

Office de la qualité et
de la responsabilité
en éducation



Analyse documentaire des données empiriques sur le lien entre l'évaluation obligatoire des compétences des enseignantes et enseignants et les résultats des élèves

Août 2019

Table des matières

| | |
|--|----|
| Sommaire..... | 3 |
| Aperçu des évaluations des compétences des enseignants passées et actuelles..... | 5 |
| Principaux objectifs de l'évaluation des compétences des enseignants..... | 5 |
| Les tests normalisés comme méthode pour reconnaître des enseignants de grande qualité | 6 |
| Les tests normalisés comme méthode pour reconnaître les enseignants de moindre qualité..... | 7 |
| Mise en place de tests normalisés pour des matières spécifiques | 7 |
| Calendrier de l'évaluation des enseignants | 7 |
| Mesure de la compétence dans les tests normalisés | 8 |
| Options pour les candidats qui ne réussissent pas les tests normalisés | 8 |
| Conception et mise en oeuvre des tests de compétences | 8 |
| Formats de tests actuels et items inclus | 9 |
| Évaluation des compétences des enseignants et résultats des élèves | 9 |
| Difficulté de mesurer la relation entre la compétence des enseignants et les résultats des élèves..... | 10 |
| Lien entre l'évaluation des compétences des enseignants et les résultats des élèves | 10 |
| Effets associés à l'évaluation des enseignants..... | 13 |
| L'influence des tests normalisés sur la diversité et l'inclusion | 14 |
| Caractéristiques des enseignants et procédures d'évaluation permettant de prévoir les résultats des élèves..... | 15 |
| Conclusion | 16 |
| Références..... | 18 |

Sommaire

Le présent rapport passe en revue les recherches actuelles et les données disponibles sur le sujet de l'évaluation normalisée obligatoire du personnel enseignant et sur le lien entre ces évaluations et le rendement des élèves. Le présent rapport commence par un bref aperçu de l'historique et des méthodes actuelles d'évaluation des compétences des enseignantes et des enseignants, met en évidence les questions clés et met celles-ci en rapport avec des considérations politiques. Un aperçu des types de tests de compétences des enseignants est présenté, ainsi qu'un résumé de la conception et de l'objectif de ces tests. Le rapport résume ensuite les preuves empiriques de la relation entre l'évaluation normalisée des compétences du personnel enseignant et le rendement des élèves. Le présent résumé se concentre sur les principales sources contenant les conclusions les plus pertinentes pour les responsables des politiques publiques. Cette section comprend un aperçu de certaines des recherches sur les répercussions de l'évaluation normalisée obligatoire des compétences sur des résultats autres que le rendement des élèves. Dans la mesure du possible, le rapport met l'accent sur l'évaluation des enseignantes et enseignants en mathématiques. Le rapport se termine par un bref aperçu de la recherche sur d'autres processus d'examen du rendement et sur des programmes de perfectionnement professionnel qui ont eu un impact sur le rendement.

Il ne fait aucun doute pour les chercheurs que de meilleurs enseignants mènent à de meilleurs résultats pour les élèves (Aaronson, Barrow, Sander, 2007; Coleman, 1968; Rivkin, Hanushek et Kain, 2005; Rockoff, 2004). Cependant, à mesure que nous tentons d'établir une stratégie uniforme pour mesurer le rendement du personnel enseignant et d'établir une corrélation entre des caractéristiques « de réussite » et les résultats des élèves, les données probantes deviennent plus mitigées. Lors de l'examen des données disponibles, il convient de faire une distinction entre la connaissance du contenu, la connaissance du contenu pédagogique et le rendement du personnel enseignant aux tests normalisés pour évaluer avec précision la valeur de chacune de ces variables. D'autres questions sont soulevées au sujet de la validité, de la généralisabilité et de la robustesse de l'utilisation d'un test normalisé pour mesurer la connaissance pure ou la connaissance pédagogique du contenu.

Les données actuellement disponibles montrent que deux types fondamentaux de connaissances d'une enseignante ou d'un enseignant dans une matière influent directement sur les résultats des élèves¹. La première est une bonne connaissance du contenu (c.-à-d. une base et une compréhension du sujet) qui serait familière à quiconque travaille dans le domaine en question. La deuxième est la connaissance du contenu pédagogique, qui exige de comprendre comment enseigner efficacement la matière.

Par exemple, comprendre comment multiplier les fractions est une connaissance du contenu, alors que comprendre comment enseigner la multiplication des fractions est une connaissance pédagogique du contenu (Ball, Thames et Phelps, 2008; Rowland, Huckstep, et Thwaites, 2005). La recherche démontre que ces types de connaissance du contenu sont liés aux résultats des élèves (Carlisle et Berebitsky, 2011; Mullens, Murnane et Willett, 1996). Cependant, la connaissance du contenu pédagogique est plus fortement corrélée aux résultats positifs des élèves (Hill, Kapitula et Umland, 2011; Kelcey et Carlisle, 2013). De plus, la connaissance

¹ Aux fins du présent examen, les résultats des élèves sont mesurés en fonction de leur rendement aux tests normalisés de rendement. Cela comprend à la fois des tests de rendement internationaux comme le Programme international pour le suivi des acquis des élèves (PISA), les Tendances de l'enquête internationale sur les mathématiques et les sciences (TEIMS) et des tests de rendement propres à chaque administration comme le National Assessment of Educational Progress (NAEP) aux États-Unis ou les tests (IOWA).

pédagogique du contenu a un effet médiateur sur la connaissance du contenu pur (Baumert, Kunter, Blum, Brunner, Voss, Jordan, Klusmann, Krauss, Neubrand et Tsai, 2010; Kelcey, Hill et Chin, 2019; Kersting, Givvin, Thompson, Santagata et Stigler, 2012). La connaissance du contenu influe différemment sur les résultats des élèves selon le niveau de connaissance du contenu pédagogique.

Les conclusions qui peuvent être tirées de la recherche explorant la corrélation spécifique entre la connaissance du contenu mathématique et les résultats des élèves (tels que mesurés par des tests normalisés pour les élèves) ne sont pas claires. La majorité des études démontrent une relation positive entre la connaissance du contenu et les résultats des élèves en mathématiques (Baumert et al., 2010; Campbell, Nishio, Smith, Smith, Clark, Conant, Rust, DePiper, Frank, Griffin et Choi, 2014; Carpenter, Fennema, Peterson, Chiang et Loef, 1989; Hill, Rowan et Ball, 2005; Rockoff, Jacob, Kane et Staiger, 2011; Shechtman, Roschelle, Haertel et Knudsen, 2010; Shechtman, N, Roschelle, J., Haertel, G. et Knudsen, J. (2010). Cependant, d'autres études n'ont trouvé aucun lien significatif entre les connaissances mathématiques d'une enseignante ou d'un enseignant et les résultats des élèves (Kersting et al., 2012).

Par contre, la documentation qui examine le lien entre les connaissances mathématiques pédagogiques de l'enseignante ou de l'enseignant et les résultats des élèves est presque uniformément positive (Kelcey et Carlise, 2013; Hill et al., 2011; Kersting et al., 2012). La recherche fondée sur l'observation des enseignants s'appuie sur (1) la capacité de l'enseignante ou de l'enseignant d'élaborer des questions mathématiques efficaces (Ball et coll., 2008), (2) la présentation claire et assurée des sujets (Hill et coll., 2008) et (3) la réponse constructive de l'enseignante ou de l'enseignant aux questions des élèves et aux erreurs visibles dans les travaux en classe (Ball et coll., 2008).

De nombreuses données indiquent que la relation entre la connaissance du contenu, la connaissance du contenu pédagogique et le rendement des élèves est également influencée par d'autres facteurs. Ces facteurs peuvent inclure le matériel pédagogique (Shechtman et al., 2010), les mesures socioéconomiques (Baumert et al., 2010), le parcours scolaire et non scolaire (Baumert et al., 2010), et l'environnement scolaire et du district (Corey, Phelps, Ball, Demonte, et Harrison, 2012; Petscher et Logan 2014; Tighe et Schatschneider, 2016).

L'ensemble des données examinant le lien entre l'évaluation normalisée des compétences de l'enseignante ou de l'enseignant en ce qui a trait à la connaissance du contenu pur et à la connaissance du contenu pédagogique est mitigé. De nombreuses études révèlent que les résultats aux tests de compétences des enseignants sont en corrélation positive avec les résultats des élèves (Goldhaber, 2007; Kelcey et al., 2019; D'Agostino et Powers, 2009). Cependant, presque toutes ces analyses trouvent cette corrélation faible et non universelle. De plus, les recherches utilisant des prédicteurs multiples des résultats des élèves indiquent que les résultats des tests de compétences des enseignants sont le prédicteur le plus faible des résultats des élèves comparativement à (1) l'expérience des enseignants, (2) leurs notes dans les collèges et universités de formation des enseignants, (3) le nombre de cours de mathématiques pris au niveau postsecondaire et (4) les cours de certification professionnelle en mathématiques pour enseignants (Clotfelter, Ladd et Vigdor, 2010; D'Agostino et Powers, 2009; Ferguson et Womack, 1993; Shuls et Trivitt, 2015).

