

Pluralités culturelles et universalité des mathématiques :  
enjeux et perspectives pour leur enseignement  
et leur apprentissage

espace mathématique francophone  
Alger : 10-14 Octobre 2015



## ÉVALUATIONS NATIONALES ET INTERNATIONALES EN MATHÉMATIQUES : QUELLES ANALYSES DIDACTIQUES ?

Compte-rendu du Projet Spécial n°4

Éric RODITI\* - Caroline BARDINI\*\* - Claire VAUGELADE BERG\*\*\*

### I. BILAN ET PERSPECTIVES

Le projet spécial n°4 de la rencontre EMF 2015 à Alger a été principalement consacré aux analyses didactiques des questions – et aux questionnaires – d'évaluation. Ce projet s'inscrivait dans le prolongement des deux projets spéciaux consacrés à l'évaluation lors des rencontres EMF 2009 et 2012 qui ont respectivement développé la question des impacts politiques et curriculaires des évaluations internationales puis celle du lien entre évaluation, compétence et orientation dans les transitions entre les cycles d'enseignement. En 2015, c'est la question du regard didactique sur l'évaluation externe qui a été posée, et plus particulièrement sur l'évaluation standardisée, nationale ou internationale.

Trois types de préoccupations émergent à l'issue de la rencontre. Le premier type de préoccupations porte sur la validité des questionnaires d'évaluation c'est-à-dire sur le fait que les questionnaires aboutissent bien à rendre compte de ce qu'on veut observer. Le second porte sur les critères d'évaluation de l'activité mathématique des élèves interrogés dans les enquêtes internationales comme PISA et leur effet sur ce que ces enquêtes donnent à voir des compétences des élèves. Le dernier concerne ce qu'à l'échelle nationale, un pays peut mettre en œuvre pour piloter l'enseignement des mathématiques en fonction des données qu'il parvient à recueillir sur les acquis des élèves.

Concernant la question de la validité des questionnaires d'évaluation, le projet spécial s'est particulièrement nourri d'une étude concernant la France et ses outils d'évaluation des acquis des élèves dans le domaine numérique. La méthodologie d'analyse convoquait différentes approches disciplinaires – épistémologique, psycho-didactique et psychométrique – et les résultats obtenus ont montré la richesse et l'intérêt de croiser ces différentes approches. L'analyse épistémologique et didactique permet de contrôler le contenu de l'évaluation au

\* Université Paris Descartes – France – [eric.roditi@paris5.sorbonne.fr](mailto:eric.roditi@paris5.sorbonne.fr)

\*\* Melbourne Graduate School of Education – Australie – [caroline.bardini@unimelb.edu.au](mailto:caroline.bardini@unimelb.edu.au)

\*\*\* University of Agder – Norvège – [claire.v.berg@uia.no](mailto:claire.v.berg@uia.no)

niveau local (pour chaque tâche d'évaluation au sein du questionnaire) et global (sur le recouvrement et la variété des compétences évaluées à travers les différentes tâches du questionnaire). Il apparaît toutefois important pour la didactique des mathématiques de croiser, d'une part, ce qui apparaît comme relevant de la complexité de la tâche et qui est déterminée par l'analyse *a priori*, avec, d'autre part, l'indice de difficulté de la tâche rapporté par le taux de réussite. D'éventuelles incohérences sont autant de pistes pour une compréhension didactique renouvelée des manières dont les élèves mettent en œuvre leurs connaissances mathématiques pour résoudre un problème. Des questions similaires peuvent aussi se poser sur des items qui pourraient sembler pertinents d'un point de vue didactique, mais qui ne se révéleraient pas discriminants, c'est-à-dire qui seraient massivement réussis ou échoués. Dans tous les cas, l'interprétation d'écarts nécessite une analyse épistémologique et psycho-didactique des items à conjuguer avec des éléments relatifs aux pratiques enseignantes. On comprend ainsi qu'il ne s'agit ni d'opposer les approches ni de les mettre en œuvre de manière successive, mais bien de points de vue complémentaires à conjuguer qui renseignent différemment sur le contenu des évaluations et sur leurs résultats.

La question des critères d'évaluation des activités des élèves dans les enquêtes a complété la question précédente, en portant le regard à la fois sur les items et les démarches potentiellement mises en œuvre par les élèves pour les traiter. Les enquêtes du PISA visent un suivi des acquis scolaires des élèves de 15 ans, et, en ce qui concerne ceux de la culture mathématique, le choix de l'OCDE est d'évaluer des compétences, c'est-à-dire des capacités à mobiliser ses connaissances pour résoudre un problème en lien avec une situation de la vie réelle. Le regard didactique porté sur l'évaluation de 2012, et qui a été développé dans le projet spécial n°4 d'EMF 2015, ne peut manquer de pointer que l'OCDE ne se donne les moyens ni de recenser précisément les connaissances acquises des élèves ni d'estimer le niveau d'acquisition de ces connaissances. Une classification des items par des catégories issues de la recherche en didactique a en effet permis de distinguer les activités potentielles des élèves suivant différents « niveaux de mise en fonctionnement des connaissances mathématiques » : mise en fonctionnement directe dans des configurations usuelles ; mise en œuvre nécessitant une adaptation des données (conversion d'unités par exemple) ou des questions (changement de point de vue) ; mise en fonctionnement nécessitant l'introduction d'intermédiaires (changement de variable, introduction d'un objet géométrique supplémentaire, questions intermédiaires, etc.) ; ou enfin mise en œuvre de la compréhension qualitative d'un concept (distinction entre aire et périmètre, indépendance d'événements aléatoires, etc.). Ces niveaux permettent, d'une certaine manière, d'évaluer le niveau d'acquisition de ces connaissances. Cette nouvelle classification permet ainsi de différencier des items que les catégories définies par les experts de l'OCDE ne permettent pas de distinguer, qui requièrent pourtant des niveaux différents de mise en fonctionnement des connaissances évaluées et qui conduisent à des scores de réussite significativement différents. L'étude complète de l'ensemble des items du PISA menée à l'aune de cette nouvelle classification montre que les trois premiers niveaux de mise en fonctionnement des connaissances, qui correspondent à une exigence croissante de richesse et d'autonomie de l'activité, correspondent également, en moyenne, à un niveau de difficulté croissant pour les élèves. L'étude révèle enfin des éléments nouveaux concernant, pour le cas de la France, les inégalités de performances selon le sexe, l'origine sociale ou le retard scolaire. L'analyse s'appuyant sur la classification didactique montre que les filles sont d'autant plus pénalisées que les tâches demandent de l'initiative, mais que les élèves de milieux populaires comme les élèves en retard scolaire, ne sont pas mis davantage en difficulté lorsque les activités attendues d'eux sont plus exigeantes. Cette étude, comme la précédente, montre que le croisement d'approches – ici didactique et évaluative – sur les apprentissages scolaires en

mathématiques peuvent être fructueux, tant concernant les résultats obtenus que concernant la recherche en didactique.

Les questions de pilotage national, au vu des enquêtes évaluatives portant sur les acquis des élèves, ont été particulièrement développées en s'appuyant sur l'exemple d'une étude concernant les résultats obtenus par les élèves marocains de certains niveaux des cycles primaire et secondaire collégial dans des évaluations nationales et internationales. L'étude a permis de donner quelques interprétations quant à la faiblesse des performances constatées en mathématiques qui sont très inférieures à la moyenne internationale. L'étude se base sur une analyse de la réalité actuelle de l'enseignement marocain des mathématiques et sur l'analyse des rapports du Conseil Supérieur de l'Enseignement et de ses recommandations. Elle permet ainsi de souligner l'impact de certains facteurs institutionnels, comme ceux liés au fonctionnement du système éducatif, sur la qualité de l'enseignement des mathématiques. Ainsi, la discussion a porté sur un problème qui concerne le Maroc mais qui dépasse ce cas particulier, il s'agit de l'adéquation d'une évaluation homogène sur un territoire national alors qu'on constate une hétérogénéité territoriale du système éducatif. De manière générale, les échanges ont porté sur l'interprétation des acquis des élèves s'appuyant sur différentes facettes de la gestion du système éducatif et, au bout du compte, sur la question du pilotage du système éducatif par les résultats portant sur les acquis des élèves. Trois préoccupations majeures ont émergé de l'étude et des discussions tenues dans le projet spécial n°4 d'EMF 2015 : celle de la formation des enseignants de mathématiques, celle d'un socle de connaissances à faire acquérir à tous les élèves, et celle de l'évaluation comme moyen de régulation de l'enseignement.

Les participants au projet spécial ont maintes fois soulevé l'intérêt qu'il y aurait à développer, parallèlement aux études sur les évaluations externes nationales et internationales, des recherches portant sur l'évaluation des apprentissages dans le quotidien de la classe de mathématiques. Sans doute un espace à ouvrir pour le projet spécial « évaluation » du prochain EMF...

Pluralités culturelles et universalité des mathématiques :  
enjeux et perspectives pour leur enseignement  
et leur apprentissage

espace mathématique francophone  
Alger : 10-14 Octobre 2015



## MÉTHODOLOGIE D'ANALYSE D'ÉVALUATIONS EXTERNES

Brigitte GRUGEON-ALLYS\* – Nadine GRAPIN\*\*

**Résumé** – En France, des évaluations bilans réalisées sur échantillon sont menées tous les six ans en mathématiques, en fin d'école et de collège, pour déterminer les acquis des élèves. Après avoir présenté ce dispositif spécifique, nous présentons une méthodologie d'analyse des évaluations s'appuyant sur trois approches (épistémologique-didactique, psycho-didactique et psychométrique) et nous montrons comment ces dernières se révèlent complémentaires pour garantir la validité des résultats produits. Enfin, nous utilisons cette méthodologie pour étudier les évaluations bilans 2008 et 2014 sur deux domaines : l'arithmétique des nombres entiers en fin d'école et l'algèbre en fin de collège.

**Mots-clefs** : évaluation standardisée, validité, épistémologique-didactique, psycho-didactique, psychométrie.

**Abstract** – In France, assessments realised on sample are conducted every six years in mathematics, at the end of school and college, to determine students' knowledge. After presenting this specific device, we present an assessment analysis methodology founded on three approaches (epistemo-didactic, psycho-didactic and psychometric) and show how these prove complementary to ensure the validity of the results. Finally, we use this methodology to study the results of assessments 2008 and 2014 on two areas: arithmetic of whole numbers at the the end of school and algebra at the end of college.

**Keywords**: standardized assessment, validity, epistemo-educational, psycho-educational, psychometrics

En France, en parallèle des enquêtes internationales PISA et TIMSS, des évaluations standardisées sont menées régulièrement nationalement, sur échantillon ou sur la totalité des élèves d'un niveau scolaire donné, par l'intermédiaire de la DEPP (Direction de l'évaluation, de la performance et de la prospective) (Roditi & Chesne 2012). Alors que les résultats des élèves à ces évaluations sont utilisés pour réguler les programmes d'enseignement, voire pour orienter les politiques éducatives (Mons 2009, p.7) la didactique, en particulier la didactique des mathématiques, s'empare davantage des questions d'évaluation telles que – l'évaluation diagnostique et la régulation (Grugeon 1997, Grugeon-Allys 2012, Pilet 2012), l'évaluation de compétences (Winslow 2005 ou Schneider 2006), l'évaluation formative (El Hage & al. 2014), la comparaison des évaluations nationales ou internationales (Artigue & Winslow 2010 ou Ruminot-Vergara 2014), les perspectives à partir de données récoltées sur ces épreuves réalisées à grande échelle (Deblois, Freiman & Rousseau 2007), l'impact des évaluations standardisées sur les pratiques des enseignants (Ruminot-Vergara 2014<sup>340</sup>).

Dans la continuité des travaux de Bodin (1997), nous cherchons à étudier la validité d'un test, ce qui revient à déterminer si un test mesure effectivement ce pour quoi il a été construit et uniquement cela. Cette question apparaît comme essentielle pour les concepteurs de test et

\* Université Paris Est Créteil – ESPE de Créteil – France – brigitte.grugeon-allys@u-pec.fr

\*\* Université Paris Est Créteil – ESPE de Créteil – France – nadine.grapin@u-pec.fr

<sup>340</sup> Étude menée à partir de l'évaluation SIMCE au Chili

de nombreux travaux en psychométrie ont été menés autour de ce concept (Grégoire & Laveault 2014 p. 163), mais ici nous nous centrons principalement sur la validité de contenu qui « se réfère à la qualité des questions de l'épreuve » (De Landsheere, 1998) c'est-à-dire la représentativité des questions relativement au domaine mathématique évalué. Nous fondons notre travail sur des approches didactique, prenant en compte des aspects épistémologique et, psychologique, et psychométrique en nous situant à deux niveaux de granularité - global sur l'ensemble des items d'un domaine puis local pour chaque item. Afin d'éclairer pourquoi ces différentes approches sont nécessaires et complémentaires, nous illustrons notre questionnement en présentant d'abord un dispositif d'évaluation nationale, le bilan CEDRE (Cycle des évaluations disciplinaires réalisées sur échantillon) puis nous explicitons la méthodologie d'analyse de la validité que nous avons établie et enfin, nous l'appliquons à deux domaines distincts : celui de l'arithmétique des entiers en fin d'école et celui de l'algèbre en fin du collège.

## II. PRESENTATION DU DISPOSITIF D'EVALUATION ET QUESTIONNEMENT

### 1. *Dispositif de conception des épreuves*

En France, les acquis des élèves en mathématiques à la fin de l'école primaire (élèves âgés de 10 ans) et du collège (élèves âgés de 15 ans) sont en partie évalués par le bilan CEDRE. Cette évaluation externe est renouvelée tous les 6 ans (2008 - 2014 pour les mathématiques) et permet par conséquent une comparaison temporelle des apprentissages des élèves. Il ne s'agit pas d'établir un diagnostic des difficultés des élèves, ni de comparer des groupes d'élèves selon les établissements, mais d'établir un bilan des connaissances et des compétences des élèves au regard des programmes scolaires en vigueur (Lescure & Pastor 2012).

La conception des items et l'analyse des résultats sont menées par des personnels de l'enseignement qui définissent des items, la forme des questions et le codage des réponses puis les sélectionnent et par des statisticiens qui calculent différents indicateurs psychométriques. Selon les enjeux assignés à l'évaluation, les items du test sont répartis selon les différents domaines définis par les programmes : 385 items répartis sur 6 domaines pour le primaire (organisation des données numériques, résolution de problèmes, nombres et calcul, espace et géométrie, grandeurs et mesures) et 172 items répartis sur 5 domaines pour le collège (géométrie, nombres et calculs, organisation de données & fonctions, grandeurs & mesures). Même s'il a été demandé aux concepteurs de produire des items de difficulté variée (Brun & Huguet 2012), aucun outil d'analyse didactique ne leur a été fourni pour ce faire.

Enfin, trois formes de questions sont proposées : des questions à choix multiples (QCM) avec le plus souvent quatre choix de réponse, des Vrai-Faux et des questions ouvertes. La correction des QCM et des Vrai-Faux est faite automatiquement alors que la plupart des questions ouvertes sont corrigées par des enseignants.

Dans le dispositif CEDRE, et comme pour les évaluations internationales PISA et TIMSS, une pré-expérimentation a lieu l'année précédant l'expérimentation définitive, qui permettra aux concepteurs de choisir les items retenus définitivement pour l'évaluation. Enfin, comme il est impossible que chaque élève passe la totalité du test, un principe de cahiers tournants est mis en place (Brun & Huguet 2008) ; l'étude psychométrique rend possible d'estimer, pour tous les items du test, la probabilité de réussite de chaque élève, même s'il n'a pas été confronté à toutes les questions.

