

Réflexion : est-il possible de mesurer une compétence en contexte d'évaluation à grande échelle?¹

Sébastien Béland, Université de Montréal, Canada

Résumé : Si les autorités politiques et économiques des pays occidentaux et asiatiques appuient généralement les évaluations à grande échelle, par exemple le programme international pour le suivi des acquis des élèves (PISA) de l'OCDE, il est nécessaire de prendre un moment afin de se poser la question suivante : peut-on mesurer une compétence? Nous tenterons, dans cet article, de démontrer qu'il est possible de répondre positivement à cette question, mais que cela engendre de grands défis théoriques et pratiques.

Abstract : If the political and economic authorities of Western and Asian countries generally support large-scale assessments (e.g., Programme for International Student Assessment (PISA), OECD), it is necessary to take a moment to ask us the question: can we measure a competency? The aim of this paper is to demonstrate that it is possible to respond positively to this question, but this creates great theoretical and practical challenges.

*Mots-clés : Mesure; Compétence; Trait latent; Réalisme; Constructivisme; Enquête à grande échelle; PISA
Keywords: Measurement; Competency; Latent trait; Realism; Constructivis;= Large-scale assessment, PISA*

En science, quand le comportement humain
entre dans l'équation, les choses sont non
linéaires. Voilà pourquoi la physique est
facile et la sociologie est difficile²
(traduction libre).

Neil DeGrasse Tyson

Introduction

Le programme international pour le suivi des acquis des élèves (PISA, Organisation de Coopération et de Développement Économiques) est une initiative ayant comme objectif de comparer différents systèmes éducatifs sur la base de la mesure des compétences des élèves d'environ 15 ans dans trois matières fondamentales : la compréhension de l'écrit par la lecture, la culture mathématique et la culture scientifique. Le tableau 1 présente un survol des différentes compétences investiguées depuis 2000.

Tableau 1. Compétences évaluées à différentes passations du test PISA

	2000	2003	2006	2009	2012	2015
Lecture	Lecture	Lecture	Lecture	Lecture	Lecture	Lecture
Mathématique	Mathématique	Mathématique	Mathématique	Mathématique	Mathématique	Mathématique
Science	Science	Science	Science	Science	Science	Science

Nous observons qu'il y a une passation du test tous les trois ans et qu'un domaine principal (indiqué en caractère gras) est investigué à chacun des cycles de passation. En 2012, 65 pays ont participé au test pour un total d'environ 470 000 élèves évalués. Notons que nous serons, après la passation de 2015, à la fin de la deuxième succession des trois cycles de compétences.

¹ Cette réflexion a été bonifiée par de nombreux échanges avec Maxim Morin, Angel Arias et les participants d'un séminaire fermé portant sur la nature de la mesure, à l'Université de Montréal. Je souhaite aussi remercier Christelle Lison pour l'évaluation d'une version précédente de l'article. Enfin, il va de soi que les potentielles erreurs d'interprétation et de traduction sont l'entière responsabilité de l'auteur.

² "In science, when human behaviour enters the equation, things go nonlinear. That's why Physics is easy and Sociology is hard."

Plusieurs modélisations ont été développées pour mesurer les données d'enquêtes à grande échelle telles que celles du test PISA. Traditionnellement, ces modèles (par exemple, l'analyse en composantes principales et l'analyse factorielle) ont surtout été élaborés pour évaluer les qualités métriques de tests portant sur des connaissances unidimensionnelles. Dès lors, pour s'adapter aux exigences des nouveaux curricula, de nombreuses modélisations éducatives issues de la théorie de la réponse aux items (Bertrand & Blais, 2004; Hambleton & Swaminathan, 1986) ont été développées pour procéder à la mesure des compétences des élèves. Par exemple, le modèle mixed-coefficient multinomial logit model (Adams & Wu, 2007) a été utilisé pour analyser les données du test PISA 2012. Cette approche généralise le modèle développé par Georg Rasch (1960) qui permet de calculer la probabilité $P_i(\theta)$ qu'un élève obtienne une bonne réponse à un item en se basant sur le niveau de difficulté de l'item. Dans le cadre de cet article, x_i représente la réponse de l'élève aux items $i(i=1, \dots, I)$ et le modèle de Rasch à réponse dichotomique s'écrit :

$$P_i(\theta) = P(x_i = 1 | \theta, b_i) = \frac{\exp[(\theta - b_i)]}{1 + \exp[(\theta - b_i)]} \quad (1)$$

où θ est un paramètre d'habileté de l'élève et b_i un paramètre de difficulté de l'item. Tel que présenté à la figure 1, ce modèle présente l'avantage de mettre sur une même unité de mesure, appelée logit, le niveau d'habileté de l'élève et la difficulté de l'item :