Les tests d'évaluation des enseignants n'ont pas seulement un effet sur les résultats des élèves et peuvent influencer la filière de recrutement des enseignants et le contenu du programme d'études des enseignants en formation. Par conséquent, des changements de ces facteurs

peuvent entraîner une diminution du nombre d'enseignants en formation initiale issus de groupes ethniques minoritaires (Cobb, Shaw, Millard et Bomotti, 1999; Graham, 2013; Nettles, Scatton, Steinberg et Tyler, 2011; Petchauer, 2012; Shuls et Trivitt 2015).

Aperçu des évaluations des compétences des enseignants passées et actuelles

Afin de comprendre pourquoi les responsables des politiques et les membres du milieu de l'éducation continuent d'appuyer et d'entreprendre des tests d'enseignement normalisés, il est essentiel d'examiner l'historique de ce processus. L'utilisation de tests de compétences normalisés n'est pas un concept nouveau (Angrist et Guryan, 2008). Depuis le XIX^e siècle, l'évaluation des enseignants est surtout utilisée aux États-Unis, mais elle existe aussi dans plusieurs autres pays (Angrist et Guryan, 2008). Le recours à l'évaluation des compétences fluctue depuis un siècle en raison de facteurs tels que l'offre et la demande de personnel enseignant, la volonté politique locale et l'opinion publique sur la question (De Forest, 2006; Haney, Madaus, et Kreitzer, 1987; Santibanez, 2006).

À l'heure actuelle, les tests normalisés de compétences des enseignants sont utilisés dans la plupart des États des États-Unis (Goldhaber, 2007), en Australie et dans quelques pays d'Europe et d'Asie, comme le Royaume-Uni et l'Arabie saoudite, bien que ces deux pays envisagent de supprimer cette politique (Gibb, 2019). L'Ontario a rendu obligatoire l'évaluation du personnel enseignant entre 2002 et 2004, mais n'utilise pas actuellement de tels tests (Glassford, 2005; Glassford, 2007).

Dans le contexte de la présente analyse documentaire, il faut faire une distinction prudente entre la certification et la formation des enseignants, et l'évaluation des enseignants. Le test normalisé de compétence peut exister sous forme papier ou numérique et est habituellement à choix multiples. L'objectif principal du test est d'évaluer la compétence de base des participants dans la matière enseignée dans les écoles publiques. Ces tests sont donc conçus pour mesurer les connaissances et les compétences des futurs enseignants et enseignantes et sont utilisés par les établissements d'enseignement pour confirmer que le personnel enseignant est qualifié pour enseigner. Ces tests sont distincts des examens ou des évaluations qui font partie des programmes de formation à l'enseignement des collèges ou des universités (Goodman, Arbona et Dominguez de Rameriz, 2008).

Principaux objectifs de l'évaluation des compétences des enseignants

Bien que le gouvernement américain utilise maintenant des tests normalisés pour les enseignantes et enseignants stagiaires, l'utilisation de ces tests provient de la communauté de l'enseignement supérieur elle-même dans une tentative de mieux évaluer la performance des programmes de formation des enseignants (Pullin, 2017). Au Canada, l'évaluation obligatoire des enseignants a récemment vu le jour dans le cadre d'un souci de transparence et de responsabilité par rapport au public et pour s'assurer que le personnel enseignant est compétent pour enseigner le contenu le plus à jour possible (Glassford, 2007).

Les chercheurs postulent que l'un des objectifs fondamentaux des tests obligatoires de compétences du personnel enseignant est de fournir un indicateur valable de la préparation à l'enseignement des enseignants. L'évaluation de l'efficacité des programmes de formation des enseignants est inhérente à cet objectif. Un système éducatif qui a confiance dans le

fonctionnement d'un réseau de formation des enseignants et, par conséquent, dans les enseignants formés, pourrait ne pas exiger de tels tests normalisés (Goodman et al., 2008).

Les principales fonctions explicites des tests normalisés de compétence sont de reconnaître les enseignantes et enseignants de qualité, d'éliminer celles et ceux qui ne sont pas qualifiés et d'établir une mesure objective permettant d'identifier ces individus (Goldhaber et Hansen, 2010; National Commission on Teaching and America's Future, 1996).

Cependant, dans la littérature, il y a une discussion substantielle sur les motivations politiques implicites de ces tests. « La pertinence et l'utilité des tests PLACE ont moins à voir avec la fonction rationnelle de tri sélectif qu'ils prétendent avoir qu'avec l'histoire d'amour et le confort que le public semble avoir avec les tests pour l'obtention d'un permis, peu importe leur fonctionnalité. » (Cobb et al. 1999, p. 172).

Les tests normalisés comme méthode pour reconnaître des enseignants de grande qualité

Les informations sur l'utilisation efficace des tests normalisés pour trouver du personnel enseignant de grande qualité sont mitigées. Il existe des données qui montrent que les tests normalisés peuvent mesurer avec précision les connaissances sur le contenu pour certains candidats (Hill, Sleep, Lewis et Ball, 2007). Cependant, il y a beaucoup moins de données pour montrer que la réussite à ces tests est directement corrélée au fait que la personne qui les a réussis deviendra une enseignante ou un enseignant de qualité. Darling-Hammond (2000) le résume ainsi : « Les tests mesurent la compétence des étudiants en enseignement en évaluant leur connaissance des meilleures pratiques d'enseignement et leur connaissance des habiletés nécessaires pour être une enseignante ou un enseignant efficace. Malheureusement, les tests ne permettent pas d'évaluer dans quelle mesure les étudiantes et étudiants mettent les pratiques exemplaires en application dans des contextes concrets. » (Cité dans Goodman, Arbona et Dominguez de Rameriz, 2008, p. 24).

Dans l'ensemble, il y a peu de soutien en faveur d'une relation directe entre le rendement des enseignants aux tests normalisés et la qualité éventuelle de leur enseignement. (Hill, Umland, Litke, et Kapitula, 2012). La plupart des préoccupations au sujet des examens pour l'obtention d'un permis sont liées à des questions méthodologiques concernant la définition de la construction et la valeur prédictive (Libman, 2009). À l'heure actuelle, il n'existe aucune trace de la validation des tests Praxis, le test normalisé utilisé dans la plupart des États américains, par rapport aux performances réelles des enseignants (Hill et al., 2012).

Il semble y avoir un thème commun dans une grande partie de la recherche, à savoir que l'une des limites de l'évaluation normalisée des compétences du personnel enseignant réside dans la difficulté pour un test à choix multiples de capter les compétences et les qualités multidimensionnelles souvent difficiles à mesurer du personnel enseignant compétent. « Les enseignants de qualité ne peuvent tout simplement pas être trouvés exclusivement par leur voie d'accès au permis d'enseigner ou leurs résultats aux examens. Il semble qu'il y ait quelque chose d'inobservable chez un individu qui fait de lui un enseignant efficace » (Shuls et Trivitt, 2015, p. 665).

Les tests normalisés comme méthode pour reconnaître les enseignants de moindre qualité

Au niveau le plus élémentaire, les tests normalisés des compétences sont conçus pour garantir une qualité minimale (Shuls et Trivitt, 2015). On ne sait pas très bien si ces tests peuvent servir d'outil pour éliminer les enseignants les moins qualifiés. Une étude sur la capacité des tests normalisés de reconnaître les enseignants compétents et les personnes inaptes à l'enseignement a révélé que les tests « n'ont pas permis de reconnaître tous les enseignants ayant une pratique si médiocre » (Hill et al., 2012, p. 514).

Les universitaires qui ont examiné les résultats à des tests normalisés aux États-Unis ont déclaré : « Combien d'enseignants potentiels incompetents qui ont pu recevoir leur permis d'enseigner en raison de pratiques de notation inappropriées ou de mauvaise qualité dans leur programme de formation de l'enseignement supérieur sont en fait empêchés de recevoir un permis d'enseigner en raison de leur échec aux tests PLACE ? Bien que nous n'ayons pas de données directes pour répondre à cette question, notre intuition est qu'ils ne sont pas nombreux » (Cobb *et al.*, 1999, p. 172).

