

Sous la direction de
Micheline-Joanne Durand et Nathalie Loyer

L'INSTRUMENTATION POUR L'ÉVALUATION

La boîte à outils de l'enseignant évaluateur



marceldidier

Sous la direction de
Micheline-Joanne Durand
Nathalie Loye

L'Instrumentation pour l'évaluation

La boîte à outils de l'enseignant évaluateur

marcel**didier**

**Catalogage avant publication de Bibliothèque et Archives nationales
du Québec et Bibliothèque et Archives Canada**

Vedette principale au titre:

L'instrumentation pour l'évaluation: la boîte à outils de l'enseignant évaluateur
Comprend des références bibliographiques.

ISBN 978-2-89144-609-9

1. Apprentissage - Évaluation. 2. Évaluation en éducation. I. Durand, Micheline-Joanne, 1957- .
II. Loye, Nathalie, 1963- .

LB2822.75.I57 2014

371.27

C2014-940498-0

Les Éditions Marcel Didier bénéficient du soutien des institutions suivantes pour leurs activités d'édition: gouvernement du Canada par l'entremise du Fonds du livre du Canada (FLC) et programme de crédit d'impôt pour l'édition de livres du gouvernement du Québec.

Édition: Loïc Hervouet

Révision linguistique: Catherine Vaudry

Correction d'épreuves: Émilie Leclerc

Illustrations: Marc Mongeau

Réalisation de la maquette intérieure: Folio infographie

Conception et réalisation de la couverture: Christian Campana

Dans cet ouvrage, le masculin est utilisé sans aucune discrimination et dans le seul but d'alléger le texte.

Les auteurs remercient particulièrement les étudiants inscrits aux cours ETA 6541 – Instrumentation pour l'évaluation et ETA 6065 – Évaluation et compétences pour leur engagement dans la conception et l'expérimentation de différents outils pour l'évaluation.

Copyright © 2014, Marcel Didier inc.

ISBN 978-2-89144-609-9 (version imprimée)

ISBN 978-2-89144-655-6 (version numérique PDF)

Dépôt légal – 3^e trimestre 2014

Bibliothèque et Archives nationales du Québec

Bibliothèque et Archives Canada

Diffusion-distribution en Amérique du Nord:

Distribution HMH

1815, avenue De Lorimier

Montréal (Québec) H2K 3W6

www.distributionhmh.com

Diffusion-distribution en Europe:

Librairie du Québec/DNM

30, rue Gay-Lussac

75005 Paris FRANCE

www.librairieduquebec.fr

TABLE DES MATIÈRES

Introduction	5
---------------------------	----------

PARTIE I

DES INSTRUMENTS AUX VISÉES PLURIELLES	11
--	-----------

Chapitre 1 – De la mesure à l'évaluation.

Regard sur un domaine en turbulence	13
Les épreuves uniques menant à la sanction des études	19
L'évaluation en salle de classe	22
Conclusion	27

Chapitre 2 – L'évaluation comme aide à l'apprentissage

dans la pratique enseignante	29
Qu'est-ce que l'évaluation comme aide à l'apprentissage?	31
Quel est le rôle de l'enseignant?	32
Conditions favorables à l'évaluation comme aide à l'apprentissage	33
Stratégies de mise en pratique de l'évaluation comme aide à l'apprentissage ..	34
Conclusion	41
Références	42

Chapitre 3 – La conception des examens et des tests:

le cas de l'éducation physique	45
Introduction	46
La notion d'épreuve standardisée	47
La validité	48
Les différents formats d'items	51
Deux théories des tests	58
Conclusion	66
Références	67

Chapitre 4 – La métagrille, un outil pour évaluer la qualité des grilles

d'évaluation: exemple d'une production en histoire	69
Introduction	70
La méthodologie utilisée: l'anasynthèse	70
Applications de la métagrille en salle de classe	84
Conclusion	89
Références	91

PARTIE II

DES INSTRUMENTS UTILES EN SALLE DE CLASSE	93
--	-----------

Chapitre 5 – La planification de l'apprentissage et de l'évaluation: la mise en place de dispositifs de différenciation pédagogique

en écriture	95
Introduction	96
Nature et fonctions de la planification	96
Analyse des outils de planification globale	100

La planification spécifique dans le contexte de la différenciation pédagogique.	102
Conclusion	111
Références	112
Chapitre 6 – La cartographie conceptuelle: un exemple d'utilisation en science et technologie	113
Introduction	114
Nature et caractéristiques de la carte conceptuelle	114
Comment utiliser la carte conceptuelle dans l'évaluation?	127
Conclusion	139
Références	140
Chapitre 7 – Les échanges verbaux: le cas de l'entrevue en lecture	143
Introduction	144
Les échanges verbaux	144
La discussion	149
L'entrevue	151
Démarche d'entrevue dans le contexte de la lecture	158
Conclusion	166
Références	167
Chapitre 8 – La situation de compétences: analyse des productions proposées en mathématique	169
Introduction	170
La notion de problème liée inextricablement au concept de situation	172
Les types de situations	173
Caractéristiques des situations de compétences visant l'évaluation de l'apprentissage	175
Grille d'analyse de la pertinence d'une situation de compétences	179
Conclusion	185
Références	186
Chapitre 9 – L'évaluation des stratégies d'apprentissage: l'enseignement explicite des stratégies en anglais, langue seconde	189
Introduction	190
Les stratégies d'apprentissage	191
L'enseignement et l'évaluation des stratégies d'apprentissage dans le programme d'anglais, langue seconde: « Montre-moi comment tu fais »	193
L'évaluation des stratégies	209
Conclusion	215
Références	217
Chapitre 10 – Le portfolio d'apprentissage à la formation professionnelle: le cas des programmes du secteur de la santé	219
Introduction	220
Caractéristiques d'un portfolio	220
Pourquoi un portfolio d'apprentissage en formation professionnelle?	222
Autoévaluation et métacognition	223
Structure et contenu d'un portfolio d'apprentissage	227
Format du portfolio	233
Pour bien se préparer à l'entretien individuel avec l'élève	234
Démarche de mise en œuvre	235
Conclusion	236
Références	245
Présentation des auteures et des collaborateurs	247



INTRODUCTION

*«Lorsqu'un inspecteur [de Mesures Canada (MC)] a essayé et approuvé un instrument de mesure, il y applique un autocollant – comme on en voit souvent sur les pompes à essence ou sur les balances des épiceries – pour indiquer que l'instrument est conforme et pour autoriser son usage commercial¹. »
Qu'en est-il en ce qui concerne l'instrumentation pour l'évaluation?*

L'évaluation des apprentissages : de la collecte de l'information à la communication des résultats (2006) se voulait un guide pratique pour favoriser la compréhension des enjeux de l'évaluation des apprentissages dans un contexte authentique, tout en proposant des modalités pour mettre en œuvre la démarche d'évaluation. Sa large diffusion dans le milieu universitaire et chez les praticiens a permis de combler un besoin et de répondre à de nombreuses questions relevant du domaine de l'évaluation. En 2012, une deuxième édition s'est avérée nécessaire afin de mettre à jour les sections de l'ouvrage relatives à des pratiques évaluatives qui avaient évolué ou s'étaient transformées. Cette nouvelle édition a été réalisée d'une part en illustrant la démarche d'évaluation de nombreux exemples issus de l'enseignement secondaire et collégial et, d'autre part, en approfondissant les pratiques relevant de la régulation des apprentissages et de la différenciation pédagogique, deux pièces maîtresses dans l'évaluation **pour** l'apprentissage, ou *assessment for learning*.

