

REVUE SYSTÉMATIQUE

Les outils d'évaluation du développement des enfants de 2 à 5 ans

Sean Gayadeen M.A.P., agent de planification, de programmation et de recherche, CRDITED MCQ - Institut universitaire
Marie-Claude Lehoux, M. Sc., agente de planification, de programmation et de recherche, CRDITED MCQ - Institut universitaire

GUIDE DE PRATIQUE

L'intervention précoce auprès des
enfants de 2 à 5 ans présentant un
retard global de développement

Édition

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)
1^{re} édition – Janvier 2015

Analyse et rédaction :

Sean Gayadeen M.A.P., agent de planification, de programmation et de recherche, CRDITED MCQ – IU.

Marie-Claude Lehoux, M. Sc., agente de planification, de programmation et de recherche, CRDITED MCQ – IU.

Revue systématique sous la direction de :

Marlène Galdin, Ph. D., directrice de la qualité, de la recherche, du développement et de l'innovation, CRDITED MCQ – IU.

Chargée de projet du guide de pratique :

Joanne Larose, conseillère-cadre aux programmes et au développement des pratiques, CRDITED de la Montérégie-Est.

Correction

Nancy Bussière, agente administrative, CRDITED MCQ – IU

Marie-Ève Carpentier, adjointe de direction, CRDITED MCQ – IU

Mise en page

Productions Graphiques Michel Lamothe

Afin de faciliter la lecture du document, un seul genre est utilisé et désigne tant le féminin que le masculin.

La reproduction partielle ou complète de ce document à des fins personnelles et non commerciales est permise, à la condition d'en citer la source.

Il est recommandé de citer le document de cette façon :

CRDITED MCQ – IU. Revue systématique – Les outils d'évaluation du développement des enfants de 2 à 5 ans, Trois-Rivières, Centre de réadaptation en déficience intellectuelle et en troubles envahissants du développement de la Mauricie et du Centre-du-Québec – Institut universitaire, 2015.

© CRDITED MCQ – IU

Dépôt Légal

ISBN (version électronique) : 978-2-922227-62-8

COLLABORATEURS

Contribution à la stratégie de recherche des instruments de mesure :

Anna Gagnon, technicienne en documentation, CRDITED de Montréal.

Christine Rivest, agente de planification, de programmation et de recherche, CRDITED MCQ – IU.

Contribution à la sélection des instruments de mesure :

Marjolaine Hébert, spécialiste en réadaptation psychosociale, CRDITED Pavillon du Parc.

Katleen Picard, ergothérapeute, CRDITED de Montréal.

Annick Rajotte, chef en Réadaptions, Services aux enfants et à leur famille, équipe RGD-DI, Centre de réadaptation de l'Ouest de Montréal (CROM).

Christine Rivest, agente de planification, de programmation et de recherche, CRDITED MCQ – IU.

Révision scientifique de la revue systématique :

Francine Julien-Gauthier, Ph. D., professeure agrégée, Université Laval, faculté des sciences de l'éducation, département des fondements et pratiques en éducation, programme de psychoéducation.

Louis Laurencelle, professeur associé, Université du Québec à Trois-Rivières, département des sciences de l'activité physique.

Myriam Rousseau, Ph. D., praticienne-chercheuse, CRDITED MCQ – IU.

Comité scientifique du Guide de pratique (en ordre alphabétique) :

Anne-Marie Cloutier, conseillère à la Direction de l'organisation des services, des affaires médicales et universitaires. Association québécoise des établissements de santé et des services sociaux (AQESSS).

Dr Pierre Dagenais, MD, Ph. D., directeur du soutien à la qualité et à la méthode de l'Institut national d'excellence en santé et en services sociaux (INESSS).

Anna Gagnon, technicienne en documentation, CRDITED de Montréal.

Marlène Galdin, Ph. D., directrice de la qualité, de la recherche, du développement et de l'innovation, CRDITED MCQ – IU.

Joanne Larose, conseillère-cadre aux programmes et au développement des pratiques, CRDITED de la Montérégie-Est.

Anick Larose, représentante des parents. Conseillère aux communications, Association québécoise pour l'intégration sociale (AQIS).

Marjolaine Hébert, spécialiste en réadaptation psychosociale, CRDITED Pavillon du Parc.

Katleen Picard, ergothérapeute, CRDITED de Montréal.

Annick Rajotte, chef en réadaptation, Services aux enfants et à leur famille, équipe RGD-DI, Centre de réadaptation de l'Ouest de Montréal (CROM).

Christine Rivest, agente de planification, de programmation et de recherche, CRDITED MCQ – IU.

TABLE DES MATIÈRES

GLOSSAIRE	4
1. INTRODUCTION	5
1.1 Modèle théorique de l'intervention et questions d'évaluation	6
2. MÉTHODOLOGIE	7
2.1 Critères de sélection des documents	7
2.2 Stratégies de recherche utilisées pour l'identification des outils	9
2.3 Processus de sélection des documents	9
2.4 Évaluation de la qualité des documents retenus	9
2.5 Constitution de la liste d'outils d'évaluation	11
2.6 Extraction des données	11
2.7 Sélection des outils qui feront l'objet d'une analyse plus détaillée de leurs qualités psychométriques	11
2.8 Présentation de l'information portant sur les outils d'évaluation retenus	12
3. RÉSULTATS	12
3.1 Résultats du processus de sélection des documents	12
3.1.1 Présentation des documents retenus	13
3.1.1.1 LINKING Authentic Assessment & Early Childhood Intervention. Best measures for best practices. Second Edition (2010)	13
3.1.1.2 Avis scientifique sur le choix d'un outil de mesure du développement des enfants de 0 à 5 ans dans le cadre des Services intégrés en périnatalité et pour la petite enfance (2010)	14
3.1.1.3 Understanding and Choosing Assessments and Developmental Screeners for Young Children: Profiles of Selected Measures (2011)	14
3.1.1.4 Resources for measuring services and outcomes in head start programs serving infants and toddlers (2011)	15
3.1.1.5 A Guide to Assessment in Early Childhood; Infancy to Age Eight (2008)	16
3.2 Résultats du processus de sélection des outils d'évaluation	16
3.2.1 Résultats du processus de sélection des outils d'évaluation	17
3.3 La validité des outils d'évaluation du développement des enfants	19
3.3.1 Outils d'évaluation diagnostique	19
3.3.1.1 Preschool Language Scale (PLS-4, 2002)	19
3.3.2 Outils d'intervention et de programmation	20
3.3.2.1 Bayley Scales Of Infant and Toddler Development, Third Edition (2005)	20
3.3.2.2 Developmental Profile 3 (DP-3, 2007)	22
3.3.3 Outils d'évaluation ayant une double fonction	24
3.3.3.1 Battelle Developmental Inventory (BDI, 1984)	24
3.3.3.2 Behavior Assessment System for Children (BASC-2, 2004)	25
3.3.3.3 Brigance Inventory of Early Development II (IED-II, 2004)	27
3.3.3.4 Devereux Early Childhood Assessment for infants and toddlers (DECA - I/T, 2007)	28
3.3.3.5 MacArthur-Bates Communicative Development Inventory, 2nd Edition (CDI, 2007)	29
3.3.3.6 Peabody Picture Vocabulary Test, 4 th edition (PPVT-IV, 2007)	31
3.3.3.7 Vineland Adaptive Behavior Scales II Edition (VABS II, 2005)	32
3.4 Disponibilité des instruments validés en français et en anglais	35
3.5 Formation requise pour utiliser les outils	35
4. DISCUSSION	35
5. CONCLUSION	37

Références.....	38
Annexe A. Questions du guide de pratique.....	40
Annexe B. Liste des syndromes inclus pour l'identification de la population	41
Annexe C. Stratégie de recherche documentaire dans Internet.....	42
Annexe D. Description des documents de référence retenus.....	46
Annexe E. Description des outils d'évaluation.....	49
Annexe F. Objectifs des outils (par ordre alphabétique).....	64
Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus	78
Liste des schémas	
Schéma 1 : Modèle théorique de l'intervention.....	7
Schéma 2 : Démarche utilisée pour établir la FPS des outils d'évaluation retenus	12
Schéma 3 : Diagramme de sélection des documents	13
Schéma 4 : Diagramme de sélection des instruments de mesure	17
Liste des tableaux	
Tableau 1 : Résumé des mesures de fidélité et de validité.....	6
Tableau 2 : Critères d'inclusion et d'exclusion des documents selon le PICOTS.....	8
Tableau 3 : Critères de la grille AGREE II et modifications.....	10
Tableau 4 : Sommaire des méthodes de validité rapportées selon les catégories d'outils d'évaluation.....	18
Tableau 5 : Résumé des qualités psychométriques de 10 outils d'évaluation	34

GLOSSAIRE

Fidélité : Le degré de consistance avec lequel l'instrument mesure une variable, c'est-à-dire que l'instrument produira des résultats similaires lorsque la mesure est répétée par différentes personnes à différents moments et dans différents milieux (Institut national de santé publique, INSPQ, 2010).

Consistance interne : La fidélité d'un instrument se détermine entre autres en évaluant dans quelle mesure les items censés mesurer un même construit [p. ex. les habiletés motrices ou le développement du langage] mènent à des résultats similaires, mais également dans quelle mesure ces résultats sont consistants [c'est-à-dire qu'une personne qui passe deux fois le même test, doit obtenir deux fois le même résultat/score] (Gaul Bouchard et coll., 2009).

Fidélité test-retest (stabilité) : Il importe que l'outil soit en mesure de fournir des résultats relativement similaires entre les passations, ce qui assure que les résultats obtenus ne sont pas l'effet du hasard (Gaul Bouchard et coll., 2009).

Fidélité inter-juges (concordance) : Proportion d'accords entre deux codeurs qui utilisent un même instrument de manière indépendante au même moment et sur un même processus (adaptation d'une définition de Laurencelle, 1998).

Validité : Réfère à la capacité qu'a l'instrument de classer les objets (ou personnes) évalués en fonction de la caractéristique voulue, à sa capacité de mesurer vraiment ce qu'on veut mesurer (Laurencelle, 1998).

Validité manifeste (apparente) : L'adéquation entre les contenus apparents du test et la qualité ou caractéristique qu'on souhaite mesurer (Laurencelle, 1998) ou correspondance logique et évidente entre les items et ce que le test est censé mesurer (Bernier et Pietrulewicz., 1997).

Validité de contenu (échantillonnale) : Raffinement de la validité manifeste; pour l'établir, il faut démontrer que les contenus du test (ou du procédé de mesure) constituent un échantillon représentatif des éléments, habiletés, connaissances, dont on veut mesurer la possession par le sujet (Laurencelle, 1998). Par exemple, les concepteurs d'outils d'évaluation vont parfois demander à des experts externes si le test comprend toutes les dimensions pertinentes du concept à mesurer.

Validité conceptuelle (de construit) : autre approche de validation qui consiste à caractériser un phénomène, la réalité sous-jacente au procédé de mesure, à en découvrir et définir le concept (Laurencelle, 1998). Cette méthode de validité repose sur des stratégies de validation telles que l'analyse factorielle, qui permettent de circonscrire le phénomène ou le concept mesuré et « de prouver que la mesure covarie réellement avec le phénomène visé » (p. 105).

Validité concomitante : Corrélation qu'un test entretient avec un autre test mesurant la même caractéristique pour autant que les contenus, les formats ou les méthodes de ces deux tests soient suffisamment distincts (Laurencelle, 1998).

Validité prédictive : Consiste en la capacité qu'a le test de prédire le résultat (ou la valeur) obtenu par la personne mesurée dans une caractéristique ou une performance déterminée par la qualité mesurée au test (Laurencelle, 1998). En d'autres mots, le test permet d'anticiper une performance future ou une *performance* antérieure. Par exemple, un test de quotient intellectuel peut *prédire* la moyenne au bulletin d'un enfant, considérant la relation causale probable qui existe entre les deux mesures, la moyenne au bulletin étant « vue comme un effet (futur) de la capacité mesurée par le test de QI » (p. 107).

1. INTRODUCTION

L'évaluation des enfants est une composante majeure des services d'intervention précoce (Rousseau et coll., 2012). Plusieurs outils d'évaluation sont utilisés pour dépister, diagnostiquer ou suivre le progrès de l'enfant au cours d'une intervention ou d'un programme. Toutefois, de nombreux praticiens formuleraient des réserves à l'idée de n'utiliser que des instruments validés du développement des enfants dans les premières années de vie, puisque cela pourrait avoir des conséquences néfastes, tel le danger de stigmatisation (Rousseau et coll., 2012). De plus, d'autres soulèvent que peu d'instruments conçus pour de jeunes enfants demandent l'implication des parents, et ce même si l'importance de considérer les familles lors des différentes étapes du processus d'intervention est reconnue (Dionne et coll., 2001). Une certaine confusion existe également entre les évaluations du développement des enfants utiles à des fins de diagnostic et celles utiles pour l'intervention (Dionne et coll., 2010). De ces constats émerge l'importance d'identifier et de décrire les outils validés scientifiquement afin que les praticiens puissent, dans le cadre de leur pratique professionnelle, faire un choix éclairé.

La Fédération québécoise des centres en réadaptation en déficience intellectuelle et en troubles envahissants du développement (FQCRDITED) a mandaté un comité scientifique, appuyé par l'Institut national d'excellence en santé et services sociaux (INESSS), pour élaborer un guide de pratique (GP) à l'intention des CRDITED afin de soutenir de la manière la plus efficace possible, le développement des enfants âgés de 2 à 5 ans présentant un retard global de développement (RGD) soit des enfants de moins de cinq ans qui :

« n'évoluent pas selon les stades de développement attendus, et ce, dans plusieurs domaines de fonctionnement [ou] aux enfants dont il est impossible d'évaluer le fonctionnement intellectuel de manière systématique, c'est-à-dire les enfants trop jeunes pour qu'on utilise avec eux des outils standardisés. » (Garcin et Moxness, 2013 : traduction libre du DSM-5).

Cette revue systématique (RS) constitue l'une des trois revues qui formeront les fondements scientifiques de l'élaboration de ce GP de la FQCRDITED. Un cadre d'analyse aborde les trois thèmes suivants : 1 — les interventions destinées aux enfants et leur contexte de réalisation; 2— les outils d'évaluation validés du développement pour des enfants 2-5 ans; et 3 — les types de soutien qui sont offerts à la famille, a été élaboré par le comité scientifique afin de faciliter l'élaboration du GP.

L'objectif principal de cette RS consiste à recenser les outils d'évaluation validés du développement de l'enfant dans les différentes sphères de développement, qu'il s'agisse d'outils d'évaluation normatifs (diagnostiques), d'outils d'évaluation utilisés à des fins d'intervention ou de programmation ou d'outils qui peuvent remplir ces deux fonctions dans le but de répondre aux questions suivantes du GP :

- Comment mesurer les progrès de l'enfant?
- Quels sont les outils d'évaluation validés du développement de l'enfant dans les différentes sphères :
 - **par rapport à lui-même** : outils d'évaluation à des fins d'intervention ou de programmation.
 - **Par rapport aux enfants de son groupe d'âge** : outils d'évaluation à des fins diagnostiques.

Il est important de faire la distinction entre les outils de dépistage et les outils servant à établir un diagnostic. Bien que l'outil de dépistage permette de déterminer si les enfants se développent normalement et d'identifier ceux qui manifestent des signes d'un retard de développement ou d'un retard anormal (INSPQ, 2010), ce type d'outil ne permet pas « de dégager des conclusions cliniques ou d'établir finement le niveau de développement de l'enfant [...] [ni] de se prononcer sur la présence ou non d'un retard de développement. » (Ordre professionnel des psychoéducateurs et des psychoéducatrices du Québec, OPPQ, 2013, p. 16), tandis que :

« L'évaluation normative consiste à faire passer à l'enfant des épreuves standardisées de mesure du développement afin de le comparer à un groupe de référence à partir de normes statistiques. [...] L'outil d'évaluation [normatif] permet de mesurer le niveau de développement d'un enfant de manière fine et de le situer par rapport à une distribution normale de la population du même âge. » (OPPQ, 2013, p. 16-17).

De plus, Dionne et ses collègues (2010) spécifient que :

« l'évaluation diagnostique regroupe un ensemble d'items choisis de façon statistique afin de comparer la performance d'un enfant à un échantillon normatif. Ainsi, ces items ne sont pas nécessairement des habiletés fonctionnelles. Autrement dit, les comportements sélectionnés ne sont pas forcément des comportements utilisés dans la vie quotidienne. Il s'agit d'une caractéristique essentielle qui distingue ce genre d'instrument, des évaluations utilisées à des fins d'intervention » (p. 117).

Les évaluations utilisées à des fins d'intervention correspondent aux mesures plus fines permettant de mesurer les progrès de l'enfant par rapport à lui-même en cours d'intervention ou à la fin de celle-ci. À ces définitions, l'OPPQ (2013) ajoute que le choix d'un outil doit tenir compte des mises à jour, des traductions valides et de possibles biais culturels.

Dans le même ordre d'idées, il est important d'expliquer rapidement la relation entre la fidélité et la validité de même que certains concepts liés à la validité des outils d'évaluation, soit l'objet central de cette RS.¹ La fidélité concerne la précision avec laquelle un test mesure certaines caractéristiques en fonction d'erreurs aléatoires. Ainsi, une forte fidélité est une condition nécessaire, mais non suffisante, à une forte validité pour un test particulier dans une situation donnée; tandis que la validité concerne la qualité de ce qui est mesuré et son étude concerne les erreurs systématiques plutôt que les erreurs dues au hasard.

Tel que le fait Laurencelle (1998), il faut distinguer deux types de mesures de validité. Soit la mesure est valide de manière évidente et donc directe (p. ex. poids, taille, etc.), soit la mesure est un indice et est, par le fait même, *indirecte* (p. ex. mesures psychologiques, mesures cognitives). Laurencelle (1998) rapporte différentes sortes de démonstrations de la validité d'une mesure indirecte, qu'il regroupe en deux classes : la validité intrinsèque et la validité extrinsèque. La validité intrinsèque comprend la validité de contenu et la validité conceptuelle, tandis que la validité extrinsèque comprend quant à elle la validité concomitante et la validité prédictive. En complément de ces quatre méthodes de validité, la sensibilité et la spécificité constituent des mesures de performance clinique d'un test. La sensibilité est définie par Kisker et ses collègues (2011) comme étant la capacité d'un instrument de mesure à identifier correctement les personnes ayant un trouble comme ayant effectivement ce trouble; alors que la spécificité fait référence à la capacité d'un outil à identifier les personnes qui n'ont pas de trouble comme n'ayant effectivement aucun trouble. L'ensemble des concepts liés à la validité et à la fidélité est rapporté dans le tableau 1.

Tableau 1 : Résumé des mesures de fidélité et de validité

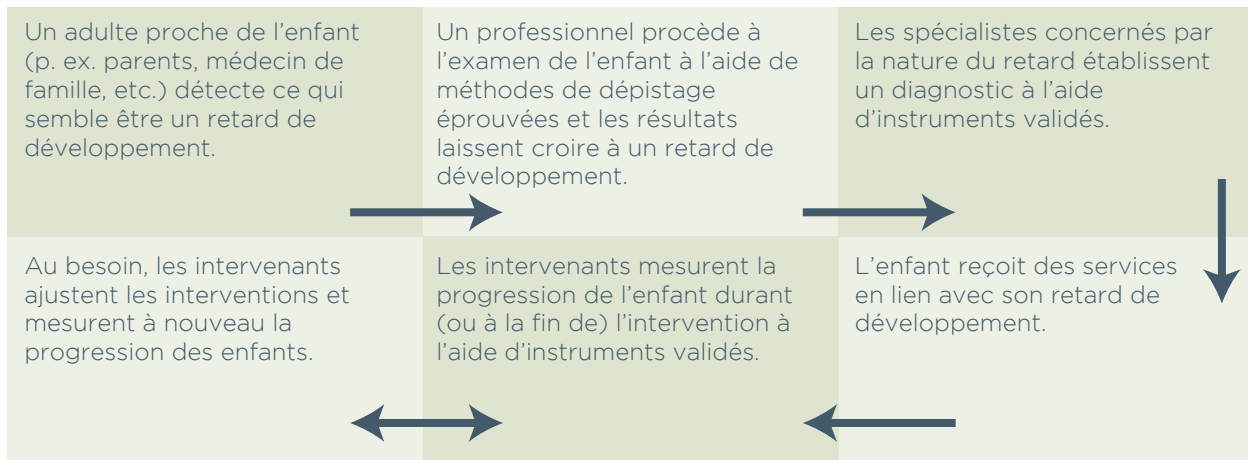
Mesures de fidélité	Cohérence interne	
	Test-retest	
	Accord inter-juges	
Validité indirecte (p. ex. mesures psychologiques, mesures cognitives)	Validité intrinsèque	Validité de contenu
		Validité conceptuelle et
	Validité extrinsèque	Validité concomitante
		Validité prédictive
Mesures de performance diagnostique	Spécificité et sensibilité	
Validité directe (p. ex. poids, taille, etc.)		

1.1 Modèle théorique de l'intervention et questions d'évaluation

Dans le cadre de cette RS, un modèle théorique de l'intervention, incluant l'activité de mesure du développement des enfants, a été conçu afin de démontrer les liens entre les différentes activités (schéma 1). Ce schéma modélise un processus d'intervention global. Comme le démontre le modèle, la mesure de la progression des enfants et l'ajustement des interventions s'inscrivent dans un processus itératif.

¹ Des définitions sont présentées dans le glossaire (p. 4), ainsi cette section permet de compléter l'information manquante.

Schéma 1 : Modèle théorique de l'intervention



Cette RS souhaite ainsi répondre à la question d'évaluation suivante :

- Pour des enfants âgés de 2 à 5 ans, quels sont les outils d'évaluation du développement ayant démontré des qualités psychométriques favorables (validité, fidélité)?²

Rappelons qu'il pourra s'agir d'outils permettant de mesurer le développement des enfants (c.-à-d. outils normatifs ou diagnostiques) ou leurs progrès dans le cadre d'une intervention visant à pallier leur retard de développement (c.-à-d. outils d'intervention et de programmation). Il s'agit là de deux catégories qui ont été déterminées par les membres du comité scientifique du GP.

De plus, des réponses à ces deux sous-questions d'évaluation suivantes permettront d'ajouter de l'information supplémentaire par rapport aux différents outils d'évaluation validés :

- Est-ce que les instruments repérés :
 - o Sont disponibles en français et en anglais?
 - o Requièrent une formation initiale spécifique ou une formation liée à leur utilisation?