## 2. *Éléments de méthodologie statistique*

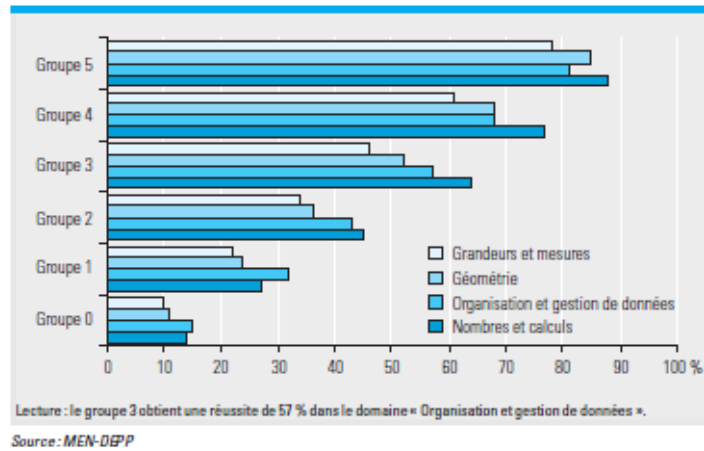
La méthodologie employée pour recueillir les résultats des élèves dans CEDRE est similaire à celle des évaluations internationales, telles que PISA. Sans revenir sur les différents modèles de mesure (théorie classique des scores et modèles de réponse à l'item) détaillés dans différents ouvrages de synthèse, tels que Grégoire et Laveault (2014, pp. 281-304) ou de façon plus synthétique dans Bottani & Vrignaud (2005, pp. 93-105), nous précisons que le modèle statistique utilisé est le modèle de réponse à l'item (MRI). Il permet la réalisation d'échelles de scores et de comparaisons temporelles (Lescure & Pastor 2012, Brun & Huguet 2008), mais repose sur deux contraintes fortes : l'unidimensionnalité du test et l'indépendance locale des items (Bottani & Vrignaud 2005 p. 100).

En psychométrie, un item est principalement caractérisé par son indice de difficulté (calculé *a posteriori* en fonction du nombre d'élèves qui l'ont réussi) et par son indice de discrimination. Les items de l'évaluation CEDRE étant codés 0 ou 1 selon l'échec ou la réussite, l'indice de difficulté correspond à la proportion des élèves qui réussissent l'item et par conséquent est compris entre 0 et 1 (Grégoire & Laveault 2014, p. 204) ; le score d'un élève correspond au nombre d'items réussis par cet élève. La discrimination de l'item est déterminée par la valeur du coefficient bisérial ( $r_{bis}$ ), défini comme « le coefficient de corrélation linéaire entre le score et une variable latente (ici, la compétence mathématique), régie par une loi normale, conditionnant la réussite à l'item » (Megherbi & al. 2009). Cet indice renseigne sur le « pouvoir discriminant » de l'item, c'est à dire « dans quelle mesure l'item s'inscrit dans la dimension générale (corrélation item-test). Il indique également la différence de performance constatée entre les individus qui réussissent l'item et ceux qui l'échouent. » (ibid, p.72). Dans des évaluations telles que CEDRE, où les résultats conduisent à la construction d'échelles, il est important de différencier entre eux les scores, et par conséquent d'avoir des items qui ont un fort pouvoir de discrimination. Par conséquent, à la suite de la pré-expérimentation, certains items sont écartés en vue de l'évaluation définitive puisqu'ils ne sont pas suffisamment discriminants (par rapport à la compétence mesurée) et n'apportent pas d'informations supplémentaires par rapport aux autres.

## 3. *Résultats produits*

Notons d'abord que les réponses des élèves ne sont pas analysées pour révéler des difficultés spécifiques (ce n'est pas l'objectif de l'évaluation CEDRE), mais uniquement traitées en tant que réponse correcte ou incorrecte ; à la différence d'évaluation telles que PISA, aucun crédit partiel n'est accordé pour des réponses à des questions ouvertes.

La mesure de la performance des élèves produite à la suite du test par le MRI permet de déterminer les compétences maîtrisées par les élèves ; ce qui se traduit pour le bilan CEDRE, par une échelle de performance qui répartit les élèves en six groupes (de 0 à 5, du plus faible au plus performant). Chacun des groupes est ensuite caractérisé par un niveau de compétence (globalement sur l'ensemble du test, mais aussi par domaine), les élèves d'un groupe maîtrisant aussi les compétences définies dans les groupes de niveaux inférieurs (Brun & Huguet 2008, Lescure & Pastor 2012). A titre d'exemple, nous présentons ci-dessous un premier graphique (Figure 1) qui représente, pour le collège, les taux de réussite par domaine suivant les différents groupes.



**Figure 1** - Groupes et taux de réussite à chacun des quatre champs (extrait de Brun & Huguet 2008)

Si la lecture directe de ce graphique permet de comparer selon les domaines, les pourcentages de réussite et de constater par exemple que pour les groupes 3-4-5, la réussite est la plus importante dans le domaine « nombres et calcul » alors que pour les groupes de plus bas niveaux, ils réussissent davantage dans le domaine « organisation et gestion de données », la question fondamentale reste posée : quelles sont les connaissances que possèdent les élèves de ces groupes ? à partir de quels items l'évaluation de ces connaissances a-t-elle été faite ?

#### 4. Questionnement

Comme nous l'avons illustré précédemment à partir de la conception des évaluations CEDRE, la détermination des compétences des élèves dans de telles évaluations repose majoritairement sur une approche psychométrique (Bottani & Vrignaud 2005, p. 101), mais comment garantir que le test est valide ? Comment s'assurer que l'ensemble des tâches permet effectivement d'évaluer les connaissances des élèves au regard des programmes ?

Il s'agit ainsi de rechercher des preuves de la validité du test en interrogeant le contenu (les items sont-ils représentatifs du domaine évalué ? Couvrent-ils le domaine évalué au regard des programmes ? ) ou en étudiant les processus de réponse afin de s'assurer que les élèves qui répondent mobilisent bien les connaissances qui sont supposées être évaluées et que chaque item possède les caractéristiques psychométriques demandées. Ces trois conditions correspondent aux trois approches qui fondent notre travail :

##### 1. préalablement à la conception des items

- approche épistémo-didactique : étude de la représentativité des items et de la couverture du domaine mathématique concerné au filtre de l'épistémologie et de la didactique des mathématiques relativement au domaine (types de tâche proposés, modalités de réponse, codage des réponses de l'élève en fonction du type de raisonnement,...),
- approche psycho-didactique : analyse des processus de réponses mis en jeu par les élèves pour produire une réponse en lien avec la forme des énoncés et des variables extra-mathématiques (contexte, découpage de l'énoncé, ..),

##### 2. a posteriori à partir des réponses des élèves

- approche psychométrique : calcul des caractéristiques statistiques des items et production de l'échelle des scores et des groupes selon leur niveau de compétence.

La question du choix des tâches dans une évaluation a déjà été soulevée par Bodin (2006), mais peut difficilement être traitée de façon effective à partir de l'analyse d'évaluations externes existantes. En effet, même si les évaluations internationales telles que PISA et TIMSS sont conçues à partir d'un cadre reposant sur des travaux en didactique, et que nous supposons que les groupes d'experts de ces évaluations s'assurent de la représentativité des tâches choisies, les chercheurs n'ont pas accès à l'ensemble des items des évaluations et ne peuvent par conséquent établir une analyse didactique de l'ensemble des tâches proposées. Deux recherches ont néanmoins été menées en ce sens : celle de Roditi & Salles (2015) sur l'évaluation PISA 2012 ou celle de Sayac & Grapin (2014) sur l'analyse du bilan CEDRE 2008 fin d'école à partir de « facteurs de complexité et de compétences ».

Pour notre recherche, nous avons eu l'opportunité d'avoir accès à l'ensemble du dispositif des évaluations CEDRE fin d'école et fin de collège de 2008 et de 2014, et nous présentons dans le paragraphe suivant la méthodologie que nous avons adoptée, puis quelques résultats sur l'analyse de ces évaluations.

### III. METHODOLOGIE D'ANALYSE

#### *1. Complémentarité des trois approches : un point de vue global et local sur le test*

L'enjeu de CEDRE étant d'évaluer les acquis des élèves au regard des savoirs visés dans les différents domaines mathématiques dans le cadre des programmes scolaires, l'ensemble des items du test doit « couvrir » les différents aspects épistémologiques des savoirs décrits dans les programmes ; par conséquent, un premier niveau d'analyse concerne la pertinence des tâches proposées dans chaque domaine mathématique et leur couverture du domaine (validité épistémologique). Pour ce faire, nous avons situé notre travail dans le cadre de la Théorie Anthropologique du Didactique (Chevallard 1999). A partir des tâches proposées dans les évaluations, il s'agit de caractériser les connaissances apprises des élèves en relation avec les savoirs à enseigner et enseignés mis en jeu dans les différentes étapes de la transposition didactique. En effet, quelle validité accorder aux inférences faites à partir des scores si les tâches présentes dans l'évaluation ne sont pas pertinentes ou « représentatives » de celles relevant du domaine évalué, au regard des programmes et de ce qui est enseigné en classe ?

De ce fait, au delà d'une étude cognitive, nous fondons l'analyse des connaissances apprises par les élèves, dans un domaine donné, sur la caractérisation d'une référence épistémologique (Bosch & Gascon 2005). La définition de cette dernière repose sur une analyse épistémologique des « savoirs savants », relativement au domaine mathématique concerné, ici, respectivement en algèbre en fin de collège (Grugeon 1997) et aux nombres entiers en fin d'école (en particulier Tempier 2013); elle conduit, relativement au domaine étudié, à une description des types de tâches, des techniques permettant de les traiter et des éléments technologiques et théoriques justifiant les techniques. Nous prenons aussi en compte, dans l'analyse des tâches, leur complexité relativement au niveau d'enseignement auquel l'évaluation a lieu ainsi que la congruence sémantique des registres de représentation sémiotique (Duval 1996).

Par conséquent, au regard de la référence, relativement au domaine considéré, nous pouvons analyser la validité d'un test à partir de preuves basées sur le contenu d'une évaluation par une étude « globale » du contenu à savoir : la représentativité des types de tâches et la couverture du domaine selon les tâches proposées dans l'évaluation, la complexité des tâches à partir de l'étude des valeurs des variables didactiques en jeu et du niveau d'intervention du type de tâche et des savoirs en jeu (Castela 2008), les différents registres sémiotiques mis en jeu.



Ce point de vue épistémo-didactique, sur la totalité des items relevant d'un domaine donné, permet d'apporter *a priori* des éléments relatifs à la validité de contenu. Mais au-delà, quelque soit la tâche proposée, il faut s'assurer, en situation, qu'elle permet bien d'évaluer ce qu'elle est sensée évaluer, lorsque les élèves vont la résoudre, et en particulier la nature du raisonnement mobilisé au regard du niveau scolaire visé. Pour ceci, il est nécessaire de s'assurer que l'élève mobilisera bien les savoirs attendus, ce qui dépend aussi d'autres paramètres liés en particulier à l'énoncé.

Développée dans le cadre des approches psycho-didactiques en évaluation, la validité d'un test, qualifiée de « psycho-didactique » Vantourout et Goasdoué (2015) s'appuie sur des preuves basées sur le « processus de réponse de l'évalué, avec le souci de comprendre avant tout son fonctionnement cognitif ». Un deuxième niveau d'analyse se pose donc au niveau local, item par item, et se révèle d'autant plus complémentaire à l'analyse épistémo-didactique développée précédemment, que bon nombre des items sont sous la forme de QCM ou de Vrai-Faux. Les études menées par Sayac et Grapin (2014) sur les stratégies employées par les élèves de fin d'école pour répondre à des questions sous forme de QCM montrent que ce ne sont pas toujours des stratégies de savoir qui sont employées et que les élèves changent de stratégies au cours d'un même test, selon leur niveau de connaissance. On peut alors raisonnablement questionner l'impact de la forme de la question sur les connaissances que mobilisera l'élève, et étudier le décalage entre ce qui est attendu et ce qui est effectif. A ce stade, nous entrevoyons comment une approche psycho-didactique, fondée sur les processus de réponse de l'élève vient en complément d'une analyse épistémo-didactique. Comment l'approche psychométrique s'articule-t-elle par la suite ?

Le troisième niveau d'analyse vise à vérifier les caractéristiques psychométriques des items comme par exemple, leur pouvoir de discrimination, leur indice de difficulté ou encore, dans le cadre de comparaison temporelle entre 2008 et 2014, les items présentant des fonctionnements différentiels (items biaisés). Pour ces items, une analyse didactique relativement aux savoirs en jeu et à leur enseignement ou une observation d'élèves en train de résoudre l'exercice peut permettre d'expliquer leurs caractéristiques statistiques ; et réciproquement, les indicateurs statistiques correspondant à ces items peuvent nous conduire à interroger les pratiques enseignantes (praxéologies enseignées) et les savoirs évalués.

Par ailleurs, en psychométrie, la difficulté d'un item se détermine *a posteriori*, selon la réussite des élèves au test. En didactique des mathématiques, la complexité d'une tâche peut être définie à partir de différents descripteurs (que nous listons dans le paragraphe suivant) lors de l'analyse *a priori* mais aussi en considérant les difficultés des élèves recensées par différentes recherches relativement à l'apprentissage d'un savoir donné. Par conséquent, il est intéressant d'étudier la corrélation entre la difficulté de l'item (*a posteriori*) et ces éléments théoriques, déterminés *a priori* : nous faisons l'hypothèse que ce double éclairage (épistémo-didactique et psychométrique) permettra ainsi de préciser l'état des connaissances des élèves en fin d'école ou en fin de collège, et réciproquement, d'apporter des éléments supplémentaires pour garantir la validité des inférences faites à partir des résultats.

## 2. Descripteurs de tâches retenus pour l'analyse

Présentons désormais plus spécifiquement la méthodologie d'analyse que nous avons retenue.

Une première étape consiste à faire, pour le domaine étudié, une analyse de toutes les tâches de l'évaluation en précisant : les types de tâches et les objets mis en jeu, les valeurs des variables didactiques associées, les techniques possibles et celles mettant en jeu le raisonnement attendu, les registres de représentations sémiotiques en jeu et leur congruence (Duval 1996). Ensuite, comme nous l'avons expliqué précédemment, nous étudions la

couverture du domaine par les tâches proposées relativement à une organisation mathématique de référence : ce qui nous permet de repérer les manques ou les redondances en termes de types de tâches, mais aussi la variété des représentations sémiotiques convoqués dans le dispositif d'évaluation.

Pour la deuxième étape, nous nous centrons plus localement sur chaque item et au delà de l'analyse *a priori*, par des observations d'élèves, étudions le processus de réponse développé par l'élève, notamment en fonction du format de question.

Ces deux analyses devraient être menées avant passation du test et permettre à la fois l'échantillonnage des tâches et le choix du format de question, pour garantir *a minima* que les distracteurs des QCM correspondent à des erreurs d'élèves et que le processus de réponse engagé pour répondre est bien celui attendu.