Ce deuxième tome met l'accent sur l'instrumentation utilisée à chaque étape de la démarche d'évaluation. Il reprend certains des outils présentés dans le premier tome sous l'angle de l'analyse conceptuelle et en propose aussi de nouveaux. Il s'appuie notamment sur plusieurs recherches récentes portant sur la manière dont l'évaluation se fait en salle de classe, et qui ont souvent mené au constat d'une compréhension ou d'une utilisation imparfaites de ces outils. Enfin, il s'intéresse à la qualité des instruments sous l'angle de leur validité, de leur fidélité, de leur faisabilité, de leur équité, de leur authenticité ou encore de leur pertinence.

Lors de la publication en 2003 de la Politique d'évaluation des apprentissages², le ministère de l'Éducation a établi que l'évaluation des apprentissages devait reposer sur des valeurs fondamentales telles que la justice, l'égalité et l'équité. À ces valeurs s'en ajoutent trois autres, dites « instrumentales » : la

1. <http://www.sciencetech.technomuses.ca/francais/collection/measurement7.cfm>

2. Ministère de l'Éducation du Québec (2003). *Politique d'évaluation des apprentissages. Formation générale des jeunes, formation des adultes, formation professionnelle*. Québec, Canada : Gouvernement du Québec.

rigueur, la cohérence et la transparence. L'ensemble de ces valeurs guide les qualités recherchées en évaluation et contribue notamment à ce que cette dernière ne porte pas préjudice à l'élève. Dans cet ouvrage, nous verrons comment les valeurs instrumentales peuvent être vérifiées dans les instruments d'évaluation choisis par l'enseignant. Pour ce faire, nous avons fait appel à de nombreux experts dans le domaine, qui, chacun leur tour, présenteront la nature, les fonctions et les caractéristiques d'un outil en les contextualisant dans un champ disciplinaire spécifique et en illustrant les pratiques qui y sont associées.

Mentionnons d'abord que cet ouvrage est divisé en deux parties constituées respectivement de quatre et six chapitres.

La première partie, intitulée « **Des instruments aux visées plurielles** », s'intéresse à plusieurs enjeux soulevés par la question de l'évaluation en éducation.

Le premier chapitre amorce la réflexion sur le sujet par la contribution de Jean-Guy Blais intitulée « De la mesure à l'évaluation. Regard sur un domaine en turbulence ». Ce plaidoyer situe l'évaluation des apprentissages en éducation comme un processus en contexte qui permet de documenter l'état des connaissances, des habiletés, des compétences, des attitudes, avec des instruments et des données de qualité afin que les décisions prises soient les plus adéquates possibles.

Le deuxième chapitre, « L'évaluation comme aide à l'apprentissage dans la pratique enseignante », est rédigé par Solen Poirier. Après avoir indiqué la nature de l'évaluation formative et le rôle de l'enseignant, l'auteure s'attarde à définir les multiples stratégies qui peuvent être utilisées pour mettre en œuvre l'évaluation comme aide à l'apprentissage.

D'autres préoccupations sont à la base du troisième chapitre, « La conception des examens et des tests : le cas de l'éducation physique » de Nathalie Loye. L'auteure y décrit la façon dont les épreuves doivent être construites et les étapes de validation auxquelles elles devraient être soumises, en traitant rapidement le cas des épreuves à grande échelle, pour se concentrer sur ce qui est nécessaire et utile à l'enseignant au quotidien. Des exemples tirés de l'évaluation en éducation physique et à la santé illustrent ces propos.

Le dernier chapitre de cette première partie, « La métagrille, un outil pour évaluer la qualité des grilles d'évaluation : exemple d'une production en histoire », est rédigé par Micheline-Joanne Durand et Mireille Mouffe. Le processus d'anasyntèse de Silvern est la méthodologie privilégiée par les auteures. Cette démarche, amorcée la première fois par Durand et Trépanier en 2011³ pour le concept de situation d'apprentissage et d'évaluation, amène l'élaboration d'un prototype qui se présente sous la forme d'une grille d'analyse. Assortie de critères et de descripteurs progressant sur quatre échelons, la grille d'analyse permet de juger de la qualité des instruments élaborés, quels que soient la dis-

3. Durand, M.-J., et Trépanier, I. (2011). Validité des situations de compétence : élaboration d'une grille d'analyse. Dans G. Raïche, K. Paquette-Côté et D. Magis (dir.). *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation* (vol. 2, p. 69-89). Québec, Canada : Presse de l'Université du Québec.

cipline et l'ordre d'enseignement. Dans ce chapitre, les auteures élaborent un prototype de grille d'évaluation dont l'objectif est l'analyse de la qualité des diverses grilles d'évaluation, qu'il s'agisse de celles servant à évaluer les travaux des élèves (en histoire, par exemple) ou de celles utilisées dans d'autres outils (le portfolio, par exemple).

La deuxième partie de cet ouvrage met l'accent sur « **Des instruments utiles en salle de classe** » en illustrant chaque démarche instrumentée dans une discipline spécifique.

Le cinquième chapitre débute avec la première étape de la démarche d'évaluation soit : « La planification de l'apprentissage et de l'évaluation : la mise en place de dispositifs de différenciation pédagogique en écriture ». Marie-Hélène Asselin et Mélanie Ducharme s'intéressent à la définition et aux fonctions du concept de planification en apportant un éclairage sur ses caractéristiques et en présentant une grille d'analyse des outils de planification. Elles s'attardent ensuite au contexte des élèves ayant des besoins particuliers intégrés dans les classes du secteur régulier et pour lesquels l'enseignant se doit de mettre en place des dispositifs de différenciation pédagogique dans sa planification spécifique.