Notons enfin que cette RS adopte donc une posture plus pratique que théorique puisque le résultat de cette démarche servira à alimenter la rédaction d'un guide de pratique s'adressant aux différents praticiens travaillant avec des enfants âgés de deux à cinq ans présentant un RGD et qu'il ne s'agit pas d'un GP sur les outils d'évaluation.

2. MÉTHODOLOGIE

Cette revue systématique utilise une démarche particulière pour répondre à l'ensemble des questions d'évaluation mentionnées précédemment. Sont décrites dans cette section les méthodes utilisées pour repérer les outils validés et les données scientifiques portant sur leur validité et la démarche sur mesure qui a été mise en œuvre pour rapporter les données psychométriques des outils présentés plus en détail dans la section des résultats.

2.1 Critères de sélection des documents

Les critères d'inclusion et d'exclusion des instruments de mesure sont synthétisés dans l'acronyme PICOTS (Population, Intervention, Comparateurs, Objectifs, Temporalité, Site). Ces critères, qui découlent de la question d'évaluation et des sous-questions de recherche, ont été validés par les membres du comité scientifique. Ces critères tiennent compte de la nature spécifique de l'objet de recherche, soit des instruments validés adaptés aux enfants âgés de 2 à 5 ans avec ou sans diagnostic de RGD. En l'occurrence, les outils s'adressant aux personnes qui ne font pas partie de la population visée sont exclus, de même que les outils non validés et les outils de dépistage, ces derniers étant surtout utilisés par des intervenants de première ligne. Cela dit, tous les milieux d'intervention sont inclus afin de n'échapper aucun instrument adapté à la population visée (tous les critères d'inclusion et d'exclusion sont inscrits dans le tableau 2 à la page suivante).

² L'ensemble des questions du guide de pratique est consigné à l'Annexe A. Notons également que d'autres aspects liés à la convivialité des instruments (p. ex. l'administration, la codification ou l'interprétation) ne seront pas abordés dans cette RS.

**Tableau 2 : Critères d'inclusion et d'exclusion
des documents selon le PICOTS**

P.I.C.O.T.S	Critères d'inclusion	Critères d'exclusion
Population	Outils d'évaluation concernant : 1. Enfants de 2 à 5 ans sans diagnostic. 2. Enfants de 2 à 5 ans ayant un retard global du développement (touchés dans au moins deux sphères) : motricité globale ou fine; communication ou langage; cognition; développement personnel et social; activité de la vie quotidienne (incluant les différents syndromes présentés à l'Annexe B).	Les outils touchant seulement les personnes ne faisant pas partie de la population visée.
Intervention	Outils d'évaluation validés.	Outils non validés. Outils de dépistage.
Comparateur	Non spécifié.	
Objectif principal	Instruments validés permettant d'établir un diagnostic en lien avec l'une des cinq sphères mentionnées ci-dessus ou d'évaluer la progression de l'enfant au cours de l'intervention ou après.	
Objectifs secondaires	Langue de publication des instruments validés (français ou anglais). Formation initiale ou spécifique requise pour pouvoir utiliser l'instrument.	
Temporalité	Non spécifié.	
Site	Tous les milieux d'intervention (p.ex. milieux scolaires, service de garde de la petite enfance, centre d'intervention ou de réadaptation, à la maison...)	
Années de publication des documents	De 2003 à 2013.	
Documents admissibles	Guides de pratique, cadres de référence, revues systématiques, thèses et mémoires, manuels de référence ou recueils d'outils. Documents publiés en anglais ou en français.	Études primaires de validité des outils. Autres langues de publication.

Enfin, seuls les guides de pratique, cadres de référence, revues systématiques, thèses et mémoires, manuels de référence ou recueils d'outils publiés en anglais ou en français à partir de 2003 sont admissibles pour : 1) l'identification des outils d'évaluation et 2) l'extraction des qualités psychométriques des outils identifiés.³ Le choix de ces documents, au détriment d'études primaires de validation, est essentiellement pragmatique considérant les ressources et le temps impartis à la réalisation de cette RS. Ce choix constitue évidemment une limite puisque les données de validité proviennent uniquement de sources secondaires, ce qui pourrait engendrer un biais d'information. Idéalement, il serait préférable de s'appuyer sur les études de validation originales, ce qui permettrait d'évaluer directement la qualité des méthodes sur lesquelles repose la validité des outils. De plus, cette façon de procéder pourrait écarter, à tort, tout outil d'évaluation pour lequel aucune information de validité ne serait présentée dans les documents retenus, une information qui pourrait pourtant être publiée ailleurs, ce qui constitue une autre limite de cette méthode.

2.2 Stratégies de recherche utilisées pour l'identification des outils

Cette RS a employé trois stratégies de recherche distinctes. D'abord, les membres du comité scientifique pour l'élaboration du guide de pratique (GP) ont été sollicités pour proposer des documents qu'ils jugeaient pertinents. La deuxième stratégie repose sur une recherche effectuée dans la base de données « REGARD » par la technicienne en documentation siégeant au comité scientifique du GP (A.G.) dans le but d'identifier les outils accessibles aux différents intervenants des CRDITED. Elle a ensuite acheminé une liste d'outils aux auteurs de cette RS (S.G. et M-C.L.).

La dernière stratégie de recherche a été effectuée à l'aide de différents moteurs de recherche et bases de données sur Internet.⁴ La technicienne en documentation a effectué la recherche avec *Google Scholar* et a envoyé le résultat de la recherche aux auteurs de cette RS, tandis que ces derniers ont fait la recherche avec *Google*. De plus, la co-auteure de cette RS (M-C.L.) a consulté : six sites d'agences d'évaluation des technologies en services sociaux, les bases de données *Proquest Dissertations & Theses*, *Bielefeld Academic Search Engine* et Système pour l'information concernant la littérature grise en Europe, ainsi que les moteurs de recherche de six bibliothèques universitaires québécoises.

2.3 Processus de sélection des documents

Les documents proposés par les membres de comité scientifique ont été retenus d'emblée puisqu'ils respectent les critères spécifiés dans le PICOTS. Pour les documents issus des stratégies de recherches documentaires effectuées dans Internet, les deux auteurs de cette RS ont effectué, ensemble, une première sélection de documents en lisant uniquement les titres et les résumés et, lorsque nécessaire, en effectuant une lecture plus ciblée des documents afin de déterminer si ceux-ci respectaient les critères du PICOTS. Par la suite, l'évaluation de la qualité des documents retenus a été réalisée.

2.4 Évaluation de la qualité des documents retenus

Pour des raisons de temps et de disponibilité de ressources, l'évaluation de la qualité des documents n'a été réalisée que par l'un ou l'autre des auteurs de la RS (S.G. ou M-C.L.). Considérant que l'ensemble des documents retenus propose différents outils d'évaluation aux praticiens, ce qui pourrait s'apparenter à des recommandations d'outils, nous avons fait le choix d'évaluer la qualité méthodologique de ces documents en fonction du standard le plus élevé, soit à l'aide de la grille *Appraisal of Guidelines for Research and Evaluation* (AGREE II). Cette grille est habituellement utilisée pour évaluer la qualité méthodologique des guides de pratique; elle comprend 23 critères évalués chacun sur une échelle allant de 1 (très en désaccord) à 7 (très en accord). Ces critères sont regroupés en six domaines comprenant un nombre variable de critères dont certains ont dû être modifiés pour mieux refléter le sujet de cette RS (tableau 3, page suivante).⁵ Tous les documents évalués seront retenus pour identifier les outils d'évaluation. Cependant, seuls les documents répondants à certains critères seront utilisés pour l'extraction des informations pertinentes. Cette démarche sera expliquée au point 1.7.

³ Le catalogue REGARD a également servi à établir la liste initiale d'outils. Des précisions seront présentées sous peu.

⁴ Une présentation intégrale des stratégies de recherche effectuée sur Internet est décrite à l'Annexe C.

⁵ Le lien suivant permet d'avoir accès à l'intégralité de la grille AGREE II : http://www.agreertrust.org/wp-content/uploads/2013/06/AGREE_II_French.pdf.

Tableau 3 : Critères de la grille AGREE II et modifications

Libellé original de la Grille AGREE II	Libellés modifiés
<p>DOMAINE 1 : CHAMP ET OBJECTIFS</p> <ol style="list-style-type: none"> 1. Le ou les objectifs de la RPC sont décrits explicitement. 2. La ou les questions de santé couvertes par la RPC sont décrites explicitement. 3. La population (patients, public, etc.) à laquelle la RPC doit s'appliquer est décrite explicitement. 	<ol style="list-style-type: none"> 2. La problématique couverte par la RPC est décrite explicitement.
<p>DOMAINE 2 : PARTICIPATION DES GROUPES CONCERNÉS</p> <ol style="list-style-type: none"> 4. Le groupe ayant élaboré la RPC inclut des représentants de tous les groupes professionnels concernés. 5. Les opinions et les préférences de la population cible (patients, public, etc.) ont été identifiées. 6. Les utilisateurs cibles de la RPC sont clairement définis. 	<p>Aucune modification.</p>
<p>DOMAINE 3 : RIGUEUR D'ÉLABORATION DE LA RECOMMANDATION</p> <ol style="list-style-type: none"> 7. Des méthodes systématiques ont été utilisées pour rechercher les preuves scientifiques. 8. Les critères de sélection des preuves sont clairement décrits. 9. Les forces et les limites des preuves scientifiques sont clairement définies. 10. Les méthodes utilisées pour formuler les recommandations sont clairement décrites. 11. Les bénéfices, les effets secondaires et les risques en termes de santé ont été pris en considération dans la formulation des recommandations. 12. Il y a un lien explicite entre les recommandations et les preuves scientifiques sur lesquelles elles reposent. 13. La RPC a été revue par des experts externes avant sa publication. 14. Une procédure d'actualisation de la RPC est décrite. 	<ol style="list-style-type: none"> 7. Des méthodes systématiques ont été utilisées pour rechercher les preuves scientifiques liées à la validité des outils. 8. Les critères de sélection des outils sont clairement décrits. 9. Les forces et les limites des preuves scientifiques soutenant la validité des outils sont clairement définies. 10. Question non considérée. 11. Question non considérée. 12. Il y a un lien explicite entre <i>le choix des outils d'évaluation présentés</i> et les preuves scientifiques sur lesquelles ce choix repose.
<p>DOMAINE 4 : CLARTÉ ET PRÉSENTATION</p> <ol style="list-style-type: none"> 15. Les recommandations sont précises et sans ambiguïté. 16. Les différentes options de prise en charge de l'état ou du problème de santé sont clairement présentées. 17. Les recommandations clés sont facilement identifiables. 	<ol style="list-style-type: none"> 15. Les <i>propositions d'outils d'évaluation</i> sont précises et sans ambiguïté. 16. Les différentes options d'outils sont clairement présentées.
<p>DOMAINE 5. APPLICABILITÉ</p> <ol style="list-style-type: none"> 18. La RPC décrit les éléments facilitant son application et les obstacles. 19. La RPC offre des conseils et/ou des outils sur les façons de mettre les recommandations en pratique. 20. Les répercussions potentielles de l'application des recommandations sur les ressources ont été examinées. 21. La RPC propose des critères de suivi et de vérification. 	<p>Aucune modification.</p>
<p>DOMAINE 6 : INDÉPENDANCE ÉDITORIALE</p> <ol style="list-style-type: none"> 22. Le point de vue des organismes de financement n'a pas influencé le contenu de la RPC. 23. Les intérêts divergents des membres du groupe ayant élaboré la RPC ont été pris en charge et documentés. 	<p>Aucune modification.</p>
<p>ÉVALUATION GÉNÉRALE DE LA RPC</p> <ol style="list-style-type: none"> 1. Évaluez la qualité générale de la RPC. 2. Je recommanderais l'utilisation de la RPC. Oui/Oui, avec certaines modifications. /Non 	<p>Aucune modification.</p>

2.5 Constitution de la liste d'outils d'évaluation

Une stratégie unique a été élaborée pour sélectionner les outils d'évaluation validés. Tout d'abord, l'un des coauteurs de cette RS (S.G.) a dressé une liste initiale d'outils repérés à partir du répertoire « REGARD »⁶, liste qu'il a ensuite consignée dans une grille d'extraction (voir la description ci-dessous). Cette grille d'extraction permet de recueillir les informations pertinentes sur chacun des outils (voir la description ci-dessous). Il a ensuite vérifié et exclu tous les outils qui s'adressent à d'autres populations que celles visées par cette RS (p. ex. adultes, autistes, etc.). Par la suite, la liste a été acheminée à deux membres du comité scientifique représentant le volet clinique (A.R. et K.P.) aux fins de validation. L'exercice a consisté à : 1) écarter les outils qui ne s'adressent pas à la population visée et qui ne l'auraient pas été lors du premier tri; 2) ajouter des outils absents de la liste; 3) commencer la caractérisation des outils en fonction des champs préétablis (p. ex. s'il s'agit d'outils diagnostiques ou de mesure de l'intervention, si l'instrument est réservé à une profession en particulier, etc.).

Au besoin, ces personnes ont fait appel à leurs collègues ou se sont référées à différents documents organisationnels. Par la suite, en plus des outils ajoutés par les membres du comité, nous avons complété la liste à l'aide de trois documents sélectionnés (Bagnato et coll., 2010; INSPQ, 2010; *Washington State Office of Superintendent of Public Instruction* (WSOSPI), 2008). Cette liste complète a ensuite été réacheminée aux membres du comité scientifique représentant les intervenants cliniques (A.R., K.P., C. R. et M.H.) pour qu'elles nous assurent de son exhaustivité et qu'elles caractérisent les outils d'évaluation qu'elles connaissent. Lorsque nous avons récupéré cette liste, nous avons exclu les outils pour lesquels aucune information de validation n'était présentée dans les cinq documents retenus (Bagnato et coll., 2010; Halle et coll., 2011; INSPQ, 2010; Kisker et coll., 2011; *WSOSPI*, 2008), et ce, malgré l'existence possible de documents de validation primaire.

2.6 Extraction des données

Une grille d'extraction a été conçue à l'aide d'un logiciel de traitement de données (Excel^{MD}) afin de dresser la liste d'outils et pour y consigner toutes les informations permettant de répondre aux questions de recherche : leur nom (version et année de publication), le nom des auteurs et des éditeurs, les sphères de développement mesurées, la formation initiale ou spécifique requise pour utiliser l'outil, l'étendue de l'âge couverte par l'instrument, la langue de validation, les données liées à la fidélité et à la validité, incluant toute information liée à la performance diagnostique des instruments de mesure (c.-à-d. la spécificité et la sensibilité).

Chaque outil d'évaluation a ensuite été classé en fonction des deux catégories spécifiées par les membres du comité scientifique du GP (c.-à-d. outils d'évaluation diagnostique ou outils d'intervention et de programmation). Toutefois, une lecture des objectifs visés par les concepteurs d'outils (ceux-ci sont présentés à l'Annexe F) a permis de constater que certains instruments permettraient à la fois de poser un diagnostic et de mesurer des effets d'intervention, ce qui permet de créer une troisième catégorie d'outil qui sera dorénavant nommée outils ayant une *double fonction*. De plus, à la lecture de ces objectifs, les auteurs de cette RS ont classé chacun des outils répertoriés.

2.7 Sélection des outils qui feront l'objet d'une analyse plus détaillée de leurs qualités psychométriques

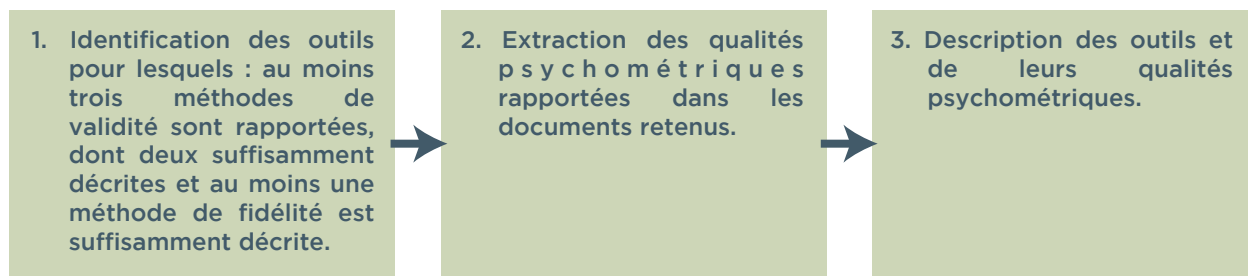
Une évaluation des qualités psychométriques des outils d'évaluation sera effectuée uniquement pour ceux dont au moins trois méthodes de validité sont rapportées dans les documents retenus pour cette RS. De plus, ces documents doivent rapporter, pour chaque outil retenu en fonction du critère précédent, des données suffisamment exhaustives pour au moins deux méthodes de validité et pour au moins une méthode de fidélité. Dans le cas contraire, les outils initialement retenus seront exclus et ne seront pas davantage décrits ni analysés. Si les données psychométriques publiées dans les documents retenus sont trop générales (p. ex. qu'une valeur se situe en dessous ou au-dessus d'un seuil prédéterminé sans précision additionnelle), ces données serviront uniquement comme complément d'information.

⁶ « REGARD » est développé par la Société GRICS. Ce produit, qui respecte les standards reconnus (format MARC et norme Z39.50) est présent dans plusieurs écoles, établissements d'enseignement supérieur, municipalités et centres de documentation d'organismes publics et privés. Plus de 2000 bibliothèques exploitent quotidiennement « REGARD ». (Source : https://grics.ca/fr/feuilles/fr_Regard.pdf).

2.8 Présentation de l'information portant sur les outils d'évaluation retenus

Autant les données de fidélité que les données de validité seront rapportées, à l'exception de la validité manifeste parce que « ce type de validité n'est pas essentiel, surtout lorsqu'on peut établir d'autres types de validité pertinente à l'usage qu'on veut en faire » (Bernier et Pietrulewicz, 1997). Cette démarche vise à faciliter le choix d'un instrument de mesure par les professionnels. La section des résultats dressera donc un portrait global de l'ensemble des outils d'évaluation retenus en fonction des critères décrits plus haut.⁷ Le schéma 2 présente une vue synoptique de la démarche utilisée.

Schéma 2 : Démarche utilisée pour rapporter les qualités psychométriques des outils répondant aux critères de sélection



3. RÉSULTATS

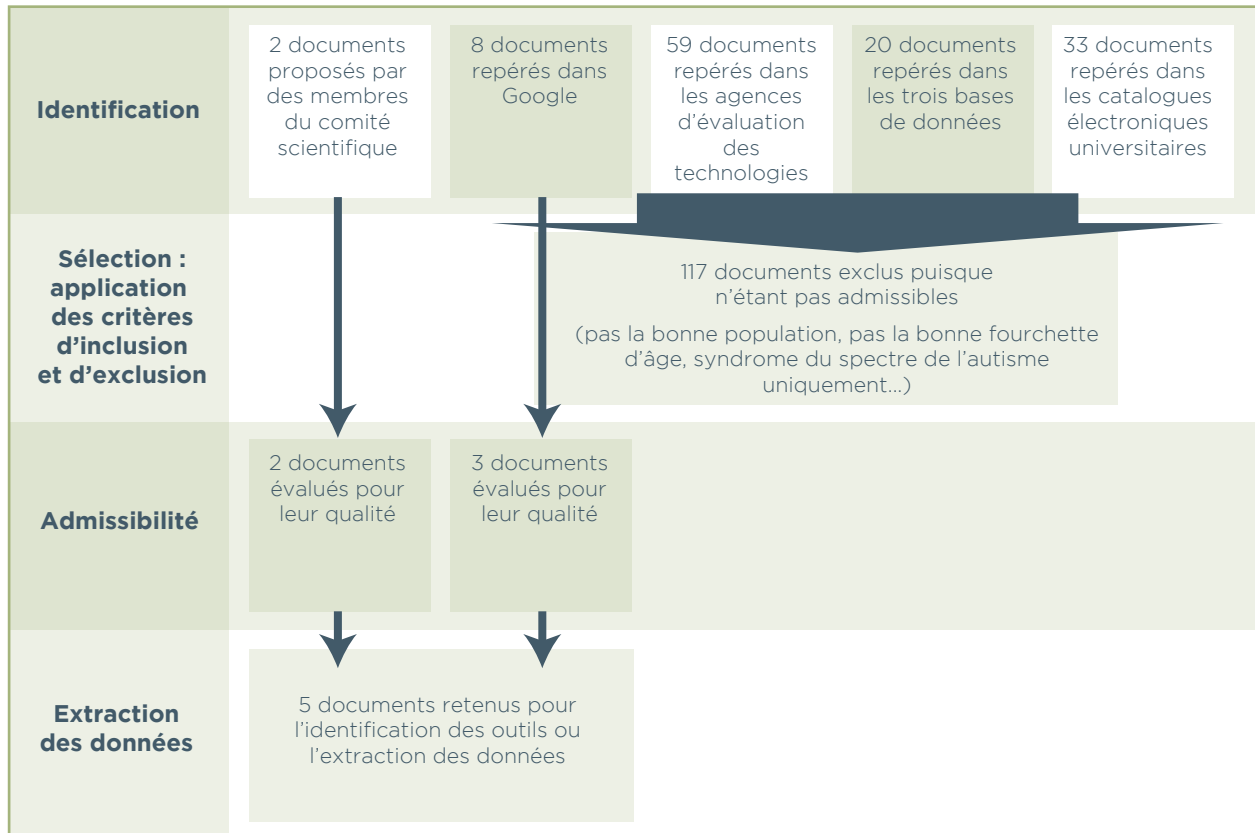
Cette section présente l'ensemble des résultats de la démarche. Sont d'abord présentés les résultats du processus de sélection des documents, ainsi qu'une brève description de ces derniers, suivi des résultats du processus de sélection des outils d'évaluation et des résultats permettant de répondre aux questions de la RS.

3.1 Résultats du processus de sélection des documents

Au total, cinq documents ont été retenus pour cette RS : deux d'entre eux ont été proposés par les membres du comité scientifique et les trois autres ont été repérés à l'aide d'une stratégie de recherche documentaire qui a permis d'identifier un total de 120 documents. Le schéma 3 (page suivante) présente le diagramme de sélection des documents. De manière générale, ces cinq documents présentent entre 18 et 130 outils d'évaluation pour des enfants âgés de 0 à 8 ans. Un seul document provient du Québec et les quatre autres documents ont été publiés aux États-Unis. Un minimum de quatre personnes a participé à l'élaboration de ces ouvrages et trois des cinq documents ont été revus par des experts externes. Somme toute, ces recueils d'outils ont obtenu une note entre 3 et 5 sur 7 sur le critère « Évaluez la qualité générale de la RPC » sur la grille AGREE II. L'annexe D présente succinctement ces documents.