Mise en place de tests normalisés pour des matières spécifiques

À l'heure actuelle, deux principaux types de tests normalisés des compétences des enseignants sont utilisés dans diverses instances : les tests de compétences de base et les tests de compétences spécialisées. Le premier type mesure les connaissances fondamentales que tous les enseignants sont censés avoir dans des domaines comme les mathématiques, la lecture et l'écriture. Les tests de compétences spécialisées visent à mesurer des connaissances spécifiques dans des domaines comme les mathématiques, l'histoire et les sciences (Cobb et coll., 1999; Angrist et Guryan, 2008).

Il y a différentes méthodes pour intégrer une matière à l'évaluation normalisée des enseignantes et des enseignants. Presque toutes les instances qui exigent l'évaluation des compétences incluent dans leur test certaines composantes d'un test de compétences de base, notamment les compétences de base en mathématiques, en écriture et en lecture. Presque toutes les instances qui exigent l'évaluation des compétences des enseignants exigent également ce test des compétences de base. De plus, les tests normalisés dans des matières spécifiques constituent une exigence supplémentaire dans de nombreuses régions. Ils sont de plus en plus fréquents chez les enseignants qui enseignent à la fin de la scolarité (c.-à-d. au secondaire) comparativement à ceux qui enseignent plus tôt (c.-à-d. à l'élémentaire ou au collège). Enfin, certaines régions exigent également un test normalisé de connaissances en pédagogie, bien qu'il s'agisse du type le moins courant de test normalisé obligatoire. (D'Agostino et Powers, 2009; Wayne et Youngs 2003; Shuls et Trivitt, 2015).

Calendrier de l'évaluation des enseignants

Le plus souvent, les enseignants sont testés pour ces compétences lors de leurs cours de formation. La plupart des personnes font ces tests normalisés des compétences avant de devenir enseignants certifiés (D'Agostino et Powers, 2009). Toutefois, de nombreuses régions exigent un test normalisé des compétences avant même l'admission à des programmes de formation à l'enseignement (Libman, 2009; Wayne et Youngs, 2003).

Mesure de la compétence dans les tests normalisés

Très peu d'instances qui ont recours à l'évaluation obligatoire des compétences des enseignantes et des enseignants publient les taux de réussite à l'examen la première fois. Les quelques endroits où ces détails sont publiés présentent des données d'un niveau de variation si élevé qu'il est difficile de se faire une idée claire d'un taux de réussite valide. Par exemple, le taux de réussite à l'examen de mathématiques Praxis II à l'échelle d'un État peut varier de 52 % à 93 %, sept États seulement ayant répondu. (Putman et Walsh, 2019). Le gouvernement australien rapporte que le taux de réussite aux tests normalisés des enseignants est supérieur à 90 % pour les personnes qui font le test pour la première fois. Par conséquent, il n'existe aucun précédent ou indicateur dans les pratiques d'évaluation actuelles qui permettrait d'obtenir un taux de réussite applicable à divers endroits.

En général, les instances qui utilisent ces tests de compétences normalisés pour déterminer la réussite ou la non-réussite ne donnent pas les notes individuelles. La note de passage requise varie même pour les tests normalisés les plus utilisés aux États-Unis (p. ex. les tests Praxis) et elle est déterminée à l'échelle de l'État (Shuls, et Trivitt, 2015). Les recommandations sur ces notes proviennent le plus souvent d'un groupe d'experts qui combinent des informations sur les items et les caractéristiques des candidats (Hill et al., 2012).

Les chercheurs ont étudié l'impact potentiel des mesures qui tentent d'améliorer la qualité du personnel enseignant en augmentant la note de passage requise. Les données produites montrent qu'une augmentation de la note requise pour réussir pourrait améliorer la qualité du personnel enseignant, telle que mesurée par les résultats des élèves aux tests normalisés. Toutefois, ce changement entraînerait presque certainement une réduction du nombre d'enseignants issus de groupes minoritaires et entraînerait des résultats défavorables dans les écoles défavorisées (Shuls et Trivitt, 2018). Les données montrent également que l'augmentation du score nécessaire pour réussir les tests normalisés des enseignants éliminerait 0,2 % des enseignants les plus inefficaces tout en excluant 7 % des enseignants efficaces (Goldhaber, 2007).

Options pour les candidats qui ne réussissent pas les tests normalisés

Dans la plupart des provinces et territoires où le test est obligatoire, il n'y a pas de limite quant au nombre de fois qu'une personne peut essayer de faire le test et le réussir (Libman, 2009).

De nombreuses régions qui pratiquent l'évaluation obligatoire des enseignants, comme la Caroline du Nord, la Floride et le Royaume-Uni, permettent aux diplômés d'un programme agréé de formation des enseignants qui ne réussissent pas les examens de travailler pendant plusieurs années avec un permis d'enseigner temporaire (Goldhaber, 2007; Libman, 2009).

Conception et mise en œuvre des tests de compétences

La documentation montre que, même si c'est le gouvernement qui met en œuvre la nécessité de rendre obligatoire l'évaluation des compétences du personnel enseignant, la conception et l'administration de ces tests proviennent d'organismes non gouvernementaux (Glassford, 2005; Libman, 2009; Wang et al., 2003). Aux États-Unis, la plupart des tests sont administrés par le National Evaluation System (NES), une filiale de la société à but lucratif Pearson Education ou, par l'Educational Testing Service (ETS), un organisme privé sans but lucratif. (Donovan et

Cannon, 2018; Libman, 2009). En Australie, l'Australian Council for Educational Research (ACER) est une organisation indépendante qui se consacre à la recherche en éducation et qui est chargée de faciliter un test normalisé pour les enseignantes et enseignants en formation initiale dans des centres de test via une plate-forme en ligne.

Formats de tests actuels et items inclus

Les tests normalisés de compétences modernes pour le personnel enseignant sont principalement des tests à choix multiples, mais ils comprennent aussi quelques modules de réponses courtes. Ils sont effectués sur papier, ou par voie électronique. Les tests sont présentés dans ce format afin de réaliser des économies de temps et d'argent et ce format améliore la possibilité d'évaluer des connaissances spécifiques dans un vaste domaine (D'Agostino et Powers, 2009).

Les tests d'aptitudes de base courants contiennent de 50 à 55 questions pour chaque domaine, habituellement les mathématiques, la lecture et l'écriture. La plupart des tests contiennent également des questions à réponses courtes. Plusieurs tests (p. ex. le California Basic Educational Skills Test [CBEST]) contiennent un petit nombre de questions de dissertation dans les modules de lecture et d'écriture.

Le format à choix multiples de presque tous les tests normalisés de compétences en mathématiques est un sujet de discordance pour de nombreux chercheurs. Heather C. Hill, une chercheuse de premier plan dans le domaine de l'élaboration et de la compréhension des mesures des compétences en mathématiques des enseignantes et enseignants, écrit : « Nos résultats indiquent également que la compétence en enseignement étant multidimensionnelle, il est peu probable qu'un seul test de certification à choix multiples permette de capter cette multi dimensionnalité » (Hill et al., 2012, p. 515).

Évaluation des compétences des enseignants et résultats des élèves

Un grand nombre de preuves empiriques démontrent une relation positive entre la qualité du personnel enseignant et les résultats des élèves (Aaronson et al., 2007; Coleman, 1968; Rivkin et al., 2005; Rockoff, 2004). Toutefois, les données disponibles sur la façon de mesurer la qualité du personnel enseignant et de mettre à jour tout lien de cause à effet entre la qualité du personnel enseignant et le rendement des élèves sont beaucoup moins concluantes. Par exemple, plusieurs études ont révélé que les titres de compétence des enseignants (p. ex., niveau du diplôme, statut de l'autorisation d'enseigner) ne permettent pas de prédire le rendement des élèves de façon statistiquement significative (p. ex., Aaronson et coll., 2007; Goldhaber et Brewer, 2000; Harris et Sass, 2011). Aaronson et coll., 2007; Monk et King, 1994). Les chercheurs n'ont pas non plus observé un rapport entre le type de diplôme ou d'examen d'entrée au collège (p. ex., Kane et Staiger, 2008) et le rendement des élèves en mathématiques. L'analyse documentaire présente des résultats mitigés en ce qui concerne l'association entre test normalisé de compétences du personnel enseignant et rendement des élèves.

Difficulté de mesurer la relation entre la compétence des enseignants et les résultats des élèves

L'évaluation de la relation entre le test normalisé des compétences du personnel enseignant et le rendement des élèves est complexe. L'un des défis que pose la mesure de cette relation est qu'il y a très peu de points de comparaison. Les chercheurs ne peuvent pas comparer les enseignants qui réussissent les tests de compétences avec ceux qui ne les réussissent pas car ceux-ci ne restent pas dans le système éducatif (Shuls et Trivitt, 2015).

