que se inscriben a primer año y que se estima que lograrán cumplir con todos los requisitos para certificar el nivel.

La fuente para los dos últimos indicadores es el Relevamiento Anual. Varía en 0 y 100.

f : es un factor de expansión para el índice varíe entre 100 y 500, siendo 100 la peor situación posible y 500 la ideal.

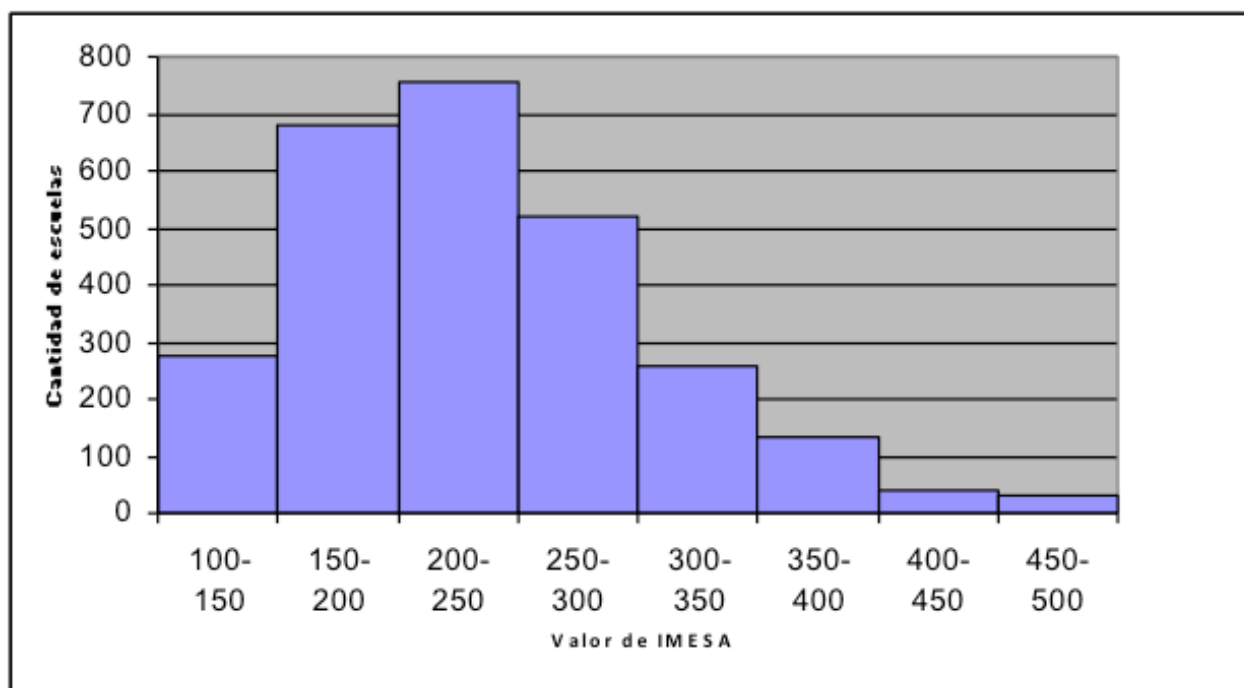
Para alcanzar el valor máximo del índice todos los alumnos de la escuela deben egresar, sin repetir ningún año de estudio y obteniendo las mejores notas en las pruebas de evaluación de la calidad. Es así que dicho valor máximo debe entenderse como un valor ideal que difícilmente ocurra en la práctica.

Ejemplo:

Se presenta a continuación la distribución de frecuencias del índice obtenido para las escuelas participantes de la primera etapa del IMESA que fueron seleccionada por contar con Planes de Mejora⁹ Institucional en 2010.

8 - Nota técnica de consulta en web http://download.inep.gov.br/educacao_basica/portaI_ideb/o_que_e_o_ideb/Nota_Tecnica_n1_concecaoIdeb.pdf.

9 - Los Planes de Mejora institucional son elaborados por las instituciones educativas con asistencia técnica y financiera del Ministerio de Educación de la Nación y el seguimiento y monitoreo de equipos técnicos de las jurisdicciones Abordan problemáticas institucionales de la enseñanza, los aprendizajes y los indicadores de rendimiento. Surgieron en el año 2010 para un tercio de las escuelas secundarias públicas y en la actualidad son una acción universal.



Cantidad de escuelas según valor de IMESA Fuente DINIECE.

Usos del IMESA

- A nivel de las jurisdicciones educativas, por sus características el IMESA entra en relación con las Metas de Eficiencia Interna de los sistemas educativos jurisdiccionales acordados a través de los convenios bilaterales¹⁰, toda vez que se constituyen en punto de llegada indicativos. Cabe recordar aquí cuáles son esas metas acordadas federalmente en a nivel país para el año 2016¹¹ :

Secundaria Ciclo Básico: llevar el abandono al 4% en cada jurisdicción y disminuir la repitencia en un tercio.

Secundaria Ciclo Orientado: disminuir el abandono en un 50% y la repitencia en un 25% en cada provincia

Niveles de desempeño marcados por el ONE: mejorar en dos puntos porcentuales los niveles medio y alto en todas las áreas básicas evaluadas, en 9° y último año de secundaria.

- A nivel de las escuelas, el IMESA puede direccionar los proyectos de mejora institucional posibilitando a directores y supervisores "complejizar las lecturas sobre las trayectorias escolares de sus estudiantes y el trabajo de sus docentes". De igual modo es un complemento instrumental de gran valor pedagógico para el sostenimiento de una alerta permanente sobre necesidad de visibilizar los procesos de inclusión/exclusión que acontecen en el cotidiano escolar y las consecuencias de los mismos en relación al a la aspiración de una buena educación para todos.