⁷ L'Annexe E présente tous les outils d'évaluation par catégorie, qu'il s'agisse d'outils diagnostiques, d'intervention et programmation ou d'outils qui ont cette double fonction. L'annexe F présente les objectifs de chacun de ces outils.

Schéma 3 : Diagramme de sélection des documents



3.1.1 Présentation des documents retenus

3.1.1.1 LINKING Authentic Assessment & Early Childhood Intervention. Best measures for best practices. Second Edition (2010)

Ce livre publié aux États-Unis a été suggéré par les membres du comité scientifique représentant le volet clinique. Les auteurs, Bagnato, Neisworth et Pretti-Frontczak (2010), ont voulu créer un répertoire d'outils de mesure en situation authentique (c.-à-d. des outils qui évaluent l'enfant dans son milieu de vie) pour aider les personnes intéressées, qu'il s'agisse de praticiens, de gestionnaires, de chercheurs ou de parents, à effectuer un choix éclairé sur le type d'outil d'évaluation à utiliser. Le document présente 87 instruments de manière plus ou moins exhaustive. Ces outils ont été retenus par Bagnato, Neisworth et Pretti-Frontczak (2010) parce qu'ils sont adaptés pour de jeunes enfants âgés de 0 à 8 ans et accessibles au grand public américain et parce qu'ils contiennent un nombre suffisamment élevé d'items, ce qui permettrait de déterminer des objectifs individuels et pour identifier des stratégies d'intervention.

Le groupe de développement du document (les trois auteurs) a collaboré avec 1 087 utilisateurs (c.-à-d. enseignants de la petite enfance, prestataires, thérapeutes et autres membres de l'intervention précoce) et sept experts nationaux afin de qualifier chaque outil d'évaluation sur différentes dimensions telles que l'authenticité et l'utilité, pour ne nommer que celles-là. Les répondants ont rempli un questionnaire accessible en ligne. Le document a fait l'objet d'une révision externe.

La principale limite de ce manuel réside dans la précision des données rapportées et dans la rigueur de la démarche que les auteurs ont utilisée pour recueillir les informations auprès des utilisateurs : on ne connaît ni le nombre de personnes ni la composition exacte des groupes qui se sont prononcées sur chaque outil. Par le fait même, les données psychométriques des outils présentés sont fragmentaires. Par exemple, certains outils sont présentés de manière abrégée souvent sans présenter de données psychométriques. En conséquence, dans le cadre de cette RS, ce document n'a été utilisé que pour identifier les outils d'évaluation.

3.1.1.2 Avis scientifique sur le choix d'un outil de mesure du développement des enfants de 0 à 5 ans dans le cadre des Services intégrés en périnatalité et pour la petite enfance (2010).

Ce document publié par l'Institut national de santé publique du Québec (INSPQ) a également été suggéré par les membres du comité scientifique. Ce document s'adresse principalement aux directions régionales de santé publique pour les aider à choisir un outil d'évaluation. L'objectif est d'analyser les outils validés les plus utilisés auprès des enfants âgés de 0 à 5 ans vivant en contexte de vulnérabilité. Les outils admissibles devaient couvrir quatre sphères de développement, être disponibles en français et ils devaient pouvoir être codés et interprétés par des professionnels ayant diverses formations; ainsi les auteurs (Beauregard, Comeau et Poissant, 2010) ont écarté les outils qui requièrent une formation spécialisée pour l'administration, la codification ou l'interprétation. De plus, ils ont exclu les outils qui évaluent le développement d'enfants déjà identifiés comme ayant un retard de développement.

Le groupe de développement du document comprend sept personnes, soit les auteurs nommés ci-dessus et des répondants régionaux qui ont pu mettre à profit leurs connaissances des enjeux sur le terrain et leur compréhension des besoins des intervenants. Ces répondants ont également été sollicités pour participer : 1) aux décisions quant à la démarche à suivre pour ce document, 2) à l'analyse des outils et 3) au processus de validation de l'avis scientifique. De plus, six experts externes ont révisé le document, certains pour valider les informations descriptives rapportées par le groupe de travail, d'autres pour les aspects méthodologiques du document.

Il est important de préciser que les critères d'inclusion de l'avis sont très restrictifs, ce qui a pour conséquence d'écarter plusieurs outils d'évaluation qui pourraient être pertinents pour la pratique dans un contexte comme celui d'un CRDITED. Au total, les auteurs ont révisé 63 outils d'évaluation et ils en recommandent cinq qu'ils ont décrits plus en détail. Au sujet des qualités psychométriques rapportées pour ceux-ci, des tableaux sommaires indiquent, pour chaque méthode de fidélité ou de validité si les données : 1) se situent *Sous le seuil acceptable*, 2) sont *Acceptables* ou 3) sont *Bonnes*. Pour chaque méthode de fidélité et de validité, les auteurs de l'Avis ont consulté différentes sources d'information pour déterminer leurs seuils d'interprétation. Aussi, une colonne de commentaires permet d'ajouter de l'information. Par exemple, pour le Brigance, la fidélité test-retest est à la fois *sous le seuil, acceptable* et *bonne* et le commentaire indique que les valeurs varient en fonction des études. En somme, l'utilisation de ce document est très limitée et, dans le cadre de cette RS, ce document a uniquement été utilisé pour identifier et décrire des outils d'évaluation et pour vérifier la cohérence du corpus, car l'information sur les qualités psychométriques est trop générale.

3.1.1.3 Understanding and Choosing Assessments and Developmental Screeners for Young Children: Profiles of Selected Measures (2011).

L'objectif principal des auteurs (Halle, Zaslow, Wessel, Moodie et Darling-Churchill) est de soutenir les programmes *Head Start*, aux États-Unis, qui sont dans un processus de sélection ou de changement d'outils d'évaluation ou de dépistage. En particulier, ils désirent améliorer le niveau de compréhension des gestionnaires de ce programme au sujet des mesures de fidélité et de validité des outils d'évaluation et de dépistage conçus pour des enfants âgés de 3 à 5 ans. Le recueil présente 18 outils d'évaluation validés qui sont les plus susceptibles d'être utilisés dans le programme *Head Start*. De plus, le choix de ces outils repose sur les critères suivants : ils doivent couvrir un vaste éventail de domaines, ils peuvent être utilisés à grande échelle et il existe de l'information sur leur fidélité et sur leur validité.

Le groupe de développement du document (Halle, Zaslow, Wessel, Moodie et Darling-Churchill, 2011) s'est appuyé sur les manuels d'utilisateur, les sites Internet des développeurs, ou de l'information transmise directement par les concepteurs pour obtenir l'information de chacun des outils présentés. Les auteurs (Hall et coll.) ont également demandé aux concepteurs d'outils d'examiner l'exactitude et l'exhaustivité des informations rapportées sur leurs outils d'évaluation. Enfin, ils mentionnent que ce compendium d'outils ne devrait pas être utilisé pour recommander ou exiger l'utilisation d'outils en particulier.

Les limites sont surtout que le document n'a pas fait l'objet d'une révision externe par des experts et que les outils d'évaluation étaient retenus s'ils sont utilisés à grande échelle (*available for widespread use*), un critère qui manque de clarté.

Au sujet des qualités psychométriques rapportées pour ceux-ci, des tableaux sommaires indiquent, pour chaque méthode de fidélité, si les données sont *Acceptables*, *Faibles* ou si la méthode n'a pas fait l'objet de mesures. Pour les méthodes de validité, on indique pour la validité de contenu si ce dernier a été révisé ou non par des experts; pour la validité prédictive s'il existe des preuves quant à une performance future (there is evidence of prediction to later achievement); et pour la validité conceptuelle et concomitante les données sont interprétées comme étant *Fortes*, *Modérées*, *Faibles* ou si la méthode n'a pas fait l'objet de mesures. Les auteurs (Halle et coll.) indiquent également que leurs critères d'interprétation s'appuient sur des standards généralement acceptés et qu'ils ont consulté différentes références en l'absence de tels standards (p. ex. des articles scientifiques portant sur l'évaluation des enfants, des documents portant spécifiquement sur le développement de mesures, des recommandations émises par des organismes tels que le *American Academy of Pediatrics*).

Des fiches descriptives de chacun des outils présentés ajoutent des informations par rapport à chaque méthode de fidélité ou de validité qu'il s'agisse des caractéristiques des échantillons (p. ex. nombre et groupes d'âge) et des modalités opérationnelles de passation des tests (p. ex. le temps moyen écoulé entre deux passations du test pour mesurer la fidélité test-retest est de 11 jours). Cela dit, ces informations sont trop générales et renseignent peu sur les qualités psychométriques des outils présentés. En somme, dans le cadre de cette RS, ce document a uniquement été utilisé pour identifier et décrire des outils d'évaluation et pour vérifier la cohérence du corpus.

3.1.1.4 Resources for measuring services and outcomes in head start programs serving infants and toddlers (2011).

Ce document a pour objectif de soutenir les programmes *Head Start*, aux États-Unis, qui offrent des services aux femmes enceintes et aux familles ayant de jeunes enfants. Le document de Kisker et ses collègues (2011) présente des informations sur 44 outils de dépistage ou d'évaluation qui ont été développés spécifiquement pour les enfants de trois ans et moins et leur famille. Les auteurs, un groupe de développement comprenant 15 personnes, considèrent que cette information permettra aux personnes œuvrant au sein du programme *Head Start* de développer des plans de mesure de la performance et d'effectuer des collectes d'informations qui soutiendront leurs efforts d'amélioration continue. Afin d'être retenus, les outils d'évaluation devaient être postérieurs à 1987 et avoir une forte probabilité d'être utilisés dans le cadre du programme *Head Start*.

Le groupe de développement (Kisker et coll., 2011) affirme avoir consulté plusieurs sources d'informations pour identifier les outils et rassembler des renseignements pertinents : les manuels d'utilisateur, le site Internet des développeurs, des communications directement avec les concepteurs et même les articles scientifiques desquels proviennent les données de validité des outils d'évaluation. Ce groupe a également évalué la qualité méthodologique des articles scientifiques et a parfois analysé certains outils d'évaluation.

Ce recueil d'outils n'a pas été révisé par des experts externes, mais rapporte plusieurs informations pertinentes sur la validité, la fidélité, la manière d'interpréter les résultats, etc. Comme pour le document de Halle et ses collaborateurs (2011), cette publication présente des tableaux sommaires et des fiches descriptives de chacun des outils présentés, mais de manière beaucoup plus exhaustive, ce qui permet d'obtenir des informations sur les qualités psychométriques des outils présentés. Dans le cadre de cette RS, ce document a été utilisé pour : l'identification des outils d'évaluation, la description des outils retenus et l'extraction de leurs qualités psychométriques.

3.1.1.5 A Guide to Assessment in Early Childhood; Infancy to Age Eight (2008)

Le Washington State Office of Superintendent of Public Instruction (WSOSPI, 2008) a publié un recueil comprenant 130 outils qui peuvent être utilisés avec des enfants âgés de 0 à 8 ans, selon l'outil. Ce recueil s'adresse aux différents intervenants (p. ex. aux gestionnaires du domaine de l'éducation préscolaire, aux spécialistes de l'intervention précoce, aux psychologues scolaires, etc.) qui travaillent avec de jeunes enfants ayant des retards de développement ou qui ont un développement typique, apprenant l'anglais ou évoluant dans un environnement comprenant des facteurs de risque. Les outils présentés devaient être bien connus et postérieurs à 1992. Le groupe de développement du document comprenant 14 personnes : trois auteurs (Slentz, Early et McKenna, 2008) et 11 spécialistes en éducation spécialisée ont présenté les dernières versions des outils seulement.

L'information sur les instruments de mesure incluant les résultats de fidélité et de validité provient uniquement des guides d'utilisateur. Le document a toutefois été révisé par sept experts externes. La démarche manque également de clarté en ce qui concerne le processus de sélection des outils d'évaluation. Au sujet des qualités psychométriques, un score général de fidélité est rapporté et interprété selon une échelle à trois niveaux (*Élevée* > 0,80; *Adéquate* 0,65 – 0,79; ou *Faible* < 0,65). Pour chaque outil, le score de fidélité rapporté dans cette publication correspond à la méthode de fidélité ayant obtenu le score le plus élevé selon les données fournies par les concepteurs de l'outil, en priorisant les scores totaux. Lorsque les mesures portent sur différentes sous-échelles seulement, les auteurs (Slentz et coll., 2008) ont rapporté la moyenne des scores ou le score médian. Cette démarche a également été employée pour la validité concomitante qui est aussi rapportée à l'aide d'un score général et interprétée selon une échelle à trois niveaux (*Élevée* > 0,70; *Adéquate* 0,50 – 0,69; ou *Faible* < 0,50). En somme, l'utilisation de ce document est très limitée et, dans le cadre de cette RS, ce document a uniquement été utilisé pour identifier et décrire des outils d'évaluation et pour vérifier la cohérence du corpus, car l'information sur les qualités psychométriques est trop générale.

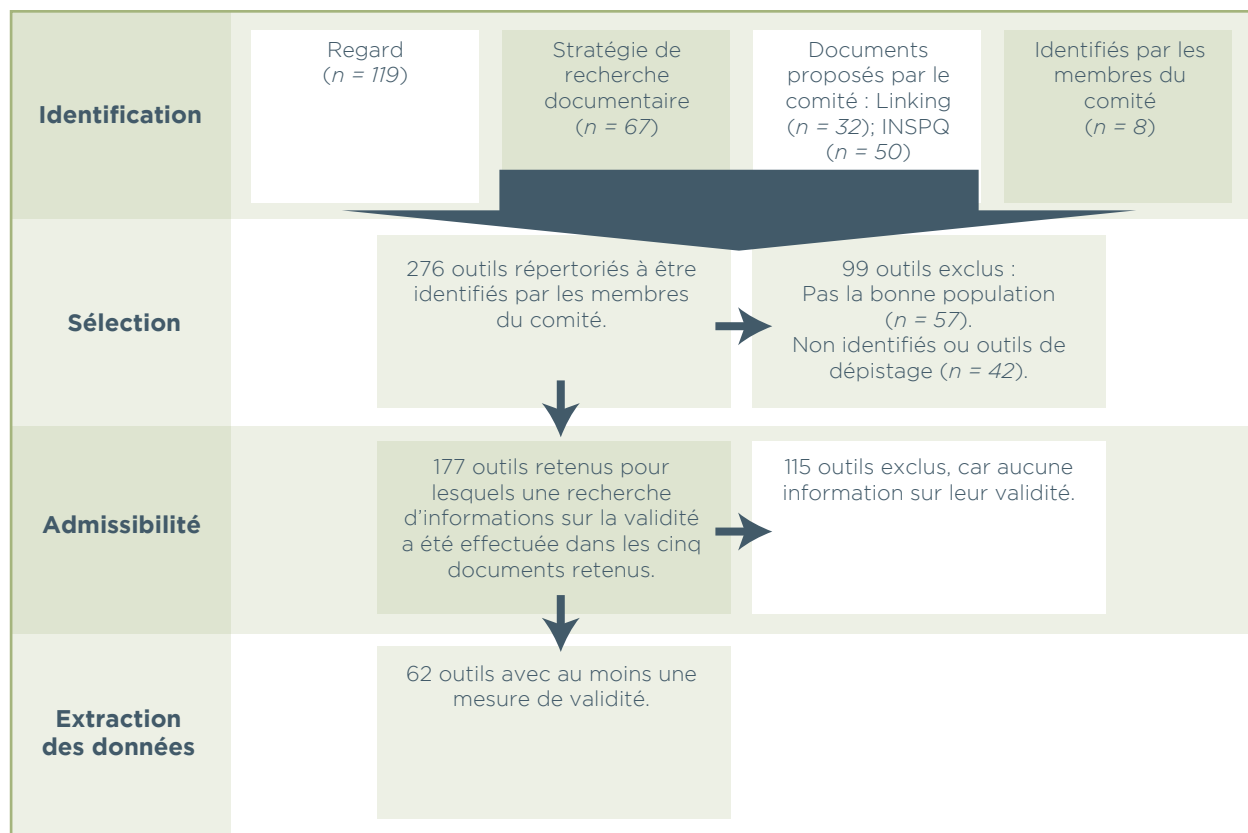
En résumé, quatre documents obtiennent un score d'au moins 4 sur 7 sur le critère « Évaluation générale de la qualité de la RPC » de la grille AGREE II, l'autre obtenant un score de 3 sur 7 (tableau 9, page suivante). Tous les documents seront utilisés pour identifier et décrire les outils d'évaluation, mais seul le document (*Resources for measuring services and outcomes in head start programs serving infants and toddlers*, Kisker et ses collègues, 2011) sera utilisé pour extraire les données psychométriques des outils, et trois autres documents seront utilisés à des fins de validation de l'information, le cas échéant, ce qui permettra de vérifier la cohérence du corpus.

3.2 Résultats du processus de sélection des outils d'évaluation

La liste initiale comprenait 276 outils provenant de la stratégie de recherche dans le catalogue « Regard » ($n = 119$), des trois documents retenus suite à la stratégie de recherche documentaire ($n = 67$), des deux documents proposés par les membres du comité ($n = 82$) ainsi que de leurs connaissances tacites ($n = 8$). De ces outils, 57 ont été exclus parce qu'ils ne s'adressent pas aux enfants de 2 à 5 ans et 42 ont été exclus parce qu'ils n'ont pu être identifiés par les membres du comité scientifique ou parce qu'il s'agissait d'outils de dépistage. Enfin, 115 outils ont été exclus, car aucun des documents retenus pour cette RS ne présente d'information sur leur validité, et ce, malgré l'existence possible de documents de validation primaire.⁸ La liste finale comprend 62 outils d'évaluation. Le schéma 4 présente le diagramme de sélection des instruments de mesure.

⁸ Cette liste pourra être fournie sur demande en communiquant avec les auteurs de cette RS.

Schéma 4 : Diagramme de sélection des instruments de mesure



INSPQ : Institut national de santé publique du Québec. WSOSPI : Washington State Office of Superintendent of Public Instruction.

Les 62 outils d'évaluation sont classés de la manière suivante : 14 outils diagnostiques, 27 outils d'intervention et de programmation et 21 outils qui ont une double fonction. Ce classement a ensuite été validé par une psychologue, une ergothérapeute et une orthophoniste du CRDITED MCQ-IU qui ont également bonifié l'information présentée aux annexes E et F.

3.2.1 Résultats du processus de sélection des outils d'évaluation

Des 276 outils répertoriés à l'aide des stratégies de recherche, les cinq documents de référence retenus ne présentaient des mesures de validité que pour 62 d'entre eux : 38 mesurent plus d'une sphère de développement, mais pas nécessairement l'ensemble des cinq sphères du développement, et 24 n'évaluent qu'une seule sphère de développement (douze la communication, cinq la sphère socioémotionnelle, trois l'alphabétisation, deux la motricité, un la cognition et un le concept des nombres). L'étendue d'âge des différents outils sélectionnés diffère également de manière importante. Par exemple, huit outils s'adressent à des individus âgés de 0 à plus de 90 ans tandis que 41 outils s'adressent spécifiquement aux nourrissons et aux tout-petits.

Notons aussi qu'environ le tiers des 62 outils (*n* = 22) prévoient une participation des parents lors de la collecte de données et que plus de la moitié reposent sur un seul « type d'évaluation » : évaluation directe individuelle de l'enfant par un intervenant qualifié (*n* = 31), questionnaire complété par le parent (*n* = 7), observation des comportements par un intervenant qualifié (*n* = 4) et questionnaire complété par l'enseignant (*n* = 2). Les autres outils (*n* = 18) comprennent au moins deux types d'évaluation (p. ex. observation et évaluation directe individuelle de l'enfant ou observation et questionnaire complété par le parent).

Dans un autre ordre d'idées, il est possible de regrouper les outils en fonction des objectifs spécifiques nommés par les concepteurs.⁹ Par exemple, des tests peuvent être utilisés :

- À des fins scolaires, soit pour mesurer la maturité scolaire, la performance ou la réussite scolaire, les difficultés académiques ou pour aider les enseignants à adapter et planifier leur enseignement en leur permettant de développer des curriculums (p. ex. le BDI-2, le CIBS-R, le YCAT ou le OWLS 1995).
- Pour déterminer l'admissibilité des enfants à un service d'intervention précoce (p. ex. l'ABAS-II, le BASC-2, ou l'ITSEA).¹⁰
- Pour déterminer des objectifs d'intervention, créer des programmes individualisés d'intervention ou pour évaluer les effets d'une intervention et suivre les progrès des enfants.

Pour ces 62 outils d'évaluation, les documents de référence retenus pour cette RS rapportent souvent qu'une seule méthode de validité (tableau 4, page suivante).¹¹ Cela dit, trois méthodes de validité sont rapportées pour dix outils et quatre méthodes de validité le sont pour trois autres outils d'évaluation. Toutefois, les données psychométriques de trois outils d'évaluation (Mullen Scales of Early Learning, 1995; Learning Accomplishment Profile, 2005; Infant Toddler Social and Emotional Assessment, 2005) ne sont pas suffisamment explicites, c'est pourquoi ils ne seront pas davantage décrits. Bref, la section suivante présentera une description plus exhaustive de dix outils d'évaluation, de même qu'une évaluation plus approfondie de leurs qualités psychométriques.¹²

Tableau 4 : Sommaire des méthodes de validité rapportées selon les catégories d'outils d'évaluation

Méthodes de validité rapportées dans les documents retenus	Outils diagnostiques (n)	Outils d'intervention et de programmation (n)	Outil ayant une double fonction (n)	Total
Une	9	22	12	43
Deux	2	2	2	6
Trois	3	1	6	10
Quatre	0	2	1	3
Total	14	27	21	62

⁹ L'Annexe F présente les objectifs spécifiques de chacun des outils.

¹⁰ Il faut cependant mentionner que plusieurs outils sélectionnés pour cette RS proviennent de documents américains élaborés spécifiquement pour faciliter le choix des professionnels travaillant dans le cadre du programme Head Start dans le but d'identifier les enfants admissibles à ce programme plus spécifiquement.

¹¹ L'Annexe G présente l'ensemble des méthodes de validité présentées dans les documents retenus.

¹² Rappelons que l'information présentée provient essentiellement de Resources for measuring services and outcomes in head start programs serving infants and toddlers (Kisker et coll., 2011).

