

Accord interjuges des nouveaux domaines, la littératie et la numératie, de l'AEPS/EIS

Inter-rater agreement of the news areas, literacy and numeracy of the AEPS/EIS

Colombe Lemire, Carmen Dionne and Suzie McKinnon

Volume 44, Number 1, 2015

URI: <https://id.erudit.org/iderudit/1039271ar>

DOI: <https://doi.org/10.7202/1039271ar>

[See table of contents](#)

Publisher(s)

Revue de Psychoéducation

ISSN

1713-1782 (print)

2371-6053 (digital)

[Explore this journal](#)

Cite this article

Lemire, C., Dionne, C. & McKinnon, S. (2015). Accord interjuges des nouveaux domaines, la littératie et la numératie, de l'AEPS/EIS. *Revue de psychoéducation*, 44(1), 63–81. <https://doi.org/10.7202/1039271ar>

Article abstract

This article presents the degree of interjudge reliability between two raters in regard to the new literacy and numeracy domains found in the *Programme EIS Évaluation Intervention Suivi*, which is the French translation of the *Assessment, Evaluation, and Programming System* (AEPS). For several years now, actors working alongside children with disabilities have been unsatisfied with conventional tests evaluating the child's development due to the rigid nature of the administrative protocols and the inconsistency of the assessment exercises when comparing them to the natural behavior of the child, two issues that are remedied with the AEPS. This tool for the assessment of the development of the child aims to introduce the domains of literacy and numeracy to its latest edition. The research questions tackled in the study are concerned with the interjudge reliability between two raters tasked with evaluating both domains as well as the objectives that constitute them. The sample is composed of 32 children who have been evaluated in their natural living environment. The results show an excellent interjudge reliability for literacy and a substantial one for numeracy. In conclusion, as far as their fidelity is concerned, the new domains present some interesting psychometric qualities. While this study contributes to the elaboration process surrounding AEPS's domains of literacy and numeracy, more studies are required in order to further validate the new edition of the program, as a whole.

Accord interjuges des nouveaux domaines, la littératie et la numératie, de l'AEPS/EIS

Inter-rater agreement of the news areas, literacy and numeracy of the AEPS/EIS

C. Lemire¹
C. Dionne¹
S. McKinnon¹

¹ Université du Québec à
Trois-Rivières, Département
de psychoéducation

Résumé

Cet article présente les degrés d'accord interjuges entre deux évaluateurs pour les nouveaux domaines, la littératie et la numératie, du Programme EIS Évaluation Intervention Suivi, une adaptation francophone de l'Assessment, Evaluation, and Programming System (AEPS). Depuis plusieurs années, les tests d'évaluation conventionnels du développement sont source d'insatisfaction pour les intervenants œuvrant auprès d'enfants présentant des incapacités, en raison des protocoles d'administration rigides et des exercices d'évaluation éloignés des comportements naturels de l'enfant, ce que pallie l'AEPS/EIS. Ce dernier, un outil d'évaluation du développement de l'enfant, vise dans sa nouvelle édition l'introduction des domaines de la littératie et de la numératie. Les questions de recherche de la présente étude concernent les degrés d'accord interjuges entre deux évaluateurs pour chacun des deux domaines et pour chacun des buts qui les composent. L'échantillon est constitué de 32 enfants, évalués dans leur milieu de vie. Les résultats montrent un degré d'accord excellent pour la littératie et substantiel pour la numératie. Seul un but en littératie présente un degré d'accord faible. De même, seule une absence d'accord est notée pour un but de la numératie. En conclusion, les nouveaux domaines présentent des qualités psychométriques intéressantes sur le plan de la fidélité. Cette étude est un apport au processus d'élaboration des domaines de la littératie et de la numératie de l'AEPS/EIS. Néanmoins, la poursuite de la validation de l'ensemble de la nouvelle édition demeure incontournable.

Correspondance :

Colombe Lemire
Chaire de recherche du Canada
en intervention précoce
Université du Québec à
Trois-Rivières
Département de psychoéducation
3351, boul. des Forges, C.P. 500
Trois-Rivières (QC) G9A 5H7
coland2010@gmail.com

Mots-clés : Évaluation authentique, développement de l'enfant, fidélité, test conventionnel, littératie, numératie.

Abstract

This article presents the degree of interjudge reliability between two raters in regard to the new literacy and numeracy domains found in the Programme EIS Évaluation

Intervention Suivi, which is the French traduction of the Assessment, Evaluation, and Programming System (AEPS). For several years now, actors working alongside children with disabilities have been unsatisfied with conventional tests evaluating the child's development due to the rigid nature of the administrative protocols and the inconsistency of the assessment exercises when comparing them to the natural behavior of the child, two issues that are remedied with the AEPS. This tool for the assessment of the development of the child aims to introduce the domains of literacy and numeracy to its latest edition. The research questions tackled in the study are concerned with the interjudge reliability between two raters tasked with evaluating both domains as well as the objectives that constitute them. The sample is composed of 32 children who have been evaluated in their natural living environment. The results show an excellent interjudge reliability for literacy and a substantial one for numeracy. In conclusion, as far as their fidelity is concerned, the new domains present some interesting psychometric qualities. While this study contributes to the elaboration process surrounding AEPS's domains of literacy and numeracy, more studies are required in order to further validate the new edition of the program, as a whole.

Keywords: Authentic assessment, children development, fidelity, conventional test, literacy, numeracy.

Introduction

Déjà dans les années 1970, des professionnels œuvrant auprès d'enfants présentant un risque de retard développemental ou des incapacités identifient le besoin de solutions de rechange aux tests d'évaluation conventionnels (Bricker, 2002). Ils font part de leur insatisfaction sur le plan du contenu, des procédures et de l'utilité de ces derniers (Bagnato, Neisworth et Pretti-Frontczak, 2010). Ils souhaitent des outils d'évaluation du développement de l'enfant utiles pour l'intervention, valides, fonctionnels et plus adaptés à cette clientèle.

En intervention précoce, les tests conventionnels ont pour objectif de déterminer un score normatif à des fins diagnostiques et n'ont pas été développés pour soutenir l'intervention (Bagnato et al., 2010). Ces tests sont administrés selon un protocole strict dans un environnement contrôlé par des évaluateurs, en général inconnus de l'enfant (Bagnato et al., 2010). Les insatisfactions ainsi que les lacunes dont les écrits scientifiques font généralement mention recouvrent quatre catégories : le contenu du test, son utilité pour l'intervention, ses procédures d'administration et l'implication du réseau social de l'enfant dans l'évaluation.

En ce qui concerne le contenu, Chaytor et Schmitter-Edgecombe (2003) ainsi que Thurman et McGrath (2008) mettent en évidence que les items des tests conventionnels reflètent pas ou peu les exigences de la vie quotidienne. Par conséquent, ils témoignent peu des habiletés que les enfants ont dans le cadre d'une interaction avec leur environnement humain ou physique (Gao et Grisham-Brown, 2011). L'utilité de ces tests est limitée, car les résultats de l'évaluation sont difficilement transférables en buts et objectifs d'intervention (Macy, Bricker et Squires, 2005). Sur le plan des procédures, les conditions de passation strictes ne permettent pas d'adapter le matériel ou les activités d'évaluation pour les enfants qui présentent des limitations (Bagnato et al., 2010). Par ailleurs, les tests conventionnels ont été développés pour être réalisés de façon individuelle avec l'apport d'un professionnel; l'implication de personnes significatives pour l'enfant est ainsi exclue (Bagnato et al., 2010). Ceci fait en sorte que le portrait de l'enfant, bien

que valide par rapport à l'identification de repères développementaux, se révèle partiel. En effet, le fonctionnement de l'enfant en interaction avec les différentes personnes de son entourage et dans différents contextes est peu documenté.