La troisième étape d'analyse correspond à la mise en perspective des caractéristiques statistiques des items (difficulté de l'item, indice de corrélation) avec les analyses épistémologique et psycho-didactique : y a-t-il une cohérence localement, globalement ? Si oui, quel éclairage apporte l'approche didactique ? si non, quelles modifications apporter pour enrichir l'évaluation ? Quelles conclusions apporter sur l'état des connaissances des élèves selon les groupes de l'échelle de performance ? Quelles hypothèses formuler alors sur la raison de ces difficultés en lien avec les programmes d'enseignement existants et les pratiques des enseignants ?

#### IV. ANALYSE DES EVALUATIONS EXISTANTES

##### 1. Observations générales et évolution 2008-2014

Sur les évaluation CEDRE fin d'école et fin de collège, nous observons que les questions relatives au contenu, que ce soit dans sa globalité ou sur un domaine donné, mais aussi plus localement, item par item (choix des distracteurs pour les QCM ou modalités de correction pour les questions à réponse ouverte) restent insuffisamment prises en compte, notamment parce qu'elles ne font pas l'objet d'une analyse *a priori* (avant expérimentation) ou *a posteriori* (après expérimentation). Parallèlement, les questions ouvertes notamment en résolution de problèmes ont des consignes de correction qui ne permettent pas de prendre en compte différents types d'erreurs, la réponse étant considérée comme correcte ou incorrecte au regard du résultat attendu.

La question du choix des distracteurs dans les QCM est importante puisqu'elle impacte sur la réussite de l'élève. Même s'il est recommandé (Leclercq 2000) de choisir des distracteurs qui correspondent à des erreurs d'élèves correspondants à des conceptions ou des démarches erronées, nous avons pu constater, notamment dans l'évaluation 2008, que ce n'était pas forcément toujours le cas, et que par conséquent ces items ne se révélaient pas toujours valides d'un point de vue épistémologique et psycho-didactique.

Nous présentons ci-après, pour les deux domaines, des résultats synthétiques de l'analyse que nous avons menée selon les deux premières étapes : une analyse épistémologique et psycho-didactique.

##### 2. L'arithmétique des entiers en fin d'école

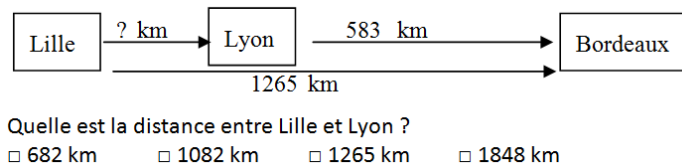
Nous entendons par « arithmétique des nombres entiers » le domaine couvert par la numération (écriture et lecture des nombres dans différents systèmes sémiotiques), le calcul

posé et mental (addition, soustraction, multiplication et division euclidienne) ainsi que la résolution de problèmes additifs et multiplicatifs.

En ce qui concerne les registres sémiotiques mis en jeu, une évolution est constatée entre 2008 et 2014 : alors qu'ils étaient absents en 2008, des items mobilisant des écritures avec des unités de numération et des représentations symboliques avec du matériel de numération ont été insérés dans l'évaluation 2014. Les autres registres (écriture chiffrée, droite graduée, décomposition additive ou en puissances de dix) figurent parmi les exercices proposés, mais les écritures en lettres (et plus largement la numération parlée) sont très peu présentes. Par ailleurs, la plupart des nombres mobilisés sont de l'ordre du millier ou de la centaine, mais « les grands nombres » sont peu présents (une quinzaine d'items pour chacune des évaluations).

Certains manques constatés dans l'évaluation 2008, en termes de types de tâches ont pu être comblés en 2014, en particulier celles portant sur les conversions de registres (par exemple, le passage d'une écriture en chiffres à une écriture en lettre n'était pas évalué en 2008). Enfin, aussi bien en 2008 qu'en 2014, les quatre opérations (maîtrise des techniques opératoires) ont fait l'objet de nombreuses tâches similaires (ce qui interroge leur redondance et l'équilibre global), mais on observe une répartition équilibrée des types de problèmes proposés (additifs et multiplicatifs).

Dans un deuxième temps, si on s'intéresse plus localement à certains items, on peut constater par exemple un effet du format de la question sur la réussite de l'élève. L'item suivant (Figure 2) a été proposé sous forme ouverte et sous forme de QCM :



**Figure 2 - Item extrait de CEDRE fin d'école 2008**

Dans le cas du QCM, les choix proposés étaient : 682 km (la bonne réponse), 1082 km (pas d'explication reposant sur un type d'erreur reconnu), 1265 km (recopie du nombre le plus grand) et 1848 km (la somme des deux nombres). Alors que l'item est réussi en ouvert à 62 %, il est réussi en QCM à 73 % ; le constat de cet écart de réussite (11 %) cible bien la question de ce qui est évalué en fonction de la forme de la question et l'impact de cette dernière sur l'activité de l'élève. Dans ce type de cas, une première analyse épistémologique permettra de mettre en avant les procédures de résolution possibles, celles attendues, les erreurs possibles et conduira à trouver des distracteurs « pertinents » pour le QCM ; dans cet item, nous interrogeons justement le choix du distracteur « 1082 » qui aurait pu avantageusement être remplacé par 1322 correspondant à une mauvaise maîtrise de la technique opératoire de la soustraction (soustraire le plus grand du plus petit systématiquement sur chaque colonne). Par ailleurs, l'écart de réussite avec la question sous forme ouverte peut s'expliquer par l'effet rétroactif du QCM (si l'élève, après calcul, ne trouve pas une des réponses proposées, il peut vérifier ou modifier son raisonnement), mais aussi par la possibilité d'utiliser les choix de réponse et les tester un par un (calculer  $682 + 583$  et constater que la somme est égale 1265). Mais, c'est en observant l'activité de l'élève lors de la résolution effective du problème que l'on pourra savoir s'il mobilise effectivement les procédures attendues ou s'il met en place des stratégies autres, en particulier dans le cas de QCM.

### 3. *L'algèbre en fin de collège*

Nous entendons par algèbre, en fin de collège, le domaine couvert par la résolution de différents types de problèmes intra ou extra mathématiques et dans différents cadres - problèmes de généralisation (expressions littérales), problèmes de modélisation (formules), des problèmes de mise en équation (équations ou inéquations), problèmes de preuve - ainsi que le calcul sur les expressions littérales, les équations, mobilisant des propriétés d'ordre syntaxique et sémantique (propriété de distributivité, conservation de l'égalité, mais aussi équivalence des expressions) et des modes de représentation associés à différents registres sémiotiques de représentation.

Pour la première étape d'analyse, les tâches mises en jeu dans l'évaluation CEDRE en 2008 et en 2014 relèvent bien du domaine algébrique mais ne couvrent tous les types de tâches du domaine.

Dans CEDRE 2008, au-delà des tâches de calcul habituelles bien représentées - substituer, développer ou factoriser une expression algébrique, tester si un nombre est solution d'une équation ou inéquation du premier degré, résoudre une équation produit, un système de deux équations du premier degré à deux inconnues - certaines, plus complexes, permettent d'étudier la capacité des élèves à articuler calcul algébrique et numérique (figure 3). Cette tâche n'a pas été reprise en 2014.

<p><math>a</math> et <math>b</math> sont deux nombres tels que <math>a + b = 5</math> et que <math>a - b = 3</math>.</p> <p>Quelle est la valeur de <math>a^2 - b^2</math> ?</p>	<p>En 2008, Chantal fête ses 53 ans et sa fille Sophie, ses 24 ans.</p> <p>En quelle année l'âge de Chantal sera-t-il le double de celui de sa fille Sophie ?</p>
--	---

**Figure 3** - Item extrait CEDRE fin de collège 2008

**Figure 4** - Item extrait CEDRE fin de collège 2008

En 2014, la résolution d'une équation du premier à une inconnue «  $8x - 3 = 5x + 30$  », sous forme ouverte, permet de repérer les techniques utilisées par les élèves. Une tâche « développer » de format V/F, repris en 2014, permet d'aborder la problématique de l'équivalence des expressions algébriques. Du côté objet, globalement les items en 2008 sont représentatifs du domaine, l'aspect calcul intelligent étant même présent via l'item de la figure 3 et les exercices V/F.

En 2008, au delà des tâches de production de formules, les problèmes numériques proposés, au vu de leur structure et des valeurs numériques retenues, peuvent être résolus par des démarches numériques et ne nécessitent pas la mise en équation, la résolution algébrique du problème s'avérant en plus difficile (réponse décimale) (cf. Figure 4). On peut donc remettre en cause la validité de cet item, au vu de l'objectif visé « mettre en équation ». Cet item a été repris en 2014, mais l'objectif visé est devenu « résoudre un problème par diverses démarches ». Aucun problème de généralisation, de modélisation fonctionnelle, de mise en équation ou de preuve ne permet d'étudier si les élèves rentreraient d'eux mêmes dans une démarche algébrique pour les résoudre. Des tâches de généralisation ont été ajoutées en 2014, sans prise en charge de la mobilisation des lettres, mais des types de tâches restent absents.

En ce qui concerne la deuxième étape d'analyse, on peut s'interroger sur le codage des réponses, limité à correct / incorrect et le choix des distracteurs pour certains problèmes. Dans le cas des questions ouvertes, ce codage ne permet pas de distinguer les types de technique utilisés par les élèves conduisant à une réponse correcte, ni les types d'erreurs. Ces techniques relèvent-elles de l'algébrique (Cf. figure 4) ? En 2014, davantage d'énoncés à « question ouverte » ont été proposés mais quelle exploitation en sera faite ?

En bilan, l'analyse des items du test, au niveau global (couverture des items relativement au domaine) et local (choix des distracteurs ou modalités de correction pour les réponses ouvertes), nous amène à interroger partiellement la répartition de la population sur l'échelle de performance, réalisée *a posteriori* à partir de la réussite aux exercices.

## V. CONCLUSION ET PERSPECTIVES

La méthodologie d'analyse proposée ici et illustrée à partir de quelques exemples montre la richesse et l'intérêt de croiser différents champs de recherche (épistémologique, psychodidactique et psychométrique). Si une analyse épistémologique et didactique permet de contrôler le contenu de l'évaluation au niveau local (par tâche) et global (sur le recouvrement et la variété), il s'agit aussi de mettre en perspective la complexité de la tâche déterminée par l'analyse *a priori* avec l'indice de difficulté de l'item pour repérer d'éventuelles incohérences. Des questions similaires peuvent aussi se poser sur des items qui pourraient sembler pertinents d'un point de vue didactique, mais qui ne se révéleraient pas discriminants après calcul du  $r_{bis}$ . Dans les deux cas, une analyse de la validité psycho-didactique des items, ou la formulation d'hypothèses relatives aux pratiques enseignantes permettraient d'interpréter ces écarts. On comprend ainsi qu'il ne s'agit pas de trois approches successives, mais bien de points de vue complémentaires, parfois sur un même item, qui renseignent différemment sur le contenu des évaluations et sur leurs résultats.

Au-delà de l'analyse des évaluations existantes, les résultats obtenus devraient ainsi contribuer à perfectionner la conception des évaluations externes existantes. Signalons pour conclure que les évaluations CEDRE en mathématiques entre 2008 et 2014 ont vu leurs dispositifs de conception évoluer grâce à deux types de travaux en didactique : l'intégration des exercices extraits du dispositif Pépite (Grugéon 1997) pour évaluer les compétences des élèves en algèbre (fin de collège) et l'exploitation de l'outil d'analyse « facteurs de complexité et de compétence » développé par Sayac & Grapin (2014) pour équilibrer les items de l'évaluation CEDRE fin d'école.

## REFERENCES

- Artigue M., Winslow C. (2010) International comparative studies on mathematics education: a view from the anthropological theory of didactics. *Recherches en didactique des mathématiques* 30(1/3), 47–82.
- Bodin A. (2006) L'évaluation du savoir mathématique. *Recherches en didactique des mathématiques* 17(1/3), 49–96.
- Bosch M., Gascon J. (2005) La praxéologie comme unité d'analyse des processus didactiques. In Mercier A., Margolinas C. (Eds.) *Balises pour la didactique des mathématiques* (pp. 197 – 122). Grenoble : La Pensée Sauvage.
- Bottani N., Vrignaud P. (2005) La France et les évaluations internationales. Rapport établi à la demande du Haut Conseil de l'Évaluation de l'École, 16. Paris : DEP/Bureau de l'édition.
- Brun A., Huguet T. (2008) Les compétences des élèves en mathématiques en fin de collège, *Note d'information*, 10-18. Direction de l'Évaluation de la Prospective et de la Performance.
- Castela C. (2008) Travailler avec, travailler sur la notion de praxéologie mathématique pour décrire les besoins d'apprentissage ignorés par les institutions d'enseignement. *Recherches en didactique des mathématiques* 28(2/3), 135–182.
- Chevallard Y., Feldmann S. (1986) *Pour une analyse didactique de l'évaluation*. IREM d'Aix-Marseille.

- Chevallard Y. (1999) L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en didactique des mathématiques* 19 (2/3), 221 – 266.
- De Landsheere (1988) *Faire réussir, faire échouer*. Paris : PUF
- Duval R. (1996) Quel cognitif retenir en didactique des mathématiques ? *Recherches en didactique des mathématiques* 16(3/3), 349–380.
- El Hage S., Le Hebel F., Coppé S., Tiberghien A. (2014) Identifier l'évaluation formative en classe. In *Cultures et politiques de l'évaluation en éducation et en formation, Actes du 26ème colloque ADMEE-Europe*. Marrakech.
- Grégoire J., Laveault D. (2014) *Introduction aux théories des tests en sciences humaines* Bruxelles : De Boeck Université.
- Grugeon B. (1997) Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire. *Recherches en didactique des mathématiques* 17 (2/3), 167 – 210.
- Grugeon-Allys B., Pilet J., Chenevotot-Quentin F., Delozanne E. (2012) Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire. In Coulange L., Drouhard J.P., Dorier J.L., Robert A. (Eds.) *Recherche en Didactique des Mathématiques, Enseignement de l'algèbre élémentaire, Bilan et perspectives*, Hors-série (pp. 137-162). Paris : La pensée sauvage.
- Leclercq D. (1986) *La conception des questions à choix multiple*. Bruxelles : Labor.
- Lescure S. & Pastor J-M. (2012) *Mathématiques en fin d'école primaire. Le bilan des compétences*. Paris : Scéren.
- Megherbi H., Rocher T., Gyselink V., Trosseille B., Tardieu H. (2009) Évaluation de la compréhension de l'écrit chez l'adulte. *Économie et statistique*, 424-425, 63-86.
- Mons N. (2009) Les effets théoriques et réels de l'évaluation standardisée. Compléments à l'étude Eurydice. Réseau Eurydice.
- Pilet J. (2012) *Parcours d'enseignement différencié appuyés sur un diagnostic en algèbre élémentaire à la fin de la scolarité obligatoire : modélisation, implémentation dans une plateforme en ligne et évaluation*. Thèse de doctorat Paris : Université Paris-Diderot.
- Roditi E., Chesne J.-F. (2012). Un point de vue didactique sur les questions d'évaluation en éducation. In Lattuati M., Penninckx J. & Robert A. (Eds.) (pp. 279–292) *Une caméra au fond de la classe*. Besançon : Presses universitaires de Franche-Comté.
- Roditi E., Salles F. (2015) Nouvelles analyses de l'enquête PISA 2012 en mathématiques, un autre regard sur les résultats. *Revue Education et Formations* 86 – 87, 235-255.
- Ruminot-Vergara C. (2014) Effets d'un système d'évaluation sur l'enseignement des mathématiques : le cas de SIMCE au Chili. Thèse de doctorat. Paris : Université Paris Diderot.
- Sayac N., & Grapin N. (2014) Evaluer les capacités des élèves à résoudre des problèmes dans le cadre d'une évaluation externe en France ; la spécificité des QCM. *Education et francophonie* XLII :2.
- Schneider M. (2006) Comment des théories didactiques permettent-elles de penser le transfert en mathématiques ou dans d'autres disciplines ? *Recherches en didactique des mathématiques* 26(1), 9 – 38.
- Tempier F. (2013) La numération décimale de position à l'école primaire. Une ingénierie didactique pour le développement d'une ressource. Thèse de doctorat. Paris : Université Paris Diderot (Paris 7).
- Vantourout M., Goasdoué R. (2015) Approches et validité psycho-didactique des évaluations. *Revue Éducation et Formations* e302, 139-156.
- Winslow C. (2005) Définir les objectifs de l'enseignement mathématique : la dialectique matières - compétences. *Annales de didactique et de sciences cognitives* 10, 131 – 155.