3.3 La validité des outils d'évaluation du développement des enfants

Dans la catégorie des outils diagnostiques, seul le *Preschool Language Scale* (PLS-4, 2002) sera décrit. Les outils d'intervention et de programmation qui seront décrits sont le *Bayley Scales Of Infant and Toddler Development, Third Edition* (2005) et le *Developmental Profile 3* (DP-3, 2007). Enfin, ces huit outils qui ont une double fonction, selon notre propre typologie, seront aussi décrits :

- *Battelle Developmental Inventory* (BDI, 1984);
- *Behavior Assessment System for Children* (BASC-2, 2004);
- *Brigance Diagnostic Inventory of Early Development II* (IED-II, 2004);
- *Devereux Early Childhood Assessment for infants and toddlers* (DECA – I/T, 2007);
- *MacArthur-Bates Communicative Development Inventory 2nd Edition* (MCDI, 2007);
- *Peabody Picture Vocabulary Test, 4th edition* (PPVT-IV, 2007);
- *Vineland Adaptive Behavior Scales II Edition* (VABS, 2005).¹³

3.3.1 Outils d'évaluation diagnostique

3.3.1.1 Preschool Language Scale (PLS-4, 2002)

Description

Le *PLS-4* est un outil américain d'évaluation diagnostique mesurant le développement du langage d'enfants âgés de 0 à 6 ans et 11 mois. Plus spécifiquement, il permet d'identifier un retard de langage chez l'enfant et de cerner la source du problème (audition, expression orale ou les deux). Les scores et le rang centile obtenus permettent par la suite d'évaluer la sévérité du retard et de cibler les objectifs d'intervention. En fait, le *PLS-4* se distingue du *PLS-3* (1992) de par l'élargissement de la couverture linguistique. Il cible précisément les compétences langagières réceptives et expressives dans ces domaines : attention, jeux, gestuelle, développement vocal, communication sociale, concepts de vocabulaire, structures du langage, intégration des compétences langagières et conscience phonologique. Cet instrument peut être utilisé par des professionnels détenant une maîtrise en psychologie, en éducation, en ergothérapie ou en orthophonie. Il peut également être employé par d'autres professionnels qui ont une expérience et une formation en évaluation.

Contenu de l'outil

L'instrument comprend deux échelles principales qui mesurent la compréhension auditive et la communication expressive. De plus, le *PLS-4* inclut trois mesures supplémentaires dont l'administration est facultative : un tableau d'articulation qui examine l'articulation des phonèmes en position initiale et finale, une liste de contrôle évaluant l'expression spontanée de l'enfant, et un questionnaire parental destiné aux personnes ayant en charge un enfant âgé de 0 à 2 ans et 11 mois. Les items du *PLS-4* sont organisés par tranches d'âges et par ordre de difficulté croissante. Le test contient 16 tranches d'âges réparties entre 0 et 83 mois.

Passation du test

La passation du test est individuelle et la durée varie en fonction de l'âge de l'enfant. Le temps moyen est d'environ 20 à 45 minutes.

Évaluation et interprétation

Un livret de passation indique les critères de réussite de chaque item. Lorsque l'enfant atteint le critère de réussite, l'intervenant inscrit « 1 » au livret à côté de l'item, dans le cas contraire il inscrit un « 0 ». Une fois la passation terminée, un score brut est calculé pour chaque échelle. Le manuel explique par la suite comment interpréter les scores afin de déterminer le niveau de sévérité du trouble et indique si l'enfant a besoin de services particuliers.

¹³ Des informations sur l'ensemble des outils d'évaluation sont présentées dans les Annexes E et F.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon global de 1 534 enfants (75 à 110 enfants pour chaque groupe d'âge). Les coefficients de consistance interne du PLS-4 varient de $\alpha = 0,73$ à $0,97$ selon les échelles et les groupes d'âge et de $\alpha = 0,81$ à $0,97$ pour ce qui est des scores totaux. La fidélité test-retest repose sur des mesures prises auprès de 218 enfants répartis en deux groupes d'âge et évalués dans un intervalle de 2 à 14 jours. Tous les coefficients de corrélation sont supérieurs à $0,80$. La fidélité inter-juges repose sur l'évaluation, par 15 enseignants, de 100 protocoles sélectionnés parmi les protocoles standardisés de la partie expressive du test. Chaque protocole a été évalué par deux enseignants. Cette méthode a généré un coefficient de $r = 0,99$.

En ce qui concerne la validité de contenu du test, Kisker et collègues (2011) rapportent que les développeurs ont effectué une recension de la littérature scientifique et qu'ils ont recueilli des commentaires d'experts. À la suite de ces démarches, les concepteurs ont révisé ou supprimé des sous-tests ou des sous-items. Aucune information ne porte sur la validité conceptuelle du test, mais il est indiqué que la corrélation entre les parties Auditive et Expressive du test est de $r = 0,74$. Les mesures de validité concomitante de l'instrument proviennent d'une comparaison avec la version antérieure du test, soit le PLS-3 (1992), auprès d'un échantillon de 104 enfants. La corrélation entre les parties Auditive de ces deux tests est de $r = 0,65$, alors qu'elle est de $r = 0,79$ entre les parties Expressive.

En résumé, la consistance interne du test est généralement bonne à excellente, à l'exception de l'échelle *Communication auditive* qui paraît moins cohérente pour certains groupes d'âge. Notons cependant la taille de l'échantillon n'est pas précisée pour cette méthode de fidélité. De plus, le test démontre un bon niveau de stabilité, quoique l'intervalle de temps entre les deux passations paraît court. Bien que le niveau d'accord entre les codeurs soit presque parfait, la démarche utilisée par les concepteurs est perfectible, par exemple il aurait été préférable que tous les enseignants évaluent les mêmes enfants. Bref, l'information rapportée par Kisker et ses collègues (2011) tend à démontrer la fidélité de ce test, ce qui est d'ailleurs corroboré par le WSOSPI (2008). De plus, le contenu du PLS-4 semble avoir été validé par des experts et les corrélations avec la version précédente de l'instrument sont assez élevées, mais il aurait été intéressant de rapporter des corrélations avec d'autres outils mesurant des construits semblables. Enfin, aucune information n'est rapportée pour la validité prédictive ni pour la validité conceptuelle. D'ailleurs, la corrélation élevée entre les deux échelles soulève des questions quant à leur capacité respective à évaluer des construits distincts, soit la communication réceptive et la communication expressive.

3.3.2 Outils d'intervention et de programmation

3.3.2.1 Bayley Scales Of Infant and Toddler Development, Third Edition (2005)

Description

Le Bayley-III est un outil américain qui évalue le développement physique, cognitif, communicationnel, socioémotionnel et adaptatif des enfants âgés de 1 à 42 mois. Le test permet : d'identifier les sphères pour lesquelles une évaluation plus approfondie est requise; aux parents et à la personne offrant des services à l'enfant de planifier des stratégies d'intervention; de suivre les progrès de l'enfant au cours de l'intervention. Il est recommandé que l'évaluateur soit formé en évaluation développementale et en interprétation des résultats, qu'il soit familier avec ce test et qu'il soit en mesure d'établir et de maintenir des relations avec les nourrissons, les tout-petits et les parents. Cet outil est disponible en anglais et en espagnol et aucune information ne permet de savoir s'il est disponible en français.

Contenu de l'outil

Le test comprend cinq échelles distinctes : *Cognitive* (91 items), *Language Composite* (97 items), *Motor Composite* (138 items), *Social-Emotional* (35 items) et *Adaptive Behavior*.¹⁴

¹⁴ Notons que l'échelle Adaptive Behavior correspond à l'ABAS-II (2003) et que l'échelle Social-Emotional, complétée par les parents, correspond à une version adaptée du Greenspan Social-Emotional Growth Chart.

Passation du test

La passation du test est individuelle. Le temps moyen est d'environ 50 à 90 minutes et varie en fonction de l'âge de l'enfant.

Évaluation et interprétation

Le formulaire d'évaluation inclut : les titres de chaque item, le matériel nécessaire pour chaque item, les critères de cotation et un espace libre pour les commentaires de l'évaluateur. Chaque item obtient un score de « 1 » (réussi) ou un score de « 0 » (échoué). Le manuel présente des tables qui permettent de convertir les scores bruts en un score composé et de calculer le rang centile et l'intervalle de confiance.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon global de 1 700 enfants (100 enfants pour chaque groupe d'âge). Pour la consistance interne du test, les corrélations pour les cinq échelles du test varient de $r = 0,70$ à $0,99$. Pour l'échelle Adaptive Behavior (soit l'ABAS-II) les corrélations d'un sous-groupe de 246 enfants ayant des retards de développement, des déficiences motrices, des retards de langages, et des facteurs de risque biologique varient de $r = 0,90$ à $0,99$, tandis que celles de l'échantillon standardisé varient de $r = 0,70$ à $0,98$ selon les groupes d'âge et les échelles de mesure. D'ailleurs, les corrélations pour un sous-groupe de 668 enfants avec un diagnostic clinique varient de $r = 0,90$ à $0,99$ pour l'échelle Cognitive, de $r = 0,74$ à $0,99$ pour l'échelle Language Composite et de $r = 0,84$ à $0,99$ pour l'échelle *Motor Composite*. La fidélité test-retest de chacune des échelles nommées ci-haut (à l'exception de l'échelle *Social-Emotional* pour laquelle aucune information n'est rapportée) porte sur les données de 197 enfants qui ont été évalués dans un intervalle moyen de six jours, sauf pour l'échelle Adaptive Behavior dont les données portent sur 207 rapports de parents d'enfants âgés de 0 à 35 mois qui ont été évalués dans un intervalle moyen de 12 jours. Les coefficients de corrélation varient de $r = 0,63$ à $0,90$ selon les sous-échelles et les groupes d'âge. La fidélité inter-juges pour l'échelle *Adaptive Behavior* repose sur des mesures effectuées auprès de 56 enfants qui ont chacun été évalués par leurs deux parents. La corrélation obtenue pour le Score total de l'échelle est de $r = 0,77$ et les corrélations varient de $r = 0,69$ à $0,83$ selon les domaines.

Kisker et ses collègues (2011) ont aussi rapporté plusieurs données de validité pour le Bayley-III. Tout d'abord, le développement du test repose sur une recension de la littérature et des rencontres avec des experts, notamment. Les concepteurs ont aussi effectué une analyse factorielle confirmatoire qui valide une structure latente à trois facteurs pour les échelles Cognitive, *Language Composite* et *Motor Composite*, mais aucune information n'est présentée pour les deux autres échelles du test. Ajoutons que des corrélations de $r = 0,71$ entre l'échelle Language Composite et chacune des sous-échelles (Receptive et Expressive) indique que chaque sous-échelle a un poids équivalent dans le score composé, ce qui est également le cas pour l'échelle Motor Composite et ses deux sous-échelles *Fine Motor* ($r = 0,71$) et *Gross Motor* ($r = 0,69$). D'ailleurs, les corrélations seraient plus élevées entre l'échelle Cognitive et les sous-échelles du *Language Composite*, qu'entre l'échelle Cognitive et les sous-échelles du *Motor Composite*, ce qui démontre que les construits plus proches sont davantage corrélés que les construits moins apparentés, cependant aucune donnée n'est présentée.

La validité concomitante rapportée repose sur des comparaisons avec plusieurs instruments mesurant des dimensions similaires ou différentes :

- L'échelle Cognitive du Bayley-III a été comparée avec le BSDI-II Mental Scale ($r = 0,60$), avec le WPPSI-III ($r = 0,72$ à $0,79$) et avec le PLS-4 ($r = 0,57$).
- L'échelle *Language Composite* du Bayley-III a été comparée avec le BSDI-II Mental Scale ($r = 0,71$), avec le WPPSI-III ($r = 0,71$ à $0,83$) et avec le PLS-4 *Composite* ($r = 0,66$).
- L'échelle Composée motrice du Bayley-III a été comparée avec le PDMS-II Total Motor Quotient ($r = 0,57$), avec le PDMS-II fine motor ($r = 0,59$) et avec le PDMS-II gross motor ($r = 0,57$).

Ces résultats reposent en grande partie sur des mesures effectuées auprès d'un échantillon variant entre 50 et 100 enfants âgés de 0 à 42 mois. En ce qui concerne les corrélations entre des échelles mesurant des sphères différentes, tous les coefficients de corrélations sont d'au plus $r = 0,55$.

De plus :

- L'ABAS-II a été comparé avec le VABS (Interview Edition) auprès d'un échantillon de 45 enfants âgés de 1 à 69 mois ayant un développement typique ($r = 0,70$) et avec le SIB-R auprès d'un échantillon de 34 enfants âgés de 2 à 23 mois ayant un développement typique ($r = 0,18$).
- Le *Social-Emotional Scale* a été comparé avec BSDI-II Mental ($r = 0,25$), avec le BSDI-II Motor scales ($r = 0,24$), avec le WPPSI-II ($r = 0,43$), avec le PLS-4 Composite ($r = 0,23$) et avec le PDMS-II Total Motor score ($r = 0,25$).

Finalement, Kisker et ses collègues (2011) rapportent uniquement la validité prédictive de la version originale du test (BSDI, 1969). Les échelles du BSDI (1969) ne sont pas aussi prédictives que les sous-échelles spécifiques (c.-à-d. intellectuelle, langage, performance de réalisation) et plus les scores du BSID sont obtenus tard lors de la période préscolaire (au-delà de deux années), plus ils sont prédictifs du fonctionnement de l'enfant, cependant aucune mesure n'est présentée.

En résumé, la consistance interne est généralement élevée à très élevée, mais l'échantillon utilisé pour l'échelle *Social-Emotional* est inconnu. La stabilité du test varie généralement d'acceptable à bonne, à l'exception de l'échelle *Language Composite* qui présente des coefficients discutables et de l'échelle *Social-Emotional* pour laquelle il y a absence d'information et dont l'intervalle de temps entre les deux passations paraît un peu court. Même si le niveau d'accord entre plusieurs « paires de parents » varie de modéré à élevé, ce qui démontre que les parents ont tendance à évaluer leur enfant de manière similaire, cette méthode ne permet pas d'attester véritablement de la fidélité inter-juges de l'échelle *Adaptive Behavior* et aucune information n'est présentée pour les autres échelles. En somme, l'information rapportée par Kisker et ses collègues (2011) tend à démontrer la fidélité de ce test, ce qui est corroboré par le WSOSPI (2008). De plus, le contenu du Bayley-III semble avoir été validé par des experts et les construits sont validés à l'aide d'une analyse factorielle confirmatoire pour trois des cinq échelles. D'ailleurs, les corrélations avec d'autres tests évaluant des construits différents (p. ex le langage avec le moteur) sont généralement faibles, ce qui semble démontrer que chaque échelle mesure réellement ce qu'elle est censée mesurer. Les corrélations avec d'autres tests évaluant des aptitudes similaires sont généralement acceptables. Enfin, seule la validité prédictive de la première édition du test est présentée, mais aucune mesure n'est présentée, ainsi il n'est pas vraiment possible de juger de la validité prédictive du Bayley-III.

3.3.2.2 Developmental Profile 3 (DP-3, 2007)

Description

Le *DP-3* est un outil d'évaluation américain qui mesure le développement moteur, cognitif, communicationnel, émotionnel et social des enfants âgés de 0 à 12 ans. Le test peut être utilisé auprès d'enfants ayant un développement typique et pour évaluer rapidement la probabilité d'un retard d'un enfant. De plus, l'instrument permet de suivre le progrès de l'enfant au cours de l'intervention et de recueillir des informations pour la recherche. L'outil comprend deux volets. Le volet *Interview* du test devrait être complété par un professionnel familier avec l'évaluation pédagogique ou psychologique ou par une personne qui travaille habituellement à soutenir ce type d'activité professionnelle. Le volet *Parent/Caregiver Checklist*, devrait être complété par une personne qui a complété sa scolarité primaire et qui connaît bien l'enfant. Toutefois, il est recommandé que ce soient des professionnels cliniques ayant une formation en développement de l'enfant ou en éducation (p. ex. médecins, psychologues, infirmières, enseignants) qui interprètent les scores. Cet instrument de mesure est disponible en anglais et en espagnol et n'est pas disponible en français selon le site Internet de l'éditeur (*Western Psychological Services*).

Contenu de l'outil

Le *DP-3* comprend cinq échelles : Physique (35 items), Comportements adaptatifs (37 items), socioémotionnels (36 items), Cognition (38 items), et Communication (34 items). Comme mentionné précédemment, deux volets peuvent être utilisés pour évaluer l'enfant : entrevue ou *liste de vérification*.

Passation du test

La passation du test est individuelle et la durée est d'environ 20 à 40 minutes.

Évaluation et interprétation

Pour chacune des échelles, les items sont placés en ordre croissant de difficulté. La méthode de pointage consiste à indiquer si l'enfant maîtrise ou non l'habileté en sélectionnant entre oui ou non. Les réponses sont par la suite colligées afin de calculer un score brut qui sera ensuite comparé à l'échantillon normatif de 2 216 individus afin d'obtenir un score standardisé.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 2 216 enfants répartis en trois groupes d'âge. Au niveau de la consistance interne de l'*Interview Form*, les corrélations varient de $r = 0,71$ et $0,93$ selon les sous-échelles et les groupes d'âge, et de $r = 0,92$ à $0,97$ pour le *Composite General Development* (cinq échelles), alors que les corrélations du *Parent/Caregiver Checklist*¹⁵ varient de $r = 0,71$ à $0,96$ pour les sous-échelles, et de $r = 0,97$ à $0,98$ pour le *Composite General Development*. Pour le test-retest, les concepteurs ont interviewé 66 parents deux fois à l'intérieur d'un intervalle moyen de deux semaines. Les corrélations varient de $r = 0,81$ à $0,88$ selon les sous-échelles et le score est de $r = 0,92$ pour le *Composite General Development*.

Pour la validité de conceptuelle, Kisker et ses collègues (2011) rapportent que le concepteur de l'outil (Alpern, 2007) a effectué une analyse factorielle qui a permis de valider un facteur. De plus, le modèle d'analyse de Rasch a démontré que la gamme de capacité de l'enfant et la difficulté des items pour chaque échelle étaient similaires : pour toutes les échelles, l'étendue des capacités d'une personne est légèrement sous la difficulté des items, démontrant ainsi que les items mesurent fidèlement le développement des enfants dans la plage de compétence ciblée. De plus, toutes les échelles ont une corrélation plus élevée avec le *General Development score* qu'avec l'une ou l'autre des autres échelles (p. ex., elle est de $r = 0,79$ entre les échelles Cognitive et le *General Development score*) et la corrélation de $r = 0,30$ entre les échelles Physical et Communication indique que ces deux échelles mesurent bel et bien des construits différents. En ce qui concerne la validité concomitante, chaque échelle a été comparée à des instruments mesurant des construits comparables :

- Les échelles du *Parent/Caregiver Checklist* ($n = 99$ parents) ont été comparées à celles du Vineland-II ($r = 0,61$ à $0,78$) et à celles du ABAS-II ($r = 0,45$ à $0,87$).
- Les échelles du *Interview Form* ($n = 398$ enfants) ont été comparées à celles du Vineland-II ($r = 0,68$ à $0,85$), du DAYC ($r = 0,64$ à $0,72$); avec le PDMS-II ($r = 0,56$ à $0,71$); et avec le PLS-4 ($r = 0,64$ à $0,71$).

Finalement, la version originale du test (1972) a été utilisée pour prédire le fonctionnement de 29 enfants autistes. Les corrélations varient entre $0,43$ et $0,61$ entre les sous-échelles de ce test et un questionnaire évaluant le niveau de fonctionnement des enfants qui a été complété par les parents quatre à sept ans plus tard.

En résumé, la consistance interne des deux formulaires est généralement élevée à très élevée, cependant la taille de l'échantillon est inconnue pour le formulaire *Interview*. La stabilité des sous-échelles est bonne, et celle du score composé excellente, et l'intervalle de temps entre les deux passations est adéquat, cependant le petit échantillon restreint un peu l'interprétation de la stabilité. Aucune information n'est rapportée pour l'accord inter-juges. En somme, l'information rapportée par Kisker et ses collègues (2011) semble attester de la fidélité du *DP-3*. Ces derniers n'ont toutefois rapporté aucune méthode de validité de contenu. La validité conceptuelle repose sur une analyse factorielle qui valide un facteur, et les corrélations sont plus élevées entre les échelles qui mesurent des concepts semblables et plus faibles lorsqu'il s'agit de concepts différents, ce qui tend à démontrer que les échelles mesurent réellement ce qu'elles sont censées mesurer, cependant la taille de l'échantillon n'est pas rapportée. La validité concomitante avec d'autres tests semble acceptable. Enfin, la validité prédictive de la version originale du test semble acceptable, toutefois le résultat repose sur un échantillon de 29 enfants et aucune information n'est présentée sur le *DP-3*.

¹⁵ Les mesures du *Parent/Caregiver Checklist* reposent sur un échantillon de 432 enfants âgés de 0 à 12 ans combinant à la fois des enfants ayant un développement typique et des enfants ayant un diagnostic.

3.3.3 Outils d'évaluation ayant une double fonction

3.3.3.1 Battelle Developmental Inventory (BDI, 1984)

Description

Le *Battelle Developmental Inventory* (1984) est un outil américain qui permet d'évaluer le développement moteur, communicationnel, cognitif, émotionnel et social ainsi que la capacité d'adaptation d'enfants âgés de 0 à 8 ans. Le test permet :

- d'identifier des enfants présentant des retards de développement et d'évaluer leurs compétences fonctionnelles;
- d'évaluer la maturité scolaire des enfants ayant un développement typique;
- de planifier et dispenser des enseignements;
- d'évaluer un groupe d'enfants ayant des retards de développement.

Trois formats d'administration permettent d'évaluer l'enfant : administration structurée des items, entrevue avec les parents (ou une autre ressource) ou observations de l'enfant. Les concepteurs (Newborg, Stock, Wnek, Guidubaldi, et Svinicki, 1984) recommandent de former la personne utilisant le BDI (p. ex. enseignant, éducateur spécialisé, psychologue, etc.) pour l'administration et la notation du test. Enfin, cet instrument de mesure est disponible en anglais et en espagnol et n'est pas disponible en français, selon l'INSPQ (2010).

Contenu de l'outil

Le test comprend 341 items, incluant un test de dépistage de 96 items. Dans chacun des domaines du test, les items sont assignés à des groupes d'âge et sont organisés de manière séquentielle en sous-domaines. L'évaluateur peut administrer les items pour chaque domaine séparément ou il peut évaluer les cinq sphères de développement.