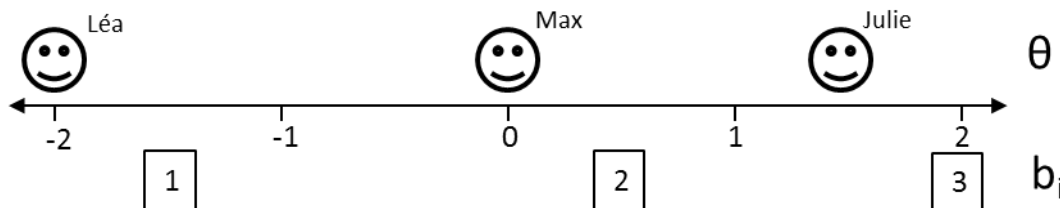


Figure 1 : Exemple théorique d'échelle logit à l'aide du modèle de Rasch

Dans ce cas-ci, les valeurs positives et élevées (la portion de droite du continuum) démontrent que l'habileté de l'élève θ et la difficulté de l'item b_i sont élevées. À l'opposé, les valeurs négatives et basses (la portion de gauche du continuum) signifient que l'habileté θ et la difficulté des items b_i sont faibles. Ainsi, nous observons que Léa a une habileté de $\theta = -2$, ce qui est inférieure à l'habileté de Max ($\theta = 0$) et à celle de Julie ($\theta = 1,5$). Puisque les trois items (représentés par les carrés) sont aussi sur une échelle logit, ils peuvent être interprétés directement avec l'habileté des trois élèves. Par exemple, Max ($\theta = 0$) a une forte probabilité d'obtenir une bonne réponse à l'item 1 ($b_i = -1,5$), mais il a une plus faible probabilité d'obtenir une bonne réponse à l'item 3 ($b_i = 2$). De son côté, Julie ($\theta = 1,5$) a une très forte probabilité d'obtenir une bonne réponse à l'item 1 ($b_i = -1,5$) et à l'item 2 ($b_i = 0,5$), mais une plus faible probabilité d'obtenir une bonne réponse à l'item 3 ($b_i = 2$).

Si le modèle éducatif présenté à la figure 1 paraît intéressant pour mesurer des compétences investiguées dans une enquête à grande échelle, il n'est pas exempt de critiques. Par exemple, Sijsma (2011) considère que l'échelle logit est différente des unités de mesure physiques telles que le mètre et le Kelvin, car « il n'y a pas de théorie spécifique qui définit les unités pour différents attributs³ [par exemple, une compétence]⁴ » (traduction libre, p. 1218). Cette idée a récemment été reprise par Bringmann & Eronen (2015) : « à notre avis, les principales différences entre les attributs physiques et les attributs psychologiques [cela est évidemment applicable au cas d'une compétence] sont que (a) les attributs psychologiques ne sont pas basés sur une théorie largement acceptée (...) et (b) il n'y a aucune base théorique solide à l'égard des unités, des ratios et des échelles de mesure concernant des attributs

³ Puisque cet article est présenté dans une revue généraliste en éducation, nous ne poserons pas de distinction entre les concepts suivants : attribut, construit et trait latent. Par contre, les psychométriciens posent une distinction entre ces trois concepts. Le lecteur intéressé pourra aller lire Markus et Borsboom (2013, p. 2-3).

⁴ «There is no underlying theory that defines the units specifically for different attributes.»

psychologiques⁵ » (traduction libre, p. 12). Ce constat est important pour nous puisqu'il met en exergue les limites de l'acte de mesurer une compétence en contexte éducatif.

Cet article se veut une réflexion quant à la possibilité de mesurer une compétence dans le cadre des enquêtes internationales. Ainsi, nous discutons, dans un premier temps, du caractère multiforme du concept de compétence. Dans un second temps, nous revenons sur la nature non observable de la compétence avant de présenter une proposition de réponse à la délicate et complexe question: peut-on mesurer une compétence? La conclusion favorisera trois avenues de recherche à adopter afin de nourrir cette réflexion.

Qu'est-ce qu'une compétence? Quelques exemples

Notons d'ores et déjà que nous ne visons pas à faire une recension exhaustive des différentes définitions du concept de compétence, mais à mettre en exergue le caractère multiforme de celle-ci à l'aide de quelques exemples pertinents (le lecteur intéressé à en savoir plus sur le concept de compétence pourra aller lire l'excellent article de Ayotte-Beaudet [2013]). Dans cette optique, voici quelques définitions retenues par des organismes ou des ministères dont la mission touche le monde de l'éducation. D'abord, le *Programme de formation de l'école québécoise* (2001) définit une compétence comme « un savoir-agir fondé sur la mobilisation et l'utilisation efficaces d'un ensemble de ressources » (p. 4). Du côté de la Belgique francophone, l'article 5 du *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre* (1997) définit une compétence comme une « aptitude à mettre en œuvre un ensemble organisé de savoirs, de savoir-faire et d'attitudes permettant d'accomplir un certain nombre de tâches » (p. 2). Enfin, l'Organisation de Coopération et de Développement Économiques (2005) présente une définition plus englobante en spécifiant que « le concept de compétence ne renvoie pas uniquement aux savoirs et savoir-faire, il implique aussi la capacité à répondre à des exigences complexes et à pouvoir mobiliser et exploiter des ressources psychosociales (dont des savoir-faire et des attitudes) dans un contexte particulier » (p. 6).

Observons aussi les définitions proposées par certains chercheurs. Selon Koeppen, Hartig, Klieme & Leutner (2008), les compétences sont conceptualisées comme des habiletés complexes qui s'insèrent dans un contexte spécifique et qui sont proches de situations que l'on retrouve dans la vie réelle. Cette définition est cohérente avec celle de Jonnaert & Masciotra, (2007) qui déclarent qu'une compétence est « le résultat du processus par lequel une personne mobilise et articule entre elles un ensemble de ressources pertinentes et traite avec efficacité une situation problème dans laquelle elle se trouve » (p. 65).