Selon Bagnato et al. (2010), l'évaluation authentique se caractérise par l'observation des comportements naturels de l'enfant dans son milieu de vie (école, maison, service de garde), dans le cadre d'activités dites fonctionnelles, c'est-à-dire des activités correspondant aux exigences de la vie quotidienne. Ce type d'évaluation se distingue également par la contribution des personnes significatives (parent ou proche) pour l'enfant. De plus, l'évaluation est associée à un curriculum d'activités (Gao, 2008). Ainsi, elle permet d'orienter et de planifier adéquatement l'intervention (Thurman et McGrath, 2008). Ce type d'évaluation n'a pas pour but de comparer l'enfant par rapport aux autres, mais plutôt de le comparer à lui-même, permettant ainsi de bien suivre les progrès de l'enfant et de réajuster l'intervention. Bagnato et al. (2010) citent d'ailleurs spécifiquement l'AEPS (Bricker, 2002) comme l'un des meilleurs outils d'évaluation authentique validé permettant d'avoir un portrait des forces et des besoins de l'enfant facilement transférables en objectifs d'intervention.

L'AEPS (Bricker, 2002) et l'EIS (Bricker, 2006)

L'AEPS a été développé afin de fournir une solution innovatrice aux tests conventionnels tout en répondant aux caractéristiques des outils d'évaluation authentique. En 1980, une première édition de l'outil voit le jour. En 2002, l'édition révisée est publiée, l'AEPS (Bricker, 2002). De 2004 à 2006, la traduction francophone est réalisée et publiée, l'EIS (Bricker, 2006). Le processus de validation de l'EIS a été effectué par un comité regroupant deux spécialistes du développement et un traducteur. Chaque item a été traduit individuellement par chacune des personnes. La mise en commun des traductions individuelles a été réalisée en comité et en présence de traductions différentes, le choix de la traduction retenue a été fait par consensus. Une adaptation linguistique a dû être réalisée par la création de nouveaux items appropriés à la langue française pour le domaine de la communication. Ces items ont été identifiés en respectant les critères d'élaboration des items de l'instrument.

Actuellement, dans un souci d'amélioration du contenu et de l'organisation de l'outil, une troisième édition de l'AEPS et son adaptation francophone, l'EIS, sont en préparation. Des versions expérimentales sont disponibles : AEPS (Bricker, 2012) et EIS (Bricker, 2013). L'AEPS/EIS est un programme destiné aux enfants de la naissance à six ans possédant des risques biologiques, environnementaux ou ayant des incapacités. L'instrument intègre l'évaluation, les buts et les objectifs d'intervention ainsi que le suivi des progrès de l'enfant. Il comprend un test qui mesure des habiletés fonctionnelles, une évaluation à réaliser par le parent, une représentation visuelle des progrès de l'enfant de même qu'un curriculum d'activités spécifiées en fonction de chaque but et de chaque objectif d'intervention. Les 464 items de l'AEPS/EIS se divisent en six domaines de développement : la motricité fine, la motricité globale, la communication et les domaines adaptatif, cognitif et social. Chaque domaine comprend des niveaux, par exemple la préécriture pour la motricité fine, qui eux, regroupent un ensemble de buts. Chaque but comprend un

ensemble d'objectifs. Avec l'AEPS/EIS, on peut planifier l'intervention, avoir accès à des activités associées aux objectifs du plan d'intervention individualisé (PII) ou du plan de services individualisé (PSI) et suivre la progression des apprentissages de l'enfant.

L'*Assessment, Evaluation, and Programming System for Infants and Children, Second Edition* (AEPS) (Bricker, 2002) et son adaptation francophone réalisée par Dionne en collaboration avec Tavares et Rivest, le *Programme EIS Évaluation Intervention Suivi* (Bricker, 2006) ont été développés dans cette optique.

Qualités psychométriques

Les propriétés psychométriques de l'AEPS font l'objet d'études depuis les années 1980. Ce programme est reconnu comme un outil présentant de solides qualités psychométriques sur le plan de la fidélité, de la validité et de l'utilité (Slentz, 2008). La fidélité interjuges est jugée d'adéquade à bonne. Les coefficients de corrélation de Pearson varient globalement dans les recherches de 0,60 à 0,97 (Bailey et Bricker, 1986; Bricker, Bailey et Slentz, 1990; Hsia, 1993; Macy et al., 2005; Noh, 2005; Slentz, 1987). Quelques études ont évalué le degré d'accord interjuges par des Kappa de Cohen; ces derniers varient de 0,60 à 0,94 (Macy, 2005; Noh, 2005). Les résultats de Grisham-Brown, Hallam et Pretti-Frontczak (2008) indiquent un degré d'accord interjuges moyen de 87 %. Bricker et al. (1990) ont évalué la fidélité test-retest; le coefficient de corrélation de Pearson pour l'ensemble du test est de 0,96. La cohérence interne est bonne, les coefficients de corrélation de Pearson varient de 0,91 à 0,99 (Bricker et al., 1990) et les alphas de Cronbach de 0,83 à 0,99 (Noh, 2005).

Sur le plan de la validité convergente, les résultats de l'AEPS sont corrélés positivement à ceux du *Bayley Scale of Infant Development* (Bayley, 1969), $r = 0,93$, $p < 0,001$, du *The Revised Gesell and Amatruda Developmental and Neurologic Examination* (Knobloch, Stevens et Malone, 1980), $r = 0,51$, $p < 0,001$ (Bricker et al., 1990), du *Gesell* (Bailey et Bricker, 1986), $r = 0,81$, $p < 0,001$, du *Battelle Developmental Inventory* (Newborg, Stock, Wnek, Guidubaldi et Svinicki, 1984, 1988), $r = 0,65$, $p < 0,05$, et du *Survey Form of the Vineland Adaptive Behavior Scales* (Sparrow, Balla et Cicchetti, 1984), $r = 0,52$, $p < 0,05$ (Macy et al., 2005). Plus récemment, Gao et Grisham-Brown (2011) ont mesuré la validité concomitante des domaines de l'AEPS associés au langage, à la littératie et aux prémathématiques; les résultats montrent que les corrélations de Pearson sont positives avec le domaine de la communication, $r = 0,60$, $p < 0,001$, et le domaine cognitif $r = 0,57$, $p < 0,001$, du *Battelle Developmental Inventory, 2nd edition* (Newborg, 2005). Sur le plan de la sensibilité de l'instrument, les résultats de Noh (2005) et de Hsia (1993) tendent à montrer que le test AEPS est sensible aux différences de performance entre les enfants de trois ans et les enfants de cinq ans, de même qu'entre les enfants au développement typique et ceux au développement atypique. L'AEPS peut donc servir à déterminer l'admissibilité aux services spécialisés (Bricker et al., 2008; Bricker, Yovanoff, Capt et Allen, 2003; Macy et al., 2005). Finalement, ces résultats permettent d'élaborer des buts et des objectifs d'intervention de qualité (Notari et Bricker, 1990; Notari et Drinkwater, 1991; Pretti-Frontczak et Bricker, 2000).