Pluralités culturelles et universalité des mathématiques :  
enjeux et perspectives pour leur enseignement  
et leur apprentissage

espace mathématique francophone  
Alger : 10-14 Octobre 2015



## ESSAI D'EXPLICATION DU NIVEAU FAIBLE DES PERFORMANCES MATHÉMATIQUES DES ÉLÈVES MAROCAINS DES CYCLES PRIMAIRE ET SECONDAIRE COLLÉGIAL

Mustapha OURAHAY\* - Somaya EL GHARRAS\*\* – Omar ROUAN\*\*\*

**Résumé :** Cette présentation s'appuie sur des résultats obtenus par les élèves marocains de certains niveaux des cycles primaire et secondaire collégial dans des évaluations nationales et internationales. Elle cherche à donner quelques explications à la faiblesse de leur performance en mathématiques qui est très inférieure à la moyenne internationale. Elle se base sur une analyse de la réalité actuelle de l'enseignement marocain des mathématiques et sur l'analyse des rapports du Conseil Supérieur de l'Enseignement et de ses recommandations. Elle permet de montrer qu'il ne faut pas sous-estimer l'impact de certains facteurs qui sont de l'ordre du fonctionnement du système éducatif sur la qualité de l'enseignement des mathématiques.

**Mots clés :** évaluation, acquis, mathématique, apprentissage, enseignement

**Abstract** -This presentation leans upon some national and international assessments results, obtained by the Moroccan students at the primary and secondary school levels. It seeks to explain the weakness of their mathematics performances, which are far under the international average. It is based on the analysis of the reality of the Moroccan mathematical teaching as well as on the reports and the recommendations of the Supreme Council of Education. It concludes that the impact of certain factors, linked to the educational system, on the quality of mathematics education should not be underestimated.

**Keywords :** assessment, acquired, mathematics, learning, teaching

Cette présentation est le fruit d'une analyse des différentes enquêtes nationales et internationales relatives à l'évaluation des acquis des élèves en mathématiques, des rapports réalisés par le Conseil Supérieur de l'Enseignement (SCE) réalisés entre 2007 et 2008, et des cadres de référence d'évaluation relatifs aux examens certificatifs des différents cycles d'enseignement et les différentes notes ministérielles qui les accompagnent.

L'évaluation des acquis des élèves constitue un outil essentiel à la fois pour l'élève, pour l'enseignant et pour le système éducatif. Elle a pour objet d'étayer un diagnostic pertinent, de permettre à l'enseignant de réguler sa pratique pédagogique afin d'optimiser son enseignement et, aux responsables du système scolaire, de prendre des décisions relatives à l'orientation et à la gestion de la carte scolaire.

\* Ecole Normale Supérieure, Université Cadi Ayyad Marrakech - Maroc - mu.ourahay@uca.ma

\*\* Ecole Normale Supérieure, Université Cadi Ayyad Marrakech – Maroc - so.elgharras@uca.ma

\*\*\* Ecole Normale Supérieure, Université Cadi Ayyad Marrakech - Maroc - omarrouan@gmail.com

Les différentes évaluations nationale (PNEA)<sup>1</sup> et internationale (TIMSS)<sup>2</sup> des acquis des élèves marocains s'accordent pour affirmer que le niveau de performance en mathématiques des élèves de la quatrième année du primaire et ceux de la deuxième année du secondaire collégial, considérées comme des années charnières, est très inférieur à la moyenne internationale.

En confrontant les résultats de ces évaluations aux notes des élèves en mathématiques obtenus au contrôle continu et aux examens de passage, nous avons constaté la présence d'une discordance. D'une part, les évaluations nationales et internationales affirment que le niveau des élèves est faible alors que les pratiques pédagogiques d'évaluation laissent croire le contraire. Comment peut-on expliquer cette discordance ? Quelles sont les raisons qui sont derrière sa présence ? Ces différentes évaluations visent-elles les mêmes objectifs, les mêmes niveaux d'habiletés et d'acquisition de connaissances et le même contenu mathématique ?

## I. PRESENTATION SYNTHETIQUE DES RESULTATS DE L'EVALUATION DES ACQUIS REALISEE PAR TIMSS

TIMSS est une enquête internationale conduite par l'Association Internationale pour l'Évaluation du rendement scolaire (AIE). Le Maroc a participé à TIMSS 2003, TIMSS 2007 et TIMSS 2011. Cette enquête cherche à évaluer l'enseignement des sciences et des mathématiques à travers plusieurs facteurs, par la mesure de la performance dans ces champs disciplinaires sur des élèves de quatrième année du cycle primaire (4<sup>ème</sup> année de scolarité) et de deuxième année du cycle secondaire collégial (8<sup>ème</sup> année de scolarité).

### 1. *Présentation de la méthodologie de TIMSS*

L'évaluation des acquis des élèves réalisée par l'intermédiaire de ces différentes enquêtes se base sur le curriculum prévu, tel qu'il est défini par les pays ou les systèmes éducatifs. Elle échantillonne par niveau scolaire et par classes entières d'élèves au sein d'établissements.

TIMSS définit quatre niveaux de performance des élèves: le niveau de performance avancé, élevé, intermédiaire et bas. Pour pouvoir situer la performance des élèves d'un pays dans chacun de ces niveaux, TIMSS utilise quatre indices sur l'échelle de rendement.

---

<sup>1</sup>Programme National d'Evaluation des Acquis PNEA 2008

<sup>2</sup>Trends in International Mathematics and Science Study



Repères internationaux de rendement	Indice	4 <sup>ème</sup> année du primaire	2 <sup>ème</sup> année du secondaire collégial
<b>Bas</b>	400	Posséder les connaissances mathématiques de base.	Avoir une connaissance des nombres entiers et décimaux, des opérations, et des diagrammes élémentaires
<b>Intermédiaire</b>	475	Appliquer les connaissances de base dans des situations simples.	Appliquer les connaissances et la compréhension dans des situations simples : résoudre des problèmes à une étape comprenant des nombres entiers et des nombres décimaux
<b>Élevé</b>	550	Appliquer les connaissances et la compréhension pour résoudre les problèmes.	Appliquer les connaissances et la compréhension dans une variété de situations relativement complexes
<b>Avancé</b>	625	Appliquer la compréhension dans des situations relativement complexes et expliquer le raisonnement.	Organiser des renseignements et en tirer des conclusions, faire des généralisations et résoudre des problèmes hors de l'ordinaire.

*Tableau 1 - Niveaux de rendement des systèmes d'enseignement des mathématiques*

TIMSS projette les items d'évaluation sur des domaines cognitifs d'activité mathématique et des domaines de contenu de chaque niveau scolaire. Ces différents domaines recouvrent le contenu mathématique enseigné ainsi que l'ensemble des processus cognitifs mobilisés dans l'activité mathématique tout au long des années scolaires primaires et secondaires. Le cadre conceptuel ainsi que la méthodologie de cette évaluation peuvent être consultés dans l'un des rapports TIMSS 2007 ou TIMSS 2011. Les tableaux suivants précisent ces différents domaines ainsi que les poids qui leurs sont associés dans les tests.

4 <sup>ème</sup> année du primaire		2 <sup>ème</sup> année du secondaire	
Domaine de contenu	Représentativité	Domaine de contenu	Représentativité
Nombre	50%	Nombre	30%
		Algèbre	30%
Formes géométriques et mesure	35%	Géométrie	20%
Traitement de données	15%	Traitement de données	20%

*Tableau 2 - Représentativité des domaines de contenu de chaque niveau scolaire*

Domaine cognitif	4 <sup>ème</sup> année du primaire	2 <sup>ème</sup> année du secondaire
Connaître	40%	35%
Appliquer	40%	40%
Raisonner	20%	25%

**Tableau 3 - Représentativité des domaines cognitifs**

## 2. Présentation des résultats

La mesure des performances des élèves s'appuie sur la technique du « scale score » qui permet de ramener tous les tests à une même échelle et sur une estimation de la moyenne par intervalle de confiance avec un degré de certitude de 95%. La performance des élèves d'un pays est ainsi représentée par deux valeurs : la valeur du score moyen estimée sur la base des résultats échantillonnaires et la valeur mise entre parenthèse qui représente la longueur de l'intervalle de confiance qui contient ce score avec un degré de certitude de 95%.

Les résultats de TIMSS 2007 et TIMSS 2011 sont regroupés dans un même tableau pour faciliter la lecture et la comparaison. Nous avons choisi de présenter les résultats de Singapour en tant que l'un des pays du monde ayant les meilleures performances pour mieux situer et estimer les résultats des élèves marocains.

## 3. Résultats des élèves de la 4<sup>ème</sup> année du cycle primaire

Pays	TIMSS	connaître	appliquer	raisonner	Moyenne
Singapour	TIMSS 2007	620(4)	590(3.7)	578(3.8)	599(3.9)
	TIMSS 2011	629(3.5)	602 (3.4)	588(3.7)	606(3.2)
Maroc	TIMSS 2007	354(4.8)	346(4.7)	Non calculé <sup>3</sup>	341(4.7)
	TIMSS 2011	320(4.2)	332(3.9)	347(4.2)	335(4.0)

**Tableau 4 - Résultats des élèves par domaine cognitif**

Le score moyen des élèves marocains est très inférieur à l'indice international du niveau "bas" (400) et à la moyenne internationale (500). On constate une diminution de la performance "moyenne" des élèves dans les domaines « savoir » et « appliquer ».

<sup>3</sup> Le rendement moyen ne pouvait pas être estimé avec précision

Pays	TIMSS	Nombre	Géométrie et mesure	Traitement de données	Moyenne
Singapour	2007	661(4.3)	570(3.2)	583(3.2)	599(3.9)
	2011	619(3.4)	589(3.6)	588(3.4)	606(3.2)
Maroc	2007	353(4.5)	365(4.3)	316(6.1)	341(4.7)
	2011	340(3.8)	350(4.0)	271(4.7)	335(4.0)

*Tableau 5 - Résultats des élèves par domaine de contenu*

On remarque une nette diminution du score moyen des élèves marocains en traitement de données. Il est passé de 316 à 271. Ce score est très inférieur à l'indice international du niveau "bas".

#### 4. Résultats des élèves de la deuxième année du cycle secondaire collégial

Pays	TIMSS	Connaître	appliquer	raisonner	Moyenne
Singapour	2007	581(3.4)	593(3.6)	579(4.1)	593(3.8)
	2011	617(3.8)	613(3.9)	604 (4.3)	611(3.8)
Maroc	2007	363(2.2)	378(1.9)	357 (2.7)	371(2.0)
	2011	365(4.4)	389(3.3)	383(3.5)	381(3.0)

*Tableau 6 - Résultats des élèves par domaine de cognitif*

Le rendement "moyen" des élèves marocains est inférieur au rendement "moyen" international. Leur rendement le plus bas est dans le domaine cognitif « connaître ». On note aussi une petite amélioration entre 2007 et 2011 dans les domaines « appliquer » et « raisonner ».

Pays	TIMSS	Nombre	Algèbre	Mesure et Géométrie	Représentation. de données	Moyenne
Singapour	2007	597(3.5)	579(3.7)	578(3.4)	574(3.9)	593(3.8)
	2011	611(3.6)	614(4.1)	609(3.9)	607(4.4)	611(3.8)
Maroc	2007	389(3.4)	362(4.0)	396(3.6)	371(3.4)	381(3.0)
	2011	379(2.6)	357(2.7)	390(2.5)	332(3.1)	371(2.0)

*Tableau 7 - Résultats des élèves par domaine de contenu*

Le rendement "moyen" des élèves marocains a connu une baisse entre 2007 et 2011, surtout dans le domaine de contenu « traitement de données ». Le domaine où ce rendement se rapproche de l'indice international du niveau bas est « Mesure et Géométrie ».

5. Performance des élèves par rapport aux repères internationaux de rendement de mathématiques

	Pays	Singapour		Maroc		Seuil international Moyen		
		TIMSS Seuil	2007	2011	2007	2011	2007	2011
4 <sup>ème</sup> primaire	Avancé		41%(2.1)	43%(2.0)	0%(0.1)	0%(0.2)	5%	4%
	Élevé		74%(1.7)	78%(1.4)	2%(0.8)	2%(0.7)	26%	28%
	Intermédiaire		92%(0.9)	94%(0.7)	9%(1.1)	10%(1.2)	67%	69%
	Bas		98%(0.3)	99% (0.2)	26%(2.0)	26% (1.5)	90%	90%
2 <sup>ème</sup> collège	Avancé		48%(2.0)	40%(1.9)	0%(0.0)	0%(0.1)	3%	2%
	Élevé		78%(1.8)	70%(1.1)	2%(0.2)	1%(0.5)	17%	15%
	Intermédiaire		92%(1.1)	88%(1.4)	12%(0.5)	13 %(1.1)	46%	46%
	Bas		99%(0.3)	97%(0.6)	36%(1.0)	41%(2.0)	75%	75%

Tableau n° 8 - taux et niveaux de performance des élèves

Ces tableaux donnent pour chaque niveau et pour chaque pays les pourcentages cumulés croissants allant des niveaux supérieurs aux niveaux inférieurs.

Au primaire les résultats de TIMSS 2007 et de TIMSS 2011 montrent que le taux des élèves marocains qui ont atteint le seuil international du niveau bas ainsi que celui des élèves du niveau intermédiaire n'ont pas changé. Ces taux sont nettement inférieurs aux taux moyens internationaux qui sont respectivement de l'ordre de 90% et de 67%. Seulement 37% des élèves marocains ont pu avoir un score supérieur ou égal à 400. Plus de 60% des élèves sont en dessous du seuil international associé au niveau bas.

Quant au collège, les taux des niveaux bas et intermédiaire ont connu une petite augmentation sauf que ces taux restent inférieurs aux taux moyens internationaux relatifs à chacun de ces deux niveaux qui sont respectivement de 75% et 45%. Plus de 40% des élèves sont au dessous du seuil du niveau bas.

Le rapport TIMSS 2011 souligne que, d'une part, plus de 25% des élèves marocains de la quatrième année du cycle primaire ont obtenu un score inférieur à 250 alors que leur score moyen est de 335. D'une autre part, plus de 25% des élèves marocains de la deuxième année du collège ont obtenu un score inférieur à 250 alors que leur score moyen est de 371. Ceci montre que la distribution des scores des élèves marocains est très hétérogène, ce qui permet de supposer que le score moyen n'a pas de signification.