Comentarios finales

- Una estrategia de evaluación nacional construye una lectura convergente y colectiva del sistema educativo. Así, la condición de integralidad se convierte en el eje central del diseño y desarrollo de las acciones dado que permite: analizar e interpretar los procesos en marcha, revisar estrategias pedagógicas, contribuir con la mejora de la enseñanza, identificar debilidades y fortalezas, y promover una necesaria retroalimentación.

¹⁰ - Los convenios bilaterales se celebran entre la Nación y las provincias para el uso de los financiamientos disponibles para el sector educativo. Se organizan por metas quinquenales y de medio tiempo.

¹¹ - En consonancia con los Convenios Bilaterales el Consejo Federal de Educación aprobó por Resolución N° 188 /12 el Plan Nacional de Educación Obligatoria y Formación Docente 2012-2016 que retoma y traduce las metas de eficiencia interna en líneas de acción específicas delimitando y responsabilidades institucionales destinatarios y tiempos de para su ejecución.

- La evaluación del sistema educativo, las instituciones y las aulas como una estrategia política sostenida en el tiempo se constituye en una herramienta que permite reafirmar o revisar prácticas; poner en evidencia necesidades de apoyo y/o fortalecimiento y enmarcar a las instituciones educativas en una agenda de mejora y formación permanente. Desde esta perspectiva los instrumentos aplicados y la información obtenida a través de ellos se convierten en insumos para un proceso individual o colectivo de gran valor formativo que supone aprender de la realidad escolar para y transformarla y enriquecerla.
- Si bien los desempeños evidenciados por los estudiantes en las pruebas estandarizadas son un indicador de la calidad de los aprendizajes que promueve y/o posibilita un sistema educativo, una mirada integral sobre su calidad requiere de la implementación de un enfoque que valore otras dimensiones constitutivas de la calidad educativa: los niveles de inclusión o exclusión que operan en el sistema, el modo en que el Estado se ocupa de garantizar la igualdad de oportunidades educativas para todos los niños/as, adolescentes y jóvenes, y el grado de acompañamiento a las trayectorias escolares reales de los estudiantes.
- En síntesis, en cada nivel educativo, la escuela puede y debe ser evaluada; y más allá de las especificidades y rasgos particulares que la evaluación asuma en cada nivel educativo, hay un núcleo de dimensiones comunes institucionales sobre las que es necesario indagar y generar saber pedagógico:
 - prácticas institucionales y pedagógicas inclusivas,
 - dinámica productiva del tiempo escolar,
 - el trabajo colectivo por la enseñanza y los aprendizajes,
 - la inserción de las escuelas en sus comunidades.

La experiencia de estos años en Argentina demuestra que en concordancia con el avance metodológico en relación a la pertinencia de los instrumentos diseñados y aplicados se impone un trabajo institucional riguroso y constante con los resultados de las evaluaciones en tanto información pedagógica a complejizar en la revisión de las prácticas docentes. Sin el compromiso y la participación del colectivo docente, los equipos de conducción y las familias, los resultados son datos que circulan, pero no interpelan ni conmueven.

Las buenas escuelas son una construcción colectiva de responsabilidades compartidas y en ella y para ellas la evaluación de la propuesta escolar en todas sus dimensiones y efectos, se impone como un desafío político y técnico ineludible para el sistema educativo de Argentina.

Bibliografía consultada:

- Ley Nacional de Educación (2006) N°26.206. Disponible en [portal.educación.gov.ar/leyes nacionales](http://portal.educación.gov.ar/leyes_nacionales)
- Consejo Federal de Educación (2009) Resolución N°84/09: "Lineamientos políticos y estratégicos de la educación secundaria obligatoria". Disponible en www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación (2009) Resolución N°88/09: "institucionalidad y Fortalecimiento de la Educación Secundaria Obligatoria, planes jurisdiccionales y Planes de Mejora". Disponible en www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación Resolución (2010) N°116/10: "Lineamientos para la evaluación de componentes de la calidad educativa". Disponible en www.gov.ar/consejo/resoluciones
- Consejo Federal de Educación Resolución (2012) N° 188/12: "Plan Nacional de Educación Obligatoria 2012-2016". Disponible en www.gov.ar/consejo/resoluciones
- IMESA (2014): Documento marco "IMESA-Plan de Mejora Institucional" disponible en mejoraeducativa.com.ar/imesa/documentosIMESA/IMESA_documento_PDF. 2014
- DINIECE (2014): Anuario Estadístico. Disponible en diniece.me.gov.ar

امتحان الكفاءة الجامعية وأثره على مؤسسات التعليم العالي الأردنية

بشير الزعبي*

نظراً للتزايد المطرد لمؤسسات التعليم العالي في الأردن بشقيها (الحكومية، والخاصة) بات من الضروري مراقبة ومتابعة هذه المؤسسات للوقوف على مخرجاتها التعليمية، والتي بدورها تؤثر على عائدات التعليم حيث أن الهدف الرئيسي من هذه المؤسسات رفد سوق العمل المحلي والاقليمي بالأفراد ذوي الكفاءات لمواكبة سوق العمل، لذا جاء امتحان الكفاءة الجامعية للوقوف على مخرجات مؤسسات التعليم العالي في الأردن، وفيما يلي أبرز الجوانب المتعلقة بامتحان الكفاءة الجامعية والتي سيتم عرضها في هذا البحث .