Les chapitres suivants présentent des outils propres à la phase de prise ou de collecte de l'information de la démarche évaluative. Le sixième chapitre, « La cartographie conceptuelle : un exemple d'utilisation en science et technologie » de Maxim Morin et France Paquette, nous entretient sur les outils de représentations graphiques permettant, entre autres, le développement d'une vision globale des connaissances et de leur organisation. Après avoir répondu à plusieurs questions d'ordres théorique, mais aussi pragmatique, les auteurs proposent deux modèles d'interprétation des cartes conceptuelles, soit un système de notation quantitative ainsi qu'une grille à échelle descriptive analytique mis à profit pour évaluer une production d'élève du secondaire en science et technologie.

Le septième chapitre porte sur « Les échanges verbaux : le cas de l'entrevue en lecture ». Micheline-Joanne Durand et Monique Pouliot présentent une méthode évaluative basée sur les échanges verbaux qui regroupe différentes pratiques comme l'observation des interactions en contexte, la discussion et l'entrevue. Dans ce dernier cas, la nature, les fonctions et les caractéristiques sont examinées, puis réinvesties dans le contexte de la lecture au secondaire.

Suit le huitième chapitre, signé par Micheline-Joanne Durand, Ibrahima Sakho et Mélanie Ducharme et intitulé « La situation de compétences : analyse des productions proposées en mathématique ». Les auteurs s'attardent à clarifier la notion de situation de compétences dans le contexte de l'évaluation de l'apprentissage (*assessment to learning*). Là encore, une grille d'analyse de la pertinence des situations de compétences est proposée afin de bien cerner les caractéristiques d'une situation dans laquelle le processus et le produit sont considérés. C'est dans cette perspective que différentes situations de compétences en mathématique sont analysées à l'aide de la grille.

Par la suite, le neuvième chapitre nous amène vers « L'évaluation des stratégies d'apprentissage : l'enseignement explicite des stratégies en anglais, langue

seconde». Marilyn Lassire et Thanh Tu Nguyen abordent d'abord la nature, les fonctions et les types de stratégies d'apprentissage, et spécifiquement celles utilisées dans le domaine des langues secondes. Elles s'intéressent ensuite au développement d'une stratégie d'enseignement des stratégies par l'entremise des leçons *ad hoc*. Elles proposent également des outils d'évaluation des stratégies d'apprentissage en anglais, langue seconde, au primaire, pour les élèves et l'enseignant.

Nous ne pouvons conclure cet ouvrage sans parler du portfolio. Carole Tremblay nous propose ainsi le dernier chapitre concernant « Le portfolio d'apprentissage à la formation professionnelle: le cas des programmes du secteur de la santé ». Les caractéristiques et les fonctions d'un portfolio d'apprentissage sont abordées pour aider à comprendre en quoi cet outil d'évaluation peut être intéressant pour les élèves de la formation professionnelle. L'auteure met l'accent sur les outils d'autoévaluation qui permettent à l'élève de développer sa capacité à porter un regard critique sur ses apprentissages. Divers instruments pouvant faire partie d'un portfolio sont proposés ainsi qu'une démarche pour la mise en œuvre d'un portfolio, quel que soit le programme d'études professionnelles visé.

Pour conclure, n'oublions pas de garder en tête que la recherche dans le domaine de l'instrumentation en évaluation est encore jeune et que tout ce qui est présenté dans cet ouvrage représente un état des lieux actuel et est sujet à évoluer.

Enfin, sur une note plus personnelle, je tiens à ajouter que ce fût un réel plaisir pour moi de collaborer avec Nathalie Loye et les nombreux experts ayant contribué à la rédaction de cet ouvrage afin de vous proposer des outils qui vous aideront à porter un jugement bien documenté et instrumenté.

J'espère qu'au terme de votre lecture, vous serez à même de bien vous instrumenter pour mieux évaluer et apposer vous aussi votre autocollant de qualité!

Micheline-Joanne Durand

Dans cet ouvrage, les pictogrammes suivants
mettent en lumière différents types d'information :



La définition de notions fondamentales



Des conseils et des éléments importants
à retenir



Des pistes d'exploration pour l'enseignant



PARTIE I

DES INSTRUMENTS AUX VISÉES PLURIELLES

CHAPITRE 1

De la mesure à l'évaluation. Regard sur un domaine en turbulence

• Jean-Guy Blais **13**

CHAPITRE 2

L'évaluation comme aide à l'apprentissage dans la pratique enseignante

• Solen Poirier **29**

CHAPITRE 3

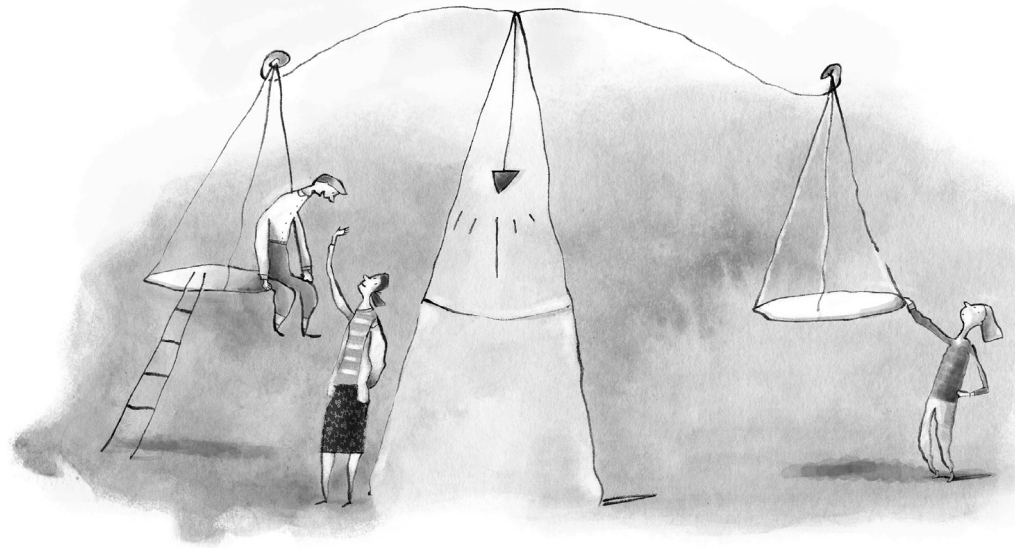
La conception des examens et des tests : le cas de l'éducation physique

• Nathalie Loye **45**

CHAPITRE 4

La métagrille, un outil pour évaluer la qualité des grilles d'évaluation :
exemple d'une production en histoire • Micheline-Joanne Durand

et Mireille Mouffe **69**



De la mesure à l'évaluation. Regard sur un domaine en turbulence

Par Jean-Guy Blais¹

« La balance et le peson, l'équerre et le compas, agissent d'une manière uniforme et invariante. Ni le peuple de Chhin ou le peuple de Chhu ne peuvent changer leurs propriétés, ni les Hu du nord et les Yueh du sud ne peuvent modifier leur forme. Ces instruments sont les mêmes pour la vie. Une seule journée les a produits; dix milles générations les ont propagés. »

Huainan Zi, II^e siècle av. J.-C.

