

Pédagogies en développement

L'évaluation à la lumière des contextes et des disciplines

Pascal Detroz
Marcel Crahay
Annick Fagnant
(sous la direction de)

L'évaluation à la lumière des contextes et des disciplines

Collection dirigée par
Jean-Marie De Ketele

Comité scientifique international
Daniel Chevolet (Rennes I) ; Claude Tapia (Tours) ; Ben Omar Boubker (Rabat) ;
André Girard (Québec) ; Claudine Tahiri (Côte d'Ivoire)

L'évaluation à la lumière des contextes et des disciplines

Pédagogies en développement

Pascal Detroz
Marcel Crahay
Annick Fagnant
(sous la direction de)

Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web : www.deboecksuperieur.com

© De Boeck Supérieur s.a., 2016
Rue du Bosquet, 7 – B-1348 Louvain-la-Neuve

Tous droits réservés pour tous pays.

Il est interdit, sauf accord préalable et écrit de l'éditeur, de reproduire (notamment par photocopie) partiellement ou totalement le présent ouvrage, de le stocker dans une banque de données ou de le communiquer au public, sous quelque forme et de quelque manière que ce soit.

Imprimé aux Pays-Bas

Dépôt légal :

Bibliothèque nationale, Paris : janvier 2017

Bibliothèque Royale de Belgique, Bruxelles : 2017/13647/016

ISSN 0777-5245

ISBN 978-2-807-30715-5

S O M M A I R E

LISTE DES AUTEURS	9
INTRODUCTION	11
Par Pascal DETROZ, Marcel CRAHAY et Annick FAGNANT	
CHAPITRE 1	
Évaluations à large échelle : prendre la juste mesure des effets de contexte	21
Par Dominique LAFONTAINE	
AXE 1 : ÉVALUATIONS ET CONTEXTES DIDACTIQUES	
Introduction à l'axe 1	55
Par Annick FAGNANT	
CHAPITRE 2	
Évaluer en mathématiques : une approche didactique et épistémologique	63
Par Brigitte GRUGEON-ALLYS	
CHAPITRE 3	
L'évaluation, un paramètre prépondérant en didactique du français langue étrangère et seconde	91
Par Fatima CHNANE-DAVIN et Jean-Pierre CUQ	

CHAPITRE 4

**Évaluation formative de l'apprentissage
par la lecture : relation complexe et dynamique
« individu-contexte »** 111

Par Sylvie C. CARTIER, Thaïs GUERTIN-BARIL, Julie ARSENEAULT
et Manal RAOUI

**AXE 2 :
ÉVALUATION ET MODÈLES COGNITIFS****Introduction à l'axe 2 : De la docimologie
à la psychologie de l'évaluation** 133

Par Marcel CRAHAY

CHAPITRE 5

**Évaluations scolaires et étude du jugement
des enseignants : pour une docimologie cognitive** .. 141

Par Rémi GOASDOUÉ et Marc VANTOUROUT

CHAPITRE 6

**Une modélisation pour appréhender
la référentialisation dans l'évaluation
des apprentissages des élèves** 169

Par Lucie MOTTIER LOPEZ

CHAPITRE 7

**Le redoublement des élèves :
des décisions prises dans l'incertitude** 193

Par Géry MARCOUX et Marcel CRAHAY

**AXE 3 :
ÉVALUATION DANS L'ENSEIGNEMENT SUPÉRIEUR,
ORIENTATIONS DISCIPLINAIRES
ET APPROCHES PÉDAGOGIQUES**

Introduction à l'axe 3	213
Par Pascal DETROZ et Nathalie YOUNES	
CHAPITRE 8	
Viser le développement des pratiques d'Assessment for Learning (AFL) dans le contexte de la 1^{re} année à l'université : cadre théorique, dispositif et initiatives pédagogiques résultant du projet « Feedbacks 1^{er} bac »	221
Par Laurent LEDUC, Audrey MOHR, Eléonore MARICHAL, Mélanie DEUM et Pascal DETROZ	
CHAPITRE 9	
Comment former à une vision élargie de l'évaluation formative en vue de réguler efficacement les apprentissages mathématiques des élèves ?	257
Par Annick FAGNANT et Isabelle DEMONTY	
CHAPITRE 10	
Regards croisés sur deux dispositifs de formation technopédagogique, leur outillage conceptuel et l'évaluation de leurs bénéfices	275
Par Catherine LOISY, Jean-François VAN DE POËL et Dominique VERPOORTEN	
LISTE DES TABLEAUX ET FIGURES	305
TABLE DES MATIÈRES	309

LISTE DES AUTEURS

Julie Arseneault

Université de Montréal

Sylvie C. Cartier

Université de Montréal

Fatima Chnane-Davin

Aix-Marseille Université

Marcel Crahay

Professeur honoraire

Universités de Genève et de Liège

Jean-Pierre Cuq

Université Nice Sophia Antipolis

Isabelle Demonty

Université de Liège

Pascal Detroz

Université de Liège

Mélanie Deum

Université de Liège

Annick Fagnant

Université de Liège

Rémi Goasdoué

Université Paris Descartes

Brigitte Grugeon-Allys

Université Paris-Est Créteil

Thaïs Guertin-Baril

Université de Montréal

Dominique Lafontaine

Université de Liège

Laurent Leduc

Université de Liège

Catherine Loisy

École normale supérieure de Lyon

Géry Marcoux

Université de Genève

Eléonore Marichal

Université de Liège

Audrey Mohr

Université de Liège

Lucie Mottier Lopez

Université de Genève

Manal Raoui

Université de Montréal

Jean-François Van de Poël

Université de Liège

Marc Vantourout

Université Paris Descartes

Dominique Verpoorten

Université de Liège

Nathalie Younes

Université Blaise Pascal de Clermont-Ferrand

Pascal DETROZ, Marcel CRAHAY et Annick FAGNANT

1. L'ÉVALUATION, CE CONCEPT QUI S'ÉTEND JUSQU'À DEVENIR INSAISSABLE

Il fut un temps où aborder le concept d'évaluation était plus simple. Un temps où l'on ne risquait pas, en embrassant cette thématique, de s'égarer dans une profusion de modèles, de postures ou encore de valeurs. Un temps où emprunter le chemin de l'évaluation n'était pas synonyme d'emprunter un « chemin sinueux, un chemin que croiseraient plusieurs sentiers propices à la distraction et parfois à la confusion » (Scallon, 2004, p. 320).

L'évaluation n'a jamais été aussi présente dans nos vies quotidiennes. En tant qu'enseignants et chercheurs, nous sommes cernés par des dispositifs d'évaluation. À certains moments, nous sommes les évaluateurs, à d'autres les évalués. À tel point que, si l'évaluation était valorisée sous forme d'actions cotées en bourse, un individu ayant misé quelques francs sur celle-ci dans les années 1960 verrait sans doute sa fortune faite aujourd'hui.

Mais paradoxalement, le concept d'évaluation a rarement été aussi peu limpide. Comme tout élément en forte croissance, l'évaluation traverse

1. Cet ouvrage fait suite au 27^e colloque de l'ADMEE Europe qui a été organisé à l'Université de Liège du 17 au 20 janvier 2015. La thématique de ce colloque est donc forcément identique à celle du présent ouvrage, à savoir « L'évaluation à la lumière des contextes et des disciplines ». En conséquence, la présente introduction s'inspire en partie de la description de la thématique telle qu'elle a été présentée dans l'appel à communication de l'ADMEE.

une crise profonde. De nombreuses voix s'élèvent contre ce « tout à l'évaluation » qui nous enferme derrière des grilles (Barbara Cassin, 2014) et qui n'est pas dénué de faces sombres (Martucelli, 2010). Certes, ces critiques ne sont pas récentes. Par exemple, Ardoino & Berger (1989) nous mettaient déjà en garde contre une évaluation « en miette » qu'ils opposaient à une évaluation « en acte ». Mais aujourd'hui, les critiques et questionnements sont prégnants. Ils questionnent les fondements mêmes de ce concept. Tardif (2006, p. 134) n'écrivait-il pas, à propos de l'évaluation dans une approche par compétences, que « Les exigences de l'évaluation des compétences sont nombreuses et elles posent des défis de taille. D'aucuns pourraient dès maintenant penser qu'une telle entreprise est impossible étant donné les théories et les instruments développés à ce jour dans le domaine de l'évaluation des apprentissages » ? Sur le même sujet, les propos de De Ketele et Gérard (2005, p. 22) étaient tout aussi peu encourageants : « Peut-on espérer un jour disposer d'une éduométrie satisfaisante pour évaluer des compétences complexes ? Rien n'est moins sûr dans l'état actuel de nos connaissances ». Citons aussi, dans le même ordre d'idées, la conclusion de l'article de Vandermaeren et Loye (2011, p. 52) qui, après une vision critique de l'évaluation des compétences, notaient dans leur conclusion : « Notre réflexion critique, très sinon trop critique selon certains collègues, ne fait que relever l'ampleur et la difficulté de leur tâche et tout le mérite qu'ils peuvent avoir dans leurs tentatives répétées, assidues, ingénieuses de relever ce défi qui, lorsque l'on combine les analyses sociologiques et psychologiques du travail scolaire, peut apparaître parfois comme une tâche impossible à surmonter. »

Nous en sommes à un stade où, à force d'être utilisé tout le temps et en tous lieux, par tout un chacun selon plusieurs sens et fonctions, le concept scientifique d'évaluation se dilue jusqu'à devenir insaisissable.