أولاً: التعريف بامتحان الكفاءة الجامعية (مفهومه، أهدافه، اغراضه، مستوياته) .
ثانياً: تطبيق امتحان الكفاءة الجامعية وتسلسل النتائج .
ثالثاً: انعكاسات امتحان الكفاءة الجامعية على الأنظمة التربوية في مؤسسات التعليم العالي .

أولاً: التعريف بامتحان الكفاءة الجامعية (مفهومه، أهدافه، اغراضه، مستوياته، تطبيقاته)

امتحان الكفاءة الجامعية:

هو عبارة عن مجموعة من الفقرات أو الأسئلة التي تقيس الكفايات العامة والمتوسطة المتوقع من الطلبة اتقانها، لطلبة الجامعات الأردنية لمرحلة البكالوريوس المتوقع تخرجهم للوقوف على مخرجات التعليم في الجامعات الأردنية الحكومية والخاصة.

أهداف الامتحان وأغراضه:

يهدف امتحان الكفاءة الجامعية إلى الوقوف على جودة مخرجات البرامج الأكاديمية لمؤسسات التعليم العالي الأردنية وليس للحكم على أداء الطلبة ومنحهم شهادات توثق الأداء الذي تحقق لهم (الطلبة وسيلة فقط وليس غاية)، وذلك تحقيقاً للأغراض الآتية:

1. رسم السياسات الخاصة بالتعليم العالي من قبل الجهات ذات الصلة بجودة التعليم العالي ومؤسساته في الأردن، إذ يمكن من خلال نتائجه (تشخيص) الوقوف على جوانب القوة وجوانب الضعف في تحقيق نتائج التعلم العامة المتوقع من الطالب الملتحق في مؤسسات التعليم العالي تحقيقها، سواء ما اتصل منها بالنتائج العامة للتعليم العالي على مستوى مؤسسات التعليم العالي أم تلك الخاصة بعائلة التخصصات للمجال المعرفي، أم تلك النتائج التخصصية التي تشكل خصوصية لكل تخصص ضمن كل عائلة من التخصصات، إضافة إلى مساعدة هذه المؤسسات على تحديد الإجراءات والسياسات الواجب اتخاذها للتغلب على نواحي الضعف وتعزيز نواحي القوة في جوانب الإعداد الذي تقدمه.

2. استخدامه كأساس للاعتماد وضمان الجودة، إذ سيشكل تحقق نتائج التعلم للبرامج والتخصصات أحد المعايير الهامة، التي سيجرى بناءً عليها استمرارية اعتماد التخصصات والبرامج لمؤسسات التعليم العالي، ورفع الطاقة الاستيعابية لها، واتخاذ قرارات حول إيقافها أو سحب الاعتماد والترخيص لها من قبل الجهات المعنية بهذه القرارات.

* رئيس هيئة اعتماد مؤسسات التعليم العالي

3. خلق روح التنافس بين البرامج والتخصصات على تحقيق الجودة في مدخلاتها وعملياتها وبالتالي مخرجاتها، مما ينعكس إيجاباً على مستوى أداء مؤسسات التعليم العالي الأردنية .
4. استخدامه كأساس للمساءلة الموجهة للقائمين على البرامج والتخصصات في داخل المؤسسة التعليمية من جهة، ومن قبل وزارة التعليم العالي وهيئة اعتماد مؤسسات التعليم العالي من جهة أخرى .
5. توفير مؤشرات كمية ونوعية مستمرة تمتاز بالدقة والثبات والصدق والموضوعية عن أداء القائمين على التخصصات والبرامج التي تقدمها مؤسسات التعليم العالي .
6. الوقوف على مدى التطوير والتحسين في مخرجات الجامعة من خلال مقارنة أداء الجامعة مع نفسها على امتحان الكفاءة الجامعية لعدة سنوات .

بنية الامتحان

إن التصور الحالي لامتحان الكفاءة الجامعية يتمثل في أنه امتحان يركز على الكفايات (Competency– Based Test) أكثر من كونه امتحاناً تحصيلياً (Achievement Test) يقيس معارف وحقائق ونظريات ومهارات تم تعلمها من قبل الطالب واختبر بها واجتازها، إذ يُركز الامتحان على نتائج التعلم العامة والخاصة التي يتوقع من الطالب اكتسابها بعد الانتهاء من دراسة البرنامج الأكاديمي الملتحق به، والتي من شأنها حمله على لعب الأدوار المهنية المتوقعة في إطار التخصص الذي درسه بفاعلية، إذ تكون المهمات الامتحانية غير موجهة لقياس مقدار الحصيله المعرفية المتحققة للطالب، بل تقيس ما تحقق للطالب من المهارات والعمليات العقلية العليا، (يتذكر، يفهم، يطبق، يحلل، يقوم، يبتكر) وقدرته على توظيفها في المواقف العملية الشبيهة بتلك المواقف التي من المتوقع منه التعامل معها في سوق العمل، وفي هذا السياق يركز الامتحان على ثلاث مستويات من الكفايات أو النتائج كما هي موضحة أدناه :

المستوى العام (Macro– Level) :