Une des améliorations importantes qui seront apportées à la nouvelle édition est la représentation adéquate de la littératie et de la numératie en les considérant comme des domaines à part entière. Dans la 2^e édition (Bricker, 2002) et l'adaptation francophone (Bricker, 2006), les habiletés émergentes en littératie et en numératie sont évaluées à même les domaines suivants : motricité fine, cognitif et communication.

Définition de la littératie et de la numératie

La littératie renvoie aux connaissances, aux habiletés de base permettant d'écrire et à celles permettant de lire pour être fonctionnel en société (Legendre, 2005). La littératie peut inclure la notion de numératie et est un concept plus large que l'alphabétisation, à savoir l'enseignement de l'écriture et de la lecture (Office québécois de la langue française, 2002, 2005). Giasson (2003) propose la définition suivante de l'émergence de la littératie : « un descripteur important englobant toutes les acquisitions en lecture et en écriture [les connaissances, les habiletés et les attitudes] que l'enfant réalise, sans enseignement formel avant de lire de manière conventionnelle » (p. 128). Hawken (2009) décrit cinq composantes du développement des habiletés en lecture et en écriture pour les enfants de maternelle, ce qui donne un portrait de l'émergence de la littératie : concept lié à l'écrit/conscience de l'écrit, la correspondance entre les unités orales et les unités écrites, l'enrichissement du vocabulaire, la compréhension en lecture et finalement, l'écriture.

La numératie est associée aux connaissances et aux habiletés nécessaires pour faire face aux exigences mathématiques de divers contextes afin d'être fonctionnel en société (Organisation de coopération et de développement économiques et Statistique Canada, 2011). L'émergence de la numératie renvoie aux précurseurs du développement des mathématiques, elle ne nécessite pas d'enseignement formel des concepts abstraits en mathématiques (Purpura, 2009). L'émergence des mathématiques peut se définir par les composantes suivantes : les nombres et la notion de quantité (réciter par cœur des nombres, compter); l'arithmétique (ajouter et soustraire de petites quantités d'objets); les formes (comparer et reconnaître des formes familières); le concept spatial (utiliser une carte simple, p.ex. : lors d'une chasse au trésor); le mesurage et la géométrie (comparer des longueurs et des espaces); les schémas ainsi que la logique (faire des suites logiques d'objets) (Greenes, Ginsburg et Balfanz, 2004; Sarama et Clements, 2009).

Pertinence de la littératie et de la numératie

La pertinence d'ajouter les domaines de la littératie et de la numératie à la troisième édition de l'AEPS (Bricker, 2012) est indéniable. Finnie et Meng (2006) soulignent le lien entre un niveau d'habiletés fonctionnelles dans ces domaines, la satisfaction sociale et le bien-être économique d'une personne. Il faut savoir qu'au Canada, c'est plus de 40 % des personnes âgées de 16 à 65 ans qui n'ont pas acquis les habiletés requises en littératie leur permettant de participer activement à la société actuelle (Organisation de coopération et de développement économiques et Statistique Canada, 2011). Lors de l'entrée à l'école, c'est 25 % des enfants canadiens qui ne possèdent pas les préalables nécessaires (connaissance des

chiffres et des lettres, le vocabulaire, la capacité de distinguer des phonèmes) au développement d'habiletés en littératie et en numératie (Conseil canadien sur l'apprentissage, 2010). L'impact de posséder de faibles habiletés en littératie et en numératie peut prendre plusieurs formes : décrochage scolaire et niveau de scolarité plus faible, difficulté à accéder au marché du travail, à des emplois à temps plein et rémunération moins élevée (Canadian Literacy and Learning Network, 2012; Knighton et Bussièrre, 2006; Maxwell et Teplova, 2008; Organisation de coopération et de développement économiques et Statistique Canada, 2005, 2011; Statistique Canada, 2013).

Pour ces raisons, porter une attention particulière au développement des habiletés émergentes en littératie et en numératie est essentiel, entre autres par des programmes d'intervention précoce. En effet, les résultats de Clements et Sarama (2007) tendent à montrer que des enfants de niveau préscolaire qui participent à un programme d'intervention précoce en numératie améliorent significativement leurs habiletés émergentes dans ce domaine. Au Canada, le Programme d'aide à l'éveil à la lecture et à l'écriture dans les milieux défavorisés (PAÉLÉ) (Myre-Bisailon et al., 2010) favorise l'ouverture aux livres chez les enfants. De plus, 54 % des parents mentionnent avoir perçu une amélioration chez leur enfant sur le plan des habiletés en lecture et en écriture après sa participation aux activités du PAÉLÉ.

Considérant l'importance et la possibilité de développer les habiletés en littératie et en numératie à la petite enfance, les intervenants et les éducateurs doivent avoir accès à des outils d'évaluation valides et fidèles pour ces domaines tel que l'AEPS/EIS. Par ailleurs, la première qualité recherchée pour un instrument de mesure est la fidélité, elle réfère à la constance et à la précision des résultats obtenus (Fortin, 2010). Si les résultats obtenus ne sont pas constants d'une fois à l'autre, il est inutile d'évaluer la validité (Salvia et Ysseldyke, 2004). La fidélité peut se mesurer à travers le temps, la stabilité temporelle (test-retest), en comparant deux éditions d'une même mesure (formes parallèles), par le degré d'homogénéité des items de l'instrument (cohérence interne) ou entre les évaluateurs (Fortin, 2010). Il est important de faire la distinction entre les termes « fidélité interjuges » et « degré d'accord interjuges ». La fidélité interjuges réfère au degré de relation qui existe entre la cotation de deux évaluateurs (Tinsley et Brown, 2000). Celle-ci peut être élevée même si les chiffres absolus de cotation diffèrent d'un évaluateur à l'autre. Le degré d'accord interjuges se définit par la proportion à laquelle les différents évaluateurs accordent exactement les mêmes notes à l'objet observé (Tinsley et Brown, 2000).

L'étude qui fait l'objet du présent article s'inscrit à l'intérieur du processus d'élaboration de la 3^e édition de l'AEPS (Bricker, 2012) et de son adaptation francophone, l'EIS (Bricker, 2013). Elle vise plus particulièrement à explorer la fidélité des nouveaux domaines : la littératie (éveil à la lecture et à l'écriture) et la numératie (éveil aux mathématiques) de l'instrument de mesure de l'EIS, version expérimentale (Bricker, 2013). Le type de fidélité mesurée est le degré de concordance entre les observations de deux évaluateurs qui utilisent les grilles d'évaluation de la littératie et de la numératie. L'étude est élaborée pour répondre aux questions suivantes : 1) quel est le degré d'accord interjuges entre deux évaluateurs pour chacun des domaines : littératie et numératie? 2) quel est le degré d'accord interjuges entre

deux évaluateurs pour chacun des groupes : typique et atypique? 3) quel est le degré d'accord interjuges entre deux évaluateurs pour chacun des buts liés aux domaines : littératie et numératie?

Méthode

Participants

L'échantillon comprend 32 enfants, dont 22 ont un développement typique et 10, un développement atypique. Le premier groupe se compose de cinq enfants de prématernelle, de 13 enfants de maternelle et de quatre enfants de première année. L'âge moyen est de 5,10 ans ($\text{ÉT} = 0,68$). Les 22 enfants qui présentent un développement typique n'ont aucun retard et ne font l'objet d'aucune demande de services (évaluation ou intervention) (Noh, 2005). Ces enfants fréquentent une classe régulière d'une école du Centre-du-Québec.