Roegiers (2000), de son côté, considère que « la compétence est la possibilité, pour un individu, de mobiliser de manière intériorisée un ensemble intégré de ressources en vue de résoudre une famille de situations-problèmes » (p. 66). Del Rey (2010) précise qu'une « compétence ne désigne pas un savoir, mais un savoir-agir qui permettrait l'assimilation de ce qu'on apprend et son utilisation efficace dans la vie quotidienne » (p.91). De son côté, Tardif (2006) déclare qu'une compétence est davantage de l'ordre du savoir-agir que du savoir-faire. Enfin, d'autres auteurs précisent qu'il existe plusieurs types de compétences. À titre d'exemple, Wolfs (1988) présente la compétence produit (qui est terminale) et la compétence processus (qui est liée aux stratégies d'apprentissage).

Quelques consensus autour du concept de compétence

Selon Legendre et Morrisette (2014), une compétence ne peut être comprise que dans le contexte où elle doit se manifester. Il y a donc techniquement autant de compétences qu'il y a de contextes différents où celles-ci peuvent se manifester. Évidemment, ce relativisme n'est pas une fatalité. Si Scallon (2004) et Blömeke, Gustafsson & Shavelson (2015) démontraient l'existence de toute une variété de définitions pour décrire une compétence, il est aujourd'hui possible de faire émerger quelques grands consensus autour de ce concept. Par exemple, nous savons qu'une compétence éducative devrait être liée au monde réel et, donc, authentique. Ensuite, une compétence est un construit multidimensionnel qui mobilise souvent plusieurs composantes cognitives inter-relées. De plus, la compétence d'un élève évolue dans le temps. Finalement, mentionnons que la compétence est un savoir-agir où l'élève sait combiner et mobiliser des ressources pertinentes.

⁵ "In our view, the main differences between physical attributes and psychological attributes are that (a) most psychological attributes have not been embedded into any successful and widely accepted theory (...), and relatedly, (b) there is no solid theoretical foundation for the units, ratios, and scales for psychological attributes."

La compétence comme trait latent

En plus de son caractère polysémique, Legendre (2008) soulève un élément essentiel de la nature d'une compétence : son caractère inobservable. Borsboom (2008) rapporte d'ailleurs cette idée de façon fort amusante: « l'intelligence générale, l'habileté spatiale, les attitudes et l'auto-efficacité ne sont pas dans la tête des gens. Lorsque nous ouvrons la tête d'une personne, il y a une sorte de gelée grise, pas des variables psychologiques comme l'intelligence générale⁶ » (traduction libre, p. 42). Ainsi, ce sont les manifestations de la compétence qui sont mesurées. Par exemple, une compétence en mathématiques doit être inférée par l'intermédiaire des réponses des élèves à un examen.

Parce qu'elles sont inobservables, les compétences sont considérées comme des traits latents dans les modèles éducatifs contemporains (les deux termes seront donc utilisés de façon interchangeable dans le reste du texte). Cette idée est illustrée à la figure 2. Le trait latent (noté par le symbole θ , qui était aussi le paramètre d'habileté dans le modèle de Rasch) est représenté par un cercle, et est inféré à partir des réponses observées à trois items, ici x_1 à x_3 , représentés par des petits carrés.

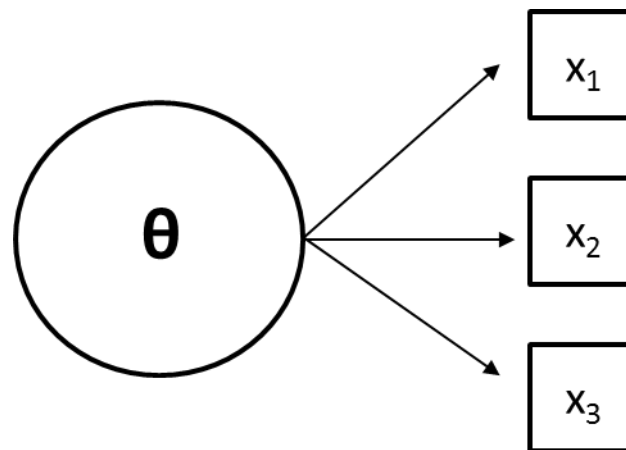


Figure 2 : Réponses observées et trait latent

Le trait latent existe-t-il?

Devons-nous croire que le trait latent existe même s'il est invisible? Selon Borsboom, Mellenbergh & van Heerden (2003), il est possible d'étudier la nature du trait latent à partir d'un vieux débat en philosophie des sciences entre les réalistes et les constructivistes (aussi appelés antiréalistes). Bien que cet échange d'idées ait principalement focalisé sur la théorie dans les « hard systems » tels que la physique et la chimie, il comporte des éléments qui sont pertinents à la réflexion de la mesure d'une compétence, qui est considérée comme un trait latent dans cet article.