حيث يُركز هذا المستوى على النتائج والكفايات العامة، التي يتوقع من جميع البرامج التي تقدمها مؤسسات التعليم العالي (بغض النظر عن تركيزاتها) أن تطورها لدى خريجها حتى يستطيعوا التمشي مع متطلبات الحياة المهنية والاجتماعية، بحيث يؤديون أدوارهم بفاعلية، وينخرطون في عملية التطوير والتعليم المستمر، حيث يضم امتحان الكفاءة الجامعية المستوى العام ورقة امتحانية محوسبة عامة تقدم لجميع الطلبة في كافة التخصصات، وهي مصممة لتقيس الدرجة التي عملت بها مؤسسات التعليم العالي -عن طريق محتوى الخطط الأكاديمية لبرامجها واستراتيجيات وأساليب التدريس المستخدمة بها- على تطويرها، لتقود إلى تطوير مثل هذه المهارات والنتائج التعليمية الهامة لإنتاج المعرفة والتنافس العالمي في عصر العولمة والاقتصاد المعرفي لدى طلبتها وهي بشكل أساسي تشمل المهارات التي يُطلق عليها الفنون العقلية (Liberal Arts) والتي تتمثل في :

1. مهارات البحث العلمي وتفسير النتائج .
2. مهارات التفكير الناقد .
3. مهارات التواصل .
4. مهارات الثقافة المعلوماتية .
5. مهارات حل المشكلات والعمل في فريق .
6. مهارات المعرفة بالحقوق والواجبات .
7. المعرفة بالقواعد والسلوكيات الأخلاقية .
8. مهارات المعالجة الرياضية .
9. مهارات المعالجة اللفظية .

المستوى المتوسط (Mid-Level) :

يركز هذا المستوى من امتحان الكفاءة الجامعية على الكفايات ونتائج التعلم التي يتوقع من عائلة التخصص (على مستوى الكلية) أن تعمل على تنميتها لدى الطلبة الخريجين، وهي تتمثل بالكفايات ونتائج التعلم المشتركة لعائلة التخصصات التي تقع ضمن الحقل المعرفي واللازمة لممارسة الأدوار المهنية المناطة بخريجي تلك العائلة، بغض النظر عن التركيزات التي ينتمي إليها البرنامج الذي تخرج منه الطالب، إذ يتم لأغراض امتحان الكفاءة الجامعية تقسيم التخصصات للبرامج التي تقدمها الجامعات الأردنية على مستوى البكالوريوس إلى أربعة عشر عائلة، وهي كما يلي:

1. الآداب واللغات وتطبيقاتها.
2. الإعلام والصحافة.
3. الشريعة.
4. العلوم الاجتماعية والخدمات.
5. العلوم الإدارية والأعمال.
6. العلوم التربوية والنفسية.
7. علوم الحاسوب وتكنولوجيا المعلومات.
8. العلوم الرياضية.
9. العلوم الزراعية وتطبيقاتها.
10. العلوم الطبية والطبية المساندة.
11. العلوم الطبيعية الأساسية.
12. علوم الهندسة والتكنولوجيا.
13. الفنون السمعية والبصرية.
14. القانون.

المستوى الدقيق (Micro-level) : نتائج وكفايات التعلم الخاصة بالتخصصات الأكاديمية

يشمل الامتحان في هذا المستوى على جميع النتائج والكفايات للتعلم التي يتوقع أن يكتسبها الطلبة ضمن تخصصاتهم الأكاديمية، وقد تم في آخر دورة عُقدت على الفصل الثاني من العام الجامعي 2015/2014 شمول بعض التخصصات الأكاديمية والبالغ عددها (52) تخصصاً وهي (الاقتصاد واقتصاد المال والإعمال، إدارة مستشفيات، تسويق، إدارة الأعمال، المحاسبة، الإدارة العامة، علوم مالية ومصرفية، نظم معلومات إدارية، أراضي ومياه وبيئة، إنتاج نباتي (بستنة ومحاصيل)، علم وتكنولوجيا الغذاء (صناعات غذائية)، تغذية الإنسان والحميات، اقتصاد زراعي، تنسيق المواقع وإنتاج الأزهار (حدائق)، إنتاج حيواني، وقاية نبات، أمن وحماية الشبكات والمعلومات، هندسة برمجيات، نظم معلومات إدارية، علم الحاسوب، نظم معلومات حاسوبية، القانون، الطب، طب الأسنان، الصيدلة، دكتور صيدلي، تمريض، علوم السمع والنطق، تكنولوجيا الأسنان، الأطراف الاصطناعية والأجهزة، العلاج الوظيفي، العلاج الطبيعي، الأشعة، هندسة معمارية، هندسة كيميائية، هندسة مدنية، هندسة حاسوب، هندسة كهربائية، هندسة صناعية، هندسة ميكانيكية، هندسة ميكاترونيكس، هندسة الإلكترونيات، هندسة الاتصالات، هندسة طبية، هندسة الطيران، علم النفس، الإرشاد النفسي، التربية الخاصة، التربية المهنية، تكنولوجيا التعليم، رياضاً لطفال، معلم صف)، وسيتم في دورات لاحقة شمول باقي التخصصات الأكاديمية في امتحان الكفاءة الجامعية المستوى الدقيق.

ثانياً: تطبيق امتحان الكفاءة الجامعية وتسلسل النتائج

يطبق امتحان الكفاءة الجامعية المستوى العام على جميع البرامج والتخصصات الأكاديمية في الجامعات الرسمية والخاصة، ويطبق امتحان المستوى المتوسط على جميع العوائل والبرامج والبالغ عددها 14 عائلة أو برنامج أكاديمي. ويطبق امتحان الكفاءة الجامعية المستوى الدقيق في آخر دورة على الفصل الثاني من العام الجامعي 2015/2014 على 52 تخصص أكاديمي، ومن المتوقع على بداية الفصل الأول من العام الجامعي 2016/2015 أن يتم وضع كفايات ونتائج لباقي التخصصات الأكاديمية للمستوى الدقيق. والجدول الآتي يبين الفصول الدراسية التي عُقد بها امتحان الكفاءة الجامعية واعداد الطلبة الخريجين المتقدمين للامتحان.