De plus, il n'existe aucune recherche qui combine des données longitudinales antérieures et postérieures à la mise en œuvre de l'évaluation obligatoire des enseignants. Par conséquent, il n'existe aucun test pour déterminer si l'existence de tests normalisés de compétences a un effet sur la qualité du personnel enseignant (Goldhaber, Gratz et al., 2017).

Il existe également une corrélation entre les autres caractéristiques des enseignants et leur rendement aux tests normalisés de compétences. Il a été démontré que les caractéristiques des enseignantes et enseignants, telles que les années d'études postsecondaires, le nombre de cours de mathématiques suivis et le parcours pour l'obtention d'un permis d'enseigner ont toutes une incidence sur le rendement des élèves et sont en corrélation avec le rendement de l'enseignante ou de l'enseignant aux tests normalisés de compétences (Seebruck, 2015). Il n'est pas impossible de démêler ces effets, mais il faut une méthodologie nuancée et des données de grande qualité (D'Agostino et Powers 2009; Seebruck, 2015).

Un autre problème tient au fait que les élèves ne sont pas inscrits dans des classes assignées au hasard aux enseignants et vice versa. Il est donc probable qu'il existe déjà une relation entre les enseignants qui obtiennent de bons résultats aux tests de compétences et le type d'élèves auxquels ils ont tendance à enseigner et le type de classes dans lesquelles ils ont tendance à enseigner (Chetty, Friedman et Rockoff, 2014; Kane et Staiger, 2008; Jackson, Rockoff et Staiger, 2014).

Lien entre l'évaluation des compétences des enseignants et les résultats des élèves

Lorsqu'on étudie les données probantes établissant un lien entre l'évaluation des compétences du personnel enseignant et les résultats des élèves, il faut faire une distinction prudente entre la certification des enseignants, les connaissances pures et simples du personnel enseignant, les connaissances pédagogiques et l'évaluation du personnel enseignant. Comme nous l'avons vu plus haut, les tests normalisés obligatoires pour les enseignants sont une forme de contrôle de leurs connaissances du contenu. Il n'y a pas de relation parfaite entre les connaissances réelles d'un enseignant et ses résultats à un test normalisé de compétence (Goodman, Arbona et Dominguez de Ramirez, 2008). De plus, il y a encore moins de lien entre les connaissances pédagogiques d'un enseignant et ses résultats à un test normalisé de compétence (Hill, 2005; Hill, 2007; Hill et al, 2008; Kelcey et al, 2019).

Bien qu'un grand nombre de données indiquent que les connaissances du contenu de l'enseignante ou de l'enseignant ont une relation favorable avec les résultats des élèves (Campbell et coll., 2014; Carlisle et Berebitsky, 2011) et que cela dépend du contexte (Kelcey, Hill et Chin, 2019), les données sur l'association entre le test normalisé de compétences et les résultats des élèves ne sont pas concluantes (Podgursky, 2005).

Ces conclusions mitigées tirées de la recherche empirique, qui montrent que les tests normalisés de compétences des enseignantes et des enseignants peuvent être statistiquement significatifs ou insignifiants pour déterminer les résultats des élèves, ne sont pas surprenantes pour plusieurs raisons. Tout d'abord - en raison des défis méthodologiques expliqués ci-dessus.

Deuxièmement, une telle variation dans l'importance de cette relation peut s'expliquer par l'utilisation individuelle de ces tests comme critère de réussite ou de non-réussite à un cours. Ces tests n'étaient pas destinés à servir de prédicteurs nuancés du rendement en classe, mais servaient plutôt à garantir une qualité minimale (Shuls et Trivitt, 2015, p. 664). Bien que plusieurs études n'établissent aucun lien entre le rendement des enseignants aux tests normalisés de compétences et le rendement des élèves (Buddin et Zamarro, 2009; Quirk, Witten et Weinberg, 1973; Wiseman et Al-bakr, 2013), la majorité des données empiriques semblent indiquer que les tests normalisés de compétences des enseignants donnent un certain signal de la qualité de ces derniers; cette association est néanmoins faible, avec de petits effets et varie également d'un contexte à l'autre (D'Agostino et Powers, 2009; Goldhaber, 2007; Kelcey et al., 2019).

Praxis II, le test de compétences des enseignants le plus couramment utilisé aux États-Unis, a été étudié pour sa relation prédictive avec les résultats des élèves dans plusieurs contextes. Shuls et Trivitt (2015) ont utilisé les données Praxis du ministère de l'Éducation de l'Arkansas et les scores des élèves au test de la troisième à la huitième année. Cet ensemble de données ne leur a pas permis d'apparier les élèves aux enseignants, de sorte que les résultats sont plutôt basés sur les corrélations entre les années d'études. Leur analyse constate une corrélation statistiquement significative entre les enseignants et les résultats des élèves aux tests. Ils font également état d'une corrélation significative entre le parcours des enseignants pour obtenir la certification et leur expérience d'une part et les résultats des élèves aux tests d'autre part. La part de l'effet de la certification des enseignants est environ le double des deux autres. (Shuls et Trivitt, 2015).

Des chercheurs du Colorado ont examiné le test de compétences de base requises dans le cadre du processus d'attribution du permis d'enseigner et ont constaté qu'il « semblait extrêmement lourd et manquait de pertinence et d'utilité, sauf dans quelques cas ». (Cobb et al., 1999, p. 173).

Une vaste étude menée au Texas a évalué la relation entre les connaissances en mathématiques du personnel enseignant, la prise de décisions en classe et le rendement des élèves et a constaté que les connaissances en mathématiques du personnel enseignant n'étaient que très faiblement liées au rendement des élèves. De plus, cette relation, lorsqu'elle existait, était non linéaire et fortement influencée par d'autres facteurs pédagogiques comme le curriculum, la technologie disponible et le milieu scolaire (Shechtman et al., 2010). Le test normalisé utilisé pour évaluer les enseignants dans l'étude du Texas était l'instrument MKT conçu par Hill et son équipe (Hill *et al.*, 2005; Hill *et al.*, 2012).

Un examen des données internationales sur le lien entre la certification du personnel enseignant et le rendement des élèves a révélé que le fait d'exiger des tests normalisés pour la certification ne permettait pas de prédire le rendement des élèves et, par association, la qualité des enseignants (Wiseman et Al-bakr, 2013).

D'autres données internationales proviennent d'une étude menée dans dix-sept pays par Ingvarson et Rowley (2017). Cette recherche internationale explore la relation entre les processus d'assurance de la qualité du personnel enseignant au niveau national et les résultats des élèves aux examens internationaux de mathématiques (PISA et TEIMS). Bien que l'étude ne

prenait pas en compte l'effet des tests normalisés isolément, elle comprenait un test de compétence des enseignants dans le cadre de l'évaluation. Les chercheurs ont évalué chaque pays en fonction de la robustesse de leurs politiques de recrutement, d'accréditation et d'éducation des enseignants, et troisièmement, de leurs politiques d'accréditation du personnel enseignant. Le test normalisé des compétences des enseignants fait partie de cette troisième catégorie. Les résultats des élèves aux deux examens de mathématiques ont été le plus fortement prédits par la deuxième catégorie, soit la capacité d'un pays à maintenir la qualité de la formation des enseignants et de l'accréditation (0,75, $p < 0,01$). Les résultats des élèves étaient également significativement liés aux politiques de certification des enseignants pour les tests du PISA seulement (0,52, $p < 0,05$) (Ingvarson et Rowley, 2017). La différence entre les résultats aux tests PISA et TEIMS est résumée par Wu (2010) : « un pays avec un score élevé au PISA montre que les élèves sont bons en “mathématiques de tous les jours”, tandis qu'un score élevé au TEIMS montre que les élèves sont bons en “mathématiques scolaire” » (p. 96).

D'Agostino et Powers (2009) ont effectué une méta-analyse approfondie pour examiner si les résultats des enseignants aux tests de compétences permettaient de prédire leur rendement en tant qu'enseignants et les résultats des élèves et ont constaté que « les résultats aux tests étaient moins liés au rendement des enseignants qu'au degré de réussite des étudiants dans les programmes de formation initiale pour lesquels les tests avaient été conçus (p. 165) ». Dans leurs vastes analyses interprovinciales, les auteurs ont comparé l'effet du rendement des enseignants dans les facultés de formation en enseignement et du rendement des enseignants aux tests de compétences sur le rendement des élèves. D'Agostino et Powers (2009) ont constaté que le rendement de l'enseignante ou de l'enseignant dans sa formation à l'enseignement ou au collège était plus fortement corrélé aux résultats des élèves que le rendement de l'enseignante ou de l'enseignant aux tests normalisés de compétences administrés immédiatement après l'obtention du diplôme. Toutefois, les résultats au test normalisé de compétences du personnel enseignant avaient une faible corrélation positive avec les résultats des élèves. La part de l'effet des tests normalisés était environ la moitié de la part de l'effet du rendement des enseignants lors des cours de formation dans les collèges (D'Agostino et Powers, 2009).