Passation du test

La passation entière du BDI est individuelle et prend environ 90 minutes à compléter selon l'âge de l'enfant.

Évaluation et interprétation

L'échelle de mesure repose sur un système de pointage : l'enfant qui complète les éléments de la tâche correctement obtient une note de 2; quelquefois une note de 1; rarement ou jamais une note de 0. Des tables sont ensuite utilisées pour convertir les scores bruts et interpréter l'information diagnostique pour chaque domaine et chaque sous-domaine. Cette information est présentée en « équivalent d'âge » des enfants, en rangs centiles et en scores standardisés (c.-à-d. quotient développemental, score-z, score-T et courbe normale d'équivalence).

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 800 enfants. Les coefficients de corrélation du test-retest, dont l'intervalle de passation est de quatre mois, sont de $r = 0,98$ et $r = 0,99$ selon les groupes d'âge.

Kisker et ses collègues (2011) indiquent que la validité de contenu repose sur une vérification par des experts. La structure conceptuelle du BDI serait quant à elle soutenue par une analyse factorielle et des corrélations entre les domaines et les sous-domaines qui la soutiennent. La validité concomitante repose sur des comparaisons avec trois autres tests :

- Avec le *Vineland Social Maturity Scale* et le *Developmental Activities Screening Inventory* les coefficients varient de $r = 0,78$ à $0,94$.
- Avec le *Stanford-Binet Intelligence Scale* les coefficients varient de $r = 0,40$ à $0,61$.
- Avec le *BDI Screening Test*, le coefficient est de $r = 0,99$.

En résumé, la stabilité du test est presque parfaite, mais l'intervalle de temps entre les deux passations (quatre mois) semble un peu long. Bien que Kisker et ses collègues (2011) ne rapportent pas d'information sur la consistance interne ni sur l'accord inter-juges. Toutefois, Halle et ses collègues (2011) indiquent que la consistance interne est élevée et le WSOSPI (2008) rapporte que ce test rapporte de bons scores de fidélité. Bref, ces informations nous amènent à interpréter les données de fidélité de cet outil avec prudence. Le contenu de l'outil semble avoir fait l'objet d'une révision par des experts, ce qui est également corroboré par Halle et ses collègues (2011). Pour la validité conceptuelle, l'analyse des facteurs et les corrélations entre les domaines et les sous-domaines soutiennent la validité factorielle et la structure conceptuelle du BDI, mais la taille de l'échantillon n'est pas précisée. Aussi, les mesures rapportées indiquent une validité concomitante plus ou moins acceptable avec le *Stanford-Binet Intelligence Scale* alors qu'elle est élevée avec deux autres instruments, cependant la taille de l'échantillon est inconnue. De son côté, le WSOSPI (2008) rapporte une validité concomitante élevée. Enfin, la validité prédictive n'est pas rapportée.

3.3.3.2 Behavior Assessment System for Children (BASC-2, 2004)

Description

Cet outil d'évaluation américain emploie une approche multidimensionnelle reposant sur plusieurs méthodes pour évaluer le comportement et la perception de soi des personnes âgées de 2 ans et 6 mois à 25 ans. Le BASC-2 a été conçu pour faciliter le diagnostic différentiel d'une variété de troubles affectifs et comportementaux et pour aider à l'élaboration de plans d'intervention. L'éditeur (Pearson) indique que les personnes utilisant l'outil devraient détenir au moins une maîtrise en psychologie ou en éducation, ou détenir un permis de pratique dans un domaine pertinent de l'évaluation décerné par un organisme provincial ou national. Le BASC-2 est disponible en anglais et en espagnol et n'est pas disponible en français, selon l'INSPQ (2010).

Contenu de l'outil

Le test comprend seize échelles, sept échelles optionnelles et les cinq échelles composées suivantes : *Adaptive Skills*, *Behavioural Symptoms Index*, *Externalizing Problems*, *Internalizing Problems*, *School Problems*. Le nombre de sous-échelles utilisées dépend de l'âge de l'enfant étant donné que certaines d'entre elles ne sont applicables qu'à certains groupes d'âge (p. ex. l'échelle des activités de la vie quotidienne ne s'applique pas aux adolescents). Le BASC-2 comprend plusieurs questionnaires, dont quatre pertinents pour les enfants âgés de 2 à 5 ans :

- Le *Teacher rating scale* (TRS) et le *Parent rating scale* (PRS) comprennent des questions fermées portant sur des comportements typiques lors d'activités quotidiennes courantes.
- Le *Structured developmental History* (SDH) comprend plusieurs questions ouvertes pour obtenir de l'information médicale et les dates des étapes clés du développement de l'enfant.
- Le *Student observation system* (SOS) permet d'obtenir de l'information sur le fonctionnement de l'enfant en classe.

Passation du test

La passation du test dépend de l'échelle d'évaluation utilisée et du nombre d'enseignants ou de parents impliqués pour rapporter les comportements de l'enfant. Selon le manuel, le TRS prend environ 10 à 15 minutes à compléter, le PRS 10 à 20 minutes et le SOS, 15 minutes.

Évaluation et interprétation

Les informations sur les comportements des enfants sont triangulées en tenant compte de différentes perspectives : celle de la personne évaluée et celles des enseignants et des parents. L'évaluateur utilise une table qui permet de l'orienter sur la force des résultats (faible ou élevée) et sur leur signification. Les concepteurs (Reynolds & Kamphaus, 2004) de l'instrument proposent d'utiliser les scores absolus (c.-à-d. toujours ou jamais) pour identifier les comportements à travailler lors des interventions.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 13 000 enfants âgés de 2 à 18 ans (400 ont 2 ou 3 ans et 650 ont 4 ou 5 ans). Pour la consistance interne du TRS, les coefficients alpha varient de $\alpha = 0,75$ à $0,96$ selon les échelles et les groupes d'âge; ces coefficients varient de $\alpha = 0,70$ à $0,93$ pour le PRS. Pour le TRS, la double passation (test-retest) a été réalisée dans un intervalle de 8 à 65 jours par un enseignant qui effectué deux mesures auprès de 69 enfants : les coefficients varient de $r = 0,83$ à $0,92$ pour les échelles composées et de $r = 0,72$ à $0,92$ pour les échelles primaires. Pour le PRS la double passation (test-retest) a été réalisée dans un intervalle de 9 à 70 jours par le parent ou le gardien de l'enfant pour 87 enfants : les coefficients varient de $r = 0,79$ à $0,86$ pour les échelles composées et de $r = 0,66$ à $0,88$ pour les échelles primaires. Pour le TRS, l'estimé médian ajusté de la concordance entre les codeurs est de $0,65^{16}$ (74 enfants) et varie de $0,40$ à $0,80$ selon les échelles et pour le PRS, l'estimé médian ajusté de la concordance entre les codeurs est de $0,77$ (87 enfants) et varie de $0,72$ à $0,86$.

La validité conceptuelle du TRS repose sur une analyse factorielle confirmatoire démontrant des coefficients de saturation élevés entre les niveaux d'âge pour l'échelle *Externalizing Problems factor* et pour les *Adaptive Skills factor*. Les coefficients de saturation de l'échelle *Internalizing Problems* se situent quant à eux dans une étendue considérée *moyenne à élevée* et il y a plus de variations entre les groupes d'âge. La validité conceptuelle du PRS repose sur une analyse de la structure de covariance (covariance structure analysis) et les coefficients de saturation varient considérablement entre les échelles et entre les groupes d'âge, sauf pour l'échelle *Externalizing Problems composite*. De plus, les corrélations entre les échelles équivalentes du TRS et du PRS varient :

- entre $r = 0,17$ et $0,52$ selon les domaines de l'échelle *Composite*;
- entre $r = 0,17$ et $0,43$ selon les domaines de l'échelle *Clinical*;
- entre $r = 0,28$ et $0,59$ selon les domaines de l'échelle *Adaptive*.

Enfin, la validité concomitante repose sur une comparaison avec l'ASEBA Scale (1999). Les coefficients de corrélation avec ce dernier varient de :

- $r = 0,64$ à $0,82$ pour le TRS *Composites* ($n = 46$);
- $r = 0,02$ à $0,90$ pour le TRS *Clinical et Adaptive* ($n = 46$);
- $r = 0,68$ à $0,83$ pour le PRS *Composites* ($n = 53$);
- $r = 0,32$ à $0,79$ pour le PRS *Clinical et Adaptive* ($n = 53$).

Enfin, les corrélations avec la version antérieure du test sont surtout $r \geq 90$, hormis sept d'entre elles qui se situent entre $r = 0,76$ et $0,89$.

En résumé, la consistance interne de l'instrument varie d'acceptable à excellente selon les sous-échelles, autant pour le formulaire TRS que PRS, mais la taille de l'échantillon est inconnue, ce qui limite quelque peu l'interprétation. La stabilité du test varie d'acceptable à excellente (à l'exception du PRS qui présente des scores discutables pour les échelles primaires) et l'intervalle de temps entre les deux passations est adéquat, cependant les échantillons sont un peu petits. En ce qui concerne l'accord inter-juges, le niveau d'accord entre enseignants (TRS) apparaît faible, tandis que celui des parents semble acceptable, mais la méthode employée pour attester de la fidélité inter-juges du PRS n'est pas explicitée. Ces résultats tendent tout de même à démontrer la fidélité du BASC-2, ce qui est aussi corroboré par le WSOSPI (2008).

Toutefois, aucune méthode de validité de contenu n'est rapportée, mais la validité conceptuelle du test repose sur une analyse factorielle (TRS) et sur une analyse des structures de covariance PRS qui démontrent des coefficients de saturation se situant dans un intervalle allant de moyen à élevé, selon les échelles. Ainsi, l'échelle *Internalizing Problems* du TRS et toutes celles du PRS à l'exception du *Externalizing Problems composite* seraient moins bien définies. Ajoutons que les corrélations entre les échelles équivalentes du TRS et du PRS varient de faible à modérées, mais que cela nous renseigne peu sur la fiabilité inter-juges, car cela peut autant vouloir indiquer que les enfants ont des comportements différents et à l'école que les parents et les enseignants ont une perception différente du comportement de l'enfant. La validité concomitante repose sur des tests effectués avec un petit échantillon pour ce qui est de l'ASEBA, alors que la taille de l'échantillon est inconnue pour les tests faits avec la version antérieure du test. Les corrélations avec cette dernière sont généralement élevées alors qu'elle est nettement plus variable avec l'ASEBA, selon les échelles comparées. Ajoutons que le WSOSPI (2008) considère que la validité concomitante de ce test est élevée. Enfin, la validité prédictive n'est pas rapportée.

¹⁶ Il n'est pas précisé s'il s'agit de pourcentages d'accords ou de corrélations.

3.3.3.3 *Brigance Inventory of Early Development II (IED-II, 2004)*

Description

L'*IED-II* est un outil d'évaluation américain permettant de mesurer les forces et les faiblesses des enfants âgés de 0 à 7 ans dans les cinq sphères de développement suivantes : développement physique, développement du langage, calcul de base et aptitudes en littératie, développement socioémotionnel, et activité de la vie quotidienne. Cet instrument est utilisé pour : établir un diagnostic, suivre les progrès de l'enfant au cours d'une intervention, évaluer en classe et évaluer un programme. Des intervenants qualifiés appartenant à différentes professions (p. ex. enseignants, thérapeutes, infirmières ou médecins) peuvent utiliser l'*IED-II* et aucune formation spécifique à l'instrument n'est requise. Cet outil est disponible en anglais et la première version de l'outil est disponible en français sous le nom « Inventaire du développement de l'enfant (1997) ».¹⁷

Contenu de l'outil

L'*IED-II* comprend cinq domaines couvrant dix sous-domaines, huit scores composés, 44 évaluations et 706 items. Les évaluations peuvent être utilisées sans avoir à respecter une séquence précise et il est possible d'utiliser uniquement certains items précis du test en fonctions des besoins d'information. Pour plusieurs items, l'information peut être recueillie à l'aide d'une entrevue avec un parent ou un enseignant.

Passation du test

Le test prend de 20 à 55 minutes à compléter selon l'âge de l'enfant.

Évaluation et interprétation

Les scores sont calculés en attribuant un pointage à chacune des compétences évaluées et le total des points peut ensuite être converti en : quotient, rang centile, équivalent d'âge ou plage d'enseignement selon le type de résultat désiré. Le profil développemental de l'enfant est aussi colligé dans un carnet de route (*Record Book*) et l'évaluateur interprète les résultats de l'enfant en se référant aux scores moyens.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 1 171 enfants âgés de 0 à 5 ans et deux tiers de l'échantillon à 3 ans ou moins. Les coefficients de consistance interne du test dépassent $\lambda = 0,85$ ¹⁸ pour chaque score composé, sous-domaine et domaine, alors que le coefficient total du *Adaptive Behavior* (somme des cinq domaines) est de $\lambda = 0,99$. En ce qui concerne sa stabilité (test-retest), les coefficients varient de $r = 0,68$ à $0,99$ en fonction des groupes d'âge et des domaines, tandis que les scores du *composite Total Adaptive Behavior* varient de $0,89$ à $0,95$ selon les groupes d'âge. Enfin, le pourcentage d'accord entre les deux enseignants qui ont évalué 20 enfants varie de 80 % à 96 % selon les domaines.

La validité de contenu s'appuie sur un examen approfondi de la littérature et la validité conceptuelle repose sur une analyse factorielle (rotation varimax) entre les sous-domaines qui valide trois facteurs : *Understanding and Expressing, Movement and Social Activity, and Academic/Preacademic*. D'ailleurs, les coefficients de corrélation varient de $r = 0,36$ (entre les échelles *Social and Emotional Development et Writing*) à $r = 1,0$ (entre le *Total Physical Development* et le *Total Motor*). La validité concomitante repose sur des mesures prises auprès de 484 enfants à l'aide de plusieurs tests évaluant les mêmes domaines que ceux du *IED-II*. Six mesures sont rapportées par Kisker et ses collègues (2011), les coefficients de corrélation varient de $r = 0,51$ entre le *IED-II Social and Emotional Development* domaine et les scores regroupés de trois instruments à $r = 0,88$ entre le *IED-II Cognitive domain* et les scores regroupés de trois instruments. Enfin, Kisker et ses collègues (2011) ne rapportent aucune donnée en lien avec la validité prédictive de l'outil, mais le concepteur (Brigance, 2004) du test affirme que les corrélations sont « élevées » entre le *Brigance Screens* (un autre test) et des mesures de performance scolaire, d'intelligence ou de développement du langage, par exemple, qui ont été prises six mois à deux ans plus tard.

¹⁷ Selon l'INSPQ (2010), la version française du test a fait l'objet d'un processus d'adaptation rigoureux et d'une validation par un groupe d'experts crédibles, mais aucune étude psychométrique ou normée avec des échantillons francophones n'a été effectuée.

¹⁸ Le coefficient Lambda de Guttman s'interprète comme un coefficient alpha de Cronbach.

En conclusion, la consistance interne est bonne, mais la taille de l'échantillon est inconnue. Notons toutefois que le coefficient lambda du score total du *Adaptive behavior* dépasse $\lambda = 0,95$, et qu'il est également possible que certains domaines ou certaines échelles dépassent ce seuil. Par contre, la stabilité du test varie de discutable à excellente selon les domaines et les groupes d'âges, alors qu'elle est excellente pour ce qui est du score total du comportement adaptatif. Ajoutons que l'échantillon est adéquat, mais que l'intervalle de temps entre les deux passations du test de l'échantillon original de l'année 1991 est inconnu, alors qu'il est d'une semaine pour le groupe de 36 enfants testés en 2003. L'accord inter-juges est très bon, mais le nombre d'évaluateurs et d'enfants est inconnu. En somme, l'information rapportée par Kisker et ses collègues (2011) tend à démontrer la fidélité de l'instrument, ce qui est corroboré par le WSOSPI (2008). Le contenu de l'outil a aussi fait l'objet d'une recension des écrits. Une analyse factorielle a permis de valider trois facteurs et les corrélations sont plus faibles entre les échelles mesurant des construits différents et plus fortes entre des échelles mesurant des construits semblables, mais la taille de l'échantillon n'est pas mentionnée. La validité concomitante est élevée pour quatre domaines, tandis que les corrélations frôlent le seuil de l'acceptabilité pour deux domaines. De son côté, le WSOSPI (2008) considère que la validité concomitante du test est élevée. Enfin, même si le concepteur du test affirme que les corrélations sont « élevées » entre le Brigance Screens et des mesures de performance scolaire, d'intelligence ou de développement du langage cela ne permet pas de statuer sur la validité prédictive du IED-II.

3.3.3.4 Devereux Early Childhood Assessment for infants and toddlers (DECA - I/T, 2007)

Description

Le *DECA - I/T* est un instrument de dépistage et d'évaluation mesurant le développement socioémotionnel des enfants âgés de 1 à 36 mois. Le test permet de mesurer les progrès au fil du temps et de s'assurer que les enfants atteignent les normes du développement social et émotionnel établi pour leur âge. Cet instrument peut être utilisé par des personnes possédant une scolarité équivalant à une 6^e année du primaire et qui ont eu la chance d'observer les comportements de l'enfant au cours des quatre semaines précédant le test, et ce au moins deux heures par jour, deux jours chaque semaine. Cet outil est disponible en anglais et en espagnol et n'est pas disponible en français, selon le site Internet de l'éditeur (Kaplan Early Learning Company).

Contenu de l'outil

L'instrument comprend deux formulaires. Le *DECA-I* possède deux échelles (Initiative [18 items] et Attachement, [15 items]) et est conçu pour les nourrissons âgés de 1 à 18 mois et le *DECA-T* a trois échelles (attachement [18 items], initiative [11 items] et autorégulation [7 items]) et s'adresse plutôt aux tout-petits âgés de 18 à 36 mois.

Passation du test

La passation du test est individuelle et prend environ 20 minutes.

Évaluation et interprétation

La notation permet de comparer les informations afin de déterminer les similitudes et les différences entre deux environnements, soit à la maison ou dans un centre de la petite enfance. Cette information est ensuite utilisée dans l'élaboration des plans d'intervention déployés par la famille et les éducateurs. Plus précisément, les scores bruts sont convertis en rang centile, en score-T, ou en courbe normale équivalente, ce qui permet d'interpréter les résultats et de les comparer avec l'échelle des scores typiques. Les scores-T pour le *DECA-I/T* varient de 28 à 72 et sont interprétés comme suit : ≤ 40 = besoin d'aide; 41 à 59 = développement typique; > 60 = force.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 2 183 enfants âgés de 4 semaines à trois ans. L'échantillon est composé à 45 % de nouveau-nés (enfants, c.-à-d. des enfants âgés de 4 semaines à 18 mois) et à 55 % de tout-petits (toddlers, c.-à-d. des enfants âgés de 18 à 36 mois). Les coefficients de consistance interne du *DECA-I* et du *DECA-T* pour les scores totaux varient de $\alpha = 0,90$ à $0,95$ que ce soient les parents ou les enseignants qui évaluent les enfants, et ils varient de $\alpha = 0,79$ à $0,94$ entre les sous-échelles. Les coefficients du test-retest dans un intervalle d'un à trois jours varient de $r = 0,84$ à $0,99$ pour les deux formulaires et de $r = 0,72$ à $0,99$ entre les sous-échelles, que ce soit les parents ou les enseignants qui ont évalué les enfants. La concordance entre les codeurs (fidélité inter-juges) varie de $r = 0,68$ à $0,74$ pour les scores totaux et de $r = 0,59$ à $0,76$ entre les sous-domaines.

La validité de contenu repose sur une revue de la littérature et des échanges lors de groupes de discussion avec des parents, enseignants et professionnels de la santé mentale œuvrant auprès des enfants. La validité conceptuelle s'appuie sur une analyse factorielle qui valide deux facteurs pour les nourrissons et trois facteurs pour les tout-petits. De plus, les concepteurs du test (Mackrain, LeBuffe, & Powell, 2007) mentionnent que les conclusions sont cohérentes avec le construit théorique du *DECA I/T* : pour les enfants à risque faible et élevé, les scores plus élevés de l'échelle *Protective factor* sont associés avec une meilleure santé sociale et émotionnelle que les scores plus faibles. Les concepteurs (Mackrain et coll., 2007) du *DECA I/T* ont aussi analysé la sensibilité et la spécificité¹⁹ de leur outil afin d'être en mesure de classer les enfants selon les catégories suivantes : besoin d'aide, développement typique ou démontrant des forces. Selon les échelles, la sensibilité varie de 41 à 57 % et la spécificité varie de 80 à 87 %. L'étude de validité concomitante repose sur comparaison avec la version antérieure de l'outil (*DECA*) auprès d'un échantillon de 35 enfants. Les coefficients de corrélations varient de $r = 0,83$ à $0,91$ selon les échelles.

En résumé, les coefficients de consistance interne sont excellents pour les scores totaux et ils sont minimalement bons en ce qui concerne les sous-échelles où nous avons constaté un peu plus de variabilité chez les parents comparativement aux enseignants, cependant la taille de l'échantillon est inconnue. Les coefficients de stabilité de l'outil (test-retest) varient d'acceptables à excellents, quoique l'intervalle de temps entre les deux passations paraît un peu court. Notons que les coefficients entre les sous-échelles du *DECA-T* indiquent plus de variabilité lorsque ce sont les enseignants qui complètent le test. Le niveau d'accord entre les juges est acceptable. Toutefois, la méthode utilisée pour déterminer le niveau d'accord entre les parents (paires de parents) est perfectible et l'échantillon d'enfants évalués par les deux enseignants est inconnu. En somme, l'information rapportée par Kisker et ses collègues (2011) tend à démontrer la fidélité du *DECA-I/T*. Ajoutons que les concepteurs ont validé le contenu à l'aide d'une recension des écrits scientifiques et les construits avec des analyses factorielles, mais que la sensibilité de l'instrument apparaît faible, ce qui soulève des doutes quant à sa validité intrinsèque. Les corrélations de validité concomitante sont élevées à très élevées, mais reposent sur une comparaison avec le *DECA* seulement à partir d'un échantillon assez restreint, alors qu'il aurait été souhaitable de comparer le *DECA-I/T* à d'autres outils mesurant les mêmes concepts. Enfin, aucune information n'est rapportée pour la validité prédictive.