Il faut noter que, dans l'évaluation des acquis, la mesure est plus efficace quand il existe une correspondance raisonnable entre le niveau d'habiletés mathématiques de l'élève à évaluer et la difficulté des éléments d'évaluation. Plus il y a décalage, plus il devient difficile d'obtenir une évaluation fiable. En particulier, lorsque les tâches d'évaluation sont beaucoup trop difficiles pour la plupart des élèves, ces derniers auront tendance à réagir d'une manière

hasardeuse, et dans ce cas, il est extrêmement difficile d'atteindre la qualité acceptable de l'évaluation.

## II. ÉVALUATION DES ACQUIS DES ELEVES PAR PNEA

Le Programme National d'Évaluation des Acquis (PNEA) a été piloté en 2008 par l'Instance Nationale d'Évaluation du Système d'Éducation et de Formation (INESEF) auprès du Conseil Supérieur de l'Enseignement (CSE), en collaboration avec le Centre National des Examens et d'Évaluation relevant du Ministère de l'Éducation Nationale. Il s'est assigné pour objectif de mesurer les acquis scolaires des élèves marocains en langues arabe et française, en mathématiques et en sciences.

Ce programme vise une évaluation diagnostique des apprentissages et porte sur les programmes scolaires officiels. Pour chacune des matières ciblées, l'accent a été mis sur les connaissances et compétences effectivement acquises par les élèves, en regard des curricula et programmes prescrits. Les niveaux scolaires visés sont la 4<sup>ème</sup> et la 6<sup>ème</sup> année de l'enseignement primaire et la 2<sup>ème</sup> et la 3<sup>ème</sup> année de l'enseignement secondaire collégial.

### 1. Présentation de la méthodologie de PNEA

La méthodologie adoptée pour évaluer les acquis des élèves par le PNEA est fondée sur des tests disciplinaires (Arabe, Français, Mathématiques et Sciences) administrés à un échantillon représentatif d'élèves mais aussi sur des questionnaires "Élève", "École", "Enseignant" et "Parents".

Les différents questionnaires ont pour but de recueillir de l'information permettant l'interprétation des performances des élèves. Les variables retenues par PNEA renvoient aux différents facteurs de la réussite scolaire exprimés dans Boufrahi & al. (2003) qui sont :

Les facteurs associés aux caractéristiques personnelles des élèves, et à leurs conditions socio-économiques;

Le rôle, le suivi et la participation des ménages dans la vie scolaire ainsi que leurs conditions socio-économiques;

Les caractéristiques personnelles des enseignants, leurs aspects professionnels et l'exercice du métier en relation avec la discipline enseignée ;

L'environnement de l'établissement scolaire et sa gestion pédagogique et administrative.

Les tests disciplinaires de rendement scolaire sont formulés sur la base de cadres de références et portent sur les apprentissages relatifs au niveau visé. Dans chaque cadre de référence on détermine le degré d'importance de chaque domaine d'apprentissage ainsi que les principaux objectifs d'apprentissage qui lui sont associés, et on élabore des tableaux de spécification pour chaque test dans lesquels on établit le pont entre les items, les domaines d'apprentissages et les domaines cognitifs.

Cette méthodologie a été développée par le Centre National des Examens et d'Évaluation relevant du Ministère de l'Éducation Nationale. Elle s'appuie sur les cadres de référence des examens des cycles primaire et secondaire collégial. En mathématiques, les objectifs d'apprentissage sont répartis selon trois domaines de contenu, à savoir les « activités numériques », les « activités géométriques » et les « activités de mesure ». Les domaines cognitifs d'activité mathématique retenus sont « connaître », « appliquer » et « résoudre ». Les tableaux ci-dessous montrent comment les objectifs d'apprentissage sont répartis selon

les domaines de contenus et les domaines cognitifs. Notons que dans cette évaluation, le domaine cognitif « connaître » n'apparaît pas au niveau 6<sup>ème</sup> année primaire.

Domaine/objectif	4 <sup>ème</sup> primaire				6 <sup>ème</sup> primaire		
	Connaître	Appliquer	Résoudre	Total	Appliquer	Résoudre	Total
Domaine numérique	4%	32%	5%	41%	33%	6%	39%
Domaine géométrique	4%	18%	5%	27%	22%	6%	28%
Domaine de la mesure	4%	23%	5%	32%	7%	6%	33%
<b>Total</b>	12%	73%	15%	10%	82%	18%	100%

Tableau n° 9 - Tableau de spécification des objectifs du cycle primaire

Domaine/objectif	2 <sup>ème</sup> collège				3 <sup>ème</sup> collège			
	Connaître	Appliquer	Résoudre	Total	Connaître	Appliquer	Résoudre	Total
Domaine numérique	12%	31%	5%	48%	11%	18%	7%	35%
Domaine géométrique	11%	27%	4%	42%	15%	25%	10%	50%
Domaine de la mesure	3%	6%	1%	10%	5%	8%	3%	15%
<b>Total</b>	26%	64%	10%	100%	30%	50%	20%	100%

Tableau n° 10 - Tableau de spécification des objectifs du cycle secondaire collégial.

## 2. Présentation et analyse des résultats

Le codage retenu pour le traitement des productions des élèves consiste à affecter pour les domaines « Connaître » et « Appliquer », 1 point pour une bonne réponse et 0 point pour une mauvaise. Quand à « Résoudre », le codage adopté consiste à affecter 2 points aux réponses complètes, 1 point aux réponses partielles et 0 aux réponses fausses.

Que ce soit au collège ou au primaire, les objectifs d'apprentissage du domaine cognitif « résoudre » ne dépassent pas 20%, alors qu'ils dépassent 80% pour les deux autres domaines. Cette répartition des domaines cognitifs et de contenus adoptée par le PNEA est similaire à celle utilisée par TIMSS. Dans ce sens, nous dirons que ces deux évaluations adoptent la même méthodologie et elles diffèrent au niveau du traitement des données. TIMSS vise l'évaluation des acquis alors que PNEA vise l'évaluation de la performance du système éducatif.

Chez PNEA, le traitement des performances des élèves porte sur une analyse descriptive des scores des élèves par domaine de contenus et par domaine cognitif et ce, selon le genre, le milieu et la région.

En l'absence d'une échelle répartissant les élèves selon des seuils de performance, les résultats de chaque catégorie d'élèves sont donnés sous forme d'un score moyen calculé sur une échelle de 100. Par ce traitement, le PNEA cherche à relier les scores moyens des différentes catégories aux facteurs choisis comme étant ceux de la réussite scolaire. Il vise surtout l'analyse du rendement du système éducatif en relation avec les facteurs de la réussite scolaire retenus, au lieu de l'analyse de la représentativité des différents seuils de performances.

Nous retenons des résultats présentés dans le rapport PNEA, que la performance moyenne des élèves est de 34% en quatrième année du cycle primaire et de 25% en deuxième année du cycle secondaire collégial. Cette performance est faible et très inférieure au taux moyen requis.

Selon ce rapport, le facteur genre n'est pas significatif ce qui lui a permis de conclure que l'école joue en faveur de l'égalité des chances entre les deux sexes.

PNEA	Primaire		Secondaire	
	4 <sup>ème</sup> année primaire	6 <sup>ème</sup> année primaire	2 <sup>ème</sup> année collège	3 <sup>ème</sup> année collège
<b>Enseignement privé</b>	57%	68%	53%	65%
<b>Enseignement public urbain</b>	38%	48%	26%	31%
<b>Enseignement public rural</b>	31%	39%	22%	22%

*Tableau n°11 - Performance des élèves en mathématique, PNEA 2008*

Les résultats de ce tableau montrent que les élèves qui étudient dans des écoles privées ont de meilleures performances que ceux des écoles publiques et que les élèves des écoles publiques en milieu urbain ont de meilleures performances que ceux des écoles publiques en milieu rural.

Selon PNEA, l'analyse des résultats des différents questionnaires a permis de relever des indices qui permettent d'expliquer à la fois la faible performance des élèves en mathématiques ainsi que l'écart entre les performances des différentes catégories d'élèves. Parmi ces facteurs, on peut citer:

La majorité des élèves de l'enseignement privé ont profité, dans leur majorité, d'un enseignement préscolaire d'au moins deux ans. De plus, les élèves de l'enseignement privé font généralement l'objet d'une sélection, par des tests appropriés, avant leur admission.

Les différences au niveau de la réussite scolaire entre les deux milieux sont dues aux disparités du développement social et humain entre les zones rurales et les zones urbaines. L'offre d'enseignement préscolaire, plus importante en milieu urbain, est encore au stade embryonnaire dans le milieu rural. Aussi, les conditions de travail des enseignants en ville sont plus favorables que celles de leurs collègues du milieu rural.

L'analphabétisme touche plus les parents du milieu rural. Ceci se répercute sur l'intérêt qu'ils portent à la formation et à l'instruction de leurs enfants.

L'adoption des programmes officiels comme référence à l'évaluation des acquis scolaires des élèves peut être un frein à l'obtention de résultats performants. En effet il s'est avéré que, souvent, ces programmes ne sont pas réalisés complètement et que, lorsqu'ils le sont, les

enseignements sont conduits dans des conditions qui ne permettent pas aux élèves de maîtriser suffisamment les compétences visées.

### 3. *Commentaires*

Ce programme national d'évaluation des acquis ne cherche pas à catégoriser les élèves selon leurs performances. Il cherche surtout à évaluer la performance du système éducatif en fonction de certains facteurs de réussite scolaire. Il n'a pas fourni d'éléments explicatifs du faible rendement de la totalité du système scolaire.

Rappelons que le rapport TIMSS 2011 montre que, d'une part, les résultats des élèves marocains ne sont pas homogènes, d'autre part, le traitement de données adopté par le PNEA n'a pas tenu compte de l'hétérogénéité et de la dispersion des performances des élèves au sein d'une même catégorie. Ceci permet de supposer que le score moyen des élèves marocains peut être biaisé et peut ne pas avoir de signification.

## III. ANALYSE ET DISCUSSION

Les deux évaluations se complètent dans le sens où elles ont adopté la même ingénierie méthodologique. Toutefois, elles ne visent pas les mêmes objectifs dans le sens où TIMSS essaie de comparer les différentes performances des élèves des différents pays tandis que PNEA s'intéresse au rendement du système éducatif et à l'impact de facteurs dominants de la réussite scolaire.

Ces différentes évaluations des acquis se sont toutes appuyées sur le curriculum, tel qu'il est défini par le système éducatif et par conséquent sur les connaissances et compétences que l'apprenant devrait acquérir. Elles attestent qu'il y a présence d'un écart considérable entre ce qui est visé comme acquis scolaire et ce que les élèves ont acquis réellement sans fournir d'éléments explicatifs.

Le rapport de PNEA soutient les résultats de TIMSS dans leur globalité et permet d'apporter des explications à la disparité des scores des élèves marocains en se référant à des facteurs de réussite comme le milieu et le type d'enseignement. La disparité soulignée par PNEA concerne la performance des différentes catégories d'élèves. Mais il n'a pas abordé la disparité des performances des élèves dans une même catégorie que nous pensons être plus importante que celle entre les catégories d'élèves.

Pour compléter ces deux analyses, nous essayons de déterminer certains facteurs qui semblent agir sur la totalité du système éducatif et qui pourraient expliquer ce faible rendement. Pour déterminer ces facteurs explicatifs, nous nous sommes basés sur nos échanges avec différents acteurs et instances pédagogiques et sur la réforme du système éducatif de 2002, qui a apporté des changements au niveau des pratiques pédagogiques et de l'évaluation des acquis soutenant l'enseignement des mathématiques. Il s'agit de suppositions plutôt que d'affirmations.

### 1. *L'abolissement du redoublement*

Dans le rapport<sup>4</sup> sur l'état et les perspectives du système éducatif marocain et de la formation, le Conseil Supérieur d'Enseignement affirme que le taux élevé du redoublement que connaît le système éducatif devient un obstacle majeur à la réforme. En 2005-2006 les redoublants

---

<sup>4</sup> Volume 2 Rapport analytique, 2008 réalisé par l'Instance Nationale d'Évaluation du système d'Éducation et de formation (une structure au sein de conseil supérieur de l'enseignement).



représentaient 13% des effectifs du primaire, 16% de ceux du secondaire collégial et 18% de l'ensemble des lycéens<sup>5</sup>. Il ajoute que : « Le redoublement réduit de façon substantielle la capacité de notre système éducatif à généraliser la scolarité obligatoire et à améliorer la qualité de l'enseignement. » p.57 et qu'« Il est donc nécessaire que le système d'évaluation et d'examen tienne compte des dangers liés au redoublement massif des élèves et ses répercussions négatives sur l'atteinte des objectifs de la Charte en matière de généralisation de la scolarisation obligatoire et l'extension de l'enseignement. » p.58.

Pour palier à ce problème de redoublement, le système éducatif a pris la décision d'abolir le redoublement en cycle primaire, sauf pour des cas exceptionnels, sans aucun accompagnement pédagogique auprès des élèves en difficulté d'apprentissage et sans aucune formation continue pour les enseignants dans le but de les accompagner dans ce changement de paradigme.

Il faut reconnaître que ni le passage automatique ni le redoublement ne peuvent à eux seuls résoudre les problèmes des élèves en difficulté d'apprentissage. D'un côté, s'ils sont promus sans avoir acquis le minimum de connaissances et d'habiletés exigibles, ils risquent d'avoir autant de mal à la poursuite de leurs études. D'un autre côté, s'ils redoublent, ils auront peu de chances de réussir avec la reprise à l'identique d'une année scolaire. Dans les deux cas, il serait plus efficace de prendre des mesures d'accompagnement pour répondre aux besoins spécifiques de chaque élève d'autant plus que le redoublement est étroitement lié à un même rythme d'apprentissage imposé à l'ensemble des élèves, regroupés au sein d'un même groupe pédagogique, la classe. Ils doivent assimiler le même ensemble de connaissances dans le cadre d'une planification scolaire annuelle, rigide et identique.

## 2. *Les pratiques pédagogiques et l'évaluation*

Les pratiques pédagogiques se caractérisent par le peu de place allouée à l'évaluation formative. Par « manque de formation » des enseignants, comme l'affirme le rapport du CSE<sup>6</sup>, l'évaluation sommative est le mode le plus utilisé. Par conséquent, le seul feedback retourné aux élèves prend généralement la forme d'une note.

Dans le cadre de la réforme du système éducatif marocain entamé en 2002, le ministère de l'éducation nationale a produit des cadres de référence des examens de fin de cycle et des notes ministérielles relatives à l'évaluation continue. Ces différents documents permettent de cadrer les pratiques évaluatives, d'explicitier les niveaux d'habiletés et de connaissances à évaluer, et de fixer le minimum exigible à la réussite scolaire.

Notons que la correspondance entre les niveaux d'acquisition du savoir et de développement d'habiletés exigées aux élèves pour réussir et celles exigées pour poursuivre les études est conditionnée par :

- 1) Le rôle de régulateur de l'enseignant qui doit optimiser son enseignement en fonction des résultats d'évaluation des acquis de ses élèves,
- 2) La présence d'une concordance entre l'offre et la demande scolaire du système éducatif qui se traduit par la carte scolaire.

Une analyse des pratiques évaluatives en enseignement des mathématiques pourrait nous offrir des éléments de réponses au rendement faible des élèves en mathématiques. Le rapport du CSE souligne que les enseignants du primaire manquent de compétences professionnelles pour assurer un enseignement de mathématiques de qualité. Ce manque de compétence

<sup>5</sup> Volume 2 Rapport analytique, 2008, Paragraphe « Le redoublement est pédagogiquement inefficace », p. 55

<sup>6</sup> État et perspectives du système d'éducation et de formation Volume 4 : Métier de l'enseignant, CSE 2008.

impose une diminution des exigences de réussite, ce qui permet d'assurer, par voie de conséquences, d'une part, une fluidité dans la gestion de la carte scolaire, mais d'autre part, mène à une inadéquation entre la gestion de la carte scolaire et le minimum des acquis exigible à la poursuite des études.