الفصل الدراسي	اعداد الطلبة الخريجين المتقدمين للامتحان	عدد مستويات الامتحان المطبقة في الفصل	عدد الامتحانات المحوسبة	عدد الامتحانات للبرامج الأكاديمية والتخصصات المشمولة في الامتحان
الفصل الثاني من العام الدراسي 2013/2012	8582	2 (عام+متوسط)	17164	4
الفصل الأول من العام الدراسي 2014/2013	7251	2 (عام+متوسط)	14502	8
الفصل الثاني من العام الدراسي 2014/2013	6542	1 متوسط	6542	12
الفصل الأول من العام الدراسي 2015/2014	19762	2 (عام+متوسط)	39524	15
الفصل الثاني من العام الدراسي 2015/2014	42685	3 (عام+متوسط+دقيق)	112748	77

الطلبة المستهدفون

يطبق امتحان الكفاءة الجامعية على جميع طلبة البكالوريوس المتوقع تخرجهم كل فصل دراسي، ويتم إبلاغ الجامعات بأسمائهم قبل اليوم المحدد لجلوسهم للامتحان، إذ لا يتطلب الامتحان أي نوع من الإعداد المسبق، فهو يقيس مهارات وكفايات تراكمية تحققت للطلبة خلال التحاقه بالجامعة.

إجراءات سير امتحان الكفاءة الجامعية

يتم تنفيذ الامتحان بُنيته المقترحة من قبل المركز الوطني للاختبارات التابع لهيئة اعتماد مؤسسات التعليم العالي، بالتعاون مع مؤسسات التعليم العالي. إذ يتم الاستفادة مما يتوفر لهذه الجهات من بنى تحتية وكوادر بشرية في المساهمة في الإشراف على تطبيق الامتحانات، بالتعاون مع الكوادر البشرية العاملة بالهيئة إلى جانب اللجان الفنية للامتحان، ويكون الامتحان محوسباً وله نماذج متعددة، للمساهمة في تسهيل مهمة تطبيقه ومرونتها والتغلب على محاولات الغش، أو يتم عقد الامتحان لجميع المعنيين بالجلوس له في الوقت نفسه.

إجراءات تسلسل سير نتائج امتحان الكفاءة الجامعية.

1. اطلاع مجلس التعليم العالي والبحث العلمي على نتائج امتحان الكفاءة الجامعية.
2. اطلاع مجالس الامناء للجامعات على نتائج امتحان الكفاءة الجامعية.
3. عقد مؤتمر صحفي ونشر نتائج امتحان الكفاءة الجامعية في المواقع الرسمية.
4. تزويد كل مؤسسة من مؤسسات التعليم العالي بتقرير مفصل عن نتائج امتحان الكفاءة الجامعية للكفايات والنتائج الرئيسية والفرعية المتقنة وغير المتقنة.

5. الطلب من الجامعات اعداد خطط تحسينية بناءً على نتائج امتحان الكفاءة الجامعية .
6. حصول الهيئة على تغذية راجعة من الجامعات، من خلال قيام بعض الجامعات بتعديل خططها الدراسية حسب نتائج امتحان الكفاءة، ومراجعة محتويات المواد التي تدرس للطلبة والتأكد من مدى مطابقتها مع وصفها .
7. متابعة نتائج امتحان الكفاءة الجامعية مع الجامعات لدراسة خطط الجامعات التحسينية للوقوف على مدى التطوير والتحسين في مخرجات الجامعة من خلال مقارنة أداء الجامعة مع نفسها على امتحان الكفاءة الجامعية لعدة سنوات .
8. يتم اعتماد على نتائج امتحان الكفاءة الجامعية كمعيار في اعتماد وجود مخرجات البرامج الاكاديمية وزيادة الطاقة الاستيعابية للبرامج والتخصصات الاكاديمية .

التصور الجديد للبنية المقترحة لامتحان الكفاءة الجامعية

إن التصور الحالي لامتحان الكفاءة الجامعية يتمثل في أنه امتحان يركز على ثلاث مستويات هي (المستوى العام، والمستوى المتوسط، والمستوى الدقيق)، إذ يُركز الامتحان بمستوياته الثلاث على نتائج التعلم العامة والخاصة التي يتوقع من الطالب اكتسابها بعد الانتهاء من دراسة البرنامج الأكاديمي الملتحق به، ونظراً لإجراء امتحان الكفاءة الجامعية بمستوياته الثلاث (العام، المتوسط، الدقيق) خلال الفصل الدراسي على ثلاثة أسابيع متتالية، ونظراً لكثرة الامتحانات المحوسبة التي تعقدتها الجامعات خلال الفصل، ولالتزام الطلبة الخريجين بمحاضراتهم اليومية ولكثرة الامتحانات المترتبة عليهم، فإن التصور الجديد المقترح لامتحان الكفاءة الجامعية بأن يصبح امتحاناً واحداً يشمل الكفايات والنتائج للمستويات الثلاث لامتحان الكفاءة الجامعية، بحيث جزء من فقرات الامتحان تغطي كفايات ونتائج المستوى العام، وجزء من فقرات الامتحان تغطي كفايات ونتائج المستوى المتوسط، وجزء من فقرات الامتحان تغطي كفايات ونتائج المستوى الدقيق .

ثالثاً : إنعكاسات امتحان الكفاءة الجامعية على الأنظمة التربوية في مؤسسات التعليم العالي الأردنية .