D'une part, les réalistes croient que la réalité est indépendante de la conscience humaine. Ces derniers stipulent d'ailleurs que des entités inobservables qui sont expliquées par une bonne théorie peuvent exister dans la réalité. Hood (2013) parle alors du principe du « réalisme épistémique minimal » qui stipule qu'il est, en principe, possible de justifier l'hypothèse de l'existence d'entités inobservables. Un exemple classique en physique est l'électron ou le trou noir, ceux-ci n'ayant jamais été observés à l'œil nu. En ce qui concerne le domaine de l'éducation, une approche réaliste consisterait à supposer qu'un trait latent (par exemple, une compétence en mathématiques) existe dans la mesure où nous avons accumulé des preuves empiriques suffisantes à son sujet.

D'autre part, l'approche constructiviste (ou antiréaliste) postule l'existence « d'une différence entre la connaissance théorique et la connaissance empirique » (Hawking & Mlodinow, 2011, p. 56). Ainsi, il n'y aurait que les variables observées qui existeraient. Une variable qui ne serait pas observée aurait donc de

⁶ "General intelligence, spatial ability, attitudes, and self-efficacy are not in people's heads. When we open up a person's head, we find a sort of gray jelly, not psychological variables like general intelligence."

fortes probabilités d'être une construction de l'esprit et rien ne nous obligerait à supposer son existence. Selon van Fraassen (1980), « la science vise à nous donner des théories qui sont empiriquement adéquates⁷ » (p.12) et « accepter une théorie est (pour nous) de croire qu'elle est empiriquement adéquate - que ce que la théorie dit à propos de ce qui est observable (par nous) est vrai⁸ » (traduction libre, p.18). La question, pour les constructivistes, est alors de savoir quand il est possible de considérer qu'une observation n'est plus directe. Comme le souligne Maxwell (1962) : lorsque l'on regarde à travers une fenêtre? Lorsque l'on regarde à travers des lunettes avec des verres épais? Lorsque l'on regarde avec des jumelles de chasse? Lorsque l'on regarde avec un microscope? La distinction devient, à l'évidence, difficile à faire. Et si, comme le précise Van Fraassen (1980), la question de l'observation est entourée d'un flou important, le défi pour les phénomènes non-observés prend encore plus d'ampleur.

Revenons à notre question de base: le trait latent existe-t-il? La réponse à cette question n'est pas si simple. Selon Hood (2008, 2013), on doit présupposer l'existence d'un trait latent avant de prétendre pouvoir le mesurer. De son côté, Rossiter (2011, p. 13) mentionne : « un construit est une définition et cette définition peut être jugée comme raisonnable ou déraisonnable, mais pas comme étant vrai ou fausse⁹ ». Malheureusement, il n'existe pas de preuve absolue de l'existence d'un trait latent : on doit donc faire comme si celui-ci existait. Cette posture est foncièrement réaliste : « estimer l'habileté d'un élève nécessite à la fois le réalisme méthodologique et une reconnaissance de l'existence des traits latents¹⁰ » (traduction libre, Hood, 2013, p. 755).

Si le trait latent existe, peut-il être mesuré? Quelques défis importants...

Dans son sens le plus général, Mari (2013) souligne que la mesure « est une activité qui comporte de nombreuses facettes et cette multiplicité est irréductible¹¹ » (traduction libre, p. 2889). Néanmoins, plusieurs auteurs opposent une vision traditionnelle de la mesure, qui est plus près de la physique, à une vision qui se retrouve dans les « soft systems » (Finkelstein, 2005), soit des disciplines comme la psychologie et les sciences de l'éducation. Ainsi, Luce (2005) oppose la mesure comportementale (*behavioral*) à la mesure physique (*physical*), alors que Michell (1999) pose une distinction formelle entre la mesure en science naturelle (qui est plus traditionnelle) et la mesure en psychologie. De son côté, Boumans (2013) considère qu'il y a encore beaucoup de travail à faire afin de bien comprendre la nature des modèles de mesure utilisés à l'extérieur du laboratoire (par exemple, en sciences sociales et, nous pouvons le présumer, en éducation).

Sachant que la mesure en éducation se distingue de celle des domaines scientifiques traditionnels, nous allons maintenant discuter d'un minimum de cinq défis à relever pour améliorer notre compréhension de la mesure de traits latents en contexte d'évaluation à grande échelle :

1. mieux comprendre la nature de ce qui doit être mesuré;
2. mieux interpréter le score de l'outil de mesure utilisé;
3. maîtriser plus finement l'erreur de mesure engendrée par le trait latent;
4. trouver une alternative à la définition de la mesure de Stevens (1946), qui est présentement la plus utilisée en éducation; et
5. accentuer l'effort afin de mieux comprendre la nature de la mesure dans le domaine spécifique de l'éducation.

Un premier défi consiste à bien comprendre une compétence avant de formellement la mesurer. Pour paraphraser René Thom : « si l'on n'a pas le concept d'un objet, on ne le reconnaîtra pas » (traduction libre, cité par Perdijon, 2012, p. 77). Ainsi, il est nécessaire de bien prendre la température avant de développer un outil pour tenter de la mesurer (voir l'exemple du thermomètre dans Bringmann & Eronen [2015]). Soulignons que la conceptualisation en vue d'opérer l'acte de mesurer est un grand défi en psychologie et en éducation. Prenons l'exemple de l'un des concepts les plus étudiés dans notre domaine : le quotient intellectuel (Q.I.). En 2005, Finkelstein déclarait qu'il n'existe toujours pas de définition satisfaisante du

⁷ "Science aims to give us theories that are empirically adequate."