L'évaluation des acquis des élèves en dernière année du cycle primaire (6<sup>ème</sup> année) est soutenue par un cadre de référence qui permet de définir à la fois les orientations générales soutenant les pratiques évaluatives et les deux niveaux d'habiletés à évaluer, à savoir, le niveau des applications directes et celui de la résolution de problèmes. Cette évaluation donne une importance considérable aux applications directes avec un taux de 80%. Ce premier niveau des habiletés renvoie à l'activité mathématique opératoire qui consiste à effectuer une opération de calcul, à construire une figure géométrique ou à convertir une mesure, domaines cognitifs "connaître" et "appliquer". Quant au deuxième niveau des habiletés, il représente 20% et il renvoie à la compréhension et à la construction du sens des notions et des opérations mathématiques, domaine cognitif "raisonner" ou "résoudre".

Selon ce cadre de référence d'évaluation des acquis en mathématiques, le seuil de réussite est fondé sur le premier niveau des habiletés alors que le seuil de poursuite des études renvoie au deuxième niveau des habiletés qui fait appel à la compréhension, à la communication et à la résolution en mathématique.

Nos contacts et échanges avec les enseignants de mathématiques laissent savoir que la majorité de ces derniers déclarent avoir des classes d'élèves dont la majorité d'entre eux ont des difficultés majeures en mathématiques dues aux lacunes cumulées dans les niveaux inférieurs. Ce cumul de lacunes est en grande partie due à l'écart qui se creuse entre le minimum des acquis nécessaire à la réussite et celui qu'exige la poursuite des études. Ces enseignants sont ainsi pris entre un niveau initialement faible des élèves et une programmation rigide d'exécution du programme planifiée sous forme d'un calendrier à respecter dans lequel le nombre d'heures par thème et le nombre de semaines par domaine de contenu sont prédéfinis dans une progression linéaire.

La réduction du nombre de redoublants, la planification pédagogique annuelle et rigide de l'enseignement des mathématiques et l'absence de soutien pédagogique aux élèves en difficultés d'apprentissage favorisent le cumul de lacunes chez ces derniers et creusent l'écart entre leurs acquis et le minimum exigible à la poursuite de leurs études. Ceci amène les enseignants à mettre l'accent sur la maîtrise des techniques et sur la préparation aux épreuves de l'évaluation sommative en leur donnant des activités et des problèmes similaires aux épreuves d'examen, c'est-à-dire à conditionner leurs pratiques pédagogiques par l'évaluation pour permettre à leurs élèves d'avoir les notes qui leur permettraient de réussir. Les enseignants se sentent souvent obligés d'enseigner « pour l'examen » et les élèves sont encouragés à réussir aux examens aux dépens des véritables objectifs d'apprentissage. De plus, les pratiques évaluatives pourraient conduire les élèves à concevoir leurs évaluations comme des moments stressants, en raison de leur importance pour leur avenir scolaire et social.

Ce glissement pédagogique des pratiques des enseignants est l'une des causes de l'écart qui existe actuellement entre le niveau des acquis des élèves et le niveau minima exigible pour la poursuite des études. Ce cadre de référence d'évaluation associé aux pratiques pédagogiques de la majorité des enseignants encouragent les élèves à maîtriser des techniques opératoires sans comprendre leurs sens et sans pouvoir les mobiliser dans des situations problèmes contextuelles.

#### IV. CONCLUSION

Nous pensons que le problème de l'adéquation de l'évaluation des acquis à la carte scolaire se traduit actuellement par une baisse de niveau des acquis des élèves et qu'il résulte des facteurs qui sont de l'ordre de la gestion du système scolaire qui dépasse le cadre de l'enseignement des mathématiques.

En y regardant de près, ces différentes pistes d'explication ont un point commun : elles sont toutes des facettes de la gestion du système éducatif. Pour cette raison, nous encourageons l'amorce d'une réflexion importante sur cette problématique, dans laquelle la question du pilotage du système éducatif par les résultats devrait tenir une place centrale.

Malheureusement, pour diverses raisons, dans le pilotage quotidien du système éducatif, cette dimension des acquisitions des élèves n'est pas directement prise en compte. Cette situation contribue à expliquer l'écart qui se creuse entre ce qui est prescrit comme performances en mathématiques et les niveaux d'acquisitions atteints.

Tout enseignement des mathématiques doit être soutenu par un minimum d'exigences telles que:

Les enseignants doivent avoir une certaine maîtrise de la discipline à enseigner et une compétence professionnelle qui leur permettraient d'assurer leur fonction d'enseignant

Les élèves doivent avoir un minimum de pré requis qui leur permettraient de poursuivre un apprentissage en mathématiques.

L'évaluation sommative et/ou certificative doit avoir pour objectif de valider les acquis des élèves et d'assurer le niveau minimal d'acquisition de connaissances et de développement de compétences nécessaire à la poursuite des études au niveau supérieur.

Quand l'une de ces trois exigences n'est pas satisfaite, les difficultés d'apprentissage commencent à se cumuler et peuvent devenir insurmontables.

#### REFERENCES

##### A. RAPPORTS D'ENQUÊTES

Programme National d'Évaluation des Acquis PNEA 2008, Rapport analytique version en langue française réalisé par le Conseil Supérieur de l'Enseignement du Maroc publié en 2009

Programme National d'Évaluation des Acquis PNEA 2008, Rapport synthétique en langue française réalisé par le Conseil Supérieur de l'Enseignement du Maroc publié en 2009

État et perspectives du système d'éducation et de formation Volume 3 : Atlas du système d' »éducation et de formation, publié par le Conseil Supérieur de l'Enseignement du Maroc, l'Instance Nationale d'Évaluation du système d'Éducation et de formation, rapport annuel 2008.

État et perspectives du système d'éducation et de formation Volume 2 : Rapport analytique, publié par le Conseil Supérieur de l'Enseignement du Maroc, l'Instance Nationale d'Évaluation du système d'Éducation et de formation, rapport annuel 2008.

État et perspectives du système d'éducation et de formation Volume 4 : Métier de l'enseignant, publié par le Conseil Supérieur de l'Enseignement du Maroc, l'Instance Nationale d'Évaluation du système d'Éducation et de formation, rapport annuel 2008.

*TIMSS (2007) International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, Mullis, I.V.S., Martin M.O., Foy P. (with Olson J.F., Preuschoff C., Erberber E., Arora A., Galia J.) (2008) Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS (2011) International Results in Mathematics, Ina V.S. Mullis, Michael O. Martin, Pierre Foy, and Alka Aroram, TIMSS&PIRLS International Study Center, Boston College

#### B. NOTES MINISTÉRIELLES

Arrêté ministériel Numéro 2383.06 (16/09/2006) relatif à l'organisation d'examen d'option du certificat des études primaires

Note ministérielle N° 46-2006, Cadre de référence d'évaluation des matières d'examen normalisé pour l'obtention du certificat des études primaires

Note ministérielle N°43 - 2006, organisation des études en enseignement secondaire

Note ministérielle N° 28-2010, Cadre de référence d'évaluation des matières d'examen normalisé pour l'obtention du certificat des études secondaire collégiales

Note ministérielle, note142-8, 2007 relative à l'évaluation en mathématiques du cycle secondaire qualifiant

#### C. DOCUMENTS PEDAGOGIQUES OFFICIELS

Ministère de l'Éducation Nationale de l'Enseignement supérieur et de la recherche scientifique, Guide pédagogique de l'enseignement primaire, Maroc, 2009

Le livre blanc volume 2, curricula des cycles primaires, Maroc, 2002

Les programmes de mathématiques de l'enseignement primaire

Les programmes de mathématiques de l'enseignement primaire 2010

Les programmes de mathématiques de l'enseignement secondaire collégial 2009

Les programmes de mathématiques de l'enseignement secondaire qualifiant 2007



## DIDACTIQUE ET ÉVALUATION : UN NOUVEAU REGARD SUR LE PISA 2012

Eric RODITI\*

**Résumé** – Les enquêtes du PISA visent à mettre au jour les acquis des élèves de 15 ans. Ce que le PISA nous apprend dépend strictement de ce qui est mesuré. En mathématiques, un éventail de compétences diverses est évalué, mais les critères ne tiennent pas compte du degré d’initiative demandée à l’élève. Une catégorisation des items est proposée dans cette communication, elle distingue les compétences en fonction de différents niveaux de mise en fonctionnement des connaissances mathématiques. Elle conduit à un nouveau regard sur le test PISA ainsi que sur les résultats de la France.

**Mots-clefs** : Didactique des mathématiques, Évaluation, PISA, ANR.

**Abstract** – The inquiries of the PISA aim at indicating the knowledge of the 15-year-old pupils. What the PISA teaches us depends strictly on what is measured. Different mathematical skills are assessed, but the criteria do not take into account the level of initiative attempt from the pupil. A new categorization of items is proposed in this lecture, it distinguishes the skills according to various levels of using mathematical knowledge. It leads to a new comprehension on the test PISA as well as on the results of France.

**Keywords**: Mathematics education, Assessment, PISA, ANR

Les enquêtes du PISA (Programme International de Suivi des Acquis des élèves) visent à mettre au jour les acquis des élèves de 15 ans – c'est-à-dire ce qu'ils sont capable de faire avec ce qu'ils ont appris –, avec quelles différences suivant les pays comme au sein de chacun d'eux. Ces enquêtes conduisent à un classement international, mais aussi à rendre compte du niveau des élèves les plus performants comme de celui des plus faibles, à pointer les inégalités entre les filles et les garçons ou selon les catégories professionnelles et sociales. Les connaissances produites par le PISA dépendent strictement de ce qui est mesuré. En mathématiques, par exemple, on ne demande pas aux élèves de restituer leur connaissance des définitions ou des règles. On évalue leur compétence, c'est-à-dire ce qu'ils mobilisent pour comprendre et résoudre un problème. Les organisateurs du PISA cherchent à évaluer un éventail de compétences diverses : les questions ont différents niveaux de difficulté, elles font appel à différents processus psycho-cognitifs, elles correspondent à plusieurs domaines mathématiques et elles sont posées dans des contextes diversifiés de la vie réelle. Néanmoins, même s'il apparaît clairement que les questions posées tiennent compte, en amont, du degré d'initiative nécessaire, les grilles de codages – et donc les analyses qui en résultent en aval –

\* Sorbonne Paris-Cité, Université Paris Descartes, Laboratoire EDA – France – eric.roditi@paris5.sorbonne.fr

ne différencient pas, par exemple, les questions qui nécessitent l'application directe d'une règle mathématique de celles qui exigent une prise d'initiative de la part de l'élève<sup>1</sup>.

Une nouvelle catégorisation des items est proposée dans cette communication ; elle distingue les compétences en fonction de différents niveaux concernant les activités mathématiques requises. Sur deux items, des informations présentées dans le rapport de l'OCDE et des analyses auxquelles conduit cette nouvelle catégorisation sont mises en regard. Puis une étude des données concernant la France conduira à une réinterprétation des connaissances apportées par le PISA quant aux difficultés en mathématiques, aux inégalités filles-garçons, à l'effet des différences sociales ou du retard scolaire sur les acquis en mathématiques. Cette recherche s'intègre dans un projet plus large soutenu par l'ANR<sup>2</sup>

## I. DESCRIPTION DES COMPETENCES MATHÉMATIQUES : OUTILS DE L'OCDE ET CLASSIFICATION ISSUE DE LA DIDACTIQUE DES MATHÉMATIQUES

Piloté par l'OCDE (Organisation pour la Coopération et le Développement Économique), le PISA fait référence quant à l'évaluation des acquis des élèves à la fin de la scolarité obligatoire. En 2012, comme pour la dernière fois en 2003, le domaine majeur de l'évaluation PISA fut la culture mathématique. Mais qu'évalue vraiment le PISA en mathématiques ? Qu'est-ce que la culture mathématique et comment est-elle représentée dans les items du test ?

### 1. Cadre défini par l'OCDE pour l'évaluation de la culture mathématique

Le cadre d'évaluation de la culture mathématique est fondé sur une approche psychologique de l'activité mathématique des élèves répondant aux items du test. Il a été élaboré pour l'OCDE conjointement par l'*Australian Council for Educational Research* (ACER) et par une organisation de recherche pédagogique basée aux États-Unis, *Achieve Inc.* (OCDE 2013). Indiquons-en les éléments essentiels à partir de quelques extraits de cette documentation.

L'enquête PISA se fonde sur une conception de l'évaluation des connaissances, des compétences et des attitudes qui reflète l'évolution des programmes d'enseignement : elle va au-delà des acquis purement scolaires et se concentre sur la mise en œuvre des savoirs et savoir-faire dans des tâches et des défis quotidiens, que ce soit en famille ou dans le monde du travail. [...]. L'enquête PISA cible des activités que les élèves âgés de 15 ans auront à accomplir dans l'avenir et cherche à identifier ce qu'ils sont capables de faire avec ce qu'ils ont appris [...]. Les épreuves sont conçues à la lumière du dénominateur commun des programmes scolaires des pays participants, sans toutefois s'y cantonner. Elles servent à évaluer les connaissances des élèves, certes, mais aussi leur faculté de réflexion et leur capacité à appliquer leurs connaissances et leurs expériences dans des situations qui s'inspirent du monde réel. (OCDE 2013, p. 13)

La culture mathématique est l'aptitude d'un individu à formuler, employer et interpréter des mathématiques dans un éventail de contextes, c'est-à-dire à raisonner en termes

<sup>1</sup> Le document <http://research.acer.edu.au/cgi/viewcontent.cgi?article=1003&context=pisa>, rend bien compte du fait que, selon le PISA, la question de la prise d'initiative est présente dans chacune des compétence et constitue de surcroît une compétence à part entière : « *devising strategies* ». Elle apparaît dans la construction des items, dans leur choix et se traduit dans la prédiction de difficulté de chacun des items (<http://research.acer.edu.au/cgi/viewcontent.cgi?article=1006&context=pisa>).

<sup>2</sup> L'ANR (Agence Nationale de la Recherche) soutient le projet NEOPRAEVAL sur les nouveaux outils et les nouvelles pratiques d'évaluation en mathématiques.

mathématiques et à utiliser des concepts, procédures, faits et outils mathématiques pour décrire, expliquer et prévoir des phénomènes. Elle aide les individus à comprendre le rôle que les mathématiques jouent dans le monde et à se comporter en citoyens constructifs, engagés et réfléchis, c'est-à-dire à poser des jugements et à prendre des décisions en toute connaissance de cause. (OCDE 2013, p. 27)

Afin de mesurer les acquis des élèves en mathématiques, retenons que l'OCDE catégorise les items suivant quatre grands domaines mathématiques (quantité, incertitude et données, variations et relations, espace et formes), suivant quatre types de contextes (personnel, sociétal, professionnel, scientifique) et suivant trois processus psycho-cognitifs : formuler (mathématiser les situations de vie réelle), employer (travailler au sein du modèle mathématique) et interpréter/évaluer (mettre un résultat mathématique à l'épreuve d'une situation réelle). Les experts s'accordent pour penser qu'il ne serait pas pertinent de construire un dispositif d'évaluation portant, pour toutes les questions de ce dispositif, sur chacun des processus psycho-cognitif. Il est ainsi fréquent, dans les items de l'enquête PISA, que plusieurs d'entre eux soient déjà pris en charge dans l'énoncé et que l'élève qui cherche à résoudre le problème n'en ait qu'un ou deux à mettre en œuvre (OCDE 2013, p. 28).