1. Jean-Guy Blais est professeur titulaire à l'Université de Montréal au Département d'administration et fondements de l'éducation, spécialiste de la mesure et de l'évaluation.

Le mot «évaluer» vient du latin *valutare*, qui signifie «déterminer la valeur de». À l'époque où le latin était une langue de premier plan en Europe, cette expression prenait tout son sens, principalement dans certains échanges commerciaux. Lors de l'achat de biens ou de nourriture, il existait, comme aujourd'hui, des conventions sociales permettant un arbitrage sur la quantité. Toutefois, ces échanges n'étaient jamais totalement indépendants des personnes impliquées et du contexte propre à leur époque. Ainsi, lorsqu'une personne voulait acheter des olives au marché, la valeur de la marchandise était fonction de sa quantité, certes, mais aussi de sa qualité, de sa provenance, de sa rareté, de son coût de production, de la marge de profit visée, mais aussi de l'honnêteté du marchand, de la tête du client et de sa capacité à négocier ou à intimider. Il fallait donc d'abord déterminer le poids ou le volume de la marchandise et, en fonction de la valeur accordée à une unité de poids ou de volume, arrêter un prix pour la quantité désirée par un client, le tout en fonction de plusieurs critères plus ou moins prévisibles liés au contexte particulier de l'échange. D'autres biens étaient vendus à l'unité plutôt qu'au poids ou au volume. Ainsi, il y avait un prix fixe pour un melon et le prix de plusieurs melons était déterminé par une simple multiplication. Le plus souvent, un rabais proportionnel au nombre d'unités achetées ou résultant des qualités de négociateur du client était consenti.

Dans un premier temps, l'attribut à mesurer est défini, puis sa quantité est mesurée avec un instrument étalonné, comme une balance, qui produit une donnée stable dans le temps et l'espace. Déterminer la quantité d'un attribut est une chose, mais en arrêter la valeur, le prix, en est une autre. Comme il a été souligné ci-dessus, plusieurs critères entrent en interaction lorsqu'il s'agit de déterminer la valeur de ce qui fait l'objet des échanges commerciaux. C'était vrai à une époque lointaine et ce l'est encore de nos jours.

Il faut donc d'abord qu'il y ait une forme de consensus sur ce que représente l'objet de la quantification (le poids par exemple), et que l'instrument utilisé pour déterminer la quantité soit calibré et en donne une indication précise et fiable, reconnue par les deux parties. Il s'agit d'une donnée tangible et objective, qui permet aux protagonistes de s'entendre sur la quantité de l'attribut et sur ce qu'elle représente, le tout dans un climat de confiance où celui qui détermine la mesure de la quantité, c'est-à-dire le propriétaire du contenant ou de la balance, ne triche pas. Cette confiance doit être réciproque, et il est également supposé que, lors de la transaction, le client est de bonne foi, en ne payant pas avec de la fausse monnaie par exemple.

Ensuite, un critère d'évaluation plus subjectif est celui de la qualité. Lorsque la qualité entre en compte pour déterminer la valeur d'un objet, il faut faire appel à des critères spécifiques permettant de situer celui-ci par comparaison avec d'autres objets similaires. Si celui qui désire faire l'acquisition d'une marchandise ne dispose pas des bons principes pour juger si elle répond au critère de qualité, ou s'il n'a jamais pu faire de comparaison avec une marchandise équivalente, mais d'une qualité distincte, il ne peut réellement en établir la qualité. Celle-ci se détermine par la prise en compte d'une variété de critères

internes établis par les expériences et les observations d'une ou plusieurs personnes aptes à porter un jugement éclairé en la matière.

La rareté est un autre critère qui peut être pris en compte pour déterminer la valeur d'une marchandise. À titre d'exemple, le prix élevé des épices en Angleterre au XVI^e siècle était justifié par leur rareté sur le marché, elle-même attribuable à la distance parcourue — on les importait de lointaines contrées d'Asie — et à la dangerosité des routes empruntées par les caravanes pour les obtenir. Mais la rareté peut être une caractéristique moins objective, influencée par des critères plus relatifs et individuels, comme le prestige associé à l'objet ou le désir qu'il suscite chez l'acheteur. Elle peut aussi dépendre de contextes politiques, socio-économiques ou climatiques particuliers, comme la guerre ou la sécheresse. La rareté est donc une donnée supplémentaire dont il faut tenir compte, en gardant toutefois à l'esprit qu'elle est fortement influencée par le contexte.

Finalement, deux autres éléments peuvent être considérés quand il s'agit de déterminer la valeur d'un objet : l'honnêteté du marchand et l'équité dont il fait preuve à l'égard de ses clients. D'abord, on espère que le marchand ne fausse pas les données en utilisant des instruments qui ne seraient pas valides pour mesurer une quantité comme le volume ou le poids. Ensuite, on espère que les clients soient traités de manière équitable et qu'il n'y ait pas de favoritisme lié à l'amitié, à des faveurs espérées ou à la crainte de représailles. On espère donc qu'il n'y a pas d'effet de halo ou de discrimination basée sur l'origine du client. Les données que l'on peut recueillir sur l'honnêteté et l'équité d'un marchand sont du même ordre que celles associées à la qualité ou à la rareté. Elles ne sont pas aussi facilement mesurables que le poids ou le volume, car elles sont intimement liées au comportement et à l'éthique des personnes dont elles émanent.

Ce détour à propos des échanges commerciaux permet une comparaison intéressante avec la démarche d'évaluation des apprentissages, car, dans les deux cas, il s'agit de statuer sur la valeur de quelque chose (marchandise ou apprentissage) en tenant compte du contexte et des objectifs, puis d'obtenir des données valides pour accompagner le jugement sur les éléments constitutifs de la valeur. On peut recourir à divers types de données dans le processus d'évaluation de la valeur : des données objectives provenant d'une mesure valide et fiable d'une ou plusieurs quantités étalonnées, et des données plus subjectives liées au contexte de l'évaluation, comme la qualité, la rareté, l'honnêteté et l'équité.