Considérons à titre d'exemple (et pour s'en tenir uniquement au domaine de l'enseignement²) les dispositifs contemporains suivants : un professeur de médecine questionnant le raisonnement clinique de ses étudiants, des études internationales de type PISA, un test d'entrée donnant accès aux études en sciences appliquées, un enseignant observant un élève en prise avec un problème mathématique pour lui donner un feedback précis, les étudiants notant la qualité de l'enseignement qu'ils ont reçu, un inspecteur étudiant la posture d'un enseignant face à sa classe, un chasseur de têtes faisant un screening du profil d'un candidat à un poste à haute responsabilité dans une administration scolaire, des experts internationaux auditant une Faculté de philosophie et lettres d'une université quelconque, un responsable administratif analysant le rendement de l'un des enseignants de son

2. Car on pourrait également évoquer l'évaluation dans les entreprises, l'évaluation des services et des produits par les clients, etc. Bref, l'évaluation est partout.

département, un élève soumis à un test adaptatif en anglais, le classement « Shanghai » des universités...

Si l'on analyse ces démarches évaluatives et que l'on essaie de déterminer leur(s) point(s) commun(s), la tâche est complexe. Hormis le fait qu'elles traitent toutes, d'une manière ou d'une autre, d'évaluation, il n'est pas aisé de trouver d'autres similarités. En effet, les méthodes, les publics visés, les parties prenantes, les modèles sous-jacents, les finalités, les temporalités, etc. sont tellement divergents qu'il est difficile de considérer ces dispositifs comme faisant partie d'un même ensemble unifié. Le lecteur pourrait argumenter en disant qu'il s'agit dans tous les cas de prendre de l'information et d'effectuer un jugement afin de prendre une décision³, mais peut-on encore se contenter d'un cadre conceptuel aussi large pour répondre à des critiques⁴ de plus en plus précises, instruites et... virulentes.

2. UNE COMPLEXITÉ QUI S'EST DÉVELOPPÉE AU FIL DU TEMPS

Aujourd'hui, la complexité est une facette inhérente à l'évaluation. Il n'en a pas toujours été ainsi... En effet, il fut un temps où « évaluation » était quasi synonyme de « mesure ». On sait aujourd'hui que c'est une chimère, mais les edumétriciens de l'époque cherchaient à obtenir des mesures de qualité similaires à celles qu'obtenaient les psychométriciens qui, eux-mêmes, visaient à obtenir des mesures de qualité identiques à celles des physiciens. La primauté était ainsi donnée à la qualité de la mesure. Pour l'anecdote, on prête à Alfred Binet, cofondateur du célèbre test Binet-Simon, la tirade suivante, lorsqu'on lui demanda de définir l'intelligence : « l'intelligence, c'est ce que mesure mon test ». Même si cette réplique semble apocryphe, elle est signifiante. Nous ne dirions pas que le construit mesuré était à l'époque sans importance, mais la qualité statistique de la mesure passait avant toute autre considération.

Plus tard, les référentiels ont fait leur entrée remarquée dans le domaine de l'évaluation. Le contexte est lié à l'organisation du travail selon les principes du taylorisme, la psychologie est avant tout behavioriste et Tyler (1935) modifie durablement les conceptions pédagogiques en proposant l'approche par objectifs. La conjonction de ces phénomènes va profondément influencer l'évaluation scolaire. Elle va se construire à partir d'un référentiel de ce que les étudiants devraient avoir acquis à un moment donné de leur formation. C'est l'âge d'or des taxonomies d'objectifs. L'idée,

3. Ce qui correspond à la définition de l'évaluation par Stufflebeam *et al.* (1980).

4. Voir par exemple Del Rey (2013), *La tyrannie de l'évaluation* ou encore Miller et Milner (2004), *Voulez-vous être évalué ?* en plus des références citées en amont.

très présente dans la perspective de la mesure, de comparer les étudiants entre eux – le mythe de la courbe de Gauss – est ici supplantée par l'idée de comparer la performance d'un étudiant par rapport à un attendu (référentiel) clairement défini. On mesure des acquis d'apprentissage.

L'évaluation est ensuite considérée comme un outil de gestion. Si l'on arrive à mesurer les acquis des étudiants, on est à même de déceler des décalages par rapport aux attendus. S'il y a un décalage négatif entre ce que l'étudiant sait ou sait faire et ce qui est prescrit dans le référentiel, il s'agit alors d'une erreur menant à une régulation de type cybernétique.

L'avènement du cognitivisme et du socioconstructivisme, associé à l'explosion des savoirs et aux besoins de flexibilité constante en entreprise (qui prend le contre-pied du taylorisme et conduira notamment à vouloir, dans une certaine mesure et avec des désaccords parfois marqués, « remplacer » les objectifs par des compétences), modifie encore une fois la posture des évaluateurs. L'évaluation devient plus dynamique. Elle est porteuse de sens pour les différentes parties prenantes. L'idée est d'éclairer les situations d'enseignement/apprentissage, de les rendre intelligibles pour chacun en vue de construire du sens. Les modèles cognitifs sont construits à partir du comportement des apprenants, mais aussi à partir des stratégies mises en place par les évaluateurs.

Esquissé à la grosse louche et de manière parfois caricaturale⁵, cette vision évolutive du concept d'évaluation poursuit la finalité suivante : mettre en évidence le fait qu'aujourd'hui, divers modèles et postures en lien avec l'évaluation se côtoient, se juxtaposent ou s'intègrent dans nos systèmes éducatifs, rendant complexes la lisibilité et la compréhension même de cette notion d'évaluation. Les modèles s'entrechoquent et ceux dont l'évaluation est le métier ne peuvent plus s'abstenir d'empoigner cette problématique à bras le corps, même si la tâche est ardue. Car le concept d'évaluation semble *a priori* polysémique et flou. Sa définition est large et à le définir de façon trop générale, on court le risque de le vider de sa substance. À l'inverse, en entrant dans sa complexité, on court le risque opposé de produire un discours qui ne soit audible que par quelques initiés.

3. VERS UNE PRISE EN COMPTE DES CONTEXTES

Le parti pris des éditeurs scientifiques du présent ouvrage n'est pas de plaider en faveur d'un modèle – ou d'une logique – d'évaluation plutôt qu'un autre ; il s'agirait plutôt de les mettre en débat. En tant que

5. Pour un historique complet, voir Bonniol et Vial (1997).

spécialistes de l'évaluation, nous sommes autant passionnés par des études psychométriques présentant de nouveaux indices permettant de détecter la réponse au hasard dans les questionnaires standardisés que par des études beaucoup plus écologiques concernant l'utilisation du feedback en sciences par les étudiants de troisième secondaire des écoles liégeoises en Belgique francophone.

Il nous semble par ailleurs stérile d'opposer diverses postures épistémologiques en évaluation qui nous semblent *de facto* plus complémentaires qu'inconciliables... Par exemple, lorsqu'un enseignant met en place un dispositif d'évaluation formative dans le cadre de *l'Assesment for learning* – et que, dès lors, il est surtout très préoccupé par la qualité de son feedback – n'est-il pas nécessaire qu'il se préoccupe aussi de la qualité de l'information ou de la mesure qu'il récolte à travers son dispositif ? Peut-il faire abstraction de la validité de la mesure sous prétexte qu'il vise avant tout son caractère diagnostique ? Poser la question, c'est sans doute y répondre.