Le deuxième groupe se compose d'enfants au développement atypique, dont quatre jeunes dits à risque et six, présentant des incapacités diagnostiquées. L'âge moyen des enfants de ce groupe est de 5,09 ans ($\text{ÉT} = 0,79$). Les enfants à risque sont ceux pour lesquels on soupçonne un retard de développement et qui font l'objet de demandes d'évaluation et d'intervention (Hsia, 1993). Les enfants avec des incapacités ont été évalués et sont admissibles à des services spécialisés, ont un diagnostic ou sont en attente de diagnostic (Noh, 2005). Les diagnostics rencontrés sont les suivants : trouble du spectre de l'autisme (TSA), déficience intellectuelle (DI), trouble du langage, trouble du déficit de l'attention avec ou sans hyperactivité, retard de développement global. Les enfants de ce groupe sont recrutés via le Centre de réadaptation en déficience intellectuelle et en troubles envahissants du développement de la Mauricie et du Centre-du-Québec Institut universitaire (CRDITED MCQ IU) et une école du Centre-du-Québec.

Cette étude est approuvée par le Comité d'éthique de la recherche conjoint destiné aux centres de réadaptation en déficience intellectuelle et en troubles envahissants du développement (CÉRC/CRDITED) ainsi que par le Comité d'éthique de la recherche de l'Université du Québec à Trois-Rivières.

Instruments de mesure

Les instruments de mesure utilisés sont les grilles d'évaluation (Bricker, 2013), *Domaine de la littératie* et *Domaine de la numératie et des mathématiques*, traduites en français de l'AEPS, Experimental Version (1.0) (Bricker, 2012). Le *Domaine de la littératie* est composé de 39 items (11 buts et 28 objectifs) répartis en cinq niveaux. Ces cinq niveaux sont « Concepts de l'écrit imprimé », « Conscience phonologique », « Connaissance de l'alphabet », « Vocabulaire et compréhension » et « Écriture ». Le *Domaine de la numératie et des mathématiques* comprend 30 items (10 buts et 20 objectifs) répartis en trois niveaux qui sont « Compter », « Relations quantitatives » et « Relations spatiales ». Pour chacun des items des grilles d'évaluation, trois cotations sont possibles : (0) *aucune performance*, (1) *performance émergente* et (2) *performance maîtrisée*. La cotation *aucune performance* signifie que l'enfant ne démontre pas l'habileté selon les critères, et ce, même s'il reçoit de l'aide. Une performance en émergence indique que les

critères de l'item sont atteints en partie ou que l'enfant a eu besoin d'aide pour démontrer l'habileté attendue. Finalement, une performance qui est maîtrisée sous-entend que l'habileté est réalisée par l'enfant telle que spécifiée dans le critère, et ce, sans aide de l'adulte.

Déroulement

Le recrutement est réalisé auprès de deux organisations, soit une commission scolaire du Centre-du-Québec et le CRDITED MCQ IU. La majorité des enfants est évaluée (28/32) en milieu scolaire, un environnement naturel et familial pour eux. Deux enfants sont vus en milieu familial. Un enfant est évalué au CRDITED MCQ IU et un autre à l'UQTR, à la demande des parents. La chercheuse principale présente un scénario d'activités à l'enfant qui a pour thème l'organisation d'un magasin de jeux d'école. L'enfant est invité à devenir un employé de ce magasin et à tester des jeux scolaires. Chacun des jeux permet d'évaluer une ou des habiletés en littératie et en numératie. Le scénario d'activités élaboré pour la présente étude a été testé en préexpérimentation auprès de six enfants de quatre à six ans au développement typique et atypique. Toutes les préexpérimentations et les évaluations ont été filmées. Les évaluations ont été réalisées par la chercheuse principale; les enfants sont vus deux fois 30 minutes pour un total d'une heure, et ce, dans un intervalle de temps inférieur à deux semaines. Comme la durée des évaluations est courte et que l'ordre de présentation des jeux peut différer d'un enfant à l'autre, la concentration et l'intérêt des enfants demeurent.

Les films sont visionnés par trois évaluateurs formés à l'utilisation de l'AEPS/EIS, dont la chercheuse principale. Les trois évaluateurs ont préalablement visionné et coté les vidéos des préexpérimentations. Les évaluateurs ont comparé leurs cotations et se sont assurés d'une compréhension commune des buts et des critères afin d'atteindre un degré d'accord satisfaisant, supérieur à 80 % (Bachelor et Joshi, 1986). Ils ont également précisé certains critères entre eux. À titre d'exemple, en littératie, pour le but B1 « Identifie les mots qui riment », les évaluateurs se sont entendus pour que l'enfant doive identifier « deux » paires de mots afin de considérer la performance comme maîtrisée. Le degré d'accord interjuges de la préexpérimentation se situe au-dessus de 80 %; il a été calculé en multipliant par 100 le nombre d'accords divisé par le nombre d'accords additionné au nombre de désaccords (Salvia et Ysseldyke, 2004).

Résultats

Afin de calculer le degré d'accord interjuges, un pairage de deux évaluateurs pour chacun des enfants évalués est réalisé avec les trois évaluateurs (A, B et C). La paire A-B ainsi que la paire B-C évaluent chacune 11 enfants et la paire d'évaluateurs A-C évalue 10 enfants. Le degré d'accord interjuges est calculé entre deux évaluateurs indépendants avec trois catégories de choix de réponses : (0) *aucune performance*, (1) *performance émergente* et (2) *performance maîtrisée*. Les données manquantes sont inexistantes, car si la compétence n'est pas observée, elle est cotée (0) *aucune performance*. Le coefficient Kappa de Fleiss (Fleiss, 1971) est l'indice statistique approprié; il permet d'évaluer le degré d'accord interjuges en présence d'un nombre d'évaluateurs fixe et supérieur à deux, trois évaluateurs pour la

présente étude. Il calcule le degré de concordance de classification par rapport à ce qui pourrait être attendu si elle était faite au hasard (Gwet, 2012). Comparativement au Kappa de Cohen (1960) qui considère que les deux mêmes évaluateurs classent un ensemble de sujets, le Kappa de Fleiss prend en compte que différents sujets peuvent être classés par différentes paires d'évaluateurs. Il faut souligner que le Kappa de Fleiss ne prend pas en considération la nature hiérarchique des données ordinales (Gwet, 2012). Aux fins de l'étude, bien que les choix de réponse soient des données ordinales, elles sont considérées mutuellement exclusives, à savoir qu'une habileté peut être classée par les évaluateurs dans une seule catégorie. Chacun des désaccords est considéré comme ayant la même valeur; un désaccord entre (0) *aucune performance* et (1) *performance émergente* a la même valeur qu'un désaccord entre (0) *aucune performance* et (2) *performance maîtrisée*. La notion d'ordre n'est pas prise en compte, d'où le choix de cet indice statistique. L'interprétation des Kappas est faite selon Emam (1999), soit accord faible (< 0,44), accord modéré (0,44, 0,62), accord substantiel (0,62, 0,78), accord excellent (> 0,78).