L'évaluation des apprentissages est une pratique sociale et un produit social placés sous la responsabilité d'êtres humains², que ce soit pour l'élaboration des instruments ou l'analyse et l'interprétation des données utilisées à cette fin. Cette pratique et ce produit ont varié à travers les époques et les sociétés, et

2. C'est encore le cas actuellement, mais, ces dernières années, plusieurs recherches ont fait état, dans le cadre d'opérations à grande échelle, du recours à l'informatique pour la génération automatisée de questions et pour la correction automatisée de textes courts ou longs. Il n'est donc pas impossible dans ce contexte que, d'ici une dizaine d'années, la contribution humaine dans l'évaluation des apprentissages soit réduite.

elles ont subi des influences politiques, idéologiques et scientifiques au gré des changements politiques, démographiques, économiques et technologiques. Ainsi, lorsqu'il faut évaluer un apprentissage et qu'il faut tenter d'en déterminer la valeur, le processus adopté n'est jamais indépendant de l'objectif de l'opération et du contexte, en plus d'être intimement lié à l'objet (c'est-à-dire à ce qui est évalué). Il dépend aussi de la situation dans laquelle l'évaluation a lieu, de la personne qui en a la charge, de la manière dont elle est menée, de l'instrument utilisé pour collecter des données, et, évidemment, de la clientèle visée.

Il serait possible de remonter assez loin dans l'histoire et de faire état de pratiques très organisées d'évaluation comme la sélection des mandarins dans la Chine du II^e siècle, ou encore de pratiques locales comme l'évaluation des apprentissages à l'Université Harvard au XVI^e siècle. Il n'en demeure pas moins que ce n'est que très récemment, à partir de la fin du XIX^e siècle, avec la massification de l'éducation dans les sociétés industrialisées, que les diverses conceptions modernes de la connaissance et de l'apprentissage se sont précisées et affrontées dans le débat public³, et que les instruments accompagnant les processus d'évaluation attenants se sont développés, affinés et imposés.

Le début du XX^e siècle a ainsi vu l'avènement de la psychométrie, une discipline s'inspirant des récents progrès de la physique et des travaux de la psychophysique. Celle-ci avait pour objectif de quantifier ou de classer des aspects intangibles de l'esprit humain en faisant ressortir les différences individuelles liées à des attributs psychologiques, comme les capacités intellectuelles ou les composantes de la personnalité, qui ont été étiquetées comme des « traits latents » parce qu'elles ne pouvaient être observées directement (de nos jours, c'est l'expression « variables latentes » qui est surtout utilisée en psychométrie). Dans cette discipline, les chercheurs ont développé des instruments pour collecter des données permettant de situer l'état des individus dans le continuum de ces traits latents. À cet effet, au moins deux hypothèses étaient acceptées d'emblée: d'abord, même si on ne pouvait l'observer directement, l'attribut théorisé, le trait latent, existait vraiment et il était possible d'en observer certaines manifestations; ensuite, l'intensité de l'attribut variait selon les individus. La première hypothèse nous ramène à ce que l'on a défini par la suite comme la « validité du construit », et la deuxième hypothèse, à l'existence d'une quantité que l'on peut mesurer. Pour faire une analogie avec l'utilisation de la balance dans les échanges marchands, on peut dire que le fait de vouloir déterminer le poids de quelque chose implique d'abord de définir ce que l'on entend par « poids » et, ensuite, de montrer que l'on peut produire des unités équivalentes de cet attribut permettant de distinguer des choses ayant des poids différents. En psychométrie, le construit est ainsi une théorisation élaborée à partir des manifestations observables d'un attribut humain, une sorte de fiction utile, qui, selon la force des éléments de preuve, fait consensus dans la communauté scientifique ou professionnelle où il est valorisé. L'étiquette accolée à cette

3. Un exemple typique et d'actualité est le débat sur la position prépondérante des connaissances et des compétences dans le cycle d'apprentissage, ou encore sur la valeur des arguments socioconstructivistes en éducation.

théorisation, le mot «intelligence» par exemple pour désigner les capacités intellectuelles, a facilité les échanges grâce au recours à un terme ayant le même référent pour tous.

En psychométrie, l'instrument privilégié pour collecter des données est appelé un «test» et, le plus souvent, ce dernier est composé d'un certain nombre de petites unités, les items, qui sont soit des rappels de connaissances, soit de courts problèmes à résoudre. Ces items peuvent aussi prendre la forme d'une série d'énoncés pour lesquels le candidat doit préciser un degré d'accord ou de désaccord. Dans beaucoup de tests, le candidat se voit proposer un choix de réponses pour chaque item (questionnaire à choix multiples [QCM] ou échelle de type Likert), mais cette approche n'est pas exclusive et d'autres types de supports pour la réponse ont été proposés et utilisés. Traditionnellement, les tests sont passés sur papier, mais, avec l'émergence de la technologie, les tests électroniques vont croissant. Les réponses des candidats sont ensuite transformées en «scores» par un processus en deux étapes: une première étape de notation et une deuxième étape de modélisation mathématique plus ou moins complexe. Par exemple, une approche classique dans un test de capacités intellectuelles consiste à donner une note quelconque pour une bonne réponse à un item, 1 par exemple, et à construire un score pour le test en comptant toutes les notes associées à une bonne réponse, c'est-à-dire en déterminant leur fréquence. Ce modèle simple est plus ou moins réaliste, car il considère que les items sont de même valeur et il ne pondère pas les notes attribuées selon leur difficulté ou leur puissance de discrimination⁴.

Aux États-Unis, le berceau de la psychométrie moderne, les tests ont d'abord été utilisés pour des besoins militaires et pour le classement des nouveaux immigrants, mais, de nos jours, ils le sont largement pour la sélection du personnel en entreprise ou pour l'orientation sur le marché du travail, par exemple. Ainsi, tout un domaine de pratiques, le *testing*, et de modélisations mathématiques — les théories et modèles de la mesure — a d'abord fleuri presque exclusivement en psychologie, avec une terminologie technique bien précise, les tests et les items, et des concepts incontournables comme la validité et la fidélité. Très tôt cependant, les pratiques de la psychométrie ont essaimé dans des champs d'application connexes. Plusieurs pratiques associées à l'évaluation des apprentissages se sont ainsi développées dans l'ombre de la psychométrie pour ensuite s'en affranchir en partie. En éducation, l'influence de cette théorie a donc donné naissance à un domaine organisé de pratiques que nos collègues anglo-saxons ont appelé l'*educational assessment*. Les tests y sont devenus des examens, des épreuves ou des questionnaires; les items, des questions ou des énoncés; et les scores, des notes chiffrées, des résultats ou des cotes littérales (A, B, C, D, E). L'évolution de la réflexion sur les apprentissages réalisés par les individus dans les sociétés organisées, sur le rôle de la connaissance et sur sa structuration a permis à l'évaluation des apprentissages en éducation de suivre une voie qui lui est propre, sans pour autant s'affranchir totalement de certaines

4. Voir à ce sujet le chapitre 3 du présent ouvrage.

conceptions et pratiques empruntées au domaine de la psychométrie. Ainsi, les expressions « tests » et « scores » sont bien présentes en éducation et elles sont souvent utilisées pour désigner des instruments et des données dont les caractéristiques peuvent être éloignées de celles qu'ont ces termes en psychométrie. Au-delà des équivalents du test que seraient l'examen ou l'épreuve et qui marquent toujours la proximité avec cette discipline, d'autres types de tâches et d'instruments pour collecter des données ont été développés en éducation. Les travaux longs, seul ou en équipe, les démonstrations, les projets ou les grilles de correction et d'observation sont autant d'exemples d'une pratique distincte et moins standardisée que ce qui entoure la mise au point de tests en psychométrie.