Ainsi, bien que tous les problèmes soient posés dans un contexte de vie réelle, ce dernier n'a pas toujours d'influence effective sur l'activité de l'élève traitant l'un des items liés à un problème. En outre, même lorsque l'élève doit *formuler* en langage mathématique la situation issue de ce contexte de la vie réelle et/ou *interpréter/évaluer* les résultats obtenus par rapport à ce contexte, il reste toujours une partie de l'activité de résolution qui consiste à *employer* des connaissances mathématiques. C'est donc la subdivision des étapes de la résolution du problème en différents items qui permet aux experts du PISA de classer chaque item dans une et une seule de ces trois catégories ; et ce classement témoigne, non pas d'une seule activité, mais plutôt de l'activité dominante. Il reste que tous les items ne sont pas équivalents quant au niveau de mise en fonctionnement des connaissances mathématiques, et qu'ils ne reflètent donc pas le même niveau d'acquisition. C'est justement pour se donner les moyens de distinguer ces niveaux qu'une étude a été menée en 2013. D'autres études complémentaires avaient d'ailleurs déjà été menées, en France, après le PISA 2003, pour étudier la correspondance entre les items du questionnaire et les pratiques usuelles d'enseignement en fonction des programmes scolaires en vigueur dans notre pays (Bodin 2009).

## 2. *Différents niveaux de mise en fonctionnement des connaissances mathématiques dans les items de mathématiques du PISA 2012*

L'étude complémentaire dont il est question a été réalisée par un groupe d'experts de la DEPP, c'est elle en effet qui, en France, administre le test PISA. Le groupe était composé d'enseignants, de formateurs, d'inspecteurs et d'un professeur des universités, didacticien des mathématiques et auteur de cette communication.

Les analyses produites se fondent sur des apports de la recherche en didactique des mathématiques et conduisent à des interprétations nouvelles des résultats du PISA. La nouvelle classification s'applique à tous les items du questionnaire, elle repose sur une analyse de l'énoncé visant à déterminer la nature de la mise en fonctionnement de connaissances mathématiques nécessaire pour répondre à la question posée dans l'item<sup>3</sup>. Elle

---

<sup>3</sup> L'étude porte donc sur tous les items de toutes les questions « papier-crayon » du PISA 2012. Le travail ne peut être montré que partiellement, faute de place et pour des raisons de confidentialité, ce qui biaise parfois peut-être le regard que peut porter le lecteur sur les analyses produites des items analysés dans ce texte. Des analyses complémentaires figurent dans une publication liée à cette recherche (Roditi & Salles, 2015). En outre, d'autres questions ont été soumises à des élèves, le travail s'effectuant par ordinateur cette fois. Leur analyse n'a pas été

est par conséquent indépendante des passations préalables qui permettent de déterminer la difficulté relative des items et leur pouvoir discriminant, en référence à la théorie de la réponse à l'item utilisée par les experts du PISA.

Les didacticiens se sont encore peu consacrés aux questions d'évaluation hormis pour l'analyse des productions d'élèves en situations scolaires afin de comprendre les conceptions que produit l'enseignement quant aux notions dont les recherches sont l'objet (Roditi 2012). Il est intéressant toutefois de tirer profit des travaux produits en didactique pour l'analyse de situations d'enseignement afin d'étudier des questions d'évaluation. Les items du PISA peuvent ainsi être différenciés selon deux premières catégories de mise en fonctionnement des connaissances mathématiques : ceux pour lesquels la réponse repose uniquement sur la compréhension qualitative de contenus – concepts, théorèmes, etc. –, sans mise en fonctionnement de la part de l'élève, et ceux qui nécessitent la mise en œuvre d'une procédure reposant sur des contenus mathématiques. Les questions posées dans le PISA émergent toujours de situations liées à la vie réelle. Les items de la première catégorie évaluent ainsi la compréhension d'un savoir mathématique en contexte, mais seulement en tant qu'objet, les élèves n'ayant pas à l'utiliser. Les items de la seconde catégorie évaluent, en revanche, l'acquisition de ces savoirs en tant qu'outils, c'est-à-dire la capacité à les mettre en œuvre pour résoudre un problème. Cette distinction entre les caractères objet et outil des savoirs mathématiques avait été effectuée par Douady (1986), didacticienne, pour rendre compte de la dynamique à l'œuvre lors de la construction de nouvelles connaissances mathématiques, ces deux caractères entretenant une relation dialectique au cours de l'activité.

Ainsi, certaines questions d'évaluation portent sur des contenus mathématiques pour attester de leur compréhension pour eux-mêmes, elles visent le caractère objet de ces contenus. C'est le cas des exercices classiques d'entraînement de calcul numérique ou algébrique où les élèves attestent de leur capacité à effectuer une opération sans même que soit interrogée l'opportunité de poser cette opération dans un problème. Comme cela a déjà été expliqué, il n'y a pas d'items de la sorte dans le PISA. Il y a, en revanche, des items qui portent sur le caractère objet d'un concept, et où les élèves doivent témoigner d'une compréhension de ce concept sans avoir à le mettre en œuvre, ce que certains auteurs appellent une compréhension conceptuelle (Kilpatrick & al. 2001). Ce serait le cas, par exemple, d'un item demandant si un enfant qui jette un dé qui est tombé sur 6 la première fois, possède plus ou moins de chance d'obtenir 6 la deuxième fois. Il s'agirait seulement d'exprimer par une réponse sa compréhension de l'indépendance des événements aléatoires. Les savoirs ainsi évalués dans le PISA concernent souvent la probabilité, la notion de moyenne, les fonctions et les grandeurs. Nous avons réuni ces items dans une même catégorie appelée *Compréhension qualitative de concepts* ou plus simplement *Concept*.

D'autres questions évaluent le caractère outil des savoirs, l'élève doit alors mettre une connaissance mathématique en fonctionnement après s'être assuré de la pertinence de cette connaissance pour traiter la question posée dans le contexte indiqué. Nous distinguons ces mises en fonctionnement suivant qu'elles sont plus ou moins suggérées par l'énoncé, suivant aussi le degré d'initiative demandée à l'élève. Cela correspond en effet, selon nous, à différents niveaux d'acquisition des connaissances. En nous inspirant de travaux déjà effectués sur ce sujet en didactique (Robert 1998), nous considérons trois niveaux de mise en fonctionnement des contenus mathématiques.

Le premier d'entre eux est celui où l'élève effectue une tâche courante et obtient directement le résultat attendu par la mise en œuvre d'une procédure, souvent unique, qui est



indiquée ou suggérée par l'énoncé, et dont les programmes scolaires permettent de penser qu'elle est automatisée pour les élèves. Dans les items du PISA, de tels items conduisent généralement à l'application d'une propriété géométrique, d'une règle de calcul, d'une lecture graphique directe, etc. Les items correspondant à ce premier niveau de mise en fonctionnement sont regroupés dans une catégorie appelée *Mise en fonctionnement directe d'une procédure* ou plus simplement *Directe*.

Les items qui relèvent du second niveau nécessitent que l'élève adapte ou transforme l'énoncé – les données ou la question posée – avant d'appliquer ses connaissances. La transformation peut prendre la forme d'une transformation d'information : convertir, par exemple, une donnée dans une autre unité de mesure. Il peut s'agir d'un changement de point de vue sur des objets mathématiques ou sur une relation entre des objets : isoler, par exemple, une figure plane d'une figure de l'espace ; ou, ayant à établir que trois points sont alignés, considérer la droite qui passe par les deux premiers et montrer que le troisième appartient à cette droite. L'élève peut aussi avoir à changer de cadre (Douady 1986) ou de registre de représentation (Duval 1995) : passer, par exemple, dans le cadre graphique pour résoudre un problème numérique ; convertir une procédure indiquée dans le registre langagier en un calcul appartenant au registre numérique ou algébrique. Tous ces items ont été regroupés dans une catégorie appelée *Mise en fonctionnement d'une procédure avec adaptation de l'énoncé* ou plus simplement *Adaptation*.

Dans les items du troisième niveau, la mise en fonctionnement des contenus nécessite que l'élève, de manière autonome, introduise un ou plusieurs intermédiaires. Ils peuvent concerner le processus de résolution lui-même : décomposer une question en plusieurs étapes ; introduire une notation pour traiter le problème, par exemple en attribuant une lettre à différentes variables ; etc. Il peut s'agir également d'intermédiaires ajoutés aux données : considérer une nouvelle variable combinant, par exemple, deux variables déjà explicitées dans un problème numérique ; utiliser un nouvel objet géométrique pour résoudre un problème, par exemple en considérant une droite ou un cercle qui n'apparaît pas dans l'énoncé ; introduire une fonction là où deux variables étaient indiquées avec une relation numérique les reliant ; etc. Parce qu'il s'agit d'un intermédiaire qui n'est pas suggéré par l'énoncé, l'introduction correspond à une initiative de la part de l'élève, elle est totalement à sa charge. Les items de ce type sont regroupés dans une catégorie appelée *Mise en fonctionnement d'une procédure avec introduction d'intermédiaires* ou plus simplement *Intermédiaires*.

La distinction de ces quatre catégories de mises en fonctionnement des connaissances mathématiques, qui portent sur leur dimension objet comme sur leur dimension outil, conduit à poser un nouveau regard sur les items du PISA ainsi que sur les résultats produits par ce programme. Le fait que les items du PISA ne soient pas tous libérés nous interdit d'en montrer l'analyse exhaustive à l'aide de cette nouvelle classification. Deux exemples sont néanmoins proposés qui montrent son intérêt pour apprécier un item. Puis nous développerons ce qu'apporte ce nouveau regard sur les résultats du PISA 2012 en mathématiques.

## II. APPORTS DE LA NOUVELLE CLASSIFICATION A L'ANALYSE D'UN ITEM DE MATHÉMATIQUES DU PISA 2012

Illustrons l'intérêt de cette nouvelle classification pour l'étude de deux items du questionnaire de culture mathématique du PISA 2012. Comme tous les items du PISA, ces deux items sont liés à des contextes de la vie réelle, néanmoins, comme nous l'avons déjà signalé, ce contexte n'a pas toujours d'influence sur l'activité de l'élève. Les deux items analysés sont regroupés dans la catégorie « employer » selon le cadre du PISA. C'est bien le cas dans l'exemple ci-dessous (figure 1) où l'étude d'une roue de manège est proposée mais où l'activité de l'élève

porte essentiellement sur la figure géométrique représentée dans l'illustration figurant dans l'énoncé.

Une grande roue est installée sur les rives d'un fleuve.  
En voici un dessin et un schéma :

Le diamètre externe de la grande roue est de 140 mètres et son point le plus élevé se situe à 150 mètres au-dessus du lit du fleuve. Elle tourne dans le sens indiqué par les flèches.

---

**Question : LA GRANDE ROUE**

La lettre  $M$  dans le diagramme indique le centre de la roue.

À combien de mètres (m) au-dessus du lit du fleuve se trouve le point  $M$  ?

Réponse : ..... m

Figure 1 – Item libéré du PISA 2012

Dans cet item en effet, l'élève peut appliquer les propriétés du diamètre et du rayon d'un cercle à une figure où les mesures sont de simples nombres entiers puis tenir compte de la longueur séparant le bas de la roue avec le lit du fleuve. La figure proposée rend possible une utilisation implicite de ces connaissances puisque le point  $M$ , défini comme le centre du cercle, est placé au milieu du segment  $[PR]$  qui en est un diamètre. Différentes procédures directes et équivalentes sont envisageables : 1°) calculer la moitié de 140 et ajouter 10 ; 2°) enlever la moitié de 140 à 150 ; etc. Ce qui est important ici, c'est de noter que l'activité de l'élève porte seulement sur la figure géométrique donnée dans l'énoncé, c'est pourquoi l'item est associé au processus psycho-cognitif *employer* et au domaine mathématique *espace et formes*. C'est de remarquer ensuite que la résolution de l'exercice nécessite une application *directe* des connaissances mathématiques.

Examinons maintenant un autre item (Figure 2) portant lui aussi sur une situation du champ géométrique et ne demandant pas non plus à l'élève d'utiliser le contexte dans son activité de résolution du problème posé.

Voici le plan du magasin de glaces de Marie, qu'elle est en train de rénover.

La zone de service est entourée d'un comptoir.

Remarque : Chaque carré de la grille représente 0,5 mètre sur 0,5 mètre.

---

**Question 1 : CHEZ LE GLACIER**

Marie veut installer une nouvelle bordure le long de la paroi extérieure du comptoir. Quelle est la longueur totale de bordure dont elle a besoin ? Montrez votre travail.

Figure 2 – Item libéré du PISA 2012

Après lecture de l'énoncé et identification de la paroi extérieure du comptoir sur le plan, plusieurs méthodes de résolution sont possibles pour un élève en fin de scolarité obligatoire en France. Par exemple une méthode par mesure et application d'échelle est possible : déterminer l'échelle en mesurant à la règle graduée la longueur de deux carreaux sur le dessin qui représentent un mètre dans la réalité ; mesurer la longueur sur le plan de la paroi extérieure ; appliquer enfin l'échelle précédemment déterminée pour calculer la longueur réelle recherchée. Il est à noter qu'une valeur approchée de la réponse sera acceptée par le correcteur, les consignes de correction internationales étant appliquées. L'élève peut également travailler tout autrement et mener un raisonnement géométrique fondé sur le théorème de Pythagore après avoir introduit sur le plan un triangle rectangle dont l'hypoténuse est la partie oblique du comptoir. Le contexte du magasin de glace n'intervenant pas sur l'activité permettant de trouver la bonne réponse, cet item est, comme le précédent, associé au processus psycho-cognitif *employer* et au domaine de connaissances mathématiques *espace et formes*. La résolution de l'exercice, quelle que soit la méthode, nécessite l'introduction d'intermédiaires et cette introduction est laissée à l'initiative des élèves.

Dans les deux items que nous venons d'examiner, on ne peut être certain de la connaissance mathématique mobilisée pour répondre aux questions posées. Le PISA ne

cherche pas à connaître précisément les connaissances mathématiques acquises par les élèves, mais seulement leur domaine parmi les quatre qui sont distingués pour cette discipline. Il ne vise pas non plus à rendre compte des différentes modalités d'expression des connaissances mathématiques acquises par les élèves. Les deux items se retrouvent en effet classés dans les mêmes catégories *employer* et *espace et formes* par les experts du PISA, ils requièrent pourtant des modalités d'expression des connaissances mathématiques très différentes. Dans le premier, l'élève effectue un calcul explicitement demandé dans la consigne, ce calcul portant sur deux longueurs clairement indiquées sur une figure elle-même fournie dans l'énoncé. Dans le second, l'emploi d'un calcul d'échelle ou d'un raisonnement basé sur le théorème de Pythagore nécessite des étapes qui, n'étant absolument pas induites par l'énoncé, sont entièrement à la charge de l'élève. Ces deux items évaluent donc bien la capacité à *employer* des connaissances dans des situations géométriques déjà mathématisées, mais ils ne sont absolument pas équivalents quant au niveau de mise en fonctionnement de ces connaissances. L'échec ou la réussite à ces deux items ne témoigne donc pas du même niveau d'acquisition. C'est ce dont la classification que nous proposons permet de justement rendre compte.