⁸ "[...] to accept a theory is (for us) to believe that it is empirically adequate—that what the theory says about what is observable (by us) is true."

⁹ "A 'construct' is a definition [and a] definition can be judged as reasonable or unreasonable but not as true or false."

¹⁰ "Estimating ability requires both methodological realism and a commitment to the existence of latent traits."

¹¹ "Is a many-faced activity, and this multiplicity is some-how irreducible."

concept multidimensionnel d'intelligence. Ainsi, la simple utilisation d'un outil de mesure ne nous semble pas suffisante pour croire que nous mesurons bien le Q.I.

Le deuxième défi est lié au problème découlant de l'interprétation du score émanant de l'outil de mesure. Observons la définition de la mesure selon Sydenham (traduction libre, 1973¹²) : « la détermination quantitative d'une grandeur physique comparée à une norme ou obtenue à l'aide d'un appareil calibré. Le résultat d'une mesure est donc une valeur numérique qui exprime le ratio entre une grandeur étudiée et une grandeur standardisée considérée comme une unité.¹³ » Le concept d'unité de mesure est relativement aisé à comprendre. Dans le cadre d'une analyse selon le modèle de Rasch, nous avons déjà dit que l'unité de mesure est le logit. Dans ce cas-ci, on peut facilement poser des comparaisons qualitatives entre les élèves mesurés. Par exemple, à la figure 1, nous observons que Léa a une habileté égale à $\theta = -2$ qui est inférieur à celle de Max ($\theta = 0$) et à celle de Julie ($\theta = 1,5$). Par contre, les choses deviennent plus compliquées lorsque nous souhaitons interpréter le ratio de l'unité de mesure. Il est pertinent de reprendre l'exemple du Q.I. pour illustrer cette idée: « les échelles de tests d'intelligence sont ordinales, mais le fait d'utiliser un nombre peut donner l'apparence qu'elles sont sur une échelle de ratio. Or, rien ne permet d'affirmer qu'une personne ayant obtenu un score de 120 à un test d'intelligence est 20% plus intelligente qu'un individu ayant obtenu un score de 100¹⁴ » (traduction libre, Finkelstein, 2005, p. 271). Force est de constater que l'interprétation du ratio lors d'une prise de mesure éducatrice est différente de celle en physique.

Voici un troisième grand défi: puisque le trait latent est une entité qui n'est pas directement observée, il peut y avoir de l'erreur au moment de procéder à l'action de mesurer. Les modèles éducatifs postulent généralement que les erreurs de mesure sont aléatoires. Ainsi, lorsqu'un chercheur tente de faire des inférences sur le score d'un élève au test PISA, ces inférences sont, dans le meilleur des cas, probables plutôt que certaines. Survient alors une question importante : l'hypothèse des erreurs de mesure aléatoires est-elle plausible dans le cadre des enquêtes à grande échelle? Pas systématiquement si l'on réfère aux nombreux écrits portant sur le problème du « person-fit » (Meijer & Sijtsma, 2001) et qui s'intéressent, entre autres, à la détection de la réponse au hasard, de la tricherie ou de l'inattention. Ainsi, il faudra probablement se pencher plus sérieusement sur la modélisation de ces comportements afin de raffiner la compréhension de l'erreur de mesure, qui n'est pas uniquement aléatoire en éducation.

Le quatrième défi est lié au fait que nous référons généralement à la définition de Stevens (1946) pour définir la mesure en sciences de l'éducation : « mesurer est le fait d'assigner des nombres à des objets ou à des événements en fonction d'une règle¹⁵ » (traduction libre, cité par Michell (2004), p. 11). Or, cette définition a été grandement critiquée au cours des cinquante dernières années. Par exemple, Michell (1999) souligne que, selon la vision traditionnelle de la mesure, les nombres ne sont pas assignés, mais découverts. De plus, cette définition est incomplète puisqu'elle ne soulève pas l'importante question des ratios propres aux mesures physiques (voir, peu plus haut, la définition de Sydenham [1973]). Stevens a explicitement une vision pragmatique de la mesure. Selon lui, il suffirait d'utiliser un outil de mesure pour prétendre mesurer quelque chose. Or, nous croyons que la mesure de compétences devrait aussi être le fruit d'un grand effort de conceptualisation, un élément absent de la vision de cet auteur. Pour toutes ces raisons, nous croyons que la conception de la mesure de Stevens est inappropriée et qu'elle devra être remplacée dans les cours de méthodologie en sciences de l'éducation.

Le cinquième défi est lié au précédent : nous savons que nous devons remplacer la définition de Stevens (1946), mais par quelle définition? À notre humble avis, il n'y a pas encore de proposition satisfaisante pour le cas spécifique de l'éducation. Il est, d'abord, important de définir plus adéquatement la mesure dans un contexte qui n'est pas celui des sciences traditionnelles, soit, celle qui est effectuée hors du laboratoire. Par exemple, la conception de la mesure de Michell (1999), qui est davantage traditionnelle, porte essentiellement sur le formalisme mathématique qui sous-tend une opération de mesure, mais elle est très peu utile pour caractériser et définir un construit équivalent à celui d'une compétence. Par chance,

¹² Note de l'auteur : au moment de la publication de l'article, il était impossible de retracer le numéro de page de cet extrait.