Mesurer ou collecter des données ?

Un instrument de mesure est avant tout une technologie visant à recueillir des données. Cette technologie est sensible aux variations entre les données et intègre un modèle mathématique (simple ou complexe) permettant de transformer les données en unités égales (ou mesures) et pouvant être combinées selon des mécanismes de concaténation (l'addition par exemple). Les mesures ainsi produites sont indépendantes du moment et du lieu de leur collecte. Ainsi, le simple fait d'utiliser un instrument pour collecter des données et de les codifier avec des nombres ne signifie pas qu'on est en train de *mesurer* quelque chose. Donc, est-ce qu'un test, un examen, un questionnaire ou une grille de correction sont des instruments de mesure ? Ce sont bien des outils permettant de recueillir des données, mais ils sont potentiellement instables et fortement dépendants du contexte de collecte. Est-ce que ces données peuvent être transformées en mesures grâce à un modèle mathématique ? Peut-être, mais il faudrait le démontrer, ce qui est rarement fait lorsqu'on évalue les apprentissages avec les outils traditionnels dont nous disposons. Les données recueillies avec ces outils ne deviennent donc pas des mesures par la magie de leur seule représentation par des nombres entiers (0, 1, 2, 3, etc.), mais plutôt lorsqu'elles sont mises en contexte et qu'elles deviennent des informations utiles pour la prise de décision, ce qui reste l'objectif ultime de l'évaluation des apprentissages en éducation.



En éducation, il est possible de définir au moins **deux contextes⁵ d'évaluation des apprentissages**. Ils ont des objectifs distincts, font appel à des instruments différents (mais pas toujours) et ne sont pas sous la responsabilité

5. Pour le besoin de l'exercice, seuls deux contextes sont présentés, mais il serait possible d'être plus nuancé et d'en évoquer d'autres.

des mêmes personnes. Les personnes visées au premier chef, celles qui sont appelées à démontrer leurs apprentissages, restent cependant toujours les mêmes : les élèves. Le premier contexte est celui des opérations à grande échelle que sont les épreuves ministérielles et dont les résultats sont pris en compte pour la sanction des études. Lors de ces opérations, on cherche principalement à établir un constat sur la performance individuelle, mais on utilise aussi des résultats agglomérés (la moyenne ou le taux de réussite par exemple) pour obtenir un regard macroscopique sur le système. Au Québec, les épreuves sont sous la responsabilité de professionnels et sont élaborées en fonction du Programme de formation de l'école québécoise en vigueur. La passation d'une épreuve est ponctuelle et accaparante, en ce sens qu'elle a lieu à un moment bien précis et qu'elle demande une organisation particulière qui perturbe les activités éducatives du quotidien. Le deuxième contexte est celui où l'évaluation des apprentissages se vit au jour le jour dans une interaction continue. Il s'agit de l'évaluation en salle de classe. Il peut y avoir des variantes selon les juridictions et les pays, mais l'évaluation des apprentissages en salle de classe au quotidien est directement sous la responsabilité de l'enseignant (ou d'une petite équipe d'enseignants). Ce dernier en choisit les stratégies et les instruments, en détermine le moment, en gère la passation, recueille et analyse les données, pose un diagnostic et, seul ou après consultation de ses pairs, prend les décisions qui s'imposent.

Ces deux contextes d'évaluation ont des objectifs différents, mais ils peuvent tous deux être comparés à une opération d'évaluation du domaine marchand, car ils font appel à des critères similaires quand vient le temps de déterminer la valeur. Ainsi, dans chacun de ces contextes, on constate que la valeur d'un apprentissage, son évaluation donc, implique une quantité étalonnée et des instruments de mesure valides et fiables. On peut aussi se demander comment se manifestent les critères associés à la qualité et à la rareté, et ceux associés à l'honnêteté et à l'équité.

LES ÉPREUVES UNIQUES MENANT À LA SANCTION DES ÉTUDES

Au Québec, les épreuves uniques du ministère de l'Éducation, du Loisir et du Sport (MELS) de la fin du secondaire sont des exemples typiques de ce qui a cours dans un contexte certificatif d'évaluation des apprentissages à grande échelle. Elles se déroulent selon des modalités bien arrêtées : tous les candidats doivent passer la même épreuve dans les mêmes conditions et les critères qui mènent à l'attribution de la note doivent être appliqués uniformément. Il s'agit d'un contexte très standardisé où l'importance accordée à la validité et à la fidélité de l'épreuve se rapproche de ce qui a cours lors de la validation de tests en psychométrie. Les épreuves ont lieu à la fin d'un parcours et, comme elles mènent, en partie, à l'obtention du diplôme, les résultats qu'un candidat y obtient sont déterminants pour la reconnaissance sociale des apprentissages réalisés. Elles fournissent aussi des données qui permettent de brosser un portrait d'ensemble,

quoique ponctuel, des apprentissages prescrits dans les programmes pour différentes cohortes d'élèves. Les données prennent alors la forme de taux de succès ou de moyennes des résultats et sont examinées dans une perspective transversale avec une fonction systémique de régulation et de contrôle.

Les responsables ministériels pourraient essayer de montrer que les notes attribuées sont des mesures de « quelque chose », des mesures qui seraient du même ordre que ce qu'on obtiendrait avec une balance, mais ce serait très difficile d'y arriver et ce n'est pas l'objectif. Une épreuve scolaire, aussi uniforme ou unique soit-elle, n'est pas une balance, et les résultats chiffrés des élèves n'ont pas les propriétés du kilogramme. Le modèle de mesure le plus souvent utilisé est simple et consiste en l'addition pondérée ou non des cotes numériques attribuées aux différentes composantes de l'épreuve. Cependant, étant donné les mécanismes de production⁶ d'une épreuve unique, les notes ont un caractère objectif et permettent de comparer les individus. Elles sont aussi des indicateurs de l'évolution transversale du système d'éducation. Ainsi, les données recueillies et transformées en notes peuvent certainement informer les différents intervenants qui devront prendre des décisions sur l'attribut qui est visé par l'opération d'évaluation, par exemple la compétence en français écrit en 5^e secondaire ou en mathématique en 4^e secondaire. Malgré tout, le résultat à une épreuve ministérielle unique typique reste un portrait pris à un moment précis et établi en une durée plutôt limitée (trois heures par exemple). Les responsables de cette épreuve présument que cet instantané représente un portrait relativement valide et fiable des apprentissages réalisés. Évidemment, un laps de temps aussi court conditionne les instruments qu'il est possible d'utiliser et explique pourquoi, dans le contexte des épreuves uniques à grande échelle, les préoccupations au sujet de la validité et de la fidélité se rapprochent davantage de ce qui a cours en psychométrie et peuvent s'éloigner de préoccupations liées à l'authenticité de la tâche, à la collaboration entre les élèves ou encore au développement des compétences.