Les études qui ont comparé l'impact relatif des caractéristiques des enseignants, notamment leurs résultats aux tests normalisés, montrent parfois une corrélation positive avec le rendement des élèves. Cependant, ces études ne sont jamais les plus prédictives. Par exemple, Begle (1979) a signalé que l'accumulation de crédits dans les cours de méthodologie des mathématiques était corrélée au rendement des élèves. De même, Ferguson et Womack (1993) ont laissé entendre que le montant de cours complétés par les enseignants représentait plus de quatre fois la variance du rendement des enseignants (c.-à-d. 16,5 %) comparativement à toute mesure de connaissances du contenu (c.-à-d. moins de 4 %).

Une étude utilisant des données de la Caroline du Nord montre que les tests normalisés pour les enseignants permettent de prédire l'efficacité des enseignants, telle que mesurée par le rendement des élèves aux tests normalisés. Ces chercheurs concluent que la réussite des enseignants au test Praxis II augmente le rendement des élèves avec un écart-type supérieur de 3 à 6 % en mathématiques en comparaison des enseignants qui ne réussissent pas le test (Goldhaber, 2007).

Clotfelter, Ladd et Vigdor (2010) ont également utilisé des données de la Caroline du Nord relatives au test Praxis pour évaluer les facteurs déterminants le rendement des élèves et établir une relation positive entre les résultats aux tests de compétences des enseignants et le

rendement des élèves aux tests normalisés de mathématiques. L'étude de Clotfelter (2010) a exploré les multiples caractéristiques des enseignantes et des enseignants et a constaté que leur score au test de compétence était le moins prédictif des résultats des élèves avec un effet de 0,01. Le nombre d'années d'expérience des enseignantes et enseignants avait un effet de 0,05, le type de permis d'enseigner avait un effet de 0,06 et l'enseignante ou l'enseignant détenant la certification pour enseigner les mathématiques avait un effet de 0,12.

Certaines données indiquent que l'importance relative des connaissances du contenu du personnel enseignant et des tests normalisés connexes pourrait augmenter à mesure que la complexité de la matière augmente dans les années d'études supérieures (Appleton, 2013). Une étude utilisant les données du test ABCTE de la Floride révèle que les enseignants qui réussissent à la fois le test de connaissances professionnelles et le test sur la matière sont associés à de meilleurs résultats chez les élèves et que la relation est plus forte lorsqu'on ne considère que le rendement des élèves de la sixième à la dixième année (Goldhaber, Gratz et al., 2017; Sass, Semykina, et Harris, 2014).

Les chercheurs ont utilisé les données de l'État de Washington et se sont penchés sur la relation entre les résultats aux tests des connaissances du contenu des enseignants et le rendement des élèves en mathématiques au palier intermédiaire et celui en algèbre en neuvième année. Ils ont constaté qu'il existe une relation modérément positive entre les résultats aux tests de compétences en mathématiques des enseignants et le rendement des élèves en mathématiques au niveau intermédiaire et aucune relation pour l'algèbre de neuvième année (Goldhaber et al., 2017).

Cela dit, les données empiriques suggèrent que les tests normalisés à choix multiples sur les compétences des enseignants peuvent être faiblement à modérément prédictifs de la pratique en classe et du rendement des élèves, particulièrement au palier intermédiaire ou secondaire (Hill et al., 2012). Toutefois, les résultats aux tests normalisés sont beaucoup moins liés aux résultats des élèves que la certification des enseignants, tant générale que spécifique à une matière, leur expérience et d'autres facteurs contribuant à l'efficacité du personnel enseignant à l'échelle de la classe et du district.

Effets associés à l'évaluation des enseignants

Bon nombre des études qui mesurent l'association entre l'évaluation du personnel enseignant et les résultats des élèves discutent des répercussions de l'évaluation du personnel enseignant si on ne prend pas en compte le rendement des élèves aux tests normalisés. Cobb et ses collaborateurs (1999) soulignent qu'étant donné qu'une si faible proportion d'enseignants potentiels ne réussissent pas le test des compétences de base, cela signifie « ... qu'une proportion extrêmement élevée d'étudiants doivent faire les tests avec peu de pouvoir discriminatoire associé à ce temps et à ces efforts » (p. 173). Il est suggéré que le contenu des tests est, pour la plupart, équivalent aux tests que les étudiants ont déjà passés afin d'être admis dans l'enseignement supérieur (Cobb et al., 1999).

Il est également prouvé que l'évaluation obligatoire des compétences des enseignants a une incidence sur la réserve de personnel enseignant, les effectifs, la formation du personnel enseignant et la diversité de la population enseignante. « Les coûts de transaction peuvent empêcher des personnes compétentes d'envisager une carrière en enseignement » (Shuls et Trivitt, 2015, p. 647).

Les résultats de l'étude d'Ingersoll (2007) suggèrent que des sélections uniques comme les tests normalisés, pourraient être perçues comme des obstacles à la compétence et au professionnalisme, à moins qu'elles ne soient accompagnées de la prise en compte d'autres facteurs comme la rémunération et les conditions de travail.

De plus, les données disponibles indiquent que certains des tests normalisés de compétences des enseignants peuvent dissuader des personnes ayant du potentiel d'entrer dans la profession plutôt que de dissuader des enseignants objectivement non qualifiés. On estime que les tests qui prétendent à la supériorité dans la sélection des enseignants diplômés des programmes de certification ont un taux de faux-rejet de 80 % et un taux de faux positifs de seulement 20 % (D'Agostino et Powers, 2009, p. 151). Les données empiriques montrent que les organismes de réglementation des États ne sont peut-être pas les mieux placés pour évaluer la qualité de l'enseignement. Podgursky (2005) suggère que la qualité de l'enseignement pourrait être mieux prise en compte au niveau local par les directions d'école, signalant que celles-ci savent reconnaître les enseignantes et les enseignants dont les méthodes contribuent à améliorer les résultats des élèves.

L'influence des tests normalisés sur la diversité et l'inclusion

La mise en œuvre de tests normalisés pour les enseignants, en restreignant l'accès à la profession, a un impact sérieux sur la diversité raciale au sein de la réserve d'enseignants. Il est de plus en plus évident que l'évaluation obligatoire des compétences des enseignants a un effet limitatif sur la diversité et l'inclusion du personnel enseignant. Angrist et Guryan (2008) montrent que lorsqu'ils analysent l'intersection entre la race et les tests normalisés d'enseignement, les évaluations obligatoires se révèlent être une barrière subjective plutôt qu'un test de sélection de la qualité. D'autres recherches montrent qu'un système d'examen normalisés à « enjeux élevés » porte préjudice aux groupes sociaux ou aux minorités les plus faibles (Libman, 2009, p. 11). De nombreuses études ont montré que le taux de réussite aux tests normalisés courants de compétences des enseignants est significativement plus faible chez les personnes de couleur que chez les candidats blancs (Cobb et al. 1999; Graham, 2013; Nettles et al., 2011); Petchauer, 2012). En outre, il est prouvé que ces types de tests manquent d'objectivité à l'encontre de presque toutes les classes vulnérables d'enseignants potentiels, autres que les femmes (Angrist et Guryan, 2008). Les données empiriques montrent que les étudiants noirs (Petchauer, 2012) et hispaniques (Cobb et al., 1999) sont beaucoup plus susceptibles d'échouer aux tests normalisés que leurs pairs blancs.

Les chercheurs ont utilisé des concepts comme la théorie du capital social, la menace du stéréotype ou le biais d'évaluation comme raisonnement potentiel pour expliquer pourquoi les populations racialisées échouent de façon disproportionnée aux tests normalisés d'enseignement comme les tests Praxis (Graham, 2013). Bien qu'il n'existe pas de conclusion définitive quant au facteur qui a le plus d'influence, il est continuellement prouvé que ces tests ont un impact négatif sur les futurs enseignants en formation initiale. Graham (2013) met en lumière les travaux de recherche de D^{re} Dara Wakefield, qui a abouti à la conclusion que le test Praxis « plus que tout autre critère, réduit à lui seul le nombre de minorités qui entrent dans la profession enseignante » (Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing (Présélection des candidates et candidats à l'enseignement : problèmes avec les tests à enjeux élevés). The Educational Forum, 67, 380-388).