En tant que spécialiste de l'évaluation, notre rôle est d'éclairer les enjeux et de dénoncer les éventuelles incohérences que l'on peut trouver dans les dispositifs d'évaluation auxquels nous sommes confrontés. En effet, s'il ne nous semble pas opportun de défendre systématiquement l'un ou l'autre modèle, il est impératif de repérer et d'améliorer des démarches qui ne semblent pas cohérentes. Pour reprendre l'exemple cité en amont, un enseignant qui vise une évaluation formative dans le cadre de *l'Assesment for learning* ne peut résolument pas se contenter de donner le score brut d'un test à un étudiant. Idéalement, s'il souhaite atteindre son objectif, il devra au contraire analyser finement les réponses offertes par l'étudiant afin qu'il puisse se construire une représentation des éventuelles difficultés de ce dernier. Cette analyse devrait donner lieu à un feedback détaillé qui sera le point de départ d'un dialogue collaboratif entre l'enseignant et l'étudiant pouvant déboucher sur des pistes très concrètes de remédiation. Les problèmes de concordances et/ou de tensions entre les finalités poursuivies (ou annoncées) par l'évaluation d'une part et le dispositif réellement mis en place d'autre part ne sont pas rares. Citons à titre d'exemple le fait de réaliser des évaluations formatives, mais qui comptent pour un pourcentage de la note finale, évaluer de manière formative la qualité des enseignements universitaires, mais acheminer les rapports de résultats aux autorités académiques...

En tant que scientifique dans le domaine de l'évaluation, il nous semble absolument fondamental de mener des recherches s'intégrant dans des contextes spécifiques, se déroulant dans des écologies clairement définies. En effet, il est loin le temps où les études ne se concentraient que sur la mesure (la note), sur les différents biais (de halo, de contraste, etc.) qui l'influençaient et sur la façon de contrecarrer ces biais. Il est également loin

le temps où la problématique de l'évaluation se limitait à l'évaluation scolaire (des élèves) qui était abordée essentiellement dans le cadre théorique de la psychométrie. Du fait de la diversification des fonctions de l'évaluation, de l'élargissement des publics évalués, de son inscription dans de nouveaux modèles et de l'évolution des valeurs la concernant, des champs de recherche nouveaux ont émergé, des questions nouvelles se sont posées.

Il est devenu impossible de traiter scientifiquement l'ensemble de la complexité qui résulte de cette évolution. Les chercheurs doivent, dès lors, nous semble-t-il, assumer ce fait, et proposer des études dans lesquelles le contexte est un élément prépondérant. Il faut également qu'ils le fassent de manière audible pour qu'un public de non-initiés puisse bénéficier de leurs conclusions. Les différents chapitres de cet ouvrage cherchent à remplir cette double contrainte. Nous les avons regroupés en trois axes thématiques qui étaient, par ailleurs, trois des axes du colloque de l'ADMEE-Europe dont est issu le présent ouvrage.

4. STRUCTURE DE L'OUVRAGE

Cet ouvrage débute par un texte introductif, correspondant à la conférence inaugurale de ce colloque et traduisant sa problématique centrale : l'importance de la prise en compte des contextes lors des évaluations.

Dans les lignes qui suivent, nous présentons brièvement ce texte introductif, ainsi que les raisons ayant présidé aux choix des trois axes thématiques. En vue de donner une vue d'ensemble de l'ouvrage, le contenu de chacun des axes est également très brièvement évoqué.

Notons toutefois qu'une description plus complète des thématiques traitées par chacun des axes sera présentée en introduction de ceux-ci. Un court résumé de chacun des chapitres constitutifs y sera également présent. Nous renvoyons donc le lecteur intéressé à ces présentations liminaires.

4.1. L'importance de la prise en compte du contexte lors des évaluations

Dans son texte intitulé « Évaluations à large échelle : prendre la juste mesure des effets de contexte » (chapitre 1), Dominique Lafontaine part du principe selon lequel, « sans appréhension rigoureuse des effets de contexte, les enquêtes internationales sont sans intérêt, sans pertinence, voire risquent d'induire en erreur... » (p. 22 du présent ouvrage). Le ton est donné d'emblée et l'analyse rigoureuse présentée tout au long du texte convainc rapidement le lecteur, à la fois du bien-fondé de ces enquêtes, mais aussi de toute la prudence qu'il convient de conserver lors de l'interprétation des résultats. Appuyant ses propos sur l'évolution des cadres de références

(*frameworks*) développés dans le cadre de l'enquête PISA, elle montre comment l'impact des contextes est pris en compte en amont du processus d'évaluation, au moment du test et en aval de celui-ci. Ces « contextes » se déclinent selon deux dimensions : une « diversité des contextes » d'abord, caractérisée par les particularités linguistiques et culturelles des pays, mais aussi par l'organisation même du système éducatif ; un « emboîtement des contextes » ensuite, dans la mesure où les individus dont on mesure les performances sont situés dans des classes, elles-mêmes situées dans des filières, des programmes d'études, des régions et des pays. Le texte décrit très précisément les avancées scientifiques réalisées ces dernières années, tout en présentant des illustrations permettant au lecteur de comprendre les enjeux liés à cette « juste » prise en compte des effets de contexte. Si la porte d'entrée choisie est celle des enquêtes internationales à large échelle, Dominique Lafontaine explique aussi que certains éléments, comme les « biais de réponse » ou le « *big-fish-little-pond-effect* », sont également à prendre en considération dans des études à plus petite échelle ou dans des études plus qualitatives. Pour conclure, elle se tourne alors vers l'avenir, en pointant les acquis, les points d'attention et les défis qu'il reste encore à relever.

4.2. Axe 1 : Évaluations et contextes didactiques

Cet axe était incontournable à partir du moment où la prise en compte des contextes était la thématique centrale du colloque liégeois. En effet, une des manières de proposer une contextualisation accrue des recherches en évaluation est de passer par une « didactisation » de l'évaluation. La didactique des disciplines a fait la preuve qu'elle était très complémentaire avec les sciences de l'éducation en ce qu'elle apporte l'éclairage résolument disciplinaire que ces dernières n'ont pas toujours. Or, s'il paraît aujourd'hui évident que l'on n'enseigne pas le français langue étrangère comme l'on enseigne les mathématiques ou les sciences, il semble tout aussi évident que l'on n'évalue pas de la même manière les étudiants dans ces disciplines. C'est d'autant plus vrai quand les méthodes d'enseignement et d'évaluation se fertilisent mutuellement comme c'est le cas dans *l'Assessment for Learning*. Dès lors, il est non seulement pertinent, mais aussi nécessaire, d'augmenter les synergies entre ces deux champs théoriques. Il est en effet probable que les pratiques et modèles de l'évaluation d'une part, les connaissances didactiques inhérentes à certaines disciplines d'autre part, puissent s'interféconder, ouvrant ainsi de nouveaux champs d'investigations prometteurs.

Cette partie de l'ouvrage comprend trois chapitres portant sur les liens entre **évaluations** et, respectivement, **didactique des mathématiques** (chapitre 2 : Brigitte Gugeon-Allys – Évaluer en mathématiques : une approche didactique et épistémologique), didactique du français langue

étrangère et seconde (chapitre 3 : Fatima Chnane-Davin et Jean-Pierre Cuq – L'évaluation, un paramètre prépondérant en didactique du français langue étrangère et seconde ») et **apprentissage par la lecture** dans trois contextes disciplinaires différents (chapitre 4 : Sylvie C. Cartier, Thaïs Guertin-Baril, Julie Arseneault et Manal Raoui – Évaluation formative de l'apprentissage par la lecture : relation complexe et dynamique « individu-contexte »).

4.3. Axe 2 : Évaluations et modèles cognitifs

Lorsque les organisateurs du colloque liégeois ont proposé cet axe, ils envisageaient de recevoir des contributions exclusivement en lien avec les modèles cognitifs de l'apprentissage. Ils ont été pris à contre-pied : il s'avère que des communications signifiantes portaient sur les aspects cognitifs liés aux pratiques d'évaluation. Notons que cette thématique illustre à merveille la vision évolutive du concept d'évaluation que nous avons brièvement décrite en amont. En effet, après avoir mis l'emphase sur la mesure (la note) produite par les enseignants – et sur les biais éventuels l'entachant –, après avoir mesuré l'effet de procédure destiné à contrer ces biais (par exemple les grilles critériées d'évaluation) des chercheurs se donnent pour objectif premier de comprendre l'activité évaluative telle qu'elle se déploie dans la réalité complexe de la vie des classes. Renonçant à la posture normative et prescriptive des pionniers (de Landsheere, 1972 et 1976), les chercheurs ayant contribué à l'axe 2 du présent ouvrage veulent comprendre les raisonnements évaluatifs des enseignants, avant de les conseiller ou leur faire des suggestions.