¹³ "The quantitative determination of a physical magnitude by comparison with a fixed magnitude adopted as the standard, or by means of a calibrated instrument. The result of measurement is thus a numerical value expressing the ratio between the magnitude under examination and a standard magnitude regarded as a unit."

¹⁴ "The scales of the intelligence tests are ordinal only, though the number may give the appearance that they are on a ratio scale. There is no basis of stating that an individual with a score of 120 in an intelligence test is 20% more intelligent than an individual with a score of 100."

¹⁵ "Measurement is the process of assignment of numerals to object or events according to rule."

d'autres auteurs ont tenté de dépasser la conception de la mesure au sens traditionnel du terme. Par exemple, Mari (2013) a étudié toute une panoplie de définitions du concept de mesure. Cet exercice important a démontré que la mesure peut porter autant sur le processus que sur le produit; ou encore que sur la comparaison à une unité de mesure ou non. Il intègre aussi l'importance de l'objectivité du mesureur face à l'activité de mesure et le caractère intersubjectif de la mesure, à savoir que différentes personnes doivent interpréter de la même façon l'information tirée de l'acte de mesurer. Dans la poursuite de sa réflexion, Mari considère d'ailleurs qu'il n'est pas approprié de discuter de *la* mesure, mais plutôt *des* mesures.

Selon nous, les travaux les plus intéressants pour définir la mesure en éducation viennent de la collaboration de ce même Mari avec des chercheurs de l'Université de Berkeley, en Californie. D'abord, Wilson, Mari, Maul & Irribarra (2015) ont démontré qu'il existe plusieurs similitudes entre la mesure dans le domaine de la physique et celui des sciences sociales et, nous pouvons le présumer, de l'éducation. Ensuite, Mari, Maul, Irribarra & Wilson (2013) déclarent que la mesure doit être basée sur la structure du processus plutôt que sur le produit (par exemple, en déclarant que la propriété est une quantité). Ce faisant, ils mettent l'accent sur quelque chose de fort important: la quantification n'est pas nécessaire pour procéder à l'acte de mesurer, en éducation.

Revenons à la question que nous posons dans le titre de cette section: si le trait latent existe, peut-il être mesuré? Force est de constater que mesurer une compétence est une opération complexe comportant de nombreux défis. Néanmoins, il nous semble pertinent de stipuler que le trait latent peut être mesuré conditionnellement à une compréhension adéquate d'une compétence et à l'adoption d'une posture épistémologique réaliste.

Conclusion

Cette réflexion focalise sur la possibilité de mesurer une compétence dans le cadre des épreuves d'évaluation à grande échelle. Certains auteurs tels que Scallon (2004) croient qu'une compétence doit être évaluée plutôt que mesurée. Cette idée est aussi rapportée par Newton & Shaw (2014): « certains diront que les attributs en éducation [par exemple, une compétence] ne sont tout simplement pas le genre de choses qui peuvent être mesurées. Ils peuvent être évalués, parce que l'idée de l'évaluation est moins contraignante que celle de la mesure, de sorte que rien n'est perdu; mais l'idée de la mesure réfère à une structure que les attributs en éducation (...) ne peuvent pas posséder¹⁶ » (traduction libre, p. xxi). De leur côté, les promoteurs du test PISA font l'hypothèse que la compétence peut être mesurée. Ils utilisent d'ailleurs le logit du modèle mixed-coefficient multinomial logit model (qui est une généralisation du modèle de Rasch) comme unité de mesure des compétences en lecture, en mathématique ou en sciences.

Pour notre part, nous avons déjà soulevé qu'il est possible de mesurer une compétence si nous adoptons une posture réaliste. Par contre, il ne faut pas perdre de vue les nombreux défis à relever lors de la mesure de celle-ci. Nous discutons d'ailleurs, dans la section 3.2, de cinq défis à relever afin d'améliorer notre compréhension de la mesure de traits latents en contexte d'évaluation à grande échelle.

En nous inspirant de cela, au moins trois avenues doivent être envisagées par les chercheurs pour développer la mesure d'une compétence en contexte d'évaluation à grande échelle. Premièrement, nous savons que le concept de compétence existe sous de multiples formes. Par contre, procéder à l'acte de mesurer sous-entend de savoir précisément quel est l'objet à mesurer. Actuellement, pareil éclairage n'est pas formellement disponible sur les compétences à mesurer dans le PISA. Néanmoins, le fait que les compétences mesurées ne soient pas totalement définies n'est pas sans intérêt. Par exemple, au sujet du Q.I., Nisbett, Aronson, Blair, Dickens, Flynn, Halpern & Turkheimer (2012) mentionnent la chose suivante:

la mesure de l'intelligence est l'une des plus grandes réalisations de la psychologie, mais aussi l'une des plus controversées. Les critiques se plaignent qu'un test unique ne peut pas saisir la complexité de l'intelligence humaine, que toute mesure est imparfaite, qu'aucune mesure n'est complètement libre de préjugés culturels en plus de la potentielle utilisation inappropriée des scores aux tests d'intelligence. Il y a un certain mérite à toutes ces critiques. Par contre, cette mesure a une valeur utilitaire, car elle est un assez bon prédicteur de résultat