La classification des items à partir d'une analyse de la mise en fonctionnement des savoirs mathématiques requise (Concept, Directe, Adaptation ou Intermédiaires) a en effet abouti à une étude systématique de l'ensemble du questionnaire afin de mieux connaître les questions posées aux élèves et de mieux comprendre les résultats de l'enquête du PISA 2012.

### III. ANALYSES COMPLEMENTAIRES DU QUESTIONNAIRE DE CULTURE MATHÉMATIQUE ET DES RESULTATS DU PISA 2012

#### 1. *Analyse complémentaire du questionnaire du PISA 2012 à l'aide de la classification des items selon les niveaux requis de mise en fonctionnement*

La répartition des 85 items de mathématiques du PISA 2012 selon le niveau requis de mise en fonctionnement des connaissances confirme que l'OCDE vise essentiellement l'évaluation de la capacité à utiliser ses acquis en tant qu'outil dans des situations issues de la vie réelle plutôt que l'acquisition de notions pour elles-mêmes en tant qu'objet d'étude. Seulement 7 items concernent en effet la compréhension qualitative d'un concept. Les 78 autres se répartissent assez équitablement selon les trois niveaux de mise en fonctionnement : on en dénombre 29 de la catégorie regroupant les items nécessitant la mise en œuvre directe d'une procédure connue, 27 exigeant une adaptation de l'énoncé et 22 nécessitant de prendre l'initiative d'introduire des intermédiaires.

Nous avons ensuite mené une analyse croisée de la répartition des items suivant, d'une part, la nouvelle catégorisation didactique et, d'autre part, une catégorie de l'OCDE : le domaine mathématique d'abord, et le processus psycho-cognitif ensuite. Le croisement avec le domaine mathématique conduit au tableau n°1 où figurent, dans chaque case, l'effectif des items, à gauche, et le pourcentage-ligne, entre parenthèses à droite.

Classifications		Didactique				
		Concept	Directe	Adaptation	Intermédiaires	Total
PISA	Espace et formes	8 (9%)	2 (10%)	7 (33%)	12 (57%)	21 (100%)
	Inconnus et données	5 (24%)	7 (33%)	7 (33%)	2 (10%)	21 (100%)
	Quantité	0 (0%)	15 (68%)	4 (18%)	3 (14%)	22 (100%)
	Variations et relations	2 (9%)	3 (14%)	9 (43%)	5 (24%)	21 (100%)
	<b>Total</b>	<b>7 (8%)</b>	<b>29 (34%)</b>	<b>27 (32%)</b>	<b>22 (26%)</b>	<b>85 (100%)</b>

Tableau 1 - Domaines mathématiques et niveaux de mise en fonctionnement

La dernière colonne du tableau montre la volonté des experts du PISA 2012 de répartir les questions mathématiques de manière équivalente suivant chacun des quatre domaines (21 ou 22 items par domaine, soit un quart des 85 items du test complet). Alors que le niveau de mise en fonctionnement des connaissances se détermine indépendamment des contenus mathématiques, les écarts qui apparaissent dans le tableau ci-dessus révèlent que ces différents niveaux ne sont pas évalués indépendamment des champs mathématiques. Inversement, notre nouvelle classification révèle que les savoirs en jeu dans les items PISA ne sont pas évalués de manière équivalente puisque les items ne conduisent pas à les mettre tous en fonctionnement aux mêmes niveaux. Par exemple, ceux du champ *quantité* sont essentiellement évalués par des tâches nécessitant la mise en œuvre directe d'une procédure connue (68% des items) alors que ceux du champ *espace et formes* le sont plus souvent par des problèmes nécessitant l'introduction d'intermédiaires (57% des items).

Une étude analogue a été menée concernant l'évaluation des processus psycho-cognitifs et des niveaux de mise en fonctionnement des connaissances mathématiques (tableau n°2).

Classifications		Didactique				
		Concept	Directe	Adaptation	Intermédiaires	Total
PISA	Employer	1 (3%)	16 (43%)	10 (27%)	10 (27%)	37 (100%)
	Formuler	2 (7%)	6 (22%)	8 (30%)	11 (41%)	27 (100%)
	Interpréter	4 (15%)	7 (26%)	9 (33%)	1 (4%)	21 (100%)
	<b>Total</b>	<b>7 (8%)</b>	<b>29 (34%)</b>	<b>27 (32%)</b>	<b>22 (26%)</b>	<b>85 (100%)</b>

Tableau 2 - Processus psycho-cognitifs et niveaux de mise en fonctionnement

Quelques écarts apparaissent qui témoignent du fait que le niveau de mise en fonctionnement des connaissances n'est pas évalué indépendamment des processus psycho-cognitifs et inversement. Ainsi, par exemple, la capacité à prendre l'initiative d'introduire des intermédiaires est davantage testée dans les items où le processus attendu est « formuler » et pratiquement jamais dans ceux où l'élève doit *interpréter*. De même, la capacité à « employer » des connaissances mathématiques est surtout évaluée par des items où c'est une mise en fonctionnement directe de procédure qui est requise.

## 2. Nouveau regard sur les résultats du PISA 2012 apporté par la prise en compte des niveaux de mise en fonctionnement des connaissances

La classification des items selon le niveau requis de mise en fonctionnement des connaissances s'effectue indépendamment de toute mesure de difficulté. Les trois niveaux concernant les items qui portent sur le caractère outil des savoirs différencient ces items selon une activité mathématique de plus en plus riche et autonome. Néanmoins, nous avons observé

que le niveau de mise en fonctionnement à lui seul ne pouvait expliquer la difficulté d'un item : de nombreux autres facteurs interviennent comme la connaissance en jeu, la familiarité avec le contexte du problème, la lisibilité de l'énoncé, etc. Ainsi, les items de chaque niveau de mise en fonctionnement se répartissent dans presque tous les niveaux de difficulté – tels qu'ils sont définis par PISA. L'étude complète du questionnaire permet de croiser le niveau croissant de mise en fonctionnement des connaissances avec le niveau croissant de difficulté des items. L'analyse révèle, d'une part, une dispersion relativement importante de la difficulté des items de chaque niveau de mise en fonctionnement, ce qui confirme que ce critère n'est pas suffisant pour prévoir la difficulté d'un item. Elle montre aussi, d'autre part, qu'en moyenne, les niveaux *directe*, *adaptation* et *intermédiaires*, qui correspondent à une exigence croissante de l'activité mathématique, correspondent également à une difficulté croissante pour les élèves. Les items de ces trois niveaux sont en effet réussis en moyenne par respectivement 59,3 %, 46,8 % et 33,9 % des élèves scolarisés en France et, de manière comparable, par 59,8 %, 45,1 % et 34,8 % des élèves scolarisés dans les pays de l'OCDE. Signalons enfin que le cas des items de la catégorie *concept* n'est pas examiné car leur effectif dans le questionnaire PISA est trop faible pour permettre des interprétations.

Cette étude globale a été complétée par une étude des sous-groupes respectivement définis selon le sexe, la catégorie socio-professionnelle et le retard scolaire des élèves de 15 ans scolarisés en France.

La publication du PISA sur les réussites aux items de culture mathématique révèle notamment que les filles scolarisées en France, en moyenne, réussissent moins bien que les garçons : la différence de réussite est de 2,5 pp (points de pourcentage) à l'avantage des garçons). L'étude des niveaux de mise en fonctionnement des connaissances apporte quelques informations supplémentaires. L'écart de performance à la faveur des garçons est de 1,5 pp pour les items qui requièrent la mise en œuvre directe d'une procédure connue et de 3,3 pp pour ceux qui nécessitent l'introduction d'un intermédiaire. Autrement dit, les filles sont d'autant plus en difficulté par rapport aux garçons que le niveau requis de mise en fonctionnement des connaissances est un niveau exigeant.

Une étude analogue a été menée concernant le lien entre les catégories socioprofessionnelles (CSP) des élèves et leur réussite. Un des constats majeurs de l'étude PISA 2012 pour la France est que notre système éducatif est fortement différenciateur : les élèves issus de milieux défavorisés obtiennent une performance moyenne de 39,4 % de réussite contre 57,4 % pour ceux de milieux favorisés, soit un écart de 18 pp. En outre, un tel écart de réussite est constaté pour tous les items, sa valeur allant de 1,9 pp pour le plus faible à 31,9 pp pour le plus élevé. L'étude complémentaire menée par la DEPP montre, contrairement à ce qui a été constaté concernant les différences entre filles et garçons, que les différences de réussite selon les CSP restent stables lorsque le niveau requis de mise en fonctionnement des connaissances augmente. Autrement dit, les élèves de milieu défavorisé n'apparaissent pas plus désavantagés que ceux de milieu favorisé par l'exigence d'autonomie mathématique requise dans les items.

Le dernier aspect étudié ici est celui du retard scolaire. L'étude du PISA révèle que les élèves ayant redoublé au moins une fois dans leur scolarité obtiennent une réussite moyenne de 28,9 % aux 85 items de mathématiques, alors qu'elle est de 56,0 % pour les autres, soit un écart moyen de 27,2 pp. Pour tous les items, les élèves à l'heure réussissent mieux que les élèves en retard, la différence de réussite allant de 3,1 pp pour la plus faible à 46,7 pp pour la plus élevée. L'étude menée par la DEPP met en lumière le fait que la différence de réussite entre les élèves scolairement en retard et les élèves à l'heure n'est pas constante lorsque varie le niveau requis de mise en fonctionnement des connaissances. Ainsi, et peut-être contre l'idée qu'on pourrait avoir *a priori*, l'écart de performance est d'autant plus faible que le

niveau de mise en fonctionnement est élevé : il est de 22,3 pp pour les tâches nécessitant l'introduction d'un intermédiaire et de 30,6 pp pour celles qui se réalisent par la mise en œuvre directe d'une procédure connue. Autrement dit, les élèves en retard sont plus souvent mis en difficulté par des tâches routinières que par celles qui nécessitent davantage d'initiative. Ici encore, ces résultats invitent à s'interroger sur le système éducatif français et les pratiques des enseignants, en particulier sur les activités proposées aux élèves ayant rencontré des difficultés qui ont conduit à un redoublement au cours de leur scolarité.

#### IV. CONCLUSION

Les enquêtes du PISA visent donc un suivi des acquis scolaires des élèves de 15 ans. En ce qui concerne ceux de la culture mathématique, le choix de l'OCDE est d'évaluer des compétences, c'est-à-dire des capacités à mobiliser ses connaissances pour résoudre un problème en lien avec une situation de la vie réelle. Un regard didactique porté sur l'évaluation de 2012 ne peut manquer de pointer que l'OCDE ne se donne les moyens ni de recenser précisément les connaissances acquises des élèves ni d'estimer le niveau d'acquisition de ces connaissances. Inversement, les didacticiens qui ont concentré leurs recherches sur les phénomènes d'enseignement et d'apprentissage des savoirs n'ont pas suffisamment développé d'outils théoriques et pratiques pour étudier la question de l'évaluation des connaissances des élèves.

Une classification des items issue de la recherche en didactique permet de les distinguer suivant différents niveaux de mise en fonctionnement des connaissances mathématiques et donc, d'une certaine manière, d'évaluer le niveau d'acquisition de ces connaissances. Cette nouvelle classification permet de différencier des items que les catégories définies par les experts de l'OCDE ne permettent pas de distinguer, qui requièrent pourtant des niveaux différents de mise en fonctionnement des connaissances évaluées et qui conduisent à des scores de réussite significativement différents.

Une étude complète de l'ensemble des items du PISA 2012 a été menée à l'aune de cette nouvelle classification. Elle montre d'une part que l'OCDE évalue peu la compréhension qualitative des concepts mathématiques. Elle montre également que les trois autres niveaux de mise en fonctionnement des connaissances, qui correspondent à une exigence croissante de richesse et d'autonomie de l'activité, correspondent également, en moyenne, à un niveau de difficulté croissant pour les élèves. Une attention particulière a ensuite été portée sur le cas de la France, et notamment les inégalités de performances selon le sexe, l'origine sociale ou le retard scolaire. L'OCDE, dans son rapport, indique une meilleure réussite des garçons ; l'analyse, en s'appuyant sur la classification didactique, montre en outre que les filles sont d'autant plus pénalisées que les tâches demandent de l'initiative. Concernant les élèves de milieux populaires comme les élèves en retard scolaire, l'étude s'appuyant sur cette même classification révèle enfin que ces élèves ne sont pas mis davantage en difficulté lorsque les activités attendues d'eux sont plus exigeantes.

Cette étude, réalisée à partir de quelques outils issus de la didactique des mathématiques, apporte des résultats qui permettent de poser un regard nouveau sur le PISA et ses conclusions. Ce croisement d'approche – didactique et évaluative – sur les apprentissages scolaires s'avère fructueux. Certains chercheurs tentent depuis quelques années d'approfondir une telle démarche (Vantourout & Goasdoué 2011 ; Sayac 2012 ; Chesné 2014 ; Sayac & Grapin 2015), gageons qu'ils ouvriront de nouvelles perspectives, pour l'enseignement des mathématiques comme pour la recherche en didactique.

## REFERENCES

- Bardini C. (2015) Computer-based assessment of Mathematics in PISA 2012. In Stacey K., Turner R. (Eds.) *Assessing Mathematical Literacy – The PISA Experience*. Springer.
- Bodin A. (2009) L'étude PISA pour les mathématiques. Résultats français et réactions. *Gazette de la SMF* 120, 53-67.
- Chesné J.-F. (2014) *D'une évaluation à l'autre : des acquis des élèves sur les nombres en sixième à l'élaboration et à l'analyse d'une formation d'enseignants centrée sur le calcul mental*. Thèse de doctorat de l'Université Paris Diderot.
- Douady R. (1986) Jeux de cadre et dialectique outil-objet. *Recherches en didactique des mathématiques* 7(2), 5-31.
- Duval R. (1995) *Semiosis et pensée humaine: registres sémiotiques et apprentissages intellectuels*. Berne: Peter Lang.
- Kilpatrick J., Swafford J., Findel B. (2001) *Adding it up: Helping children learn mathematics*, Washington: National Academy Press, pp. 115-135.
- OCDE (2013) *Cadre d'évaluation et d'analyse du cycle PISA 2012*. Paris : OCDE.
- OCDE (2014) *Résultats du PISA 2012 : savoirs et savoir-faire des élèves : Performance des élèves en mathématiques, en compréhension de l'écrit et en sciences (Volume I)*. PISA, Editions OCDE.
- OCDE (2014) *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. PISA, Editions OCDE, 98-106.
- Robert A. (1998) Outil d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en didactique des mathématiques* 18(2), 139-190.
- Roditi E. (2012) Un point de vue didactique sur les questions d'évaluation en éducation. In Lattuati M., Penninckx J., Robert A. (Eds.) *Une caméra au fond de la classe de mathématiques* (pp. 275-289). Besançon : Presses universitaires de Franche-Comté.
- Roditi E., Salles F. (2015) Nouvelles analyses de l'enquête PISA 2012 en mathématiques. *Éducation et formations* 86-87, 236-267.
- Sayac N. (2012) Évaluations nationales ou internationales : limites et perspectives. *Actes en ligne du colloque Sociologie et Didactiques*. Lausanne (Suisse).
- Sayac N., Grapin N. (2015) Évaluation externe et didactique des mathématiques : un regard croisé nécessaire et constructif. *Recherches en didactique des mathématiques* 35(1), 101-126.
- Vantourout M., Goasdoué R. (2011) Correction de dissertations en SES. *Idées* 63, 71-78.