Cet axe comprend trois chapitres, qui traitent respectivement de l'évaluation scolaire et du jugement des enseignants (chapitres 5 : Rémi Goasdoué et Marc Vantourout – Évaluations scolaires et étude du jugement des enseignants : vers une docimologie cognitive), de la référentialisation dans l'évaluation des apprentissages des élèves (chapitre 6 : Lucie Mottier Lopez – Une modélisation pour appréhender la référentialisation dans l'évaluation des apprentissages des élèves) et des facteurs influençant la prise de décision portant sur le redoublement scolaire (chapitre 7 : Géry Marcoux et Marcel Crahay – Le redoublement des élèves : des décisions prises dans l'incertitude).

4.4. Axe 3 : Évaluation dans l'enseignement supérieur, orientations disciplinaires et approches pédagogiques

La prise en compte des contextes et des disciplines est particulièrement vivace au niveau des pratiques d'évaluation de l'enseignement supérieur. D'une part parce que, comme l'ont souligné Rege Colet et Romainville (2006), ce niveau d'enseignement est en profonde mutation (dû à la massification des étudiants notamment), ce qui n'est pas sans effet sur l'évaluation. D'autre part, il y a également un ancrage disciplinaire important de l'évaluation qui mène parfois à des dispositifs très spécifiques et très originaux. C'est le cas, par exemple, des tests de concordance de scripts et des Examens cliniques objectifs structurés dans le domaine de la médecine. Par ailleurs, la qualité de l'enseignement y est un thème émergent amenant une préoccupation nouvelle pour l'évaluation, notamment à visée formative. Évaluation qui est également mobilisée dans les approches de *Scholarship of Teaching and Learning* dont la vocation est de mesurer l'impact d'innovations pédagogiques afin de favoriser un partage de pratiques entre pairs souvent d'une même discipline. Pour toutes ces raisons, l'enseignement supérieur nous semble être un contexte pertinent dans le cadre de cet ouvrage.

Comme dans les axes précédents, on trouve ici encore trois chapitres qui abordent respectivement l'évaluation dans le cadre de l'*Assesment for Learning* (chapitre 8 : Laurent Leduc, Audrey Mohr, Eléonore Marichal, Mélanie Deum et Pascal Detroz – Viser le développement des pratiques d'*Assesment for Learning* [AFL] dans le contexte de la 1^{re} année à l'université : cadre théorique, dispositif et initiatives pédagogiques résultant du projet « Feedbacks 1^{er} Bac »), la formation à l'évaluation formative de futurs titulaires d'un master en sciences de l'éducation dans le cadre des premiers apprentissages mathématiques des élèves (chapitre 9 : Annick Fagnant et Isabelle Demonty – Comment former à une vision élargie de l'évaluation formative en vue de réguler efficacement les apprentissages mathématiques des élèves ?) et l'évaluation des bénéfices de deux dispositifs de formation technopédagogique (chapitre 10 : Catherine Loisy, Jean-François Van de Poël et Dominique Verpoorten – Regards croisés sur deux dispositifs de formation technopédagogique, leur outillage conceptuel et l'évaluation de leurs bénéfices).

RÉFÉRENCES BIBLIOGRAPHIQUES

- Ardoino, J., & Berger, G. (1989). *D'une évaluation en miettes à une évaluation en actes, le cas des universités*. RIRELF : Paris.
- Bonniol, J.-J., & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles, Paris : De Boeck Université.
- Cassin, B. (2014). *Derrière les grilles. Sortons du tout-évaluation*. Mille et une nuits, Paris : Éditions Fayard.
- De Ketele, J.-M., & Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et Évaluation en Éducation*, volume 28, 3, 1-26.
- Del Rey, A. (2013). *La tyrannie de l'évaluation*. Paris : La Découverte, coll. « Cahiers libres », 143 p.
- de Landsheere, G. (1972 et 1976). *Évaluation continue et examens. Précis de docimologie*. Bruxelles : Labor.
- Martuccelli, D. (2010). Critique de la philosophie de l'évaluation. *Cahiers internationaux de sociologie*, 128-129, 27-52. DOI : 10.3917/cis.128.0027
- Meirieu, Ph., Develay, M., Durand, C., & Mariani, Y. (dir.) (1996). *Le concept de transfert de connaissance en formation initiale et continue*. Lyon : CRDP.
- Miller, J.-A., & Milner, J.-C. (2004). *Voulez-vous être évalué ?* Paris : Grasset.
- Piéron, H. (1963). *Examens et docimologie*. Paris : Presses universitaires de France.
- Rege Colet, N., & Romainville M. (2006). *La pratique enseignante en mutation à l'université*. Bruxelles : De Boeck.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par les compétences*. Bruxelles : De Boeck Université.
- Stufflebeam, D. L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M. M. (1980). *L'évaluation et la prise de décision en éducation*. Victoriaville, Canada : N.H.P.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal : Chenelière.
- Tyler, R.W. (1935). Evaluation : A challenge to progressive education. *Education Research Bulletin*, 14, 9-16.
- Van der Maren, J.-M., & Loye, N. (2011). À propos de quelques difficultés de l'évaluation des compétences. *Réflexions critiques. Éducation Sciences et Society*, 2 (2), 40-53.

Chapitre 1

Évaluations à large échelle : prendre la juste mesure des effets de contexte

Dominique LAFONTAINE

1. INTRODUCTION

En posant que les enquêtes internationales à large échelle (ci-après ELA) ont toute leur pertinence, à condition de tenir compte des effets de contexte, ce chapitre revisite un vaste corpus d'enquêtes à large échelle, y situe les principales **avancées** et **points de rupture** tant sur le plan conceptuel que méthodologique qui font en sorte que les effets de contexte sont aujourd'hui, davantage et mieux qu'hier, pris en compte dans les enquêtes comparatives à large échelle. Si la question des effets de contexte est abordée au travers du prisme des enquêtes internationales, elle ne s'y réduit toutefois pas. Plusieurs des effets de contexte décrits dans le présent chapitre, en particulier dans les sections 2 et 3, s'observent et sont à prendre en considération dans des études à plus petite échelle, voire dans des approches plus qualitatives.

2. LES ÉVALUATIONS À LARGE ÉCHELLE

Le choix des évaluations comparatives à large échelle pour étudier les effets de contexte est facile à comprendre. Celles-ci sont en effet un terrain privilégié pour étudier la prise en compte et la mesure des effets de contexte pertinents ou indésirables (biais). On peut même soutenir que sans appréhension rigoureuse des effets de contexte, les enquêtes internationales sont sans intérêt, sans pertinence, voire risquent d'induire en erreur, si leur interprétation outrepassé les limites que leur imposent précisément les contextes. Un des reproches souvent émis à l'encontre des enquêtes internationales est qu'elles « comparent des pommes et des poires ». À ce reproche relativement élémentaire, on peut opposer que la comparaison de pommes et de poires peut faire sens, pour autant qu'on s'intéresse à l'arbre, au verger, au sol, aux conditions climatiques, aux traitements phytosanitaires... qui permettent de produire des fruits divers, non formatés, dans le respect de l'environnement et des populations locales.

Dans les évaluations comparatives à large échelle, les élèves ou les individus dont on mesure les performances scolaires ou cognitives sont issus de langues, de cultures, d'horizons et de systèmes éducatifs divers (diversité de contextes). Par ailleurs, ces individus sont nichés dans des classes, des filières, des programmes d'études, des écoles, des régions, des pays (contextes emboîtés). Ces deux types de contextes doivent être pris en compte de manière adéquate, tant sur le plan conceptuel que méthodologique.

2.1. Principales études internationales à large échelle

Lorsqu'on parle d'enquêtes internationales, on entend par là les grandes enquêtes comparatives menées principalement à l'initiative de l'Association internationale pour l'évaluation du rendement scolaire (IEA, voir <http://www.iea.nl>) et de l'Organisation de coopération et de développement économiques (OCDE, voir <http://www.oecd.org>). Le tableau suivant fournit un aperçu des principales enquêtes internationales menées depuis 1960. Dans ce tableau, on peut voir qu'une majorité de ces enquêtes portent sur les acquis ou le rendement des élèves dans les principaux domaines scolaires que sont la compréhension en lecture, les mathématiques ou les sciences. Toutefois, il arrive que d'autres domaines soient investigués, tels que l'éducation civique (Cived, ICCS), les technologies de l'information (Sites, ICILS) ou encore les enseignants (Talis). Par ailleurs, si les enquêtes internationales sont aujourd'hui surtout connues *via* la fameuse enquête PISA de l'OCDE, le tableau fait clairement apparaître que les enquêtes internationales se sont déployées dès la fin des années 1960, soit près de

trente ans avant que l'OCDE prenne des initiatives en la matière, et que même si l'IEA fait l'objet de moins d'attention médiatique, elle continue à jouer un rôle important dans le champ, notamment en évaluant d'autres populations cibles ou d'autres domaines que ne le fait PISA.