En plus de réduire le bassin de candidats diversifiés, plusieurs études ont montré que les tests normalisés de compétences peuvent avoir une incidence sur la répartition du personnel enseignant qualifié. (Podgursky, 2005; Seebruck, 2015). Une répartition déséquilibrée des enseignants de qualité a pour conséquence involontaire de réduire la qualité moyenne des enseignants dans les districts les plus vulnérables, les plus pauvres ou les plus ruraux. (Podgursky, 2005; Seebruck, 2015).

Caractéristiques des enseignants et procédures d'évaluation permettant de prévoir les résultats des élèves

La documentation sur la relation entre les tests normalisés d'évaluation du personnel enseignant et les résultats des élèves comprend aussi, de façon inhérente, certaines constatations sur les caractéristiques du personnel enseignant et les procédures d'évaluation du personnel enseignant qui influent également sur les résultats des élèves. Aucune caractéristique unique n'est systématiquement considérée comme la « meilleure » qualité d'un enseignant. En général, l'expérience et la formation de haute qualité des enseignants sont considérées comme des prédicteurs positifs de la réussite des élèves (Harris et Sass, 2011). Pourtant, même ici, les résultats sont mitigés (Steinberg et Sartain, 2015).

Les recherches menées au Québec, où les résultats des élèves aux tests normalisés de mathématiques sont élevés par rapport au reste du Canada, attribuent ces résultats aux programmes d'études postsecondaires en mathématiques offerts dans la province. Ils trouvent qu'il est particulièrement important d'exiger des enseignants stagiaires qu'ils suivent davantage de cours sur la méthodologie et le contenu des mathématiques (Shulman 1987; Vashchysyn et Chernoff, 2016). À l'heure actuelle, la plupart des régions, particulièrement en Amérique du Nord, n'exigent pas du personnel enseignant du palier élémentaire qu'il consacre un grand nombre d'heures aux mathématiques pendant sa formation initiale, et les instances n'ont pas non plus de séries de questions rigoureuses dans leurs examens pour obtenir l'autorisation d'exercer afin de déterminer la compétence en mathématiques (Georges, Borman, et Lee, 2010). Ces deux facteurs ont été fortement liés à l'amélioration du rendement des élèves au Québec et ailleurs (Begle, 1979; Borman et Lee, 2010).

Plusieurs des études d'évaluation de l'efficacité des enseignants indiquent que l'expérience de l'enseignante ou de l'enseignant est un prédicteur important des résultats des élèves (Goldhaber, 2007; Huang et Moon, 2009). L'étude Huang, une analyse à plusieurs niveaux des caractéristiques des enseignants et du rendement des élèves, a révélé que « les élèves qui ont des enseignants chevronnés pendant quatre années consécutives peuvent se retrouver avec plus d'un écart-type supérieur ou environ une année d'étude de plus par rapport aux élèves qui ont des enseignants débutants » (Huang et Moon, 2009, p. 226).

D'autres chercheurs ont constaté un fort pouvoir prédictif du rendement potentiel des enseignants selon leur formation collégiale ou universitaire, surtout lorsqu'ils faisaient cours aux élèves (D'Agostino et Powers, 2009; Marder, 2016). Une étude portant spécifiquement sur les mathématiques a révélé que l'amélioration de la qualité et de la quantité des cours de mathématiques obligatoires au niveau de la formation initiale était l'une des mesures les plus utiles pour améliorer les résultats des élèves (Georges, Borman, et Lee, 2010).

L'ensemble de la recherche portant sur l'évaluation de la méthode d'évaluation des enseignants la plus prédictive de résultats positifs pour les élèves est cohérent dans ses conclusions. Plutôt

que de mettre l'accent sur les caractéristiques semi-immuables de l'enseignante ou de l'enseignant comme l'expérience antérieure, l'éducation ou les capacités cognitives, ces données suggèrent de se concentrer sur des mesures d'évaluation axées sur les processus et la pédagogie (Goe, Bell, et Little, 2008). Comme l'écrivent Clotfelter et al (2010) : « L'écart entre la variation globale de la qualité des enseignants et celle prévue par les seuls titres de compétence implique que ce serait une erreur de la part des responsables des politiques d'accorder une telle importance aux titres de compétence mesurables pour déterminer la qualité des enseignants qu'ils ignorent les autres facteurs contribuant à leur efficacité, dont beaucoup ne peuvent être déterminés que par l'observation dans les écoles ou dans les classes » (p. 676).

L'étude de Briole et Maurin (2019) révèle que l'évaluation des enseignants a une incidence importante sur les résultats des élèves et que le processus d'évaluation le plus efficace est l'évaluation en personne effectuée par un supérieur et consiste en une observation en classe suivie d'une séance de débriefing. Leurs résultats indiquent une augmentation significative et importante (d'un écart-type d'environ 4,5 %) des résultats en mathématiques des élèves. « Cet effet s'observe non seulement pour les élèves qui ont l'enseignante ou l'enseignant pendant l'année de l'évaluation, mais aussi pour les élèves qui ont la même enseignante ou le même enseignant les années suivantes, ce qui suggère que les évaluations améliorent les compétences pédagogiques de base des enseignants » (Briole et Maurin, 2019, p. 15).

Une étude similaire évaluant l'impact d'un nouveau système d'évaluation des enseignants lancé à Cincinnati en 2000 a révélé que de brèves évaluations en personne des enseignants, combinant le contenu et la rétroaction pédagogique, amélioreraient considérablement et durablement le rendement des élèves aux tests (Taylor et Tyler, 2012).

Conclusion

La qualité du personnel enseignant est considérée comme ayant une influence sur le rendement des élèves (Aaronson et coll., 2007; Coleman, 1968; Rivkin et coll., 2005; Rockoff, 2004). Pourtant, des variables individuelles (p. ex. le niveau du diplôme ou le statut d'autorisation d'exercer) semblent permettre de détecter et de prévoir des changements importants dans les résultats des élèves (Aaronson et coll., 2007; Goldhaber et Brewer, 2000; Harris et Sass, 2011).

Dans l'ensemble, les tests normalisés de compétences des enseignants, qui ont été diffusés et institués aux États-Unis et dans plusieurs endroits du monde, sont intrinsèquement liés à la perception que le public a de la robustesse et de la qualité de l'enseignement. Ces tests normalisés ont vu le jour à l'occasion d'un mouvement qui demandait une plus grande obligation de rendre compte de la qualité et de l'efficacité des enseignants locaux et des programmes de formation des enseignants. Ces pratiques sont liées à certaines recherches qui créent une voie permettant des améliorations potentielles de la qualité et du rendement du personnel enseignant en classe. Cependant, bien que ces tests puissent recevoir l'appui des responsables des politiques, des membres du milieu de l'éducation et du public, l'objectif fondamental de ces tests - améliorer l'apprentissage des élèves - n'est souvent pas atteint. Les recherches actuelles montrent que les tests normalisés pour les enseignants ne sont pas liés à un niveau de rendement uniforme qui justifie leur mise en œuvre à grande échelle à l'heure actuelle. De nombreux chercheurs recommandent de faire preuve de prudence dans le cadre de ces tests, étant donné qu'ils ne sont pas toujours associés aux avantages positifs qui sont souvent revendiqués. De plus, les effets négatifs potentiels de ces programmes, notamment les préjugés

contre les groupes marginalisés et la diminution de la disponibilité d'enseignants qualifiés, sont des répercussions plus systématiques de ces tests.

Si l'on ne cerne pas les meilleures méthodes pour accroître la réussite des élèves, la création d'un parcours restrictif pour devenir enseignante ou enseignant peut causer plus de tort que de bien. De nombreux chercheurs travaillant dans ce domaine, de concert avec l'American Psychological Association (APA), s'opposent à l'utilisation de tests normalisés comme seule mesure de l'aptitude ou du rendement (Goodman et al., 2008). Les qualités positives et le mérite des enseignants, comme nous l'avons déjà souligné dans cet article, vont au-delà de ce qui peut être mesuré uniquement par un test normalisé à choix multiples. La recherche montre que la création d'un système uniquement consacré aux tests normalisés, et en l'absence d'autres évaluations, peut avoir des répercussions négatives à la fois sur la réserve d'enseignants disponibles et sur les résultats scolaires des élèves. Tant que les tests normalisés de compétences des enseignants ne sont pas associés à des avantages constants et mesurables, ils doivent être utilisés avec prudence et en tenant compte de leur impact potentiel sur les communautés locales.

« L'écart entre la variation de la qualité des enseignants et [la qualité des enseignants] prévue par les seuls titres de compétence implique que ce serait une erreur de la part des responsables des politiques d'accorder autant d'importance aux titres de compétence mesurables pour déterminer la qualité des enseignants » (Clotfelter, Ladd, et Vigdor, 2010, p. 676).