3.3.3.5 MacArthur-Bates Communicative Development Inventory, 2nd Edition (CDI, 2007)

Description

Le *CDI* est un outil d'évaluation américain permettant d'évaluer les compétences communicationnelles des enfants âgés de 8 à 37 mois, notamment les premiers signes de compréhensions et les premiers signaux gestuels non verbaux. Le *CDI* peut être utilisé pour dépister des retards de développement du langage, identifier les habiletés plus problématiques, déterminer des stratégies d'intervention et évaluer les résultats d'intervention.

Les parents peuvent employer cet outil sans avoir à suivre une formation sur les modalités d'utilisation, mais le manuel recommande que ce soient des professionnels qui interprètent les résultats. Selon le site Internet du concepteur (Paul H. Brookes Publishing Company), l'instrument, en plus de la version anglaise, est aussi disponible en français sous le nom *Inventaires MacArthur du développement de la communication*.

¹⁹ Kisker et ses collègues (2011) ont rapporté la sensibilité et la spécificité pour cet outil seulement.

Contenu de l'outil

Le CDI comprend trois inventaires pour trois groupes d'âge distincts, et chaque inventaire est composé de plusieurs sous-tests :

- Le **CDI : Words and Gestures** (pour les tout-petits âgés de 8 à 18 mois) comprend deux parties qui évaluent la production et la compréhension de vocabulaire.
- Le **CDI : Words and Sentences** (pour les tout-petits âgés de 16 à 30 mois) comprend deux parties qui évaluent la production du vocabulaire et l'acquisition des règles de grammaire.
- Le **CDI III** (pour les tout-petits âgés de 30 à 37 mois) comprend trois parties qui mesurent le vocabulaire expressif (*expressive vocabulary*) et la grammaire.

Des versions plus courtes sont disponibles pour les deux premiers inventaires.

Passation du test

La passation du test se fait de manière individuelle et prend environ 20 à 40 minutes à compléter. Les versions les plus courtes prennent environ 10 minutes à compléter.

Évaluation et interprétation

Les questionnaires contiennent des questions fermées et ouvertes et les choix de réponse diffèrent selon les inventaires et leurs sous-échelles (p. ex., « comprend » ou « comprend et prononce », « parfois » ou « souvent »). Les ouvrages de référence indiquent la procédure à suivre pour calculer et interpréter les scores. Sauf exception, des rangs centiles distincts pour les garçons et les filles sont précisés pour presque tous les inventaires en fonction de différents groupes d'âge, ainsi la performance de l'enfant est comparée à celle d'enfants du même groupe d'âge. Lorsque le rang centile n'est pas donné, le pourcentage de bonnes réponses est calculé et comparé avec le pourcentage indiqué dans le manuel.

Les valeurs psychométriques

Pour les deux *CDI (Words and Gestures et Words and Sentences)*, les tests statistiques portent sur un échantillon de 2 550 enfants âgés de 8 à 30 mois, et sur un échantillon de 356 enfants âgés de 30 à 37 mois pour le *CDI-III*. Les coefficients de consistance interne de chaque inventaire se situent entre $\alpha = 0,86$ et $0,96$ selon les échelles. Certaines sous-catégories sémantiques affichent toutefois des scores $\alpha < 0,70$: deux dans l'inventaire *Words and Gestures (Words about Time et Question Words)* et deux dans l'inventaire *Words and Sentences (Sound Effects and Animal Sounds et Connecting Words)*. Les coefficients de consistance interne des formulaires abrégés sont de $\alpha = 0,97$ pour le *CDI : Words and Gestures* et de $\alpha = 0,99$ pour le *CDI : Words and Sentences*. La double passation du test s'est faite dans un intervalle de 1,4 mois. Les corrélations test-retest du *CDI : Words and Gestures* se situent entre $r = 0,80$ et $0,90$ pour chacune des deux dimensions (*Words produced et Words understood*) pour les enfants âgés de 0 à 12 mois ($n = 137$), alors qu'elles sont de $r = 0,60$ pour les enfants de 12 mois. Les coefficients test-retest du *CDI : Words and Sentences, vocabulary production* varient de $r = 0,62$ ($n = 228$) à $r = 0,95$ ($n = 216$). De plus, ils ont vérifié la concordance entre les deux inventaires en effectuant deux mesures auprès d'un échantillon de 217 enfants : la première entre l'âge de 10 et 16 mois et la deuxième entre l'âge de 16 et 25 mois. La corrélation entre les deux inventaires est de $r = 0,69$. Enfin, les coefficients pour les échelles des formulaires abrégés (niveaux 1 et 2) sont supérieurs à $r = 0,74$ ($n = 28$ à 50).

Au sujet du contenu du test, Kisker et ses collègues (2011) rapportent que les concepteurs (Fenson, Marchman, Dale, Reznick, Thal & Bates, 2007) du test disent avoir conçu les items de chaque sous-test à partir de la documentation sur le développement des enfants. Les corrélations entre *CDI : Words and Gestures (Total gesture scores)* et quatre échelles de cet inventaire varient de $r = 0,59$ à $0,80$, alors que les corrélations varient de $r = 0,78$ à $0,83$ entre le *CDI : Words and Sentences (Words produced)* et trois autres échelles de cet inventaire. La taille des échantillons est cependant inconnue. Pour établir la validité concomitante, les deux inventaires de même que les formulaires abrégés du CDI ont été comparés à d'autres outils mesurant les mêmes concepts. Pour le *CDI : Words and Gestures*, les coefficients de corrélation varient de $r = 0,51$ à $0,87$ selon l'échelle de l'inventaire et l'un ou l'autre des quatre tests de comparaison. Les coefficients du *CDI : Words and Sentences* varient de $r = 0,40$ à $0,88$ selon l'un ou l'autre des trois autres tests de comparaison. Notons que la taille des échantillons est inconnue pour les tests effectués avec ces deux inventaires. Pour les formulaires abrégés, les coefficients varient entre $r = 0,23$ et $0,65$ selon les instruments avec lesquels ils ont été comparés et la taille des échantillons est inconnue. Enfin, les coefficients entre le *CDI-III Vocabulary Checklist* varient de $r = 0,47$ à $0,63$ avec le PLS-3 ($n = 19$), de $r = 0,45$ à $0,63$ avec le *Peabody Picture Vocabulary Test-Revised* ($n = 22$); et deux études distinctes rapportent des corrélations avec le *McCarthy Scales* : les coefficients varient de $r = 0,44$ à $0,62$ pour la première ($n = 85$), et de $r = 0,52$ à $0,56$ pour la deuxième ($n = 113$).

En résumé, la consistance interne du test est excellente, mais il est à noter que les formulaires abrégés qui présentent des $\alpha \geq 0,97$, que certaines sous-catégories présentent des $\alpha < 0,70$ et que la taille des échantillons est inconnue, ce qui limite l'interprétation de l'information. La stabilité du *CDI : Words and Gestures* est généralement bonne pour les enfants âgés de 0 à 12 mois, cependant les corrélations sont discutables pour le groupe d'enfants âgé de 12 dont la taille de l'échantillon est inconnue. Les mesures de stabilité du *CDI : Words and Sentences* reposent sur une taille d'échantillon convenable et les corrélations sont généralement bonnes, hormis l'échelle Complexity qui affiche une moins bonne stabilité. Il est à noter que la corrélation entre les deux inventaires est plus qu'acceptable, compte tenu des concepts mesurés et de la progression des enfants qui peut varier grandement d'un enfant à l'autre. Quant aux formulaires abrégés, la fidélité test-retest varie d'acceptable à excellente. Notons qu'aucune mesure n'est rapportée en ce qui concerne l'accord inter-juges. En somme, l'information rapportée par Kisker et ses collègues (2011) tend à démontrer la fidélité du CDI. Les contenus du test reposent sur une documentation sur le développement qui est peu décrite et aucune méthode de validité conceptuelle n'est rapportée à l'exception des corrélations variant de modérées à élevées entre les échelles des inventaires. La validité concomitante entre les deux principaux inventaires et d'autres instruments mesurant la même caractéristique varie d'acceptable à forte, mais la taille des échantillons est inférieure à 30. Les mesures de validité concomitante du *CDI-III* sont plus mitigées et celles des formulaires abrégés sont inférieures au seuil jugé acceptable. Enfin, aucune mesure de validité prédictive n'est présentée pour ce test.

3.3.3.6 Peabody Picture Vocabulary Test, 4th edition (PPVT-IV, 2007)

Description

Le *PPVT-4* est un instrument d'évaluation américain normé conçu pour mesurer la compétence langagière des personnes âgées de 2 ans et 6 mois à plus de 90 ans. Le test peut aussi être utilisé pour mesurer l'apprentissage chez les personnes dont l'anglais n'est pas la langue maternelle, déterminer le niveau et les contenus pour l'enseignement et identifier les déficits linguistiques causés par une blessure ou une maladie. Cet instrument peut être utilisé par des personnes détenant un diplôme de premier cycle universitaire et formées en mesure et évaluation. Il est également suggéré que les évaluateurs aient une formation en développement de la parole et du langage. Cet outil est disponible en anglais et en espagnol. Selon l'INSPQ (2010), la version française du *PPVT-4* ne semble pas disponible, mais il existe une adaptation du *Peabody Picture Vocabulary Test-Revised* (*PPVT-R*) qui se nomme Échelle de vocabulaire en images Peabody (*EVIP*, 1993).

Contenu de l'outil

Le *PPVT-4* est séparé en deux formulaires (A et B) qui contiennent chacun des éléments de pratique conçus selon l'âge des enfants. Le test comprend 228 items répartis entre 19 groupes d'items. Dans chaque groupe, les 12 items sont placés en ordre croissant de difficulté.

Passation du test

La passation du test est individuelle. Il prend environ 10 à 15 minutes pour 60 items.

Évaluation et interprétation

Des tables permettent de convertir les scores bruts en scores normalisés (c.-à-d. niveau de développement, âge de développement ou niveau scolaire) ou bien en scores non normalisés qui permettent de mesurer les améliorations au fil du temps. Le manuel décrit brièvement chaque score ainsi que son utilité et ses limites.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 3 540 personnes âgées de 2 ans et 6 mois à plus de 90 ans (les sous-échantillons sont composés de 100 à 200 individus pour les 2 à 22 ans). Pour le formulaire A, les mesures de consistance interne sont de $r = 0,89$ à $0,97^{20}$ et de $\alpha = 0,93$ à $0,98$; tandis qu'elles sont de $r = 0,91$ à $0,97$ et de $\alpha = 0,94$ à $0,97$ pour le formulaire B. Les mesures test-retest ont été effectuées avec un sous-groupe de 340 enfants âgés de 2 à 14 ans dans un intervalle de quatre semaines. Les corrélations varient entre $r = 0,91$ et $0,94$.

²⁰ Méthode des moitiés de Spearman-Brown qui consiste à déterminer si la variabilité entre deux moitiés d'un même test est égale (Bernier & Pietrulewicz, 1997).

La validité de contenu repose sur une consultation de différents documents de référence tels des dictionnaires et des ouvrages lexicographiques. Les concepteurs (Dunn & Dunn, 2007) de l'outil ont mesuré la difficulté des items en utilisant des méthodes classiques et de Rasch.²¹ Kisker et ses collègues (2011) rapportent des résultats comparant le *PPVT-4* à cinq autres instruments mesurant sensiblement les mêmes concepts, dont sa version antérieure, le *PPVT-III*, à partir d'échantillons variant de 110 à 425 enfants âgés de 2 à 24 ans (mais plus souvent d'âge primaire ou secondaire), à l'exception du *EVT-2* dont l'échantillon est de 3 540 enfants. Les corrélations démontrent une grande variabilité allant de $r = 0,27$ à $0,79$ avec le *GRADE* (vocabulaire et score composé de la compréhension) à $r = 0,80$ à $0,84$ avec le *Expressive Vocabulary Test, Second Edition* (*EVT-2*). Tous les tests ont été administrés le même jour hormis le *PPVT-III* qui a pu être administré jusqu'à 11 jours plus tard. La validité prédictive rapportée par Kisker et ses collègues (2011) est celle de la seconde version du test (*PPVT-R*). Les résultats de six études distinctes menées auprès d'enfants du préscolaire allant jusqu'à la 5^e année démontrent des corrélations variant de $r = 0,14$ à $0,66$ entre le *PPVT-R* et différents résultats attestant soit de la performance scolaire, du langage ou d'autres résultats.

En résumé, la consistance interne apparaît être excellente. Ce test s'avère également être très stable, et l'intervalle entre les deux passations est adéquat. Aucune mesure d'accord inter-juges n'a été rapportée. La validité de contenu se limite à une consultation des ouvrages de référence et les construits ont été validés à l'aide de méthodes permettant de déterminer le niveau de difficulté des items, toutefois ni ces méthodes ni la taille de l'échantillon ne sont rapportées. Les mesures de validité concomitante reposent sur des tailles d'échantillon suffisantes, mais les corrélations varient de faibles à élevées. Enfin, l'information sur la validité prédictive porte sur une version antérieure du test.

3.3.3.7 Vineland Adaptive Behavior Scales II Edition (VABS II, 2005)

Description

Le *VABS II* est un outil d'évaluation américain standardisé conçu pour mesurer les habiletés fonctionnelles (*daily functioning*) des personnes âgées de 0 à 90 ans. Le test permet notamment de mesurer le fonctionnement quotidien de l'individu, les retards de développement, les troubles émotionnels et comportementaux ainsi que d'autres troubles mentaux. L'instrument peut également être utilisé pour effectuer une évaluation développementale, suivre le progrès, planifier un programme ou effectuer des recherches. Il est recommandé que l'utilisateur détienne un diplôme en psychologie (ou dans un domaine connexe à la relation d'aide) ainsi qu'une expérience en évaluation et en interprétation des résultats. Cet instrument de mesure est disponible en anglais et en espagnol et n'est pas disponible en français, selon l'INSPQ (2010).

Contenu de l'outil

L'instrument contient quatre formulaires distincts : trois d'entre eux s'adressent aux parents et l'autre aux enseignants. Deux des trois formulaires conçus pour les parents reposent sur des questions ouvertes (*Survey Interview Form* et *Expanded Interview Form*) alors que le *Parent/Caregiver Rating Form* utilise une échelle de réponse. Les 383 items du *VABS II* sont répartis en 11 sous-domaines appartenant à l'un ou l'autre des quatre domaines généraux : habiletés motrices, communication, socialisation et activité de la vie quotidienne. Le *Maladaptive Behavior Index* (*Internalizing, Externalizing* et autre) est optionnel et contient 50 items additionnels.

Passation du test

Le *Survey Interview Form* prend environ de 20 à 60 minutes à compléter alors que le *Parent/Caregiver Rating Form* prend environ 30 à 60 minutes.

²¹ Cette méthode permet notamment l'étude de la qualité psychométrique des items d'un test et le calibrage des items pour les tests adaptatifs (Boadé, 2013).

Évaluation et interprétation

Pour chaque item, l'évaluateur précise si la récurrence est fréquente (note de 2), peu fréquente (note de 1), si elle ne se produit jamais (note de 0), ou si l'information n'est pas connue. Certains items du *Maladaptive Behavior Index* exigent que l'évaluateur indique si le comportement est modéré ou sévère. Les scores des sous-domaines sont additionnés pour calculer les scores composés de chacun des quatre domaines et ceux-ci sont ensuite additionnés pour calculer le composé du comportement adaptatif des enfants âgés de 0 à 6 ans et 11 mois. Le manuel contient un chapitre permettant d'interpréter les scores et d'identifier les forces et les faiblesses de la personne évaluée. Les scores bruts peuvent être convertis en : scores dérivés du Vineland-II, scores standards, scores d'échelle-V, rang centile, équivalence d'âge, stannines²² et il est possible de calculer des intervalles de confiance. Les résultats sont décrits en « niveaux d'adaptation » ou en « niveaux d'inadaptation » (*Adaptive/Maladaptive levels*), ceux-ci constituant des catégories descriptives qui informent l'évaluateur au sujet des résultats du test. Les niveaux d'inadaptation sont interprétés comme étant moyens, élevés ou cliniquement significatifs.

Les valeurs psychométriques

Les tests statistiques portent sur un échantillon de 3 695 personnes âgées de 0 à 90 ans. Pour les enfants âgés de 0 à 6 ans, les coefficients de consistance interne des échelles (Spearman-Brown) sont supérieurs à $r = 0,79$ et le coefficient du *Maladaptive Behavior Index* est quant à lui de $\alpha = 0,88$ pour les 3 à 5 ans. L'intervalle moyen entre les deux passations du test est de 18 à 21 jours selon les groupes d'âge et les scores test-retest varient quant à eux de $r = 0,86$ à $r = 0,96$ selon les échelles et les groupes d'âge. Pour le *Parent/Caregiver Checklist*, la concordance entre les 152 évaluateurs (surtout des paires de parents évaluant un enfant) varie de $r = 0,61$ à $0,82$ selon les sous-échelles, tandis que la concordance varie de $r = 0,58$ à $0,82$ selon les sous-échelles du *Survey Interview Form*. Les corrélations des scores composés pour les trois formulaires (les deux précédents et le *Maladaptive Behavior Index*) sont quant à elles supérieures à $r = 0,79$.

En ce qui concerne la validité conceptuelle de l'outil, les concepteurs ont effectué une analyse factorielle confirmatoire qui permet de valider quatre facteurs pour les enfants de 6 ans et moins. De plus, les coefficients de corrélations entre les *Adaptive Behavior Domains* et le *Adaptive Behavior Composite* varient de $r = 0,82$ à $r = 0,88$ selon les domaines et les groupes d'âge. Les corrélations entre les échelles varient de $r = 0,56$ (entre les échelles *Communication et Motor Skills*) à $r = 0,73$ (entre les échelles *Daily Living Skills et Socialization*), selon les groupes d'âge. Le *VABS-II* a également été comparé à trois autres instruments, dont la première version du test de 1984. Lorsque comparé à ce dernier, les corrélations entre les scores composés sont d'au moins $r = 0,82$ tandis que, selon les échelles, les corrélations varient de $r = 0,65$ à $0,91$ pour les enfants âgés de 0 à 2 ans et 11 mois et de $r = 0,85$ à $0,94$ pour les enfants âgés de 3 à 6 ans. Les corrélations sont de $r = 0,63$ avec le *ABAS-II* et elles ne dépassent pas $r = 0,59$ pour ce qui est des différentes échelles du *BASC-2*.

En résumé, les données démontrent des coefficients élevés et très élevés pour la consistance interne du test, mais les tailles d'échantillon sont inconnues. Idem pour la stabilité du test dont l'intervalle de temps entre les deux passations est adéquat. Pour attester de l'accord inter-juges du *Parent/Caregiver Checklist* l'approche par paires d'évaluateurs, malgré le nombre élevé d'évaluateurs, est perfectible, et les méthodes pour le *Survey Interview Form* ne sont pas explicitées. De plus, aucune méthode de validité de contenu n'est rapportée, mais la validité conceptuelle repose sur une analyse factorielle confirmatoire, dont la taille d'échantillon n'est pas rapportée. Pour la validité concomitante, les corrélations sont acceptables ou fortes avec la version précédente du test et avec le *ABAS-II*, tandis qu'elles sont plutôt faibles avec le *BASC-2*. D'ailleurs, le WSOSPI (2008) considère que la validité concomitante du test est acceptable. Enfin, aucune mesure de validité prédictive n'est rapportée.

Le tableau 5 (page suivante) permet de voir en un coup d'œil l'évaluation des qualités psychométriques des 10 outils décrits ci-dessus.

²² Une méthode de mise à l'échelle des résultats des tests sur une échelle standard de neuf points avec une moyenne de cinq et un écart type de deux.

Tableau 5 : Résumé des qualités psychométriques de 10 outils d'évaluation

Outils	Échantillon	Méthodes de fidélité			Méthodes de validité			
		Consistance interne	Test-retest	Inter-juges	Contenu	Conceptuelle	Concomitante	Prédictive
Preschool Language Scale 4 (PLS-4, 2002)	1 534	++	++	?	++	?	++	?
Bayley Scales Of Infant and Toddler Development, Third Edition (2005)	1 700	++	+	?	++	++	++	?
Developmental Profile 3 (DP-3, 2007)	2 216	++	+	?	?	++	++	?
Battelle Developmental Inventory (BDI, 1984)	800	?	+	?	++	++	+	?
Behavior Assessment System for Children 2 (BASC-2, 2004)	13 000	++	+	?	?	++	+	?
Brigance Diagnostic Inventory of Early Development 2 (IED-2, 2004)	1 171	+	+	++	+	++	+	?
Devereux Early Childhood Assessment for infants and toddlers (DECA-I/T, 2007)	2183	++	+	+	+	?	+	?
MacArthur-Bates Communicative Development Inventory 2nd Edition (CDI, 2007)	2 550	+	+	?	?	?	+	?
Peabody Picture Vocabulary Test, 4th edition (PPVT-4, 2007)	3 540	+	++	?	+	+	?	?
Vineland Adaptive Behavior Scales 2 nd Edition (VABS-2, 2005)	3 695	++	++	?	?	++	+	?

Légende. Qualités psychométriques : très favorables (++), favorables (+), inconnues ou mitigées (?), inacceptables (—).

3.4 Disponibilité des instruments validés en français et en anglais

En ce qui a trait à la langue de disponibilité des 62 outils retenus pour cette RS, ils sont disponibles en anglais et seulement cinq d'entre eux semblent avoir été traduits en français. Dionne et ses collègues (2001) indiquent que les outils d'évaluation disponibles en français habituellement utilisés en intervention précoce datent de plusieurs années ou couvrent peu la petite enfance. Toutefois, dans le cadre de cette RS, nous avons identifié cinq outils traduits en français qui s'adressent aux tout-petits (< 6 ans), mais qui datent d'au plus 19 ans. Parmi ces outils d'évaluation traduits en français²³ :

- Un seul est de nature diagnostique :
 - *Infant Toddler Social and Emotional Assessment*/Évaluation sociale et émotionnelle de jeunes enfants (1995).
- Deux sont des outils d'intervention ou de programmation :
 - *Assessment, Evaluation and Programming System* / Évaluation, intervention, suivi (EIS) : 0-3 ans (2002).
 - *Social Competence and behavior evaluation*/Profil socio-affectif (1995).
- Deux outils ont une double fonction :
 - *Achenbach System of empirically based assessment* / Système d'évaluation empirique Achenbach (1999).
 - *MacArthur-Bates Communicative Development Inventory*/Inventaires MacArthur du développement de la Communication (2007).