Tableau 1.1. Aperçu des principales enquêtes internationales menées depuis 1960

	1960	1970	1980	1990	2000	2010
Langue maternelle		Reading, Literature	Composition	RLS IALS (1998)	PIRLS 2001, 2006 PISA 2000, 2003, 2006, 2009 ALL 2003, 2006	PIRLS 2011, 2016 PISA 2012, 15, 18 PIACC 2010
Mathématiques	FIMS		SIMS IAEP I	IEAP II TIMSS, TIMSS_R	PISA 2000,03, 06,09 TIMSS 2003, 07 TIMSS Ad. 08	TIMSS 2011, 15 PISA 2012, 15, 18
Sciences		FISS	SISS IAEP I	IEAP II TIMSS, TIMSS_R	PISA 2000, 03, 06, 09 TIMSS 03, 07 TIMSS Ad. 08	TIMSS 2011, 15 PISA 2012, 15, 18
Autres		English, French as foreign ; Civic	Comped Classroom environ- ment study PPP	CivED Sites-M1	Sites-M2 Teds-M (2008) ICCS 2009 Talis 2008 ESLC 2012	Talis 2013, 18 ICCS ICILS 13, 18

2.2. Prendre la juste mesure des effets de contexte

Le présent chapitre s'organise autour de la question d'une prise de mesure, considérée comme *juste*, des effets de contexte. Par « juste », il faut ici comprendre que cette prise en compte répond à une exigence de rigueur tant sur le plan conceptuel que méthodologique et que les réponses à cet impératif doivent être proportionnées et adéquates, compte tenu de l'état actuel des connaissances scientifiques. Cette prise en compte sera déclinée en trois phases : en amont des tests ou questionnaires, au moment du test et en aval du test. Ceci amènera à aborder des questions aussi diverses que les cadres de référence pour les questionnaires de contexte,

les notions d'équivalence (ou d'invariance⁶) et de stabilité cross-culturelle, les biais de méthodes/styles de réponse, ainsi que l'influence des groupes de référence et du macrocontexte « système éducatif ».

3. EN AMONT DES QUESTIONNAIRES

3.1. Cadres de référence conceptuels : penser les effets de contexte

L'analyse de l'évolution des Cadres de référence relatifs aux variables de contexte (*frameworks*) sera principalement menée au départ de l'enquête PISA. En effet, les politiques internes de gestion des enquêtes propres à l'IEA et à l'OCDE diffèrent sur plus d'un point. Sans entrer dans le détail de ces politiques, on peut souligner que l'OCDE agit de manière plus centralisée et directive, voire plus politique, au sens propre du terme ; il est donc plus aisé et plus pertinent de tenter une analyse des politiques de l'OCDE que de l'IEA. De son côté et par son statut, l'IEA est d'abord un consortium de centres de recherche et laisse davantage d'autonomie et de marge de manœuvre aux consortiums en charge de la mise en œuvre des enquêtes menées sous leurs auspices. La diversité d'approche d'une étude à l'autre y est donc plus grande et tient souvent davantage à l'empreinte que lui donne le consortium responsable plutôt qu'à une politique orchestrée par l'IEA.

La définition des fondements de l'enquête PISA, tant conceptuels que méthodologiques débute à la fin des années 1990. Un consortium de centres de recherches dirigé par l'*Australian Council of Educational Research* (ACER) se voit, après appel d'offre, confier la responsabilité de mettre en œuvre le programme PISA. Des groupes d'experts internationaux sont mis au travail et l'OCDE publie en 1999 son premier cadre de référence, intitulé *Mesurer les connaissances et compétences des élèves. Un nouveau cadre d'évaluation* (OCDE, 1999). Dans ce document fondateur, la focale est mise sur la mesure cognitive (savoirs et compétences) : une seule page (sur 85) est consacrée aux questionnaires de contexte. Certes, l'importance « d'indicateurs **contextuels** mettant en relation les performances avec les caractéristiques des élèves et des écoles » (OCDE, 1999, p. 10) est soulignée, mais aucun cadre de référence pour les questionnaires de contexte n'est produit ou rendu public. Aucune mention n'est faite du travail d'un groupe d'experts réfléchissant aux questionnaires de contexte, alors que ce groupe existe. Il travaille en quelque sorte dans l'ombre, ce qui a portée de symbole. Comme l'indique le rapport technique de *PISA 2000*,

6. Dans la littérature, on parle indifféremment d'équivalence ou d'invariance. Ici, nous utilisons le terme « équivalence » qui nous paraît plus clair.

« le cadre de référence pour les questionnaires n'a pas été publié par l'OCDE, mais est disponible sous forme de document de travail » (p. 34)⁷.

Pour les cycles 2003 et 2006, aucune évolution majeure n'est observée. Le QEG (*questionnaire expert group*) continue à travailler de manière relativement confidentielle, tandis que le travail des SMEG (*subject matter expert group*) est reconnu et mis en lumière dans les cadres de référence successifs (OCDE, 2003, 2006). Le point de rupture se produit lors du cycle PISA 2009. Un chapitre intitulé « Questionnaire framework », d'une longueur de 20 pages, apparaît dans le Cadre de référence de PISA 2009 (OCDE, 2009). Dans ce chapitre se déploie une conceptualisation systématique des constructs essentiels à prendre en considération au niveau système, école, classe et élève. En annexe, une liste exhaustive des constructs et des questions/items les mesurant dans les questionnaires est fournie.

Le fait que l'OCDE ait, pour PISA 2009, confié la tâche de développer les questionnaires de contexte au CITO⁸, consortium distinct d'ACER, n'est sans doute pas étranger à cela. L'existence d'un groupe d'experts pour les questionnaires (QEG) dont le chairman est Jaap Scheerens (Université de Twente) est officialisée à la même époque. Sans surprise, le modèle conceptuel développé dans le Cadre de référence fait la part belle aux variables liées à l'école et au courant de la *School effectiveness research*, dont Jaap Scheerens est un illustre représentant. C'est aussi à la même époque que l'OCDE se met à publier, à côté du principal rapport présentant les résultats de PISA, des rapports approfondis consacrés à des questions de politiques éducatives – efficacité des écoles, équité, efficience... – qui nécessitent un modèle conceptuel plus strict des variables conceptuelles.

L'évolution entamée en 2009 se poursuit en 2012. Plusieurs évolutions retiennent l'intérêt lors de ce cycle. Eckhard Klieme (DIPF⁹) préside désormais le QEG. On note une systématisation du cadre conceptuel au départ duquel sont pensées d'une part la constitution d'une base de données à long terme (*sustainable database*), et la sélection d'un noyau de variables contextuelles (*core content*) pensé dans la durée (OCDE, 2012). Le cadre conceptuel, organisé selon le modèle CIPO (*context-input-processes-outcomes*), s'ouvre à des variables liées à la qualité de l'enseignement et des environnements d'apprentissage (opportunités d'apprentissage, feedbacks, gestion de la classe), questions dont Eckhard Klieme est un expert reconnu. L'apparition de variables contextuelles liées à la classe n'est pas en

7. « The questionnaire framework was not published by the OECD but is available as a project working document » (*Technical report PISA 2000*, p. 34).

8. Le CITO (Central Institute for Test Development) est une importante compagnie de testing, située à Arnhem aux Pays-Bas (<http://www.cito.com>).

9. Le DIPF (Deutsches Institut für Internationale Pädagogische Forschung), branche de l'Institut Leibniz, est un important centre de recherches situé à Francfort.

soi une innovation dans le cadre des enquêtes internationales. Les premières enquêtes de l'IEA dans les années 1970 ont même été un terreau fertile pour le développement de la notion d'OTL et du modèle de Carroll. Cependant, dans l'enquête PISA, qui rappelons-le, a fait le choix d'une population d'âge (élèves de 15 ans) et ne sélectionne pas de classes entières, mais des élèves à l'intérieur des écoles, il s'agit d'une évolution notable, témoignant du fait que PISA se préoccupe beaucoup plus qu'annoncé ou que prévu d'évaluer le rendement de l'école et l'impact des facteurs scolaires sur ce rendement.