En outre, d'autres facteurs essentiels à l'efficacité des enseignants sont de nature qualitative et ne peuvent pas être mesurés par des méthodes d'évaluation normalisées, mais doivent être observés au niveau local (Clotfelter, Ladd, et Vigdor, 2010).

Références

- 2018 Teacher Prep Review. (n.d.). Retrieved July 28, 2019, from National Council on Teacher Quality (NCTQ) website: <https://www.nctq.org/publications/2018-Teacher-Prep-Review>
- Aaronson, D., Barrow, L., et Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1), 95–135. <https://doi.org/10.1086/508733>
- Angrist, J. D., et Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, 27(5), 483–503. <https://doi.org/10.1016/j.econedurev.2007.03.002>
- Appleton, K. (2013). *Elementary Science Teacher Education: International Perspectives on Contemporary Issues and Practice*. <https://doi.org/10.4324/9781315045443>
- Ball, D. L., Thames, M. H., et Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of Teacher Education*, 59(5), 389–407.
- Ball, Hill, H. C., et Bass. (2005). Knowing Mathematics for Teaching: Who Knows Mathematics Well Enough To Teach Third Grade, and How Can We Decide? *American Educator*, 29(1), 14–46.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. et Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Begle, E. G. (1979). *Critical Variables in Mathematics Education: Findings from a Survey of the Empirical Literature*.
- Blömeke, S., et Delaney, S. (2014). Assessment of Teacher Knowledge Across Countries: A Review of the State of Research. In S. Blömeke, F.-J. Hsieh, G. Kaiser, et W. H. Schmidt (Eds.), *International Perspectives on Teacher Knowledge, Beliefs and Opportunities to Learn: TEDS-M Results* (pp. 541–585). https://doi.org/10.1007/978-94-007-6437-8_25
- Briole, S., et Maurin, E. (2019). *Does Evaluating Teachers Make a Difference?* (SSRN Scholarly Paper No. ID 3390297). Retrieved from Social Science Research Network website: <https://papers.ssrn.com/abstract=3390297>
- Buddin, R., et Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115. <https://doi.org/10.1016/j.jue.2009.05.001>
- Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., et Schmidt, W. (2019). *Teaching for Excellence and Equity: Analyzing Teacher Characteristics, Behaviors and Student Outcomes with TIMSS*. <https://doi.org/10.1007/978-3-030-16151-4>
- Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., Neumayer DePiper, J., Frank, T. J., Griffin, M. J et Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, 45(4), 419–459.

- Carlisle, J. F., et Berebitsky, D. (2011). Literacy coaching as a component of professional development. *Reading and Writing*, 24(7), 773–800. <https://doi.org/10.1007/s11145-009-9224-4>
- Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., et Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499–531.
- Chetty, R., Friedman, J. N., et Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Clotfelter, C. T., Ladd, H. F., et Vigdor, J. L. (2010). Teacher Credentials and Student Achievement in High School A Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources*, 45(3), 655–681. <https://doi.org/10.3368/jhr.45.3.655>
- Cobb, R. B., Shaw, R., Millard, M., et Bomotti, S. (1999). An Examination of Colorado's Teacher Licensure Testing. *The Journal of Educational Research*, 92(3), 161–175. <https://doi.org/10.1080/00220679909597592>
- Coleman, J. S. (1968). Equality of Educational Opportunity. *Equity et Excellence in Education*, 6(5), 19–28. <https://doi.org/10.1080/0020486680060504>
- Corey, D. L., Phelps, G., Ball, D. L., Demonte, J., et Harrison, D. (2012). Explaining Variation in Instructional Time: An Application of Quantile Regression. *Educational Evaluation and Policy Analysis*, 34(2), 146–163. <https://doi.org/10.3102/0162373711431098>
- Council, N. R., Education, D. of B. and S. S. and, Education, C. for, et States, C. on the S. of T. P. P. in the U. (2010). *Preparing Teachers: Building Evidence for Sound Policy*. National Academies Press.
- D'Agostino, J. V., et Powers, S. J. (2009). Predicting Teacher Performance With Test Scores and Grade Point Average: A Meta-Analysis. *American Educational Research Journal*, 46(1), 146–182. <https://doi.org/10.3102/0002831208323280>
- Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*. Retrieved from <https://eric.ed.gov/?id=ED535859>
- Darling-Hammond, L., Berry, B., et Thoreson, A. (2001). Does Teacher Certification Matter? Evaluating the Evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57–77. <https://doi.org/10.3102/01623737023001057>
- DeForest, J. (2006). New York City's Failed Teacher Selection Project: Political Reality Trumps Educational Research, 1947-1953. *Teachers College Record*, 108(4), 726–747. <https://doi.org/10.1111/j.1467-9620.2006.00666.x>
- Donovan, M. K., et Cannon, S. O. (2018). The university supervisor, edTPA, and the new making of the teacher. *Education Policy Analysis Archives*, 26, 28. <https://doi.org/10.14507/epaa.26.2849>
- Ferguson, P., et T. Womack, S. (1993). The Impact of Subject Matter and on Teaching Performance. *Journal of Teacher Education*, 44(1), 55–63. <https://doi.org/10.1177/0022487193044001008>

- Georges, A., Borman, K. M., et Lee, R. S. (2010). Mathematics Reform and Teacher Quality in Elementary Grades: Assessments, Teacher Licensure, and Certification. *Education Policy Analysis Archives*, 18, 13. <https://doi.org/10.14507/epaa.v18n13.2010>
- Gibb, Nick Hn. (2019) Changes to the professional skills test for teachers, *UK Government, Department of Education*. <https://www.gov.uk/government/news/changes-to-the-professional-skills-test-for-teachers>
- Glassford, L. A. (2005). A Triumph of Politics over Pedagogy? The Case of the Ontario Teacher Qualifying Test, 2000-2005. *Canadian Journal of Educational Administration and Policy*.
- Glassford, L. A. (2007). Quality Assurance or Election Fodder? The Politics of Teacher Testing in Ontario, 1999 - 2003. *Historical Studies in Education / Revue d'histoire de l'éducation*, 25–52.
- Goe, L. (2007). *The Link between Teacher Quality and Student Outcomes: A Research Synthesis*. Retrieved from <https://eric.ed.gov/?id=ED521219>
- Goe, L., Bell, C., et Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.
- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765–794.
- Goldhaber, D., et Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1), 134–150.
- Goldhaber, D., Cowan, J., et Theobald, R. (2017). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education*, 68(4), 377–393. <https://doi.org/10.1177/0022487117702582>
- Goldhaber, D. D., et Brewer, D. J. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–145. <https://doi.org/10.3102/01623737022002129>
- Goldhaber, D., Gratz, T., et Theobald, R. (2017). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student STEM achievement and course-taking. *Economics of Education Review*, 61, 112–129. <https://doi.org/10.1016/j.econedurev.2017.09.002>
- Goldhaber, D., et Hansen, M. (2010). Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions. Working Paper 31. *National Center for Analysis of Longitudinal Data in Education Research*.
- Goodman, G., Arbona, C., et Dominguez de Rameriz, R. (2008). High-Stakes, Minimum-Competency Exams: How Competent Are They for Evaluating Teacher Competence? *Journal of Teacher Education*, 59(1), 24–39. <https://doi.org/10.1177/0022487107309972>
- Graham, A. (2013). Black teacher education candidates' performance on PRAXIS I: What test results do not tell us. *Negro Educational Review*, 64(1/4), 9.
- Haney, W., Madaus, G., et Kreitzer, A. (1987). Chapter 5: Charms Talismanic: Testing Teachers for the Improvement of American Education. *Review of Research in Education*, 14(1), 169–238.