Degrés d'accord interjuges (littératie et numératie, typique et atypique)

Afin de répondre aux trois questions de recherche, des analyses de Kappas de Fleiss (k_f) sont réalisées. Pour la première question de recherche, des résultats indiquent un accord excellent pour le domaine de la littératie, $k_f = 0,87$ et un accord substantiel pour le domaine de la numératie, $k_f = 0,73$. En ce qui a trait à la deuxième question de recherche, les résultats montrent également des accords substantiels pour les groupes atypique et typique, $k_f = 0,78$ et $k_f = 0,76$

Degré d'accord interjuges pour chacun des buts liés aux domaines, littératie et numératie

À la troisième question de recherche : quel est le degré d'accord interjuges entre deux évaluateurs pour chacun des buts liés aux domaines : littératie et numératie?, les résultats des analyses de Kappa de Fleiss (Tableau 1) montrent des accords excellents qui varient de 0,80 à 1,00 pour six des onze buts du domaine de la littératie.

Entre autres, un accord excellent est associé au but A1 « Participe à la lecture partagée » du niveau « Concepts de l'écrit imprimé », $k_f = 1,00$. Par ailleurs, deux buts du niveau « Conscience phonologique » présentent un degré d'accord excellent, soit le but B1 « Identifie les mots qui riment », $k_f = 0,80$ et le but B2 « Segmente des mots de deux et trois syllabes », $k_f = 0,88$.

Toutefois, des accords modérés sont présents pour les buts B3 « Segmente des mots consonne-voyelle-consonne (CVC) en sons » du niveau « Conscience phonologique », $k_f = 0,62$ et C1 « Nomme toutes les lettres en majuscule et en minuscule de l'alphabet » du niveau « Connaissance de l'alphabet », $k_f = 0,56$. L'analyse des résultats montre un accord faible pour le but E1 « Écrit des semblants de lettres » du niveau « Écriture », tel que l'indique le Tableau 1.

**Tableau 1. Degrés d'accord interjuges en fonction des buts de la littératie
(N = 32)**

Buts du domaine de la littératie par niveau	Kappa de Fleiss
Niveau A. Concepts de l'écrit imprimé	
A1. Participe à la lecture partagée avec un adulte ou un enfant plus âgé.	1,00
A2. Démontre une compréhension qu'un texte se lit dans un sens, et de haut en bas de la page.	0,82
Niveau B. Conscience phonologique	
B1. Identifie les mots qui riment.	0,80
B2. Segmente des mots de deux et trois syllabes.	0,88
B3. Segmente des mots consonne-voyelle-consonne (CVC) en sons.	0,62
Niveau C. Connaissance de l'alphabet	
C1. Nomme toutes les lettres en majuscule et en minuscule de l'alphabet.	0,56
C2. Reconnaît visuellement un texte de mots consonne-voyelle-consonne (CVC).	0,91
Niveau D. Vocabulaire et compréhension	
D1. Raconte l'histoire ou un événement simple.	0,76
D2. Associe des mots d'imprimerie à des personnes et à des objets communs ou familiers.	0,73
Niveau E. Écriture	
E1. Écrit des semblants de lettres.	0,27
E2. Écrit ou copie des mots utilisant l'orthographe conventionnelle ou approchée.	0,94
Total	0,82

En ce qui concerne le domaine de la numératie, les résultats des analyses de Kappas de Fleiss en fonction des buts (Tableau 2) indiquent des accords excellents pour trois buts du niveau « Compter », soit A3 « Démontre qu'il comprend les caractères numériques », $k_f = 0,87$, A5 « Compte jusqu'à un nombre donné ou au-delà », $k_f = 0,95$ et A6 « Résout des problèmes impliquant des nombres en comptant ou en comptant à rebours pour des sommes allant jusqu'à 20 », $k_f = 0,82$.

Quatre buts (A1, A2, A4, B2) présentent un accord substantiel, les Kappas de Fleiss variant de 0,63 à 0,77. Les résultats indiquent un accord modéré pour le but B1 « Compare des ensembles d'items en comptant » du niveau « Relations quantitatives », $k_f = 0,61$. Le but C2 « Démontre une compréhension de l'orientation spatiale en créant des cartes, des diagrammes, et des représentations visuelles »

Tableau 2. Degrés d'accord interjuges en fonction des buts de la numératie (N = 32)

Buts du domaine de la numératie par niveau	Kappa de Fleiss
Niveau A. Compter	
A1. De mémoire, récite un à un les nombres de 1 à 30.	0,70
A2. Détermine le nombre d'items dans un ensemble de 4 à 5 items.	0,77
A3. Démontre qu'il comprend les caractères numériques.	0,87
A4. Effectue la correspondance un à un pour compter de petites quantités.	0,67
A5. Compte jusqu'à un nombre donné ou au-delà.	0,95
A6. Résout des problèmes impliquant des nombres en comptant ou en comptant à rebours pour des sommes allant jusqu'à 20.	0,82
Niveau B. Relations quantitatives	
B1. Compare des ensembles d'items en comptant.	0,61
B2. Organise des groupes d'items selon une séquence respectant une règle d'organisation.	0,63
Relations spatiales	
C1. Comprend et utilise des mots reliés à la direction et à la distance.	-0,02
C2. Démontre une compréhension de l'orientation spatiale en créant des cartes, des diagrammes, et des représentations visuelles.	0,58
Total	0,73

du niveau « Relations spatiales » présente également un accord modéré à $k_f = 0,58$. Quant au but C1 « Comprend et utilise des mots reliés à la direction et à la distance », le Kappa de Fleiss est de $-0,02$.

Discussion

Tout d'abord, la discussion des résultats à l'aide des études antérieures (Hsia, 1993; Macy et al., 2005; Noh, 2005) est réalisée en regard des domaines cognitif et de la communication. En effet, la littératie et la numératie sont présentement en élaboration et ils sont inexistantes dans les versions antérieures de l'AEPS/EIS. De plus, le choix de ces domaines (cognitif et communication) s'explique entre autres par le fait que la majorité des items de la littératie et de la numératie se retrouvent intégralement ou en partie à l'intérieur du domaine cognitif des éditions précédentes de l'AEPS/EIS. Également, ces domaines sont ceux qui

mesurent le plus l'émergence de la littératie et des prémathématiques (Gao et Grisham-Brown, 2011).

Degrés d'accord interjuges (littératie et numératie, typique et atypique)

En regard de la première question de recherche, les résultats indiquent, pour le domaine de la littératie, un degré d'accord excellent et plus élevé que celui de la numératie, dont le degré d'accord est substantiel. La différence entre les deux domaines quant au degré d'accord peut s'expliquer par un manque de précision et de clarté dans la formulation de certains buts ou critères d'un des domaines (Noh, 2005); dans le cas présent, la numératie. À titre d'exemple, pour le but C2 « Démontre une compréhension de l'orientation spatiale en créant des cartes, des diagrammes, et des représentations visuelles », le nombre d'éléments de localisation que l'enfant doit incorporer dans son dessin pour que l'habileté soit maîtrisée n'est pas spécifié. À la suite de l'expérimentation, il semble qu'un minimum de trois éléments différents devrait être visible dans le dessin, dont des flèches indiquant la direction et des représentations visuelles permettant de se repérer dans l'environnement (chemin, arbre, maison, porte, escalier). Tel que le soulignent Hsia (1993) et Noh (2005), l'absence de critères clairs oblige l'évaluateur à un jugement subjectif, ce qui peut expliquer un plus grand nombre de divergences dans les cotations. Par ailleurs, les degrés d'accord pour chacun des domaines, littératie et numératie, sont supérieurs à ceux de Noh (2005). Ces résultats plus élevés peuvent s'expliquer par la différence de taille entre l'échantillon de la présente étude ($N = 32$) et celui de Noh (2005) ($N = 15$).