Dans ce même chapitre du Cadre de référence de 2012 (OCDE, 2012), ainsi que dans le Rapport technique de PISA 2012 (OCDE, 2014), on voit pour la première fois apparaître une préoccupation pour la validité et la stabilité cross-culturelle des variables contextuelles. Dans ce cas, il s'agit d'une véritable avancée, à laquelle la présence dans le QEG de Fons van de Vijver (Université d'Utrecht), expert des questions de psychologie cross-culturelle, n'est sans doute pas étrangère. La question de l'équivalence et de la validité des comparaisons est certes centrale dans les enquêtes internationales, mais jusqu'alors, le questionnement et le débat, parfois vif et polémique (Goldstein, 2004 ; Lafontaine & Demeuse, 2002 ; Romainville, 2002), portait principalement sur l'équivalence des tests (mesures cognitives). Désormais, il s'agit de s'interroger sur la stabilité et l'équivalence des variables contextuelles : lorsque l'on mesure par exemple le concept de soi en mathématiques ou la motivation pour la lecture, mesure-t-on bien la même chose dans les différents contextes, et peut-on comparer de manière valide les scores des pays sur cette échelle ? Oser poser cette question dans le cadre des enquêtes internationales constitue un véritable point de rupture.

PISA 2015 voit se poursuivre les évolutions entamées en 2009 et 2012. L'OCDE elle-même consacre l'importance des questionnaires de contexte en lançant un appel d'offres séparé pour les domaines cognitifs (tests en lecture, mathématiques et sciences) et non cognitifs (questionnaires de contexte). Le DIPF emporte ce dernier marché ; Eckhard Klieme préside à nouveau le QEG. Pour la première fois, un cadre de référence propre est élaboré pour les questionnaires de contexte dont le titre lui-même est un programme : *The PISA 2015 Framework for Context Assessment. Monitoring Opportunities and Outcomes, Policies and Practices. Modeling Patterns and Relations, Impacts and Trends in Education*¹⁰. La boucle est en quelque sorte bouclée.

De 2000 à 2015, PISA a ainsi connu une montée en puissance des questionnaires de contexte, qui passent d'un statut quasi confidentiel à

10. Ce cadre de référence n'est pas encore publié sous ce titre par l'OCDE, 2012. Seul un *Draft Questionnaire framework* est accessible en ligne sur la page <http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm>.

Liste des tableaux et figures

TABLEAUX

Tableau 1.1	Aperçu des principales enquêtes internationales menées depuis 1960	23
Tableau 1.2	Corrélations des deux facteurs OTL avec les scores en lecture (PISA 2009).....	38
Tableau 1.3	Corrélations entre les trois dimensions de l'analyse IRT et les scores en lecture (PISA 2009)	38
Tableau 1.4	Corrélations entre les variables motivationnelles et le score en mathématiques avant/après standardisation par école/filière/année d'étude (PISA 2003)	44
Tableau 4.1	Années de réalisation des situations d'APL, disciplines, enseignants et élèves visés	116
Tableau 1.	Fréquence d'utilisation des mots	216
Tableau 8.1	Facteurs reliés au développement des innovations pédagogiques en enseignement supérieur aux niveaux Département et Classe (Béchar & Pelletier, 2002, p. 4).....	227
Tableau 8.2	Modèle des six conditions de base de la pratique efficace de l'AfL dans l'enseignement supérieur (d'après Sambell, McDowell & Montgomery, 2013)	231
Tableau 8.3	Stratégies clés de l'AfL (d'après Thompson & Wiliam, 2008).....	233
Tableau 8.4	Nombre d'actions spécifiques prises en regard de chaque catégorie/étape de l'outil dynamique de catalogage des pratiques d'AfL	247
Tableau 8.5	Analyse des outputs individuels du projet « Feedbacks 1 ^{er} Bac » au travers de notre outil dynamique de catalogage des pratiques d'AfL	253

Tableau 10.1 Codage des réponses des participants..... 284

Tableau 10.2 Comparaison synoptique des dispositifs étudiés..... 296

FIGURES

Figure 1.1. Corrélation entre la moyenne par pays des résultats en sciences et du concept de soi en sciences dans PISA 2006 (Van de Gaer *et al.*, 2012)..... 30

Figure 1.2. Scores des pays sur le superconstruct de Lie et Turmo fondé sur 10 échelles d'attitudes (PISA 2003)..... 36

Figure 1.3. Corrélation entre la moyenne par pays de la performance en lecture et la moyenne de la dimension style de réponse (PISA 2009, pays de l'OCDE uniquement) 40

Figure 2.1. Processus de transposition et référence épistémologique de référence (Bosch & Gascon, 2005) 71

Figure 2.2. Items CEDRE 2008 à choix multiple..... 75

Figure 2.3. Échelle de score CEDRE 2008 86
Échelle de score CEDRE 2014 87

Figure 2.4.1. « Déterminer si une égalité est toujours vérifiée » 88

Figure 2.4.2. Justifications proposées dans le cas où l'élève a choisi « vraie »..... 88

Figure 2.4.3. Justifications proposées dans le cas où l'élève a choisi « fausse »..... 88

Figure 2.5. « Preuve et programme de calcul »..... 89

Figure 2.6. Codage des réponses au niveau technologique 89

Figure 2.7. Profil cognitif d'un élève en classe de 3^e 90

Figure 2.8. Carré bordé : les patterns selon deux groupes d'élèves 90

Figure 3.1. Dispositif d'évaluation formative en FLE/FLS 103

Figure 5.1. Modèle explicatif des comportements d'évaluation de productions scolaires (Noizet & Caverni, 1978, p. 115)..... 166

Figure 5.2. Types de pages présentées lors des situations aménagées (problème : le cycliste)..... 167

Figure 5.3.	Problème « location de cassettes vidéo » : production graphique d'un binôme d'élèves.....	168
Figure 6.1.	Unités conceptuelles qui caractérisent l'évaluation, élaborées à partir d'Allal (2008, p. 311) et Hadji (2012, p. 118)	174
Figure 6.2.	Référentialisation située.....	178
Figure 6.3.	Entête élaboré par Patricia pour évaluer les productions écrites de ses élèves, le fait divers	180
Figure 6.4.	Texte rédigé par le premier élève.....	192
Figure 7.1.	Plan d'observation de l'étude de Crahay <i>et al.</i> (en préparation)	203
Figure 8.1.	Principes d'une bonne pratique de l'évaluation et du feedback au service de l'amélioration de la qualité de l'expérience d'apprentissage en 1 ^{re} année (d'après Nicol, 2009)	234
Figure 8.2.	Outil dynamique de catalogage des pratiques d' <i>AfL</i>	240
Figure 8.3.	Niveau de prise en considération des principes de Nicol (2009) par les enseignants concepteurs des 27 outputs individuels du projet « Feedbacks 1 ^{er} Bac »	245
Figure 10.1.	Le modèle TPaCK (Mishra & Koelher, 2006).....	277
Figure 10.2.	Le visuel TPaCK réalisé à partir des données présentes dans ALOES pour les UE de A, enseignant dans le master ArchInfo.....	281
Figure 10.3.	Remplissage par C des schémas TPaCK lors de la phase d'entretien.....	285
Figure 10.4.	Le développement technopédagogique se matérialise à l'ULg dans quatre composants de formation introduisant graduellement à l'usage du numérique pour l'éducation	289
Figure 10.5.	Les quatre composants de formation de IFRES-eCampus activent essentiellement trois zones du TPaCK (volontairement élargies par rapport à l'initial)	290

Table des matières

Sommaire	5
Liste des auteurs	9
Introduction	11
1. L'évaluation, ce concept qui s'étend jusqu'à devenir insaisissable	11
2. Une complexité qui s'est développée au fil du temps	13
3. Vers une prise en compte des contextes	14
4. Structure de l'ouvrage	16
4.1. <i>L'importance de la prise en compte du contexte lors des évaluations</i>	16
4.2. <i>Axe 1 : Évaluations et contextes didactiques</i>	17
4.3. <i>Axe 2 : Évaluations et modèles cognitifs</i>	18
4.4. <i>Axe 3 : Évaluation dans l'enseignement supérieur, orientations disciplinaires et approches pédagogiques</i>	19
Références bibliographiques	20
Chapitre 1. Évaluations à large échelle : prendre la juste mesure des effets de contexte	21
1. Introduction	21
2. Les évaluations à large échelle	22
2.1. <i>Principales études internationales à large échelle</i>	22
2.2. <i>Prendre la juste mesure des effets de contexte</i>	23
3. En amont des questionnaires	24
3.1. <i>Cadres de référence conceptuels : penser les effets de contexte</i>	24
4. Dans les questionnaires	27
4.1. <i>Biais de méthode et styles de réponse</i>	27