L'objectif principal des épreuves uniques du MELS n'est pas de produire une mesure d'une très grande précision, mais plutôt de s'assurer que la réussite à ces épreuves garantit une certaine qualité des apprentissages réalisés par les élèves. D'abord et surtout parce que la société confie explicitement au gouvernement le mandat⁷ de faire en sorte que chaque élève réalise les apprentissages qu'elle considère comme importants pour une intégration sociale adéquate. Ensuite, parce que des observateurs externes — d'autres instances scolaires ou universitaires — pourraient avoir besoin de résultats relativement fiables pour effectuer des comparaisons dans le cadre d'un processus de classement ou de sélection. Les instruments de collecte des données sont en harmonie avec les programmes et les objectifs de la formation générale des élèves, et l'opération d'évaluation a un objet bien déterminé, par exemple le domaine

6. La planification, l'élaboration, la passation et la correction.

7. À travers la Loi sur l'instruction publique par exemple.

de la mathématique de la 4^e année du secondaire. Le contenu de l'épreuve est élaboré en fonction des éléments constitutifs du domaine et, idéalement, sa validité est assurée par une représentation équilibrée des tâches que l'élève a dû exécuter durant son année scolaire. Le résultat à l'épreuve constitue une représentation ponctuelle des apprentissages réalisés, une représentation limitée évidemment, mais qui joue son rôle dans un contexte où les institutions doivent veiller à la qualité des apprentissages. La qualité de chacune des réponses, des produits ou des performances est déterminée selon leur conformité à une série de critères établis par les responsables, mais qui sont le fruit d'un consensus sociétal. Cette conformité peut être établie lors d'une opération centralisée, comme pour les épreuves d'écriture en français en 5^e année du secondaire, ou lorsqu'on demande aux enseignants de corriger eux-mêmes les copies des élèves et d'appliquer une grille de correction bien précise, par exemple pour les épreuves obligatoires d'écriture en anglais, langue seconde, en 5^e année du secondaire.

Le type de validité qui importe dans les épreuves uniques est lié au contenu : celui-ci doit bien représenter le domaine d'apprentissage tel qu'il figure dans les programmes plutôt qu'un construit quelconque de l'attribut visé par l'épreuve. C'est donc en fonction de cette validité de contenu que les responsables peuvent garantir une qualité minimale, que la réussite à une épreuve unique officialise, ainsi qu'une qualité différenciée en proposant un classement des individus selon des notes. Cette qualité différenciée permet aussi de déterminer la rareté de certains apprentissages, dans une opération de sélection par exemple, en l'associant à la qualité la plus ou la moins élevée. Le fait que l'épreuve soit sous la responsabilité d'autorités ministérielles (direction de l'évaluation et direction de la sanction dans le cas du MELS) vise à garantir l'honnêteté des professionnels qui élaborent l'épreuve, ceux-ci n'ayant pas d'intérêt à la fausser. Il est préférable pour eux que le processus ainsi que le regard qu'il permet de porter sur les individus soient les plus équitables possibles. Finalement, le fait que les contextes de passation et de correction soient très standardisés garantit généralement que tous bénéficieront du même traitement. Les responsables peuvent alors faire valoir que ce processus est, autant que faire se peut, juste et équitable, sans discrimination par rapport à l'origine sociale, géographique ou ethnique de l'élève.

Toutefois, dans un contexte où les épreuves constituent des enjeux critiques, pour les élèves évidemment, mais aussi pour les écoles et les enseignants, plusieurs éléments hors du contrôle des responsables de l'épreuve peuvent faire en sorte que le portrait obtenu ne soit pas aussi précis que souhaité. Dans ces situations, on présume que les élèves ont fait étalage de toutes leurs capacités et ont bénéficié de conditions comparables. Or, des problèmes de santé, par exemple, peuvent avoir un effet non négligeable sur la performance d'un élève, qui ne sera alors plus un juste reflet de ses capacités. Cependant, ces problèmes peuvent être pris en compte par les responsables grâce à des reports ou des accommodements, car ils sont isolés, justifiables et documentés. Les problèmes

de motivation ou d'effort, qui peuvent aussi avoir un effet sur la performance, entrent quant à eux dans une autre catégorie, et en tenir compte demanderait une réflexion en profondeur. On postule donc par nécessité et pragmatisme que tous les élèves sont également motivés et qu'ils fournissent un effort maximal. Pour ce qui est des conditions similaires, les responsables s'assurent que les épreuves sont standardisées (passation et correction), mais ils ne peuvent contrôler l'intensité de la préparation des élèves, qui peut varier d'une école à l'autre et selon les enseignants. Est-ce équitable, par exemple, de faire participer à la même épreuve des élèves qui ont été entraînés intensivement dans ce but, et ce, au détriment du programme, et d'autres élèves qui n'ont eu pour entraînement que les exercices réguliers en lien avec le programme? Le bachotage ou les interventions du type *teach to the test* sont inévitables lorsque les enjeux sont cruciaux, et ils colorent différemment les résultats selon l'intensité des interventions de l'enseignant, créant *de facto* des conditions inégales. Si cette préparation se fait au détriment du programme, elle soulève une question éthique et devrait faire appel au professionnalisme (le savoir-être) des enseignants et des directions d'école. Un autre élément en lien avec les conditions égales, les avantages indus et l'éthique concerne la tricherie. Cette dernière vient le plus souvent des élèves, mais il arrive, comme on l'a vu ces dernières années avec la publication de palmarès d'écoles, qu'elle vienne aussi d'enseignants ou de directions d'école. Lorsque les enjeux d'une épreuve sont cruciaux, il arrive souvent que certaines personnes tentent de contourner les règles pour avoir accès aux avantages que confère la réussite de l'épreuve⁸, en prenant des raccourcis pour éviter de se soumettre aux mêmes conditions que les autres.