3.5 Formation requise pour utiliser les outils

Les concepteurs de plusieurs outils sélectionnés ($n = 39$) indiquent qu'une formation universitaire, une formation pointue en développement des enfants ou une grande expérience en évaluation est nécessaire pour utiliser l'instrument ou pour interpréter les résultats (p. ex. *Preschool Language Scale*, 2002; *Mullen Scales of Early Learning*, 1995). De manière générale, le diplôme universitaire exigé pour l'utilisation de ces outils correspond souvent à un champ de spécialisation tel que l'enseignement, la psychologie, l'orthophonie ou l'ergothérapie.²⁴

Des 62 outils rapportés dans cette RS, 10 peuvent être utilisés par un éducateur spécialisé, un technicien ou une personne qui a préalablement reçu une formation sur l'outil (p. ex. *Assessment, Evaluation and Programming System*, 2002 ; *Battelle Developmental Inventory*, 2004) tandis que 13 outils ne requièrent aucun diplôme particulier ni de formation spécifique à l'utilisation (p. ex. *Brigance Comprehensive Inventory of Basic Skills-Revised*, 1999; *Infant/Toddler Sensory Profile : user's manual*, 2002). Par contre, certains concepteurs d'outils indiquent bel et bien que l'administration du test ne nécessite pas de formation particulière, mais précisent que l'interprétation des résultats devrait être réservée aux personnes expérimentées ou diplômées dans un champ de spécialisation approprié : c'est le cas du *Behavior Assessment System for Children* (2004), par exemple.

4. DISCUSSION

Cette revue systématique servira à alimenter la rédaction d'un guide de pratique s'adressant aux différents praticiens travaillant avec des enfants présentant un RGD, ce qui explique la posture plutôt pratique qu'académique de ce document. Ce positionnement, ainsi que le temps et les ressources limitées impartis à la réalisation de ce projet expliquent la démarche particulière qui a été conçue spécifiquement pour cette RS, qui repose en grande partie sur de l'information publiée dans cinq documents de référence, ceux-ci constituant, en quelque sorte, des recueils d'outils d'évaluation.

²³ Des informations supplémentaires sur chacun de ces instruments sont consignées aux Annexes E et F.

²⁴ Les psychoéducateurs et psychoéducatrices pourraient faire partie de cette liste de professions. L'absence de cette profession de l'énumération est surtout attribuable au fait que les outils proviennent la plupart du temps des États-Unis et que la psychoéducation n'est pas nommée comme telle.

Cette revue systématique a permis de repérer un nombre important d'outils d'évaluation pouvant être utilisés auprès d'enfants âgés de 2 à 5 ans, dont un peu plus du tiers qui n'évalue qu'une seule sphère de développement. En conséquence, les professionnels seront sans doute appelés à recourir à plusieurs outils qui évaluent des domaines particuliers du développement (OPPQ, 2013). Il est important de préciser que les outils ne permettent pas de diagnostiquer le RGD, mais plutôt de déterminer si l'enfant a effectivement un retard de développement dans une ou plusieurs sphères. Il appartiendra donc aux praticiens de déterminer si les résultats d'un enfant à un ou à plusieurs tests l'amènent à croire que l'enfant présente un RGD ou non. L'information présentée ici vise donc à soutenir les différents professionnels lorsque survient le moment de faire un choix éclairé sur un outil à employer dans le cadre de leurs fonctions, et qu'en ce sens ce document remplit bien sa fonction.²⁵

Les outils validés

Des 177 outils répertoriés spécifiques à la population visée par cette RS, 62 sont décrits de manière succincte en annexe. De ces outils, dix ont répondu aux critères de sélection et ont, par conséquent, été décrits davantage et évalués plus en profondeur. De plus, plusieurs instruments sont fréquemment utilisés au Québec (p. ex. Brigance [2004], Bayley Scales [2005], VABS-II [2005]) (Pépin et coll., 2006; OPPQ, 2013). C'est aussi le cas de l'Évaluation, intervention, suivie (EIS) : 0-3 ans (2002) qui est couramment utilisé dans les CRDITED de la province. Toutefois, comme pour plusieurs autres instruments, cet instrument a fait l'objet d'études de validation, mais il n'a pas été décrit et évalué davantage dans le cadre de cette RS, une situation qui est attribuable à la méthodologie particulière de cette RS.

Dans l'ensemble, seulement cinq outils traduits en français s'adressant aux tout-petits (< 6 ans) ont été repérés dans cette RS, limitant ainsi le potentiel d'utilisation auprès d'une population francophone. Cette observation corrobore celle de Dionne et ses collègues (2001) indiquant que les outils d'évaluation disponibles en français habituellement utilisés en intervention précoce couvrent peu la petite enfance. Ajoutons que la majorité des 62 outils présentés s'adressent à des professionnels détenant une formation universitaire, que dix d'entre eux peuvent être utilisés par un éducateur spécialisé, un technicien ou une personne qui a préalablement reçu une formation sur l'outil et que treize autres outils ne requièrent aucun diplôme particulier ni de formation spécifique à l'utilisation. Notons également qu'environ le tiers des 62 outils prévoient une participation des parents lors de la collecte de données. Enfin, la moitié des instruments est de type évaluation directe individuelle de l'enfant uniquement, alors que 18 outils combinent plus d'un type d'évaluation (p. ex., évaluation directe individuelle de l'enfant et questionnaire complété par les parents). Afin d'éviter d'obtenir des données erronées, il pourrait s'avérer nécessaire de s'assurer que les utilisateurs, en particulier les parents, aient un niveau de littératie suffisant pour utiliser les instruments qui leurs sont destinés. Ainsi, il pourrait être nécessaire de former les utilisateurs peu familiers avec les outils d'évaluation (incluant des éducateurs, des enseignants, etc.) à employer ces derniers conformément.

Les dix outils plus abondamment décrits dans cette RS présentent des qualités psychométriques favorables. Bien que les données permettent d'attester de la fidélité de ces instruments, il s'avère souvent que les concepteurs ont utilisé des « paires de parents » pour évaluer leur enfant, ce qui constitue une méthode peu efficace pour attester de la fidélité inter-juges. De plus, seules les données test-retest étaient accessibles pour le *Battelle Developmental Inventory* (1984). La validité intrinsèque de ces dix instruments est aussi généralement démontrée. Pour quatre instruments la démonstration de validité intrinsèque repose sur des méthodes de validité de contenu et de validité de construit, pour cinq autres tests cette démonstration repose sur l'une ou l'autre des méthodes, alors que nous n'avons pas suffisamment d'information pour nous prononcer sur la validité intrinsèque du *MacArthur-Bates Communicative Development Inventory* 2nd Edition (2007). Les données de performance clinique (c.-à-d. la sensibilité et la spécificité) n'étaient quant à elles disponibles que pour le *Devereux Early Childhood Assessment for infants and toddlers* (2007), et la sensibilité de l'instrument semble faible. La validité extrinsèque des outils repose quant à elle uniquement sur la validité concomitante puisque la validité prédictive n'était pas étudiée par les concepteurs des tests ou parce que l'information portait sur une version antérieure de l'outil (qui date parfois de plus de 40 ans). Cela dit, la validité concomitante avec d'autres tests est généralement bonne, sauf pour le *Peabody Picture Vocabulary Test, 4th edition* (2007) dont les résultats sont plus mitigés. Enfin, l'absence de la taille de l'échantillon pour certains tests de fidélité ou de validité limitait parfois l'interprétation des scores présentés.

²⁵ Rappelons que l'Annexe E présente un tableau récapitulatif des 62 outils d'évaluation en indiquant les langues de publication ainsi que des précisions sur les professionnels jugés aptes par les concepteurs à les utiliser, ainsi que des indications sur des formations spécifiques requises, s'il y a lieu. L'Annexe F présente les objectifs de ces outils.

Les forces et les limites de la revue systématique

La principale force de cette revue systématique repose sur le nombre de sources utilisées pour dresser une liste exhaustive d'outils d'évaluation : le catalogue REGARD, cinq documents de référence et les connaissances inédites des membres du comité scientifique représentant le volet clinique. D'ailleurs, une collaboration étroite avec tous les membres du comité scientifique a permis d'assurer la rigueur et la pertinence de la démarche.

La principale limite de cette présente RS réside dans le fait que les données psychométriques proviennent de sources secondaires, ce qui pourrait engendrer un biais d'information, notamment parce qu'il est impossible de savoir si toute l'information est rapportée et si elle est fidèlement rapportée. D'ailleurs, les auteurs des documents retenus ne présentent aucune information au sujet de la qualité méthodologique des documents qu'ils ont eux-mêmes consultés pour établir les qualités psychométriques. Idéalement, il aurait été préférable d'effectuer des stratégies de recherches documentaires pour chacun des 177 outils répondant aux critères d'inclusion du PICOTS afin d'identifier les études de validation primaires originales. Cet exercice nous aurait sans doute permis d'évaluer directement la qualité des méthodes sur lesquelles reposent les données psychométriques des outils et sa démonstration et aurait constitué une force supplémentaire de cette RS. Toutefois, compte tenu des ressources et du temps alloués, il a été impossible de réaliser une telle démarche.

De plus, d'avoir adapté un instrument d'évaluation méthodologique (AGREE II) pour l'évaluation des cinq documents de référence sans en avoir fait une validation spécifique pour l'évaluation des documents sélectionnés constitue une limite supplémentaire de cette RS.

La provenance des documents de référence retenus constitue une autre limite. Quatre des cinq documents ont été rédigés aux États-Unis et, considérant que la plupart des outils d'évaluation proviennent souvent de ce pays, cela pourrait laisser croire que leur potentiel d'application au Québec est limité. Par exemple, certains concepteurs d'outils d'évaluation indiquent que leur instrument a été conçu pour déterminer l'admissibilité des enfants au programme d'intervention précoce *Head Start* offert aux États-Unis et que la validation a souvent été réalisée auprès d'enfants participant déjà à ce programme.

Ceci étant dit, la démarche utilisée a permis de répondre de manière adéquate à la question de recherche principale qui consistait à identifier les outils qui démontrent des qualités psychométriques favorables s'adressant aux enfants âgés de 2 à 5 ans ayant ou non un RGD.

5. CONCLUSION

L'évaluation des jeunes enfants est une composante majeure des services d'intervention précoce (Rousseau et coll., 2012). Pour cette raison, il est important d'identifier des outils d'évaluation affichant des qualités psychométriques favorables, ce que cette revue systématique a pu mettre de l'avant à l'aide d'une démarche particulière s'appuyant autant sur les suggestions des membres du comité scientifique représentant le volet clinique que sur quelques documents de référence. Ces documents ont d'ailleurs permis de caractériser ces outils en présentant leurs particularités, d'identifier les instruments de mesure qui ont été traduits en français et de déceler ceux dont l'utilisation requiert des compétences particulières. Bref, la démarche devrait permettre aux lecteurs de porter un jugement étayé lorsque vient le moment de choisir un outil d'évaluation. Cela dit, nous croyons également que :

« des travaux visant la poursuite de la validation d'instruments de mesure sont essentiels afin de s'assurer de disposer d'outils adaptés, valides et standardisés pour tous les enfants » (Dionne et coll., 2010, p. 118).

Références

AGREE Next Steps Research Consortium. Grille d'évaluation de la qualité des recommandations pour la pratique clinique (Grille AGREE II). AGREE Research Trust; 2009. Disponible à : <http://www.agreetrust.org/resource-centre/agree-ii-translations/>.

Alpern, G. (2007). *Developmental profile 3 : Manual*. Wps publish.

Bagnato, S.J., Neisworth, J.T., Pretti — Frontczak, K. (2010). *LINKING Authentic Assessment & Early Childhood Intervention*. Best measures for best practices. Second Edition. Baltimore : Paul H. Brookes Publishing Co. 380 pages.

Bernier, J-J. et Pietrullewicz, B. (1997). *La psychométrie : Traité de mesure appliquée*. Montréal : Gaëtan Morin Éditeur Ltée.

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Behavior Assessment System for Children -Second Edition (BASC2)*. Edmonton, Alberta, Canada.

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Brigance Inventory of Early Development II (IED-II)*. Edmonton, Alberta, Canada.

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Edmonton, Alberta, Canada.

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Vineland Adaptive Behavior Scales-Second Edition (Vineland-II)*. Edmonton, Alberta, Canada.

Craig A. Albers & Adam J. Grieve. (2007). Test Review : Bayley, N. (2006). Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, TX : Harcourt Assessment. *Journal of Psychoeducational Assessment*, 25, 100-190.

Dionne, C., McKinnon, S. et Squires, J. (2010). Le dépistage des retards de développement chez les jeunes enfants d'une communauté des Premières Nations. *First Peoples Child & Family Review*, 5(2),117-123.

Dionne, C., Bricker, Diane, Harguidéguy-Lincourt, M.-C., Rivest, C., Tavares, C.-A. (2001). Présentation d'un outil d'évaluation et d'intervention pour jeunes enfants : Système intégré d'évaluation de l'enfant, d'évaluation de l'intervention et de programmation (AEPS). *Revue francophone de la déficience intellectuelle*, 12(1), 21-29.

Gaul Bouchar, M-E., Fitzpatrick, E.M. & Olds, J. (2009). Analyse psychométrique d'outils d'évaluation utilisés auprès des enfants francophones. *Revue canadienne d'orthophonie et d'audiologie*, 33 (3), 129-139.

Halle, T., Zaslow, M., Wessel, J., Moodie, S., and Darling-Churchill, K. (2011). *Understanding and Choosing Assessments and Developmental Screeners for Young Children: Profiles of Selected Measures*. Washington, DC : Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. 196 pages.

Hutchins, T. (2013). MacArthur-Bates Communicative Development Inventories, Second Edition. *Encyclopedia of Autism Spectrum Disorders*, Volkmar (ed.), 1773-1779.

Beauregard, D., L. Comeau et J. Poissant (2010). *Avis scientifique sur le choix d'un outil de mesure du développement des enfants de 0 à 5 ans dans le cadre des services intégrés en périnatalité et pour la petite enfance*. Institut national de santé publique du Québec, Direction du développement des individus et des communautés. 78 pages.

Kern, S. (2006). Étalonage pilote français du P.L.S.-4 concernant la tranche d'âge 30 — 53 mois. (Thèse de maîtrise inédite). Université CLAUDE-BERNARD, Lyon, France.

Kisker, E. E., B. Kimberly, C. Cabili, C. Nagatoshi, C. Kamler, C. J. Johnson, Y. Xue, A. M. Esposito, J. Henderson, S. Asheer, C. Sciarrino, V. Jethwani, T. Zavitsky, M. Ford, J. M. Love (2011). *Resources for measuring services and outcomes in head start programs serving infants and toddlers*. Mathematica Policy Research Inc. 348 pages.

Laurencelle, L. (1998). *Théorie et techniques de la mesure instrumentale*. Québec : Presses de l'Université du Québec. 267 pages.

Pépin, G., Gascon, H., Beaupré, P., Tétreault, S., Dionne, C., Roy, S., Lapointe, C., Leroux, A et Ruel, J. (2006). Les effets des programmes d'intervention pour les enfants de la naissance à sept ans présentant un retard global de développement et recevant des services d'un centre de réadaptation en déficience intellectuelle (CRDI). *Rapport de recherche. Le Consortium National de Recherche sur l'Intégration sociale*, 137 pages.

Rousseau, M., Dionne, C. et Paquet, A. (2012). Intervention précoce et instruments d'évaluation des jeunes enfants ayant un trouble du spectre autistique. *Revue francophone de la déficience intellectuelle*, 23, 93-103.

Terwee, C.B., Bot, S., de Boer, M.R., van der Windt, D., Knol, D.L., Dekker, J., Bouter, L.M. & de Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60, 34-42.

Tyner Berls A. & McEwen, I.R. (1999). Battelle Developmental Inventory. *Journal of the American Physical Therapy*, 79(8), 776-783.

Washington State Office of Superintendent of Public Instruction (2008). *A Guide to Assessment in Early Childhood; Infancy to Age Eight*. 188 pages.

Sites Internet

Pearson.

http://pearsonassess.ca/haiweb/Cultures/fr-CA/Products/Product+Detail.htm?CS_ProductID=EVIP&CS_Category=psychologique-rendement-aptitudes-scolaires&CS_Catalog=TPC-CACatalog. Page consultée le 24 juillet 2014.

Gouvernement du Canada.

<http://www.psc-cfp.gc.ca/plcy-pltq/guides/assessment-evaluation/tips-tapf/index-fra.htm#toc1>. Page consultée le 24 juillet 2014.

Annexe A. Questions du guide de pratique

1. Comment prioriser les besoins en tenant compte des attentes des parents?
2. Comment mesurer les progrès de l'enfant?
 - 2.1 Quels sont les outils d'évaluation validés du développement de l'enfant dans les différentes sphères?
 - **Par rapport à lui-même** : outils d'évaluation à des fins d'intervention ou de programmation.
 - **Par rapport aux enfants de son groupe d'âge** : outils d'évaluation à des fins diagnostiques.
3. Comment communiquer les résultats de la progression de l'enfant aux parents?
4. Quels sont les domaines d'intervention à privilégier?
5. Quel soutien est offert à l'enfant pour stimuler son développement (interventions proposées) et pallier ses incapacités (adaptations dans l'environnement, dans les attitudes,...)?
6. Quel est le meilleur contexte d'application? (intensité : nombre d'heures/semaine, fréquence : nombre de sessions/semaine, milieu naturel vs milieu contrôlé, milieu familial vs milieu de garde) (est-ce que la littérature recommande qu'on travaille de façon plus intensive directement auprès de l'enfant? ou au contraire en soutien/coaching au parent pour que lui intervienne auprès de l'enfant)
7. Y a-t-il un moment privilégié pour commencer ou terminer l'intervention?
8. Quels sont le rôle et la participation de la famille dans l'intervention?
9. Quel est le soutien à mettre en place pour la famille?
 - 9.1 Quelles sont les meilleures pratiques pour favoriser la participation et l'engagement de la famille auprès de l'enfant?
 - Sur un plan « technique » en lui permettant de mieux comprendre et appliquer les interventions ou les façons de faire auprès de son enfant
 - Sur le plan psycho sociale : acceptation du diagnostic ou autre
10. Quels sont les rôles de chacun des intervenants : les techniciens en éducation spécialisée et les intervenants membres d'un ordre professionnel – psychoéducateur, ergothérapeute, orthophoniste, physiothérapeute, travailleur social, psychologue, neuropsychologue –, notamment lorsqu'ils interviennent directement auprès de l'enfant et lorsqu'ils sont dans un rôle de soutien auprès d'autres intervenants?

Annexe B. Liste des syndromes inclus pour l'identification de la population

Homocystéurie
Maladie d'Alexander
Monosomie 5p (Syndrome du Cri-du-chat)
Syndrome 9p (délétion du 9e chromosome)
Syndrome d'Angelman
Syndrome Bardet-Biedl
Syndrome de Bourneville
Syndrome de Carpenter
Syndrome de Coffin
Syndrome de Cohen
Syndrome de Cornelia de Lange
Syndrome du crâne en trèfle
Syndrome de Crouzon
Syndrome de Down
Syndrome de Fraser
Syndrome de Hurler
Syndrome de Lesch-Nyhan
Syndrome de Noonan
Syndrome d'Opitz
Syndrome oro-facial-digital (type I)
Syndrome Oto-palato-digital (Aussi Syndrome de Taybi)
Syndrome de Prader-Willi
Syndrome de Rett
Syndrome de Rubinstein-Taybi
Syndrome de Seckel (nanisme à tête d'oiseau)
Syndrome de Silver (Silver-Russell)
Syndrome de Soto (Gigantisme cérébral)
Syndrome de Weaver
Syndrome de Williams
Syndrome du X fragile (Syndrome de Martin-Bell)
Trisomie partielle 22q
Trisomie partielle 5p
Trisomie partielle 5q3
Trisomie partielle 8
Trisomie partielle 9p

Annexe C. Stratégie de recherche documentaire dans Internet

1. Bases de données

Date de la recherche : 25 octobre 2013		
Période couverte : 2003/01/01 – 2013/11/01		
Mots-clés : ab("assessment") AND ab("systematic review") AND ab((children OR toddlers))		
Bases de données	Articles obtenus	Articles retenus
Proquest Dissertations & Theses	10	0

Date de la recherche : 25 octobre 2013		
Période couverte : 2003/01/01 – 2013/11/01		
Mots-clés : "assessment instrument" AND children AND "systematic review"		
Bases de données	Articles obtenus	Articles retenus
Bielefeld Academic search engine (<i>BASE</i>)	10	0
Système pour l'information en littérature grise en Europe (<i>Opengrey/openSIGLE</i>)	0	0

2. Sites Internet d'universités

Date de la recherche : 25 octobre 2013		
Période couverte : 2003/01/01 – 2013/11/01		
Mots-clés : outils d'évaluation AND développement AND enfants		
Source de données	Articles obtenus	Articles retenus
Université de Montréal (http://papyrus.bib.umontreal.ca/jspui/)	0	0
Université Laval (http://theses.ulaval.ca/archimede/)	0	0
Université du Québec à Trois-Rivières (http://decouverte.uquebec.ca)	0	0
Université Sherbrooke (http://www.usherbrooke.ca/biblio/)	0	0

Annexe C. Stratégie de recherche documentaire dans Internet (suite)

Date de la recherche : 25 octobre 2013		
Période couverte : 2003/01/01 – 2013/11/01		
Mots-clés : "assessment instrument" AND children AND "systematic review"		
Sources de données	Articles obtenus	Articles retenus
Université McGill (http://digitool.library.mcgill.ca/R/)	0	0
DART – Europe E-theses Portal	33	0

3 Sites Internet d'agences d'évaluation

Date de la recherche : 28 octobre 2013		
Organisations	Articles obtenus	Articles retenus
Guidelines International Network	0	0
Centre for Reviews and Dissemination	0	0
National Institute for Health and Clinical Excellence	36	0
Haute Autorité de Santé –	23	0
International Network of Agencies for Health Technology Assessment (INESSS)	0	0
Social Care Institute for Excellence	0	0

Annexe C. Stratégie de recherche documentaire dans Internet (suite)

4. Moteur de recherche Google

Date de la recherche : 28 octobre 2013		
Mots-clés utilisés : "assessment instruments" AND (infants or toddlers) AND "best practice*" AND "global developmental delay"		
Période couverte : 2003 à 2013		
Résultat de la recherche (n = 1 120)		
Documents admissibles (n = 8)		
Note : seules les cinq premières pages ont été consultées.		
Documents admissibles	Document retenu	Motifs d'exclusion
Mots-clés utilisés : "assessment instruments" AND (infants or toddlers) AND "global developmental delay"		
Halle, T., Zaslow, M., Wessel, J., Moodie, S., and Darling-Churchill, K. (2011). Understanding and Choosing Assessments and Developmental Screeners for Young Children Ages 3 – 5 : Profiles of Selected Measures.	Oui (1)	
Grant, Gracy et Brito. <i>Developmental and Social-Emotional Screening Instruments for Use in Pediatric Primary Care in Infants and Young Children.</i>	Non	Pas une revue systématique ou un guide de pratique. Principalement des outils de dépistage pour les pédiatres.
Mots-clés utilisés : "assessment instruments" AND (infants or toddlers) AND ("best practices" OR "best practice*")		
Infant/ Toddlers development, screening, and assessment (2010)	Non	Non spécifique à la population.
Florida partnership for school readiness (2004). Birth to Three Screening and Assessment Resources Guide.	Non	Pas une revue systématique ou un guide de pratique.
Kisker, E. E. et coll. (2011). Resources for measuring services and outcomes in head start programs serving infants and toddlers.	Oui (1)	
Procedures used to Diagnose a Disability and to Assess Special Educational Needs : An International Review	Non	Pas une revue systématique ou un guide de pratique. Non spécifique à la population.
West Virginia Department of Health and Human Resources, Birth to Three. 2009. <i>Assessment instruments approved to be appropriate for screening or evaluating the needs of infants and toddlers.</i>	Non	Pas une revue systématique ou un guide de pratique.
WSOSPI (2008). A Guide to Assessment in Early Childhood : Infancy to Age Eight.	Oui	

(1) : Ces documents n'étant ni une revue systématique ni un guide de pratique clinique, ils ont tout de même servi à compléter la caractérisation des outils dans la grille d'extraction (p. ex. validation scientifique; population visée; domaines mesurés, etc.) et à déterminer la force de la preuve scientifique des outils. Ces documents ont été évalués à l'aide de la grille AGREE II.