¹⁶ "Some would say that the attributes of educational (...) simply are not the kind of things that can be measured. They might still be assessed, because the idea of assessment is far looser than the concept of measurement, so all is not lost; but the idea of measurement carries connotations of structure that the attributes of education (...) may not possess."

scolaires, de la performance au travail et de nombreux autres aspects de la réussite dans la vie¹⁷ (traduction libre, p. 131)

Les résultats des épreuves d'évaluation à grande échelle nous permettent aussi d'obtenir des informations importantes sur nos systèmes éducatifs (ce qui est mieux qu'une absence d'information). Par ailleurs, nous considérons les initiatives de chercheurs qui tentent de clarifier des concepts-clés en éducation réellement pertinentes, comme celle de Messier (2014), qui cherche à mieux comprendre les concepts apparentés aux méthodes (par exemple, stratégie pédagogique, méthode pédagogique, modèle pédagogique, etc.). Nous croyons que ce type de travail serait adéquat par rapport au concept de compétence.

Deuxièmement, Legendre & Morrisette (2014) ont mentionné qu'une compétence ne peut être comprise que dans le contexte où elle doit se manifester. Pour le dire autrement, la façon de mesurer une compétence doit tenir compte de sa spécificité. Or, les évaluateurs ont utilisé le mixed-coefficient multinomial logit model pour évaluer chacune des trois compétences au test PISA 2012. Nous retenons donc qu'il serait important de développer des modélisations applicables à des compétences particulières (par exemple, les mathématiques) afin de procéder à des mesures plus précises de celles-ci.

Troisièmement, nous considérons qu'il serait pertinent de mener une réflexion spécifique sur l'acte de mesurer en éducation. Comme nous l'avons déjà souligné, la définition de Stevens (1946), qui est largement utilisée en éducation, est incomplète et limitée. En ce sens, il nous paraît intéressant de poursuivre le travail entamé par Mari, Maul, Iribarra & Wilson (2013) ainsi que par Wilson, Mari, Maul & Iribarra (2015) afin de mieux comprendre la nature de la mesure en éducation.

En guise de commentaire final, nous savons que la communauté scientifique en éducation s'intéresse à la mesure depuis un peu plus d'une centaine d'années. Il est aussi pertinent de savoir que la mesure de la longueur physique (par exemple, le mètre) s'est développée durant plusieurs milliers d'années. Or, où en est la mesure en éducation d'un point de vue historique? Voici la réponse de Bringmann & Eronen (2015): « à bien des égards, la situation dans la pratique [de la mesure] psychologique [cela s'applique aussi à l'éducation] ressemble à la situation de la mesure de la température à la fin du 18^e et au début du 19^e siècle¹⁸ » (traduction libre, p. 6). Il ne faut donc pas prêcher par excès de pessimisme en ce qui concerne la mesure de compétences : il n'y a aucun doute que de nouvelles approches efficaces vont émerger afin d'améliorer la précision de la mesure dans le contexte des enquêtes à grande échelle. Patience!

¹⁷ "The measurement of intelligence is one of psychology's greatest achievements and one of its most controversial. Critics complain that no single test can capture the complexity of human intelligence, all measurement is imperfect, no single measure is completely free from cultural bias, and there is the potential for misuse of scores on tests of intelligence. There is some merit to all these criticisms. But we would counter that the measurement of intelligence— which has been done primarily by IQ tests— has utilitarian value because it is a reasonably good predictor of grades at school, performance at work, and many other aspects of success in life."

¹⁸ "In many ways, the situation in psychological practice resembles the situation of temperature measurement in the late 18th and early 19th century."