Puis, les résultats de Macy et al. (2005) ($N = 69$) et de Hsia (1993) ($N = 82$) montrent des degrés d'accord plus élevés que les résultats de la présente étude pour chacun des domaines, la littératie et la numératie. La taille des échantillons peut expliquer les niveaux d'accord plus faibles de la présente étude. Un autre élément qui peut avoir influencé positivement les degrés d'accord est l'utilisation d'une cotation à deux catégories (réussi ou échoué) par Hsia (1993). Une cotation à trois catégories, comme dans la présente étude, vient augmenter les risques de désaccord entre les évaluateurs (Noh, 2005).

En ce qui concerne la deuxième question de recherche, les résultats montrent un accord substantiel pour le groupe d'enfants présentant un développement typique; l'accord est également substantiel pour le groupe d'enfants présentant un développement atypique, à peine plus élevé. Un niveau d'accord plus élevé est présent pour le groupe d'enfants présentant un développement atypique dans les résultats de Noh (2005). Il suggère que les enfants présentant un développement atypique possèdent un éventail d'habiletés plus restreint, moins complexe, donc plus facile à observer.

Degré d'accord interjuges pour chacun des buts liés aux domaines, de la littératie et de la numératie

Pour le domaine de la littératie, les résultats indiquent des degrés d'accord excellents (A1, A2, B1, B2, C2 et E2), des accords substantiels (D1 et D2) et des accords modérés (B3 et C1). Un seul accord est faible, celui relié au but E1. En ce

qui concerne la numératie, les résultats montrent des degrés d'accord excellents (A3, A5 et A6), des accords substantiels (A1, A2, A4 et B2) et des accords modérés (B1 et C2). Un résultat montre une absence d'accord pour le but C1.

De façon générale, pour les deux domaines, les degrés d'accord élevés peuvent être attribués à la mesurabilité et à l'observabilité des buts, une force reconnue de l'AEPS (Slentz, 2008). La clarté et la précision des buts et de leurs critères peuvent également expliquer des niveaux d'accords supérieurs (Hsia, 1993; Noh, 2005). De plus, l'atteinte d'un degré d'accord de 80 % entre les évaluateurs lors de la préexpérimentation, tel que recommandé dans Bachelor et Joshi (1986), a probablement permis une compréhension commune des buts et de leurs critères, ce qui pourrait aussi expliquer les accords excellents ou substantiels. Lors de la préexpérimentation, les évaluateurs se sont entendus pour préciser certains critères. Ces précisions ont possiblement permis davantage de constance entre les cotations des évaluateurs. À l'inverse, l'absence de critères précis permettant de définir une performance en émergence pourrait expliquer les accords plus faibles. D'ailleurs, l'absence de critères clairs rend la cotation plus subjective, ce qui peut augmenter les désaccords (Noh, 2005).

Plus spécifiquement pour le domaine de la littératie, le degré d'accord le plus élevé de l'étude est attribué au but A1 « Participe à la lecture partagée avec un adulte ou un enfant plus âgé ». On conviendra ici que ce degré d'accord élevé peut s'expliquer par le fait que ce but est facile à évaluer vu sa simplicité. De plus, l'habileté de participation à une lecture partagée fait partie du répertoire développemental des enfants de la naissance à trois ans (Bricker, 2006). Les participants de l'étude ont plus de quatre ans et maîtrisent d'ailleurs presque tous cette habileté. Par conséquent, la cotation est plus facile pour les évaluateurs ce qui pourrait expliquer ce résultat. Pour l'accord modéré du but B3 « Segmente des mots consonne-voyelle-consonne CVC », la segmentation d'un mot en phonème est une habileté qui s'acquiert en première année du primaire (Hawken, 2009). Seuls quatre enfants de l'échantillon sont en première année et, en conséquence, un nombre très restreint d'enfants maîtrisent cette habileté. Ainsi, les évaluateurs ont dû coter ce but, performance émergente, pour un plus grand nombre d'enfants. Étant donné l'absence de critères clairs pour délimiter la performance en émergence de ce but, le risque de désaccord était plus élevé et pourrait être une explication de ce résultat modéré. Afin de pallier cette lacune, il pourrait être spécifié que lorsque l'enfant fait un ou deux sons distincts sur trois, la performance est cotée en émergence; p.ex : l'enfant fait le son « s » ou le son « a » du mot « sac ».

Quant au but E1 « Écrit des semblants de lettre », la constance des observations entre les évaluateurs est faible. Premièrement, l'exemple associé à ce but montre un dessin de semblant de lettres pouvant représenter une performance maîtrisée. Cet exemple ne montre pas de dessin pour une performance émergente ou pour l'absence de performance. Pourtant, de telles représentations graphiques pouvant servir à baliser l'évolution de l'écriture sont présentées dans Giasson (2003). Lorsque les évaluateurs réalisent les cotations de façon subjective, les risques de désaccord sont évidemment plus grands (Hsia, 1993), ce qui pourrait expliquer le faible accord. Deuxièmement, lors de l'expérimentation, le but E1 est évalué en même temps que le but E2 « Écrit ou copie des mots en utilisant l'orthographe

conventionnelle ou inventée » dans un jeu qui vise à reproduire une affiche de magasin scolaire. L'ajout au scénario d'un jeu comme écrire une carte d'anniversaire ou un journal scolaire tel que présenté dans Notari-Syverson, O'Connor et Vadasy (2007) aurait pu permettre de mieux observer les critères du but E1.

En ce qui concerne le domaine de la numératie, le but A5 « Compte jusqu'à un nombre donné ou au-delà » présente le degré d'accord le plus élevé de ce domaine. Comme dans les études de Hsia (1993) et de Noh (2005), la clarté et la précision dans la description du but semblent expliquer ce résultat. En ce sens, le but A5 est bien illustré par l'exemple : l'enseignant dit ou indique, « Donne-m'en 8 », et l'enfant en compte 8. Lors de la collecte de données, cet exemple a été reproduit lors des jeux avec différents objets communs : « Donne-moi 8 blocs de bois ». Pour les évaluateurs, le lien entre l'expérimentation et la grille d'évaluation est très clair, ce qui pourrait expliquer le degré d'accord élevé.

Finalement, le résultat pour le but C1 « Comprend et utilise des mots reliés à la direction et la distance » pourrait s'expliquer par la difficulté à réaliser la cotation, due à la complexité de ce but et de son critère qui est :

L'enfant répond à des questions, fait des commentaires, suit des consignes ou répond en utilisant des mots reliés à la direction (p. ex., dessus, dessous, sur, sous, de ce côté, gauche, droite) ou à la distance (p. ex., près, loin). (Bricker, 2013).

Plusieurs habiletés composent ce critère à savoir : répondre à des questions, faire des commentaires et suivre des consignes. Il pourrait être envisagé d'évaluer ces habiletés séparément. Une trajectoire d'apprentissage des prémathématiques telle que celle décrite par Sarama et Clements (2009) pourrait servir à baliser les buts et les critères ainsi qu'à bien définir ce qui différencie les cotes : (0) *aucune performance*, (1) *performance émergente* et (2) *performance maîtrisée*. Selon Sarama et Clements (2009), l'acquisition de ces habiletés et de ces notions (comprendre et utiliser des mots reliés à la direction ou à la distance) suit une séquence développementale qui s'échelonne de la naissance à six ans. À titre d'exemple, les premiers termes spatiaux utilisés par les enfants sont les notions « sous », « sur », et c'est seulement vers six ans que les notions « droite » et « gauche » sont acquises (Sarama et Clements, 2009).