4.2. <i>Le paradoxe attitudes-performances</i>	29
4.3. <i>Équivalence cross-culturelle</i>	31
4.4. <i>Comment prendre en compte les biais de méthode ou neutraliser les effets de contexte indésirables ?</i>	32
4.4.1. <i>Approche a priori : format d'items alternatifs et détection des styles de réponse</i>	33
4.4.2. <i>Approche a posteriori : modélisation des biais</i>	35
5. En aval du test	41
6. Conclusions	45
6.1. <i>Les acquis</i>	46
6.2. <i>Les soucis ou points d'attention</i>	46
6.3. <i>Les défis</i>	47
Références bibliographiques	49
Axe 1. ÉVALUATIONS ET CONTEXTES DIDACTIQUES	53
Introduction à l'axe 1	55
1. Préambule	55
2. Évaluations et contextes didactiques : un champ de recherche novateur ?	56
3. Présentation succincte des trois textes constitutifs de l'axe « Évaluations et contextes didactiques »	59
Références bibliographiques	61
Chapitre 2. Évaluer en mathématiques : une approche didactique et épistémologique	63
1. Introduction	63
2. Une approche multidimensionnelle de la validité d'une évaluation	65
2.1. <i>Deux dispositifs d'évaluation externe</i>	65
2.1.1. <i>CEDRE : une évaluation standardisée des acquis des élèves</i>	65
2.1.2. <i>PÉPITE : une évaluation diagnostique des compétences des élèves en algèbre en 3^e</i>	66
2.1.3. <i>Validité des dispositifs d'évaluation</i>	67
2.2. <i>Approche multidimensionnelle des validités d'une évaluation</i>	68
2.2.1. <i>Approche psychométrique</i>	68
2.2.2. <i>Approche didactique</i>	68

2.2.3. <i>Évaluation et processus d'enseignement/apprentissage</i>	69
3. <i>Évaluation et processus d'enseignement et d'apprentissage : une approche didactique</i>	70
3.1. <i>Une analyse épistémologique et didactique du savoir au service de la validité d'une évaluation</i>	70
3.1.1. <i>Évaluation du processus d'apprentissage de l'élève – sujet cognitif</i>	70
3.1.2. <i>Évaluation des praxéologies apprises par l'élève dans une institution</i>	70
3.2. <i>L'évaluation, un « aspect déterminant » pour organiser, réguler l'enseignement et optimiser le processus d'apprentissage des élèves</i>	73
3.2.1. <i>Niveau global : évaluation et moments de l'étude</i>	73
3.2.2. <i>Niveau local : évaluation et contrat didactique</i>	74
3.3. <i>Référence épistémologique en algèbre élémentaire</i>	75
4. <i>Éléments d'analyse de la validité épistémo-didactique de CEDRE</i>	75
4.1. <i>Quelques résultats</i>	75
4.1.1. <i>Au niveau local : distracteurs et erreurs envisageables</i>	75
4.1.2. <i>Au niveau local : objectif d'évaluation visé, variables didactiques et techniques</i>	76
4.1.3. <i>Au niveau global : couverture du domaine, connaissances et capacités évaluées par l'échelle</i>	77
4.2. <i>Opposition ou complémentarité des approches psychométrique et didactique</i>	78
5. <i>Un dispositif d'évaluation automatisé pour réguler l'enseignement en algèbre élémentaire en fin de scolarité obligatoire (élèves de 15 ans)</i>	79
5.1. <i>Le modèle de l'évaluation diagnostique automatisée PÉPITE</i>	79
5.2. <i>Des outils pour les enseignants pour réguler l'enseignement en fonction des difficultés des élèves</i>	81
6. <i>Conclusion et perspectives</i>	82
<i>Références bibliographiques</i>	83
<i>Annexe 1</i>	86
<i>Annexe 2</i>	88

Chapitre 3. L'évaluation, un paramètre prépondérant en didactique du français langue étrangère et seconde	91
1. Introduction	91
2. Quelques causes de l'évolution de l'évaluation en didactique du FLE/FLS	92
2.1. <i>Didactique des langues et didactique des disciplines linguistiques scolaires</i>	92
2.2. <i>La fin de l'isolationnisme du FLE</i>	93
2.3. <i>L'arrivée en nombre d'élèves allophones</i>	93
2.4. <i>Des finalités nouvelles dans l'apprentissage des langues vivantes</i>	95
2.5. <i>Le Cadre européen commun de référence pour les langues</i>	95
2.6. <i>L'évaluation : un symptôme de l'ère du soupçon</i>	96
3. Évaluer en didactique des langues, un geste professionnel qui s'apprend	97
3.1. <i>Évaluation des élèves allophones en France : des enseignants entre les prescriptions du CECR et celles du socle commun de connaissances, de compétences et de culture</i>	98
3.2. <i>Formation à l'évaluation en didactique des langues dans un pays d'Europe de l'est et d'Asie centrale</i>	100
4. Limites théoriques et pratiques de l'évaluation de la compétence en langue	104
4.1. <i>Évaluation des savoirs, évaluation de la compétence et CECR</i>	104
4.2. <i>Évaluer la tâche ?</i>	106
4.3. <i>Difficultés culturelles</i>	106
4.4. <i>Évaluation industrielle ou évaluation personnalisée ?</i>	107
5. Conclusion	108
Références bibliographiques	109

Chapitre 4. Évaluation formative de l'apprentissage par la lecture : relation complexe et dynamique « individu-contexte »	111
1. Contexte	111
2. Objectifs de l'étude	112
3. Cadre théorique	113
3.1. <i>Processus d'apprentissage par la lecture</i>	113
3.2. <i>Contexte d'apprentissage par la lecture</i>	114
3.3. <i>Évaluation formative de l'APL</i>	114
4. Aspects méthodologiques	115
4.1. <i>Participants</i>	115
4.2. <i>Outils de collecte de données</i>	116
4.3. <i>Méthodes de compilation et d'analyse de données</i>	116
5. Résultats et discussion	117
5.1. <i>Contexte d'APL en arts plastiques et processus d'APL obtenu au QAPL dans ce contexte</i>	117
5.2. <i>Contexte d'APL en français et processus d'APL obtenu au QAPL dans ce contexte</i>	120
5.3. <i>Contexte d'APL en sciences et processus d'APL obtenu au QAPL dans ce contexte</i>	123
6. Conclusion	126
Références bibliographiques	128
 Axe 2. ÉVALUATION ET MODÈLES COGNITIFS	 131
Introduction à l'axe 2 :	
De la docimologie à la psychologie de l'évaluation	133
1. La docimologie est-elle obsolète ?	133
2. La nécessité d'une analyse approfondie de l'activité évaluative des enseignants	135
3. Comprendre avant de conseiller	139
Références bibliographiques	140

Chapitre 5. Évaluations scolaires et étude du jugement des enseignants : pour une docimologie cognitive	141
1. Préambule	141
1.1. <i>Quel psychologique en évaluation ?</i>	141
1.2. <i>Vers une docimologie cognitive</i>	142
2. Regard critique sur deux résultats des travaux docimologiques	144
3. Rôle des connaissances des évaluateurs et remise en cause du modèle « comparaison référent/référé » en mathématiques	147
3.1. <i>Une double orientation didactique pour étudier l'activité des évaluateurs</i>	148
3.2. <i>Analyse de l'activité des évaluateurs : le rôle des connaissances</i>	149
3.3. <i>Analyse de l'activité des évaluateurs : quel référent pour l'évaluateur ?</i>	154
4. La correction de dissertation en sciences économiques et sociales	156
5. Des épreuves, des logiques de preuves : dépasser certains clivages pour penser la validité des évaluations	161
Références bibliographiques	164
Annexe 1	166
Annexe 2	167
Annexe 3	168
Chapitre 6. Une modélisation pour appréhender la référentialisation dans l'évaluation des apprentissages des élèves	169
1. Définition du référentiel et enjeux de la référentialisation	171
2. Modélisation en trois axes de l'activité évaluative d'un enseignant	173
2.1. <i>Les invariants de l'activité évaluative</i>	173
2.2. <i>Modèles des objets évalués</i>	175
2.3. <i>Indissociabilité entre activité, apprentissage et contextes</i>	176
3. Exemple à partir de l'analyse d'une pratique, celle de Patricia	178
3.1. <i>Entête de l'évaluation élaborée par Patricia</i>	180
3.2. <i>Jugement en acte de Patricia : référents formels et informels</i>	181