RÉFÉRENCES

- Adams, R. L., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model : a generalised form of the Rasch model. In M. von Davier & C.H. Carstensen (Eds.). *Multivariate and mixture distribution Rasch models*, (1ère éd., pp. 57-75). Berlin, Allemagne : Springer.
- Ayotte-Beaudet, J.-P. (2013). The concept of competence in the French-language education literature. *Prospects*, 43, 419-427.
- Bertrand, R., & Blais, J.-G. (2004). *Modèle de mesure: l'apport de la Théorie de la réponse aux items*. Sainte-Foy, Québec : Presses de l'Université du Québec.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3-13.
- Borsboom, D. (2008). Latent variable theory. *Measurement*, 6, 25-53.
- Borsbooms, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological review*, 110, 203-219.
- Boumans, M. (2013). The role of models in measurement outside the laboratory. *Measurement*, 46, 2908-2912.
- Bringmann, L. F., & Eronen, M. I. (2015). Heating up the measurement debate : what psychologists can learn from the history of physics. *Theory & psychology*, 25, 1-17.
- Communauté française de Belgique. (1997). Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre. En ligne : http://www.gallilex.cfwb.be/document/pdf/21557_018.pdf
- Cronbach, L. J. (1971). *Essentials of psychological testing* (3e éd). New York, N.Y.: Harper & Row.
- Del Rey, A. (2010). *À l'école des compétences: de l'éducation à la fabrique de l'élève performant*. Paris, France: La Découverte.
- Finkelstein, L. (2005). Problems of measurement in soft systems. *Measurement*, 38, 267-274.
- Gouvernement du Québec (2001). Programme de formation de l'école Québécoise. Récupéré le 21 janvier 2015. En ligne: <http://www1.mels.gouv.qc.ca/sections/programmeFormation/pdf/prform2001nb.pdf>
- Hambleton, R. K., & Swaminathan, H. (1985). *Fundamentals of Item response theory*. Newbury Park, Cal.: Sage.
- Hartig, E., Klieme, E., & Leutner, D. (Eds.) (2008) *Assessment of competencies in educational contexts: State of the art and future prospects*. Göttingen : Hogrefe & Huber.
- Hawking, S., & Mlodinow, L. (2011). *Y a-t-il un grand architecte dans l'Univers?* Paris, France : Odile Jacob.
- Hood, S. B. (2007). *Latent variable realism in psychometrics*. Thèse de doctorat inédite, Indiana University, Indianapolis.
- Hood, S. B. (2013). Psychological measurement and methodological realism. *Erkenn*, 78, 739-761.
- Jonnaert, P. & Masciotra, D. (2007). Socioconstructivisme et logique de compétences pour les programmes d'études. Un double défi. In L. Lafortune, E. Moussadak & P. Jonnaert (Eds.). *Observer les réformer curriculaires en éducation* (1ère éd., pp. 53-75). St-Nicolas, Québec : Presses de l'Université du Québec.
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie*, 216, 61-73.
- Legendre, M.-F. (2008). La notion de compétence au coeur des réformes curriculaires : effet de mode ou moteur de changement en profondeur? In F. Audigier & N. Tutiaux-Guillon (Eds.), *Compétences et contenus. Les curriculums en questions* (1ère éd., pp. 27-50). Bruxelles, Belgique: De Boeck Université.
- Legendre, M.-F., & Morrissette, J. (2014). Défis et enjeux de l'approche par compétences dans le cadre de la Nouvelle Gestion Publique. In J. Morrissette et M.-F. Legendre (Eds.), *Enseigner et évaluer : regards sur les enjeux éthiques et sociopolitiques* (1ère éd., pp. 211-245). Ste-Foy, Québec : Presses de l'Université Laval.
- Luce, D. (2005). Measurement analogies : comparisons of behavioral and physical measures. *Psychometrika*, 70, 227-251.

- Marcus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory- measurement, causation, and meaning*. New York, New York: Routledge.
- Mari, L. (2013). A quest for the definition of measurement. *Measurement, 46*, 2889-2895.
- Mari, L., Maul, A., Torres Irribarra, D., & Wilson, M. (2013) Quantification is neither necessary nor sufficient for measurement, *Journal of Physics: Conference Series, 459*, 012007
- Maxwell, G. (2009 [1962]). The ontological status of theoretical entities. In T. McGrew, M. Alspector-Kelly, and F. Allhoff (Eds.). *Philosophy of science: an historical anthology* (1ère ed., pp. 451-458). Oxford, U.K.: Blackwell Publishing.
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*, 107-135.
- Messier, G. (2014). *Proposition d'un réseau conceptuel initial qui précise et illustre la nature, la structure ainsi que la dynamique des concepts apparentés au terme méthode en pédagogie*. Thèse de doctorat inédite, Montréal : U.Q.A.M.
- Michell, J. (1999). *Measurement in psychology*. Cambridge, U.K.: Cambridge University Press.
- Newton, P.E., & Shaw, S. D. (2014). *Validity in Educational and Psychological Assessment*. London, U.K.: SAGE.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New Findings and Theoretical Developments. *American Psychologist, 67*, 130-159.
- Organisation de coopération et de développement économiques (2005). *La définition et la sélection des compétences clés- résumé*. En ligne : <http://www.oecd.org/pisa/35693273.pdf>
- Okasha, S. (2002). *Philosophy of science. A very short introduction*. Oxford, U.K.: Oxford press.
- Perdijon, J. (2012). *La mesure- Histoire, science et technique*. Paris, France: Vuibert
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago, Ill. : University of Chicago Press.
- Roegiers, X. (2000). *Une pédagogie de l'intégration: compétences et intégration des acquis dans l'enseignement*. Bruxelles, Belgique: De Boeck Université.
- Rossiter, J.R. (2011). *Measurement for the Social Sciences. The C-OAR-SE Method and Why it must Replace Psychometrics*. Berlin, Allemagne: Springer.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement, 6*, 219-262.
- Rupp, A.A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: theory, methods, and applications*. New York, N.Y.: The Guilford Press.
- Scallon, Gérard. (2004). *L'évaluation des apprentissages dans une approche par compétence*. St-Laurent, Québec : Éditions du Renouveau pédagogique.
- Sijtsma, K. (2011). Introduction to the measurement of psychological attributes. *Measurement, 44*, 1209-1219.
- Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science, 103*, 677-680.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal, Québec : Chenelière Education.
- van Fraassen, B. C. (1980). *The Scientific Image*. Oxford, U.K. : Clarendon Press.
- Wilson, M., Mari, L., Maul, A., & Irribarra, D. T. (2015). A comparison of measurement concepts across physical science and social science domains: instrument design, calibration, and measurement. *Journal of Physics: Conference Series, 588*, conference 1.
- Wolfs (1998). *Méthodes de travail et stratégies d'apprentissage: du secondaire à l'université*. Bruxelles : De Boeck & Larcier.