ساهمت التوجهات الحديثة للتقويم والاعتماد في التركيز على مخرجات التعليم . حيث تُركز على هذا العديد من هيئات الاعتماد الدولية، ويتطلب الاعتماد أن يكون لكل برنامج عملية تقويم ذاتي مستمرة يتم من خلاله تقويم مخرجاته وتطويرها . بالإضافة إلى ذلك ساهمت متطلبات العمل والمنافسة الوظيفية في ضرورة التقويم الدوري للبرامج الأكاديمية وتحسين أدائها بما يتواءم مع المستجدات الحديثة وذلك بالتركيز على تعلم الطلاب وقدراتهم للعمل بعد التخرج . ولم يعد يكتفي الاعتماد على المستوى البرامجي فحسب بل شمل المؤسسة التعليمية بجميع إداراتها التعليمية والبحثية والإدارية والمالية ونشاطاتها الطلابية و خدمة المجتمع .

بناءً على ذلك اهتمت هيئة اعتماد مؤسسات التعليم العالي بضمان جودة المؤسسات التعليمية والبرامج الأكاديمية المطروحة فيها، إيماناً منها على أهمية تأكيد جودة التعليم العالي، وخلق روح التنافس بين البرامج والتخصصات لتحقيق الجودة في مدخلاتها وعملياتها وبالتالي مخرجاتها مما ينعكس إيجابياً على مستوى أداء مؤسسات التعليم العالي الأردنية . وسعي الجامعات إلى الاستفادة من نتائج امتحان الكفاءة الجامعية والوقوف على مدى التطوير والتحسين في مخرجات الجامعة من خلال مقارنة نتائج أداء الجامعة في امتحان الكفاءة الجامعية لعدة سنوات وتوفير مؤشرات كمية ونوعية مستمرة ودقيقة عن أداء القائمين على التخصصات والبرامج التي تقدمها مؤسسات التعليم العالي . حيث يستفيد من امتحان الكفاءة الجامعية كل من :

« الطالب : تغذية راجعة لتحسين عملية التعلم .

« الدكتور : التخطيط للتدريس .

« البرنامج : تغذية راجعة لمراجعة البرنامج وتطويره .

« الجامعة : الجودة وتطوير الخطط والبرامج .

وقد سعت الهيئة الى صياغة اجراء دراسات تحليلية لنتائج امتحان الكفاءة الجامعية، ومن هذه الدراسات دراسة تحليلية للمؤسسات والبرامج المطروحة وخططها الدراسية، ودراسة لواقع المتطلبات الجامعية التي تغطي كفايات ونتائج التعلم للمستوى العام والتي يتوقع من الطالب الخريج امتلاكها لمواكبة متطلبات سوق العمل . ووفقاً للدراسة التحليلية للمؤسسات والبرامج المطروحة وخططها الدراسية فقد تم تحديد المؤشرات الكمية والنوعية في البرنامج من خلال التعرف على المؤشرات التالية :

- ما هي مكونات البرنامج الأكاديمي؟
- نسب الكفايات المختلفة التي تحقق مواصفات الخريج المخططة للبرنامج؟
- هل تم اختيار البرنامج بناء على احتياجات مجتمعية فعلية؟
- هل يوجد توصيف معتمد للبرنامج؟
- هل تم إعداد مصفوفة المعارف والمهارات والكفايات؟
- هل تتوافق المخرجات مع المعايير الأكاديمية المتبناة؟ كيف يتم التأكد؟ ومدى الفعالية والاستفادة ومجالات الاستفادة؟
- هل يوجد توصيف معتمد للمقررات الدراسية؟
- هل يتضمن التوصيف مصفوفة للمعارف والمهارات والكفايات الخاصة بكل مقرر دراسي؟
- هل توصيف المقررات يحقق المعارف والمهارات الأساسية وفقاً للمعايير الأكاديمية للبرنامج؟
- هل تتفق محتويات المقررات مع مخرجات التعلم المستهدفة لهذه المقررات؟
- ما هي مجالات التطوير والممارسات التطبيقية في المقررات الدراسية للبرنامج خلال السنوات الخمس الماضية؟
- هل يتم مراجعة البرنامج دورياً؟
- متى تتم مراجعة البرنامج؟
- هل تم الاستفادة من نتائج المراجعة في تحديث وتطوير البرنامج

وفي ما يلي عرض للملخص دراسة للكفايات الرئيسية للمستوى العام والمساقات او المتطلبات الجامعية ومتطلبات الكليات التي تغطي هذه الكفايات، والنتائج المترتبة عليها .

الكفايات ونتائج التعلم العامة لمؤسسات التعليم العالي

يشمل المستوى العام جميع الكفايات ونتائج التعلم التي يتوقع من أي طالب يتخرج من مؤسسات التعليم العالي أن يمتلكها (بغض النظر عن كليته)، حتى يتماشى مع متطلبات الحياة المهنية والاجتماعية ومتطلبات سوق العمل، والإنخراط في عملية التطوير والتعلم المستمر مدى الحياة، وهذه الكفايات العامة هي :

1. مهارات البحث العلمي وتفسير النتائج .
2. مهارات التفكير الناقد .
3. مهارات التواصل .
4. مهارات الثقافة المعلوماتية .
5. مهارات حل المشكلات والعمل في فريق .
6. مهارات المعرفة بالحقوق والواجبات .
7. المعرفة بالقواعد والسلوكيات الأخلاقية .
8. مهارات المعالجة الرياضية .
9. مهارات المعالجة اللفظية .

والجدول التالي يبين الكفايات الرئيسية للمستوى العام والمساقات او المتطلبات الجامعية ومتطلبات الكليات التي تغطي هذه الكفايات .