Limites de l'étude

Cette étude présente quelques limites. La taille restreinte de l'échantillon ainsi que la provenance des participants sont deux limites à considérer. La grande majorité des participants demeurent dans la même région et fréquentent la même école. Un échantillon plus grand et plus étendu en âge pourrait permettre d'amoinrir l'effet plafond présent pour le but A1 « Participe à la lecture partagée avec un adulte ou un enfant plus âgé » et celui de plancher pour le but B3 « Segmente des mots consonne-voyelle-consonne CVC ». Il faut également mentionner l'absence d'activité permettant d'évaluer de façon systématique le but E1 « Écrit des semblants de lettre ». Ce manque donne moins d'opportunités aux enfants de démontrer cette habileté affectant probablement la constance des résultats entre les évaluateurs.

Pour terminer, bien que les résultats sont forts utiles dans une visée d'élaboration et d'amélioration de l'outil, leur portée s'avère limitée considérant que seulement deux domaines ont été examinés sous l'angle d'une forme de fidélité, l'accord interjuges.

Conclusion

Cette étude a permis de cibler les buts pour lesquels des modifications semblent nécessaires afin d'accroître la fidélité de l'outil. Les informations recueillies serviront à améliorer les items des nouveaux domaines, la littératie et la numératie, de l'AEPS (Bricker, 2012)/EIS (Bricker, 2013). En ce sens, il apparaît pertinent d'ajouter les représentations graphiques de Giasson (2003) pour le but E1 « Écrit des semblants de lettre » afin de baliser les niveaux de maîtrise de l'habileté. De plus, le but C2 de la numératie « Démontre une compréhension de l'orientation spatiale en créant des cartes, des diagrammes, et des représentations visuelles » serait aussi accompagné de dessins qui en faciliteraient la cotation.

L'utilisation de l'AEPS/EIS pour l'évaluation d'enfants qui présentent des indices de retard de développement est recommandée par l'*Ordre des psychoéducateurs et psychoéducatrices du Québec* (OPPQ) (Auclair-Julien et al., 2013). Cette étude est une première étape pouvant appuyer le choix des intervenants d'utiliser la troisième version de l'AEPS et son adaptation francophone, l'EIS, afin d'évaluer les habiletés émergentes en littératie et en numératie auprès d'enfants dont le développement est typique ou atypique. D'autres étapes demeurent nécessaires afin de poursuivre le travail de validation de ces nouveaux domaines tels qu'évaluer d'autres formes de fidélité (test-retest, cohérence interne) et des aspects de validité (validité convergente). La reproduction de l'étude avec un échantillon plus grand, tout en intégrant les autres domaines de la version expérimentale, s'avère une suite logique afin de poursuivre la validation. L'évaluation à l'aide d'un modèle statistique de la théorie de réponse aux items se présente comme une autre avenue à considérer.

Les intervenants, dont les psychoéducateurs et les psychoéducatrices du Québec, doivent avoir accès à des outils d'évaluation du développement de l'enfant, valides et fidèles, qui soutiennent véritablement l'intervention. L'identification des bonnes cibles d'intervention est une étape primordiale ayant des répercussions sur l'ensemble du processus d'intervention. L'ajustement des stratégies d'intervention en fonction des besoins spécifiques d'un enfant à un moment précis de son développement oblige une connaissance approfondie du fonctionnement de celui-ci. Ainsi, les efforts investis pour documenter les interactions de l'enfant avec son environnement plutôt qu'une performance unique en situation d'évaluation se révèlent une opération à valeur ajoutée.

Références

- Auclair-Julien, M.-C., Lanovaz, M., Paquette, C., Pelletier, S., Thermidor, G. et Trudel, D. (2013). *Évaluation du retard de développement : lignes directrices*. Montréal, Québec : Ordre des psychoéducateurs et psychoéducatrices du Québec.
- Bachelor, A. et Joshi, P. (1986). *La méthode phénoménologique de recherche en psychologie : guide pratique*. Québec, Québec : Presses de l'Université Laval.
- Bagnato, S. J., Neisworth, J. T. et Pretti-Frontczak, K. (2010). *Linking authentic assessment & early childhood intervention best measure for best practices* (2^e éd.). Baltimore, MD: Paul H. Brookes Publishing Co.
- Bailey, E. J. et Bricker, D. (1986). A psychometric study of a criterion-referenced assessment instrument designed for infants and young children. *Journal of Early Intervention*, 10(2), 124-134. doi: 10.1177/105381518601000204
- Bayley, N. (1969). *Bayley scales of infant and toddler development*. San Antonio, TX: Psychological Corporation.
- Bricker, D. D. (2002). *AEPS Assessment evaluation and programming system for infants and children* (2^e éd.). Baltimore, MD: Paul H. Brookes Publishing Co.
- Bricker, D. D. (2006). *Programme EIS évaluation intervention suivi*. Traduction et adaptation auprès d'une clientèle québécoise sous la direction de C. Dionne en collaboration avec C.-A. Tavarès et C. Rivest. Montréal, Québec : Chenelière éducation.
- Bricker, D. D. (2012). *Assessment, evaluation, and programming system for infants and children (AEPS®): Experimental Version (1.0)*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Bricker, D. D. (2013). *Programme EIS évaluation, intervention, suivi : version expérimentale*. Traduction et adaptation auprès d'une clientèle québécoise sous la direction de C. Dionne. Document inédit.
- Bricker, D. D., Bailey, E. J. et Slentz, K. (1990). Reliability, validity, and utility of the evaluation and programming system: For infants and young children (EPS-I). *Journal of Early Intervention*, 14(2), 147-160. doi: 10.1177/105381519001400204
- Bricker, D. D., Clifford, J., Yovanoff, P., Pretti-Frontczak, K., Waddell, M., Allen, D. et Hoselton, R. (2008). Eligibility determination using a curriculum-based assessment: A further examination. *Journal of Early Intervention*, 31(1), 3-21. doi: 10.1177/1053815108324422
- Bricker, D. D., Yovanoff, P., Capt, B. et Allen, D. (2003). Use of a curriculum-based measure to corroborate eligibility decisions. *Journal of Early Intervention*, 26(1), 20-30. doi: 10.1177/105381510302600102
- Canadian Literacy and Learning Network. (2012). *Citizenship. Literacy and essential skills fact sheet*. Repéré à <http://www.literacy.ca/content/uploads/2012/02/citizenship.pdf>
- Chaytor, N. et Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests : A review of the literature on everyday cognitive skills. *Neuropsychology review*, 13(4), 181-197. doi: 10.1023/B:NERV.0000009483.91468.fb
- Clements, D. H. et Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. *Journal for Research in Mathematics Education*, 38(2), 136-163.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. doi: 10.1177/001316446002000104
- Conseil canadien sur l'apprentissage. (2010). *Bilan de l'apprentissage tout au long de la vie au Canada : progrès ou excès de confiance?* Ottawa, Ontario : Conseil canadien sur l'apprentissage.