4. Conclusion : peut-on modéliser l'activité évaluative des enseignants ?	186
Références bibliographiques	188
Annexe 1	191
Annexe 2	192
Chapitre 7. Le redoublement des élèves : des décisions prises dans l'incertitude	193
1. Un questionnement légitime	193
2. Le redoublement, une décision à colorations critériée, normative et pronostique	195
3. Les notes scolaires, un outil de gestion de la classe et une composante secondaire dans la décision de redoublement	200
4. Du repérage précoce des élèves en difficulté aux revirements possibles d'une décision finale	203
5. Les décisions de redoublement : un processus en trois phases	205
Références bibliographiques	208
Axe 3. ÉVALUATION DANS L'ENSEIGNEMENT SUPÉRIEUR, ORIENTATIONS DISCIPLINAIRES ET APPROCHES PÉDAGOGIQUES	211
Introduction à l'axe 3	213
1. Préambule	213
2. Quels sont les thèmes préférentiellement développés dans les recherches portant sur l'évaluation dans l'enseignement supérieur ?	214
3. Présentation succincte des trois textes	219

Chapitre 8. Viser le développement des pratiques d'Assessment for Learning (AfL) dans le contexte de la 1^{re} année à l'université : cadre théorique, dispositif et initiatives pédagogiques résultant du projet « Feedbacks 1^{er} bac »	221
1. Introduction	221
2. Pratiques d'Assessment for Learning et perspectives en pédagogie universitaire dans la littérature	223
2.1. <i>Notions d'Assessment for Learning et rôle central des feedbacks formatifs</i>	223
2.2. <i>La réflexion sur l'évaluation et les feedbacks au service de la réussite des étudiants de première année à l'université</i>	224
3. Conception du dispositif d'accompagnement et des principes méthodologiques du projet « Feedbacks 1 ^{er} Bac »	225
3.1. <i>Le projet « Feedbacks 1^{er} Bac » dans les faits</i>	225
3.2. <i>Une méthodologie d'accompagnement au service de l'innovation pédagogique</i>	226
4. Focale sur la phase de sélection et de conceptualisation des modèles théoriques propres à guider le développement de pratiques d'AfL et de feedbacks formatifs au service de l'autorégulation des étudiants	230
4.1. <i>Relevé de trois ressources théoriques opérationnelles</i>	230
4.1.1. <i>Le modèle de l'AfL de Sambell, Mc Dowell et Montgomery</i>	231
4.1.2. <i>Les stratégies de l'AfL minute-par-minute et jour-après-jour de Thompson et Wiliam</i>	232
4.1.3. <i>Les principes d'évaluation et de feedbacks en première année de Nicol</i>	233
4.2. <i>Conception dans le cadre du projet d'un outil dynamique de catalogage des pratiques d'AfL</i>	236
5. Bilan des initiatives pédagogiques nées de l'application du dispositif du projet « Feedbacks 1 ^{er} Bac » dans deux facultés de l'Université de Liège	242
5.1. <i>Examen des principes de bonnes pratiques de l'évaluation et du feedback de Nicol (2009) privilégiés par les accompagnés dans le cadre des initiatives pédagogiques prises</i>	244

5.2. <i>Examen du nombre d'actions spécifiques menées en regard des catégories de l'outil dynamique de catalogage des pratiques d'AfL par les accompagnés dans le cadre des initiatives pédagogiques prises</i>	246
6. En guise de conclusion et perspectives	249
Références bibliographiques	250
Annexe 1	253
Chapitre 9. Comment former à une vision élargie de l'évaluation formative en vue de réguler efficacement les apprentissages mathématiques des élèves ?	257
1. Introduction	257
2. L'analyse des productions d'élèves : une évaluation centrée sur les processus cognitifs ouvrant la voie à une vision diagnostique élargie ?	259
3. L'analyse de la mise en œuvre d'une séquence didactique : une façon de combiner analyse diagnostique et régulation interactive, ainsi qu'évaluation formative formelle et informelle ?	263
4. Pour terminer sans conclure...	268
Références bibliographiques	271
Chapitre 10. Regards croisés sur deux dispositifs de formation technopédagogique, leur outillage conceptuel et l'évaluation de leurs bénéfices	275
1. Introduction	275
1.1. <i>Le modèle TPaCK</i>	276
1.2. <i>Une perspective développementale sur la technopédagogie</i>	277
1.3. <i>Souci d'évaluation</i>	278
2. Dispositif DevSup : le TPaCK pour médier la réflexivité sur les pratiques pédagogiques	279
2.1. <i>Contexte</i>	279
2.2. <i>Recours au TPaCK pour analyser la mobilisation des connaissances technopédagogiques</i>	280
2.2.1. <i>Le TPaCK comme cadre d'analyse de traces</i>	280

2.2.2. <i>Le TPaCK dans une activité d'autoconfrontation à visée de conscientisation</i>	282
2.3. <i>Évaluation du dispositif d'accompagnement et retour sur l'instrument</i>	286
3. Dispositif IFRES-eCampus : le TPaCK pour caractériser un dispositif de formation	288
3.1. <i>Contexte</i>	288
3.2. <i>Recours au TPaCK pour caractériser un programme de formation</i>	290
3.3. <i>Évaluation du dispositif</i>	292
4. Discussion	295
4.1. <i>Usages du TPaCK</i>	296
4.2. <i>Perspective développementale</i>	297
4.3. <i>Évaluation des effets</i>	298
Références bibliographiques	299
Liste des tableaux et figures	305

L'évaluation à la lumière des contextes et des disciplines

L'évaluation n'a jamais été aussi présente dans nos vies quotidiennes. En tant qu'apprenants, enseignants, chercheurs, ou simples quidams, nous sommes cernés par des dispositifs d'évaluation. Leurs fonctions sont diverses : pour nous aider, pour nous sélectionner, pour nous accorder une promotion, etc. Les contextes dans lesquels nous sommes évalués sont multiples, de même que les disciplines mises en jeu.

Au cours d'un colloque de l'ADMEE-Europe, des chercheurs d'horizons francophones divers se sont réunis pour étudier en quoi les contextes influencent les processus d'évaluation, l'interprétation des résultats et les pratiques des acteurs.

La prise en compte effective des disciplines implique de s'intéresser aux apports réciproques entre didactiques disciplinaires et évaluations (axe 1). Prenant en considération le fait qu'évaluer revient toujours à porter un jugement, les processus psychologiques mobilisés par le jugement évaluatif sont également analysés (axe 2). Enfin, la prise en compte des contextes et des disciplines étant particulièrement vivace au niveau des pratiques d'évaluation de l'enseignement supérieur, un axe est consacré à ce niveau scolaire (axe 3).

Cet ouvrage rassemble des contributions qui, s'inscrivant dans les trois axes thématiques susmentionnés, donnent un aperçu contemporain d'un domaine riche en perspectives nouvelles, mais aussi en questionnements autour d'une problématique – évaluer – qui structure nos vies dans de si nombreuses circonstances.

Pascal Detroz est chargé de cours à l'Université de Liège (Belgique). Ses intérêts de recherche portent notamment sur les pratiques d'évaluation dans l'enseignement supérieur. Il est membre du conseil d'administration de l'ADMEE-Europe et cofondateur du réseau « Évaluation et enseignement supérieur ».

Marcel Crahay est professeur honoraire des Universités de Genève et de Liège. Parmi ses multiples intérêts de recherche et d'étude, les jugements évaluatifs des enseignants occupent une place importante, notamment dans la mesure où ceux-ci contribuent souvent à l'échec scolaire des élèves.

Annick Fagnant est chargée de cours à l'Université de Liège. Les dispositifs de développement professionnel visant à faire évoluer les pratiques d'évaluation des enseignants figurent parmi ses intérêts de recherche récents. Elle est membre du conseil d'administration de l'ADMEE-Europe et cofondatrice du réseau « EVADIDA ».

deboeck **B**
SUPÉRIEUR



ISBN 978-2-8073-0715-5
ISSN 0777-5245

www.deboecksuperieur.com