الكفايات الرئيسية للمستوى العام ومتطلبات الجامعة والكلية التي تشمل أو تغطي هذه الكفايات

الكفاية	متطلبات الجامعة	متطلبات الكليات	ملاحظات
مهارات البحث العلمي وتفسير النتائج.	لا يوجد متطلب محدد	بحث علمي، مناهج البحث العلمي، تنمية المفاهيم العلمية، مبادئ البحث العلمي، طرق البحث العلمي، التفكير العلمي، اساليب البحث العلمي، اساسيات البحث العلمي، مقدمة في البحث التربوي والاحصاء، حلقة بحث في الترجمة، مناهج البحث والاحصاء الاجتماعي،.....	اغلب الكليات في الجامعات تشمل على متطلبات للكلية تغطي كفاية مهارات البحث العلمي وتفسير النتائج.
مهارات التفكير الناقد.	لا يوجد متطلب محدد	التفكير الدقيق، تفكير علمي، التفكير والابداع، طرق التفكير، تعليم التفكير، التفكير الابداعي	مساقات مهارات الاتصال تغطي كفاية مهارة التفكير الناقد، بلاضافة فانه يوجد في العديد من الكليات متطلبات تغطي هذه الكفاية.
مهارات التواصل .	لا يوجد متطلب محدد		جميع الجامعات الرسمية والخاصة تدرس هذه المساقات
مهارات الثقافة المعلوماتية.	التربية الوطنية		جميع الجامعات الرسمية والخاصة تدرس هذا المساق
مهارات حل المشكلات والعمل في فريق.	لا يوجد متطلب محدد		جميع الجامعات الرسمية والخاصة تدرس هذه المساقات
مهارات المعرفة بالحقوق والواجبات.	التربية الوطنية		جميع الجامعات الرسمية والخاصة تدرس هذا المساق
المعرفة بالقواعد والسلوكيات الأخلاقية.	التربية الوطنية		جميع الجامعات الرسمية والخاصة تدرس هذا المساق
مهارات المعالجة الرياضية.	لا يوجد متطلب محدد		اغلب متطلبات الكليات ومتطلبات التخصصات للكليات العلمية والانسانية تغطي هذه الكفاية.
مهارات المعالجة اللفظية	مهارات اتصال (لغة عربية 101 و102)، حيث تشتمل هذه المساقات على مهارة (محادثة، تواصل، استيعاب، قراءة، كتابة).		جميع الجامعات الرسمية والخاصة تدرس هذا المساق

نتائج امتحان الكفاءة الجامعية في آخر 3 دورات .
الجدول الآتي يبين المتوسطات الحسابية للاجابات الصحيحة لامتحان الكفاءة الجامعية المستوى العام المنعقد في آخر 3 دورات لكل كفاية .

مقارنة نتائج إتقان الكفايات والنتائج في امتحان المستوى العام بين الفصل الدراسي الأول من العام الجامعي 2014/2013 والفصل الدراسي الأول والثاني من العام الجامعي 2015/2014.			
الكفايات	الوسط الحسابي للاجابات الصحيحة للفصل الدراسي الأول 2014/2013	الوسط الحسابي للاجابات الصحيحة للفصل الدراسي الأول 2015/2014	الوسط الحسابي للاجابات الصحيحة للفصل الدراسي الثاني 2015/2014
البحث العلمي وتفسير النتائج	46%	48.29%	43.74%
التفكير الناقد	42%	43.60%	52.50%
الثقافة المعلوماتية	44%	43.70%	68.03%
المعرفة بالحقوق والواجبات	53%	50.78%	62.91%
المعرفة بالقواعد والسلوكيات الأخلاقية	45%	43.00%	70.52%
المعالجة الرياضية	42%	42.57%	46.20%
المعالجة اللفظية	45%	47.84%	51.03%
مهارات التواصل	22%	21.88%	57.27%
مهارات حل المشكلة والعمل في فريق	49%	47.00%	57.29%
النسبة العامة للإتقان في المستوى العام	44%	45.00%	56.28%

تبين من خلال نتائج امتحان الكفاءة الجامعية وجود تحسن وتطور في نسبة إتقان الكفايات والمهارات ونتائج التعلم في الفصل الدراسي الثاني من العام الجامعي 2015/2014 مقارنة مع الفصول السابقة ويمكن تفسير هذه الاختلافات بدلالة التغير بتعليمات الامتحان حيث أصبح الزامي لجميع الطلبة المتوقع تخرجهم ويتضح ذلك بأن أداء الطلبة أصبح أكثر جدية ويتأكد ذلك من خلال ارتفاع نسبة الحضور وكذلك من خلال انخفاض نسبة الطلبة الغير الجادين في الامتحان ويمكن أن يعود ذلك الى تغير بنية الامتحان والمحافظة على كفايات الامتحان . وبناءً على نتائج امتحان الكفاءة الجامعية فقد اخذت الجامعات تغذية راجعة من الفصول السابقة وقامت بمراجعة محتويات المواد التي تدرس للطلبة والتأكد من مدى مطابقتها مع وصفها، وقامت بتطوير الخطط التدريسية واعادة النظر بمساقات متطلبات الجامعة للمساقات الدراسية وطرح مواد اختيارية مختلفة في جميع الاقسام، وعقد لقاءات مع الطلبة لإظهار وتوضيح أهمية مهارات وكفايات امتحان الكفاءة الجامعية في حياتهم الوظيفية العملية .

التحسينات المقترحة للبرامج الأكاديمية من قبل مؤسسات التعليم العالي بناءً على نتائج امتحان الكفاءة الجامعية .