- Harris, D. N., et Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7), 798–812. <https://doi.org/10.1016/j.jpubeco.2010.11.009>
- Hill, H., et Ball, D. L. (2009). The Curious—And Crucial—Case of Mathematical Knowledge for Teaching. *Phi Delta Kappan*, 91(2), 68–71. <https://doi.org/10.1177/003172170909100215>
- Hill, H. C. (2007). Mathematical Knowledge of Middle School Teachers: Implications for the No Child Left Behind Policy Initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95–114. <https://doi.org/10.3102/0162373707301711>
- Hill, H. C. (2010). The Nature and Predictors of Elementary Teachers' Mathematical Knowledge for Teaching. *Journal for Research in Mathematics Education*, 41(5), 513–545.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et Ball, D. L. (2008). Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An Exploratory Study. *Cognition and Instruction*, 26(4), 430–511. <https://doi.org/10.1080/07370000802177235>
- Hill, H. C., Kapitula, L., et Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. <https://doi.org/10.3102/0002831210387916>
- Hill, H. C., Rowan, B., et Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, 42(2), 371–406. <https://doi.org/10.3102/00028312042002371>
- Hill, H. C., Sleep, L., Lewis J. M., et Ball, D. L., (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In Lester, F. K., (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (p. 111-155). Charlotte, N.C: Information Age Pub. Retrieved from http://hub.mspnet.org/media/data/Hill_etal_final.pdf?media_000000005636.pdf
- Hill, H. C., Umland, K., Litke, E., et Kapitula, L. R. (2012). Teacher Quality and Quality Teaching: Examining the Relationship of a Teacher Assessment to Practice. *American Journal of Education*, 118(4), 489–519. <https://doi.org/10.1086/666380>
- Huang, F. L., et Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability*, 21(3), 209–234. <https://doi.org/10.1007/s11092-009-9074-2>
- Ingersoll, R. M. (2007). A Comparative Study of Teacher Preparation and Qualifications in Six Nations. CPRE Research Report Series. *Consortium for Policy Research in Education*.
- Ingvarson, L., et Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46(4), 177–193. <https://doi.org/10.3102/0013189X17711900>
- Jackson, C. K., Rockoff, J. E., et Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics*, 6(1), 801–825. <https://doi.org/10.1146/annurev-economics-080213-040845>

- Kane, T. J., et Staiger, D. O. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (Working Paper No. 14607). <https://doi.org/10.3386/w14607>
- Kelcey, B., et Carlisle, J. F. (2013). Learning about teachers' literacy instruction from classroom observations. *Reading Research Quarterly*, 48(3), 301-317. <https://doi.org/10.3102/0002831213492844>
- Kelcey, B., Hill, H. C., et Chin, M. J. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel quantile mediation analysis. *School Effectiveness and School Improvement*, 0(0), 1–34. <https://doi.org/10.1080/09243453.2019.1570944>
- Kersting, N., Givvin, K., Thompson, B., Santagata, R., et Stigler, J. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, 49(3), 568–589. <https://doi.org/10.3102/0002831212437853>
- Libman, Z. (2009). Teacher licensing examinations—True progress or an illusion? *Studies in Educational Evaluation*, 35(1), 7–15. <https://doi.org/10.1016/j.stueduc.2009.01.003>
- McNeal, K., et Lawrence, S. (2009). Teachers from the “Neighborhood”: Standardized Testing as a Barrier to Certification of Minority Candidates. *Online Yearbook of Urban Learning, Teaching, and Research*. Retrieved from <https://eric.ed.gov/?id=EJ859480>
- Meroni, E. C., Vera-Toscano, E., et Costa, P. (2015). Can low skill teachers make good students? Empirical evidence from PIAAC and PISA. *Journal of Policy Modeling*, 37(2), 308–323. <https://doi.org/10.1016/j.jpolmod.2015.02.006>
- Monk, D. H., et King, J. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. *Choices and Consequences: Contemporary Policy Issues in Education*, 29–58.
- Mullens, J. E., Murnane, R. J., et Willett, J. B. (1996). The Contribution of Training and Subject Matter Knowledge to Teaching Effectiveness: A Multilevel Analysis of Longitudinal Evidence from Belize. *Comparative Education Review*, 40(2), 139–157. <https://doi.org/10.1086/447369>
- Nettles, M. T., Scatton, L. H., Steinberg, J. H., et Tyler, L. L. (2011). Performance and passing rate differences of African American and white prospective teachers on Praxis™ examinations: A joint project of the National Education Association (NEA) and Educational Testing Service (ETS). *ETS Research Report Series*, 2011(1), i–82.
- Petchauer, E. (2012). Teacher licensure exams and Black teacher candidates: Toward new theory and promising practice. *The Journal of Negro Education*, 81(3), 252–267. <https://doi.org/10.7709/jnegroeducation.81.3.0252>
- Petscher, Y., et Logan, J. A. R. (2014). Quantile Regression in the Study of Developmental Sciences. *Child Development*, 85(3), 861–881. <https://doi.org/10.1111/cdev.12190>
- Podgursky, M. (2005). Teacher Licensing in U.S. Public Schools: The Case for Simplicity and Flexibility. *Peabody Journal of Education*, 80(3), 15–43. https://doi.org/10.1207/s15327930pje8003_3

- Pullin, D. (2017). What Counts? Who is Counting? Teacher Education Improvement and Accountability in a Data-Driven Era. In J. Nuttall, A. Kostogriz, M. Jones, et J. Martin (Eds.), *Teacher Education Policy and Practice: Evidence of Impact, Impact of Evidence* (pp. 3–16). https://doi.org/10.1007/978-981-10-4133-4_1
- Putman, H., et Walsh, K. (2019). *A Fair Chance: Simple Steps to Strengthen and Diversify the Teacher Workforce*. Retrieved from <https://eric.ed.gov/?id=ED593388>
- Quirk, T. J., Witten, B. J., et Weinberg, S. F. (1973). Review of studies of the concurrent and predictive validity of the National Teacher Examinations. *Review of Educational Research*, 43(1), 89–113.
- Rivkin, S. G., Hanushek, E. A., et Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2), 247–252. <https://doi.org/10.1257/0002828041302244>
- Rockoff, J. E., Jacob, B. A., Kane, T. J., et Staiger, D. O. (2010). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, 6(1), 43–74. https://doi.org/10.1162/EDFP_a_00022
- Rowland, T., Huckstep, P., et Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8(3), 255–281.
- Santibanez, L. (2006). Why we should care if teachers get A's: Teacher test scores and student achievement in Mexico. *Economics of Education Review*, 25(5), 510–520. <https://doi.org/10.1016/j.econedurev.2005.08.001>
- Sass, T. R., Semykina, A., et Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38, 9–23.
- Seebruck. (n.d.). *Teacher Quality and Student Achievement: A Multilevel Analysis of Teacher Credentialization and Student Test Scores in California High Schools*. Retrieved from https://www.mcgill.ca/msr/files/msr/seebruck_2015.pdf
- Shechtman, N., Roschelle, J., Haertel, G., et Knudsen, J. (2010). Investigating Links from Teacher Knowledge, to Classroom Practice, to Student Learning in the Instructional System of the Middle-School Mathematics Classroom. *Cognition and Instruction*, 28(3), 317–359. <https://doi.org/10.1080/07370008.2010.487961>
- Shuls, J. V. (2018). Raising the Bar on Teacher Quality: Assessing the Impact of Increasing Licensure Exam Cut-Scores. *Educational Policy*, 32(7), 969–992. <https://doi.org/10.1177/0895904816682315>
- Shuls, J. V., et Trivitt, J. R. (2015). Teacher Effectiveness: An Analysis of Licensure Screens. *Educational Policy*, 29(4), 645–675. <https://doi.org/10.1177/0895904813510777>
- Steinberg, M. P., et Sartain, L. (2015). Does Teacher Evaluation Improve School Performance? Experimental Evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, 10(4), 535–572. https://doi.org/10.1162/EDFP_a_00173

- Taylor, E. S., et Tyler, J. H. (2012). The Effect of Evaluation on Teacher Performance. *American Economic Review*, 102(7), 3628–3651. <https://doi.org/10.1257/aer.102.7.3628>
- Thibodeaux, A. K., Labat, M. B., Lee, D. E., et Labat, C. A. (2015). The Effects of Leadership and High-Stakes Testing on Teacher Retention. *Academy of Educational Leadership Journal; Arden*, 19(1), 227–249.
- Tighe, E. L., et Schatschneider, C. (2016). A Quantile Regression Approach to Understanding the Relations Among Morphological Awareness, Vocabulary, and Reading Comprehension in Adult Basic Education Students. *Journal of Learning Disabilities*, 49(4), 424–436. <https://doi.org/10.1177/0022219414556771>
- Vashchysyn, I. I., et Chernoff, E. J. (2016). A Formula for Success? An Examination of Factors Contributing to Quebec Students' Strong Achievement in Mathematics. *Canadian Journal of Education / Revue Canadienne de l'éducation*, 39(1), 1–26. Retrieved from JSTOR.
- Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing. *The Educational Forum*, 67(4).
- Wang, J., et Hartley, K. (2003). Video Technology as a Support for Teacher Education Reform. *Journal of Technology and Teacher Education*, 11(1), 105–138.
- Wayne, A. J., et Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research*, 73(1), 89–122. <https://doi.org/10.3102/00346543073001089>
- Wiseman, A. W., et Al-bakr, F. (2013). The elusiveness of teacher quality: A comparative analysis of teacher certification and student achievement in Gulf Cooperation Council (GCC) countries. *PROSPECTS*, 43(3), 289–309. <https://doi.org/10.1007/s11125-013-9272-z>
- Xu, Z., Hannaway, J., et Taylor, C. (2011). Making a difference? The effects of Teach For America in high school. *Journal of Policy Analysis and Management*, 30(3), 447–469. <https://doi.org/10.1002/pam.20585>