Donc, de prime abord, les épreuves uniques permettent d'obtenir des données qui, même si elles ne constituent pas des mesures étalonnées, peuvent jouer leur rôle en dressant un portrait assez objectif des apprentissages réalisés et de leur qualité, dans un contexte relativement équitable. Quand vient le temps de déterminer la valeur des apprentissages réalisés, on peut tenir compte de situations particulières liées à la santé. On présume par ailleurs que tous ont fait un effort semblable, que la préparation des élèves a été menée avec professionnalisme et que les résultats ont été obtenus sans tricherie. Cependant, les responsables ministériels ne contrôlent qu'une partie de l'équation et ne peuvent garantir que les conditions sont exactement les mêmes pour tous.

L'ÉVALUATION EN SALLE DE CLASSE

L'enseignant n'a pas le même regard sur les apprentissages des élèves que les responsables des épreuves ministérielles, et les choix qu'il fait et les décisions qu'il prend quant à leur évaluation au jour le jour ne sont pas du tout du même ordre. Les données recueillies sont semblables de prime abord, mais la fréquence

8. Ce comportement n'est évidemment pas exclusif au monde de l'éducation; le monde du sport, pour ne citer qu'un exemple, est assez éloquent à cet égard.

de la collecte des données, la manière dont elles sont analysées, leur mise en contexte et, surtout, les facteurs qui peuvent influencer les conditions de leur production diffèrent grandement dans ces deux contextes.

Évaluation et outils pour l'évaluation

Trop souvent en éducation, il arrive que l'on confonde l'évaluation des apprentissages et les outils mobilisés pour cette évaluation. Ainsi, on peut régulièrement lire ou entendre « l'évaluation passée par les élèves » ou encore « l'évaluation était difficile », alors qu'on devrait plutôt lire, par exemple, « l'examen passé par les élèves » et « l'épreuve était difficile ». L'évaluation des apprentissages est un processus global qui demande de préciser les objectifs, de mettre au point des outils pour collecter des données qui permettront de documenter l'atteinte de ces objectifs (par exemple, les questions ou les tâches, les critères de correction et le barème de notation), de préciser le contexte de cette collecte (le lieu, le moment, la durée et les accommodements, s'il y a lieu), de confronter les données obtenues avec ce qui était attendu (la correction) et, finalement, de qualifier le résultat en fonction du barème ou de la politique en vigueur (la notation) en vue d'une prise de décision transparente et bien documentée. L'examen, l'épreuve, le questionnaire, la grille de correction sont donc des éléments importants de l'évaluation des apprentissages puisqu'ils formalisent la prise de données, mais ils ne constituent pas en eux-mêmes l'évaluation des apprentissages. Ainsi, les élèves peuvent être soumis à des questions pour être évalués, mais les questions auxquelles ils répondent ne représentent pas à elles seules le processus d'évaluation.



Une des différences notables entre ces deux contextes réside dans la variété d'instruments que l'enseignant peut élaborer et utiliser pour recueillir les données qui lui serviront à évaluer les apprentissages de ses élèves. Alors que les responsables des épreuves ministérielles n'ont qu'une seule occasion de collecter des données et qu'ils sont très préoccupés par la standardisation des conditions de passation et de correction, l'enseignant a plus de latitude à cet égard. Pour lui, l'objectif de la collecte des données conditionne en quelque sorte l'approche et les instruments utilisés. Il peut créer ses propres instruments selon les besoins de ses élèves, en appliquant des principes de base de planification abordés lors de sa formation initiale et approfondis avec l'expérience. Il peut aussi facilement réutiliser ponctuellement des instruments qui ont fait leur preuve dans un contexte de régulation. De plus, il n'est pas soumis aux mêmes impératifs que l'épreuve unique quant au moment et à la durée de la collecte des données, et il peut le faire en tout temps et moduler la durée en fonction de la modalité.

Un contrôle ponctuel est court, un examen plus long, et la durée d'un projet varie selon les tâches demandées. En outre, il dispose d'un avantage important, car il peut observer continuellement ses élèves. En salle de classe, la diversité des modalités d'évaluation et la variabilité dans la fréquence de ces évaluations sont d'ailleurs toujours souhaitables. Cela permet à l'enseignant d'observer les apprentissages de l'élève sous plusieurs angles et ainsi de porter un jugement plus nuancé à ce propos. Une approche diversifiée donne à l'enseignant une meilleure idée de la qualité des apprentissages réalisés et, par la même occasion, lui permet de mieux adapter ses interventions. Il pourra donc réguler les apprentissages au jour le jour, mais aussi les contrôler sur une plus longue période et les certifier lorsque viendra le temps d'en faire le bilan. Cette diversité est très importante, car elle permet d'obtenir des traces variées qui ont différents rôles quant à la reconnaissance des apprentissages réalisés et qui peuvent aussi être utilisées pour informer avec transparence et à différents moments les divers acteurs intéressés : les élèves, les parents, les directions, les autorités des commissions scolaires et ministérielles. La validité du regard que pose l'enseignant sur la qualité des apprentissages est ainsi en lien direct avec la diversité des instruments et avec la fréquence de la collecte de données.

Toutefois, l'autonomie liée aux moyens utilisés et la diversité qui en découle introduisent *de facto* une première limite quant à la possibilité d'un jugement équitable des apprentissages des élèves. Lorsque chaque enseignant utilise ses propres instruments, c'est un peu comme si chacun avait en tête une balance avec des unités différentes. Les élèves d'une même classe peuvent être comparés selon les mêmes critères, mais lorsque deux enseignants utilisent des instruments différents, ces comparaisons deviennent toutes relatives. Elles dépendent alors de la nature des tâches réalisées et des critères pris en compte pour juger de l'adéquation du « produit » avec ce qui est attendu. C'est pourquoi les notes attribuées par des enseignants qui utilisent des instruments différents ne sont pas comparables et n'ont évidemment pas les propriétés des mesures stables et étalonnées d'une quantité. Ce sont des jugements relatifs et qualitatifs, basés sur des informations viables certes, mais valides dans un contexte restreint.

La validité des jugements des enseignants sur les apprentissages des élèves est aussi tributaire de facteurs divers qui prennent une importance variable chez chacun. Un premier facteur concerne leur formation en évaluation des apprentissages et leur expérience en la matière. Actuellement, la formation initiale en évaluation des apprentissages des futurs enseignants au Québec est d'environ 45 heures, concentrées dans un trimestre. Tous les futurs enseignants sont donc initiés aux principes de l'évaluation des apprentissages et à des stratégies leur permettant de mettre au point différents instruments de collecte des données. Toutefois, pour une formation initiale en enseignement de quatre ans, la place accordée à la formation spécifique en évaluation reste assez restreinte pour une tâche aussi importante et centrale dans l'accompagnement du développement des élèves. En effet, sans évaluation, il n'y a aucun moyen de savoir si les objectifs visant l'acquisition des connaissances et le développe-