- Emam, K. E. (1999). Benchmarking Kappa: Interrater agreement in software process assessments. *Empirical Software Engineering*, 4(2), 113-133. doi : 10.1023/A:1009820201126
- Finnie, R. et Meng, R. (2006). *Importance de la littératie fonctionnelle : compétences en lecture et en mathématiques et résultats sur le marché du travail des décrocheurs du secondaire*. Ottawa, Ontario: Statistique Canada.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382. doi : 10.1037/h0031619
- Fortin, M.-F. (2010). *Fondements et étapes du processus de recherche. Méthode quantitatives et qualitatives* (2^e éd.). Montréal, Québec : Chenelière Éducation inc.
- Gao, X. (2008). *Young children's accountability data on language, literacy and pre-math areas: Validating authentic assessment* (Thèse de doctorat, University of Kentucky, KY). Repéré à <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-99050-415&site=ehost-live>
- Gao, X. et Grisham-Brown, J. (2011). The use of authentic assessment to report accountability data on young children's language, literacy and pre-math competency. *International Education Studies*, 4(2), 42-53. doi: 1005539/ies.v4n2p41
- Giasson, J. (2003). *La lecture : de la théorie à la pratique* (2^e éd.). Boucherville, Québec : Gaëtan Morin éditeur.
- Greenes, C., Ginsburg, H. P. et Balfanz, R. (2004). Big Math for Little Kids. *Early Childhood Research Quarterly*, 19(1), 159-166. doi: 10.1016/j.ecresq.2004.01.010
- Grisham-Brown, J., Hallam, R. A. et Pretti-Fontczak, K. (2008). Preparing head start personnel to use a curriculum-based assessment: An innovative practice in the 'age of accountability'. *Journal of Early Intervention*, 30(4), 271-281. doi : 10.1177/1053815108320689
- Gwet, K. L. (2012). *Handbook of inter-rater reliability* (3^e éd.). Gaithersburg, MD: Advanced Analytics, LLC.
- Hawken, J. (2009). Pour un enseignement efficace de lecture et de l'écriture : une trousse d'intervention appuyée sur la recherche. Repéré à http://www.treaqfp.qc.ca/106/PDF/TROUSSE_Reseau_de_chercheurs.pdf
- Hsia, T.-H. (1993). *Evaluating the psychometric properties of the assessment, evaluation, and programming system for 3 to 6 years: Aeps test 3 to 6 years (aeps test)*. (Thèse de doctorat, Université of Oregon, OR). Repéré à <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1995-74208001&site=ehost-live>
- Knighton, T. et Bussière, P. (2006). *Liens entre les résultats éducationnels à l'âge de 19 ans et la capacité en lecture à l'âge de 15 ans*. Ottawa, Ontario : Statistique Canada.
- Knobloch, H., Stevens, F. et Malone, A. (1980). *Manual of developmental diagnosis: The administration an interpretation of the revised Gesell and Armatruda developmental and neurologic examination*. Hagerstown, MD: Harper & Row.
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3^e éd.). Montréal, Québec : Guérin.
- Macy, M. G., Bricker, D. D. et Squires, J. K. (2005). Validity and reliability of a curriculum-based assessment approach to determine eligibility for part C services. *Journal of Early Intervention*, 28(1), 1-16. doi : 10.1177/105381510502800101
- Maxwell, J. et Teplova, T. (2008). Conséquence sociale des lacunes en matière d'habiletés de langage et de littératie. *Encyclopedia of Language and Literacy Development*. Repéré à <http://www.literacyencyclopedia.ca/pdfs/topic.php?topId=34>

- Myre-Bisaillon, J., Villemagne, C., Puentes-Neuman, G., Dionne, C., Raïche, G. et Louis, R. (2010). *Évaluation des impacts du PAÉLÉ dans les milieux défavorisés : rapport scientifique intégral* (Rapport no 2007-EL-118328). Sherbrooke, Québec : Université de Sherbrooke.
- Newborg, J. (2005). *Battelle Development Inventory* (2^e éd.). Rolling Meadow, IL: Riverside Publishing.
- Newborg, J., Stock, J., Wnek, L., Guidubaldi, J. et Svinicki, J. (1984). *Battelle developmental inventory: Examiner's manual*. Dallas, TX: DLM/Teaching Resource.
- Newborg, J., Stock, J., Wnek, L., Guidubaldi, J. et Svinicki, J. (1988). *Battelle developmental inventory: Examiner's manual*. Dallas, TX: DLM/Teaching Resource.
- Noh, J. (2005). *Examining the psychometric properties of the second edition of the assessment, evaluation, and programming system for three to six years: AEPS test 2nd Edition (3-6)*. (Thèse de doctorat, University of Oregon, OR). Repéré à <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2006-99013-033&site=ehost-live>
- Notari-Syverson, A., O'Connor, R. E. et Vadasy, P. F. (2007). *Ladders to literacy: A preschool activity book*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Notari, A. R. et Bricker, D. D. (1990). The utility of a curriculum-based assessment instrument in the development of individualized education plans for infants and young children. *Journal of Early Intervention*, 14(2), 117-132. doi: 10.1177/105381519001400202
- Notari, A. R. et Drinkwater, S. G. (1991). Best practices for writing child outcomes: An evaluation of two methods. *Topics in Early Childhood Special Education*, 11(3), 92-106. doi: 10.1177/027112149101100309
- Office québécois de la langue française. (2002). *Numératie*. Repéré à http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?ld_Fiche=8363202
- Office québécois de la langue française. (2005). *Alphabétisation*. Repéré à http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?ld_Fiche=1199241
- Organisation de coopération et de développement économiques et Statistique Canada. (2005). *Apprentissage et réussite : premiers résultats de l'enquête sur la littératie et les compétences des adultes*. Paris, France : Éditions OCDE.
- Organisation de coopération et de développement économiques et Statistique Canada. (2011). *La littératie, un atout pour la vie : nouveaux résultats de l'enquête sur la littératie et les compétences des adultes : deuxième rapport international de l'ELCA*. Paris, France : Éditions OCDE.
- Pretti-Frontczak, K. et Bricker, D. (2000). Enhancing the quality of individualized education plan (IEP) goals and objectives. *Journal of Early Intervention*, 23(2), 92-105. doi: 10.1177/105381510002300204
- Purpura, D. J. (2009). *Informal number-related mathematics skills: An examination of the structure of and relations between these skills in preschool*. (Thèse de doctorat, Florida State University, FL). Repéré à <http://diginole.lib.fsu.edu/etd/502>
- Salvia, J. et Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9^e éd.). Boston, MA: Houghton Mifflin.
- Sarama, J. et Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.
- Slentz, K. (1987). *Evaluating the instructional needs of young children with handicaps: Psychometric adequacy of the evaluation and programming system: Assessment level II (EPS-II)*. (Thèse de doctorat, University of Oregon, OR). Repéré à <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1988-52788-001&site=ehost-live>

- Slentz, K. (2008). *What AEPS® users should know about reliability, validity, and utility*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Sparrow, S. S., Balla, D. A. et Cicchetti, D. V. (1984). *Vineland adaptive behavior scales: Interview edition survey form manual*. (2^e éd.). Circle Pines, MN: American Guidance Service.
- Statistique Canada. (2013). *Étude : liens entre les résultats du parcours de vie et la capacité de lecture*. Repéré à <http://www.statcan.gc.ca/daily-quotidien/110627/dq110627b-fra.htm>
- Thurman, S. K. et McGrath, M. C. (2008). Environmentally based assessment practices: Viable alternatives to standardized assessment for assessing emergent literacy skills in young children. *Reading & Writing Quarterly*, 24(1), 7-24. doi: 10.1080/10573560701753021
- Tinsley, H. E. A. et Brown, S. D. (2000). *Handbook of applied multivariate statistics and mathematical modeling*. San Diego, CA: Academic Press.