تقوم هيئة اعتماد مؤسسات التعليم العالي بتزويد كل مؤسسة من مؤسسات التعليم العالي بتقرير مفصل عن نتائج امتحان الكفاءة الجامعية، وتتابع الهيئة نتائج امتحان الكفاءة الجامعية مع الجامعات لدراسة خطط الجامعات التحسينية للوقوف على مدى التطوير والتحسين في مخرجات الجامعة من خلال مقارنة أداء الجامعة مع نفسها على امتحان الكفاءة الجامعية لعدة سنوات، ويتم اعتماد نتائج امتحان الكفاءة الجامعية كمعيار في اعتماد وجودة مخرجات البرامج الأكاديمية وزيادة الطاقة الاستيعابية للبرامج والتخصصات الأكاديمية . ومن الخطط التحسينية المقترحة للبرامج الأكاديمية ما يلي :

1. تطوير الخطط التدريسية واعادة النظر بمساقات متطلبات الجامعة للمساقات الدراسية وطرح مواد اختيارية مختلفة في جميع الاقسام .
2. ادراج الكفايات والمهارات المقاسة في امتحان الكفاءة الجامعية ضمن الاختبارات الشهرية في التخصصات ذات العلاقة .
3. عقد لقاءات مع الطلبة لإظهار وتوضيح أهمية مهارات وكفايات امتحان الكفاءة الجامعية في حياتهم الوظيفية العملية .
4. مراجعة محتويات المواد التي تدرس للطلبة والتأكد من مدى مطابقتها مع وصفها .
5. اعطاء طلبة الجامعة حرية التعبير والحوار لتشجيع التفكير الابداعي وتعزيز المواهب .
6. توفير بيئة جامعية غنية بالتجارب والوسائل التعليمية تؤدي الى ابداع اكثر واعطاء فرصة للمشاركة بنشاطات وفعاليات مختلفة من اختيارهم من خلال توفير المختبرات والاندية والمراكز في الجامعة .
7. استخدام طرق متعددة للتدريس . لان تنوع طرق التدريس واساليبها يوجه ويراعي الفروق الفردية بين الطلبة .
8. اطلاع الطلبة على الاعمال الابداعية التي يقوم بها زملائهم لان ذلك يبني عندهم الثقة والاعتزاز بالنفس لمواصلة النجاح وتحفيز مهارات التفكير الابداعي .
9. تعزيز وتطوير مساقات التدريب الميداني واعادة النظر بمحتواها واسلوب متابعتها وطرق تقييم الطالب فيها .
10. تعزيز فرص تبادل التجارب والخبرات بين اعضاء الهيئة التدريسية في القسم من حيث الابحاث والمشاريع المشتركة واعداد المحاضرات .
11. عقد دورات لتطوير اداء اعضاء الهيئة التدريسية في مختلف المهارات المتعلقة بعملية التدريس وادارتها .
12. التواصل مع شعبة الخريجين في الجامعة لمتابعة شؤون الخريجين للتأكد من مدى ملائمة مخرجات التعليم لسوق العمل
13. وضع خطط عمل في الاقسام والعمادات لتعزيز النشاطات اللامنهجية في المساقات .
14. تطوير استراتيجيات التقويم في المساقات للتضمن طرق اخرى غير الاختبارات قائمة على الواجبات والمشاريع والنشاطات اللاصفية
15. تعزيز مفهوم المواطنة والحس الوطني ووضع سبل واجراءات للمحافظة على المنظومة في مؤسسات التعليم العالي .

16. وضع منظومة من المعايير لضمان جودة التعليم العالي في الجامعة بما يتناسب مع معايير اتحاد الجامعات العربية في جودة التعليم العالي .

التوصيات بناءً على تحليل نتائج امتحان الكفاءة الجامعية.

- 1.** بناءً على تحليل نتائج امتحان الكفاءة الجامعية تبين بأن غالبية الطلبة في الجامعات الرسمية والخاصة تعاني من ضعف في مهارة البحث العلمي وتفسير النتائج والمعالجة العددية والتفكير الناقد ومهارات التواصل والمعالجة اللفظية، ويعود السبب في ذلك لعدم وجود متطلبات جامعية ومتطلبات كلية تُطرح في الجامعات تغطي هذه المهارات، وعليه يمكن التوصية الى الجامعات بإضافة متطلبات عام للجامعة تحت مسمى مهارات التفكير والبحث العلمي (يشمل : أنماط التفكير والبحث العلمي وحل المشكلات ومهارات الاتصال والتواصل) .
- 2.** التوصية للجامعات الاردنية الرسمية منها والخاصة بتطوير الخطط التدريسية واعادة النظر بالمساقات الدراسية وطرح مواد اختيارية مختلفة في جميع الاقسام تغطي كفايات المستوى العام
- 3.** عقد لقاءات مع الطلبة لإظهار وتوضيح اهمية مهارات وكفايات امتحان الكفاءة الجامعية في حياتهم الوظيفية العملية .
- 4.** تطوير الخطة الدراسية لمتطلب التربية الوطنية، وذلك لتغطية هذا المتطلب العديد من كفايات ومهارات ونتائج المستوى العام، مثل : كفاية الثقافة المعلوماتية، والمعرفة بالحقوق والواجبات، والمعرفة بالقواعد والسلوكات الاخلاقية.

**CONSEIL SUPÉRIEUR DE L'ÉDUCATION
DE LA FORMATION ET DE LA RECHERCHE SCIENTIFIQUE**
Angle Avenues Al Melia et Allal El Fassi, Hay Riad, Rabat - B.P. 6535



www.csefrs.ma