5. Moteur de recherche Google Scholar

Date de la recherche : 25 octobre 2013		
Période couverte : 2003/01/01 – 2013/11/01		
Mots-clés : (intitle:child OR intitle:preschool) (intitle:developmental OR intitle:mental OR intitle:intellectual) (psychomotor OR "communication disorders" OR "cognitive disorders" OR "daily living" OR "social develop") (tools OR instrument OR test OR scale OR assessment OR evaluation)		
Source de données	Résultat de la recherche (n)	Documents identifiés (n)
Google Scholar	233	0

Annexe D. Description des documents de référence retenus

Documents	Pays	Population cible	Outils présentés (n)	Critères d'inclusion des outils	Participants à l'élaboration	Relecture et validation externe	Qualité méthodologique (AGREE II, en %) ²⁶
Bagnato et ses collaborateurs (2010). LINKING Authentic Assessment & Early Childhood Intervention.	États-Unis	0 à 8 ans	87	1) Destiné surtout aux enfants âgés de 0 à 8 ans. 2) Accessibles au <i>grand public</i> (États-Unis). 3) Possèdent assez d'items et permettent de générer des objectifs individuels ou d'identifier des stratégies d'intervention.	1) Groupe de développement du document (trois auteurs) 2) 7 experts nationaux 3) 1083 utilisateurs des outils	Oui	Dom 1 : 100,00 Dom 2 : 83,33 Dom 3 : 50,00 Dom 4 : 100,00 Dom 5 : 8,33 Dom 6 : 0,00 Évaluation générale : 4 sur 7
Beauregard et ses collaborateurs (2010). Avis scientifique sur le choix d'un outil de mesure du développement des enfants de 0 à 5 ans dans le cadre des Services intégrés en périnatalité et pour la petite enfance	Canada (Québec)	Enfants de 0 à 5 ans	63	1) Enfants de 0 à 5 ans. 2) Couvrent 4 sphères de développement. 3) Disponibles en français; 4) Codé et interprété par des professionnels ayant des formations diverses.	Groupe de développement du document (n = 7) Révision (n = 6)	Oui	Dom 1 : 83,33 Dom 2 : 55,56 Dom 3 : 58,33 Dom 4 : 72,22 Dom 5 : 41,67 Dom 6 : 25,00 Évaluation générale : 4,5 sur 7

²⁶ Domaine 1 : Champs et objectifs; Domaine 2 : Participation des groupes concernés; Domaine 3 : Rigueur d'élaboration de la recommandation; Domaine 4 : Clarté et présentation; Domaine 5 : Applicabilité; Domaine 6 : Indépendance éditoriale.

Annexe D. Description des documents de référence retenus (suite)

Documents	Pays	Population cible	Outils présentés (n)	Critères d'inclusion des outils	Participants à l'élaboration	Relecture et validation externe	Qualité méthodologique (AGREE II, en %) ²⁷
Halle, T., et ses collaborateurs (2011). <i>Understanding and Choosing Assessments and Developmental Screeners for Young Children: Profiles of Selected Measures.</i>	États-Unis	Enfants de 3 à 5 ans	18	1) Les plus utilisés au programme <i>Head Start</i> . 2) Enfant de 3 à 5 ans. 3) Couvrent un vaste éventail de domaines. 4) Informations disponibles sur la fiabilité et sur la validité. 5) Permettent une utilisation à grande échelle.	Groupe de développement du document : 4 auteurs.	Non	Dom 1 : 94,40 Dom 2 : 33,30 Dom 3 : 20,00 Dom 4 : 88,80 Dom 5 : 87,50 Dom 6 : 50,00 Évaluation générale : 4,5 sur 7
Kisker et ses collaborateurs (2011). <i>Resources for Measuring Services and Outcomes in Head Start Programs Serving Infants and Toddlers</i>	États-Unis	Enfants de 0 à 3 ans	44	1) Les plus probables d'être utilisés par le programme <i>Head Start</i> 2) Postérieurs à 1987.	Groupe de développement du document (n = 15)	Non	Dom 1 : 94,40 Dom 2 : 11,12 Dom 3 : 33,30 Dom 4 : 66,67 Dom 5 : 79,17 Dom 6 : 0,00 Évaluation générale : 5 sur 7

²⁷ Domaine 1 : Champs et objectifs; Domaine 2 : Participation des groupes concernés; Domaine 3 : Rigueur d'élaboration de la recommandation; Domaine 4 : Clarté et présentation; Domaine 5 : Applicabilité; Domaine 6 : Indépendance éditoriale.

Annexe D. Description des documents de référence retenus (suite)

Document	Pays	Population cible	Outils présentés (n)	Critères d'inclusion des outils	Participants à l'élaboration	Relecture et validation externe	Qualité méthodologique (AGREE II, en %) ²⁸
Slentz et ses collaborateurs (WSOSPI) (2008). A Guide to Assessment to Early Childhood, Infancy to age eight.	États-Unis	Enfants de 0 à 8 ans	130	1) Outils bien connus. 2) Dernières versions uniquement; 3) Postérieurs à 1992.	Groupe de développement du document (n = 14) Révision finale (n = 7)	Oui	Dom 1 : 100,00 Dom 2 : 66,67 Dom 3 : 11,91 Dom 4 : 77,78 Dom 5 : 9,33 Dom 6 : 25,00 Évaluation générale : 3 sur 7

²⁸ Domaine 1 : Champs et objectifs; Domaine 2 : Participation des groupes concernés; Domaine 3 : Rigueur d'élaboration de la recommandation; Domaine 4 : Clarté et présentation; Domaine 5 : Applicabilité; Domaine 6 : Indépendance éditoriale.

Annexe E. Description des outils d'évaluation

1. Les outils d'évaluation diagnostiques

Note : Les outils suivis d'un astérisque ont depuis été mis à jour.

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langues	Type d'évaluation
Clinical Evaluation of Language Fundamentals, 4th Edition (CELF-4) (2003)	Semel, E., Wiig, E.H., Secord, W.A. / Pearson	Communication.	5 à 21 ans	Professionnel ayant une formation et une expérience dans l'administration et l'interprétation d'évaluations.	Anglais, espagnol	Évaluation directe individuelle
Developmental Assessment of Young Children (DAYC) (1998)	Vores, J.K., Maddox, T. / PRO-ED	Motricité, Communication, Cognition, Socio émotionnelle, Comportement adaptatif.	0 à 5,11 ans	Un diplôme de quatre ans en psychologie ou en relation d'aide ou dans un domaine relié et un cours en évaluation.	Anglais	Observation, entrevue avec le soignant (caregiver) et évaluation directe individuelle
Functional Emotional Assessment Scale (FEAS) (2001)	Greenspan, S., DeGangi, G. & Wieder, S. / The Interdisciplinary Council on Developmental and Learning Disorders	Socioémotionnelle	7 à 48 mois	Être formé et avoir de l'expérience. S'assurer d'avoir au moins 80 % de fidélité entre les scores faits en direct et les observations faites à l'aide de la vidéo avant d'évaluer directement l'enfant.	Anglais	Observation et évaluation directe individuelle
Infant Toddler Social and Emotional Assessment (ITSEA) (2005)	Carter, A.S. and Briggs-Gowan, M.J. / Pearson Assessments	Socioémotionnelle	12 à 35 mois	Peut être administré par un individu ayant une 6 ^e année. Un professionnel ayant une formation en évaluation interprète les résultats, cependant, les parents ou autres qui n'ont pas reçu de formation peuvent interpréter l'évaluation avec une supervision.	Anglais, chinois, espagnol, français	Complété par le parent ou du soignant (auto-administration) ou évaluation de l'enfant effectuée par le parent, entrevue

							structurée
--	--	--	--	--	--	--	------------

Annexe E. Description des outils d'évaluation (suite)

1. Les outils d'évaluation diagnostiques (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langues	Type d'évaluation
Infant/Toddler Sensory Profile : user's manual (2002)	Dunn, W. /Therapy Skill Builders, Psychological Corporation	Évalue les difficultés sensorielles et vise à comprendre comment le processus sensoriel affecte l'enfant dans sa performance fonctionnelle journalière.	0 à 3 ans	Aucune qualification ou formation requise.	Anglais, espagnol	Questionnaire complété par le parent indiquant la fréquence avec laquelle l'enfant répond à différentes expériences sensorielles.
Mullen Scales of Early Learning (1995)	Mullen, E.M /American Guidance Service, Inc.	Motricité, Communication, Cognition.	0 à 68 mois	Peut être administré par une personne qui a une formation universitaire et une expérience de l'évaluation auprès de jeunes enfants (0-3 ans).	Anglais	Évaluation directe individuelle (incluant une partie d'évaluation complétée par le parent)
(The) Ounce Scale (2003)	Meisels, S.J., Marsden, D.B., Dombro, A.L., Weston, D.R. and Jewkes, A.M./Pearson Assessments	Motricité, Communication, Cognition, socioémotionnelle.	0 à 42 mois	1) <i>Observation Record</i> : Parents, enseignants, employés du programme, cliniciens, spécialistes. 2) <i>Family Album</i> : Parents. 3) <i>Developmental Profiles</i> . : employés du programme.	Anglais, espagnol	Observation et complété par le parent

Preschool Language Scale (PLS-4) (2002)*	Zimmerman, I.L., Steiner, V.G., Pond, R.E. /Pearson Assessments	Communication.	0 à 6,11 ans	Diplôme de maîtrise en psychologie, éducation, ergothérapie, orthophonie ou domaines reliés et une formation en évaluation.	Anglais, espagnol	Évaluation directe individuelle
Pictorial Test of Intelligence, 2nd Edition (PTI-2) (2001)	French, J.L. /PRO-ED	Communication, Cognition, Concept des nombres (math).	3 à 8 ans	Professionnels qualifiés avec une formation en évaluation : psychologue, orthophoniste, enseignant, conseiller en relation d'aide, personne formée pour établir le niveau scolaire de l'enfant (educational diagnosticians).	Anglais	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

1. Les outils d'évaluation diagnostiques (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langues	Type d'évaluation
Qualls Early Learning Inventory (QELI) (2002)	Qualls, A.L., Hoover, H.D., Dunbar, S.B., Frisbie D.A. /Riverside publishing	Communication, Cognition, Alphabétisation, Concept des nombres (math), Approche visant l'apprentissage (<i>approaches to learning</i>)	Prématurée à 1re année	Enseignants. Formation minimale requise.	Anglais	Questionnaire complété par l'enseignant à partir d'observations effectuées en milieu naturel sur une période de temps donnée
Scales of Independent Behavior-Revised (SIB-R) (1996)	Bruininks, R.H., Woodcock, R.W., Weatherman, R.F., Hill, B.K./Riverside publishing	Motricité, Communication, Socioémotionnelle, Comportement adaptatif.	0,3 à 80 ans	L'administration de requiert pas de formation spécifique, mais l'interprétation nécessite un niveau d'expérience élevé.	Anglais	Entrevue structurée, liste de vérification
Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition [Early SB5]/ (2005) <i>Échelle d'intelligence Stanford-Binet</i>	Roid, G. H. /Riverside Publishing	Cognitif.	2,0 à 7,3 ans	Personne ayant un haut niveau de formation ou de pratique dans le développement des enfants et dans les méthodes d'évaluation.	Anglais	Évaluation directe individuelle
Wechsler Individual Achievement Test, 2nd Edition (WIAT-II) (2001)	Wechsler, D. /Pearson Assessments	Communication, Alphabétisation, Concept des nombres (math).	4 à 85 ans	Diplôme de maîtrise en psychologie, éducation, ergothérapie, orthophonie, travail social ou autres domaines reliés, plus une formation en évaluation.	Anglais	Évaluation directe individuelle
Wechsler Preschool and Primary Scale of Intelligence, 3rd Edition (WPPSI-III) (2002)*	Wechsler, D. /Pearson Assessments	Cognition, Verbal	2,6 à 7,3 ans	Doctorat en psychologie, éducation, ou domaines reliés et une formation en évaluation ou permis, brevet ou certification.	Anglais	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Assessment, Evaluation and Programming System (AEPS) ou Évaluation, suivi (EIS) : 0-6 ans (2002)	Bricker, D., Cripe, J., Slentz, K. /Paul H. Brookes Publishing Co.	Motricité, Communication, Cognition, Socioémotionnelle, Comportement adaptatif.	0 à 6 ans	Prestataire de services, intervenant, spécialiste. L'outil requiert une formation à son utilisation.	Anglais, français	Observation, évaluation directe individuelle complétée par le parent, le soignant (caregiver) ou le thérapeute
Bayley Scales Of Infant and Toddler Development (Third Edition) (2005)	Bayley, N. /Pearson	Motricité, Communication, Cognition, Socioémotionnelle, Comportement adaptatif.	1 à 42 mois	Toute personne formée à utiliser l'outil peut l'administrer et noter les résultats.	Anglais	Évaluation directe individuelle, observation et questionnaire complété par le parent
Boehm Test of Basic Concepts, Preschool, 3rd Edition (2001)*	Boehm, A.E./Pearson	Concepts de base et concepts de relation	3,0 à 5,11 ans	Orthophoniste, psychologue scolaire, éducateur, professionnels établissant des diagnostics et familiers avec des procédures d'évaluation standardisées.	Anglais, espagnol	En groupe (classe) ou individuelle
Bracken Basic Concept Scale – Revised (BBCS-R) * (1998)	Bracken, B.A./Pearson	Cognition, Communication, Concept des nombres (math).	2,6 à 8 ans	Psychologue scolaire, enseignant en éducation spécialisée. Connaissances en administration et interprétation d'évaluation.	Anglais, espagnol	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Clinical Evaluation of Language Fundamentals – Preschool, 2nd Edition (CELF-P 2, 2004)*	Semel, E., Wing, E.H., Secord, W.A./Pearson	Communication, Alphabétisation.	3 à 6,11 ans	Professionnel ayant une formation et une expérience dans l'administration et l'interprétation d'évaluations.	Anglais	Évaluation directe individuelle
Developmental Profile 3 (DP-3) (2007)	Alpern, G. D. /Western Psychological Services	Motricité, Cognition, Communication, socioémotionnelle.	0 à 12 ans	a) Entrevue : personne familière avec l'évaluation pédagogique ou psychologique ou un assistant b) <i>Parent/Caregiver Checklist</i> : personne qui connaît bien l'enfant et qui possède une 6 ^e année. c) Notation et interprétation : clinicien ou professionnel ayant une formation en développement de l'enfant, en psychologie, ou en éducation.	Anglais, espagnol	Entrevue avec les parents et auto-évaluation
Early Learning Accomplishment Profile (E-Lap) (1995)	Glover, M.E., Preminger, J.L., Sanford, A.R./Kaplan Press	Motricité, Communication, Cognition, socioémotionnelle.	0 à 36 mois	Connaissances sur la population cible et une formation de deux jours.	Anglais, espagnol	Observation
Expressive Vocabulary Test, 2nd Edition (EVT-2) (2007)	Williams, K.T./Pearson Education, Inc.	Communication.	2,6 à 90 ans	Diplôme de bachelier incluant un cours sur les principes de mesures, l'administration et l'interprétation d'outils d'évaluation, et une formation sur le développement de la parole et du langage.	Anglais	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
High/Scope Preschool Child Observation Record (COR) (1999)	High/Scope Education Research Foundation/ High/Scope Press	Cognition, langage, communication, socioémotionnelle	30 à 72 mois	Administration : parent, enseignant ou autre intervenant. Deux jours de formation donnés par High/Scope sont recommandés, mais non requis. Interprétation : compréhension des domaines.	Anglais, espagnol, selon halle et coll. (2011), serait également traduit en français, chinois et coréen	Observation
High/scope Child Observation Record for Infants and Toddlers (COR-IT). (2002)	High/Scope Educational Research Foundation/ High/Scope Press	Communication, Cognition, socioémotionnelle.	1 à 36 mois	Aucun diplôme spécifique requis. Une formation de deux jours est recommandée. Être en contact régulièrement avec les enfants.	Anglais, espagnol	Observation
Individual Growth Development Indicators (IGDIs) for Infants and Toddlers (2003)	Juniper Gardens Children's Project	Motricité, Communication, Cognition, socioémotionnelle.	0 à 3 ans	Praticiens en intervention précoce qui ont complété la formation liée à l'outil.	Toutes les langues. L'évaluateur doit parler la langue maternelle de l'enfant.	Interaction structurée avec l'enfant en contexte individuel et questionnaire complété par le parent. Mesures répétées pour suivre les progrès.
Infant-Toddler Developmental Assessment (IDA) (1995)	Provence, S., Erikson, J., Vater, S. and Palmeri, S. /Riverside Publishing	Motricité, Communication, Cognition, socioémotionnelle.	0 à 42 mois	Équipe multidisciplinaire ou clinicien expérimenté.	Anglais, espagnol	Rapport du parent et observation de l'enfant

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Kaufman Survey of Early Academic and Language Skills (K-SEALS) (1993)	Kaufman, A.S., Kaufman, N.L./AGS Publishing	Communication, Alphabétisation, Concept des nombres.	3,0 à 6,11	Enseignant, professionnel œuvrant au développement de la parole et du langage, professionnel de la santé.	Anglais	Évaluation directe individuelle
Kaufman Test of Educational Achievement, 2nd Edition (KTEA-II) * (2004)	Kaufman, A.S., Kaufman, N.L./Pearson	Communication, Alphabétisation, Concept des nombres (math).	4,6 à 90 ans et plus	Formation universitaire, plus une expérience dans l'administration et l'interprétation d'évaluations.	Anglais	Évaluation directe individuelle
Learning Accomplishment Profile — (LAP-3) (2005)	Glover, M.E., Preminger, J.L., Sanford, A.R./Kaplan Press	Motricité, Communication, Cognition.	3 à 5 ans	Enseignant, clinicien, éducateur spécialisé, orthophoniste, et autres personnes avec des connaissances sur le développement des enfants.	Anglais, espagnol	Observation
Leiter International Performance Scale-Revised (LEITER-R) (1997)	Roid, G.H. and Miller, L.J./Stoelting Co.	Cognition, socioémotionnelle. Il existe une section pour l'attention	2 à 20 ans	Administration : personne formée et qui a été supervisée. Interprétation : diplôme en psychologie, en relation d'aide ou dans un domaine connexe.	Anglais	Évaluation directe individuelle
Metropolitan Readiness Test (MRT6), 6 th Edition (1994)	Nurss, J.R., McGauvran, M.E./Harcourt Brace	Alphabétisation, Concept des nombres (math)	4 à 7 ans	Diplôme de maîtrise en psychologie, éducation, ergothérapie, orthophonie, travail social, ou autres domaines connexes et une formation en évaluation.	Anglais, espagnol	Évaluation directe individuelle et en groupe (classe)
Phonological Awareness and Literacy Screenings-PreKindergarten (PALS-Prek) (2004)	Invernizzi, M., Sullivan, A., Meier, J.and Swank, L./University of Virginia	Alphabétisation.	4 ans	Enseignants. Lecture du manuel et du guide de notation.	Anglais	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Preschool Language Assessment Instrument, 2nd Edition (PLAI-2) (2003)	Blank, M., Rose, S.A. and Berlin, L.J. /PRO-ED	Communication.	3 à 5,11 ans	Formation en évaluation. Être familier avec le domaine spécifique de cette évaluation. Les professionnels qualifiés incluent : psychologues, orthophonistes, enseignants, conseiller en relation d'aide, personne formée pour établir le niveau scolaire de l'enfant (educational diagnosticians).	Anglais	Évaluation directe individuelle
Ready to Learn: A Dyslexia Screener (2004)	Fawcett, A.K., Nicolson, R.I. and Lee, R./The Psychological Corporation	Motricité, Communication, Cognition, Alphabétisation, Concept des nombres (math)	3,6 à 6,5 ans	Psychologue scolaire, enseignant qualifié, et autres professionnels ayant une formation en évaluation.	Anglais	Évaluation directe individuelle
Receptive One-Word Picture Vocabulary Test (ROWPVT) (2000)	Brownell, R./Academic Therapy Publications	Communication. (Measures receptive vocabulary only).	2 à 18,11 ans	Administration : non-professionnel avec une formation. Interprétations : formation universitaire et expérience suffisante.	Anglais, espagnol	Évaluation directe individuelle
Social Competence and Behavior Evaluation (SCBE-Preschool Edition) (1995)	LaFreniere, P.J., and Dumas, J.E. /WPS	Socioémotionnelle.	2,6 à 6,4 ans	Enseignants de l'enfant. Aucune formation spécifique sur l'outil.	Anglais, français	Réponses de l'enseignant
Test of Early Language Development, 3rd Edition (TELD-3) (1999)	Hresko, W.P., Reid, D.K. and Hammill, D.D./PRO-ED	Communication.	2,0 à 7,11 ans	Clinicien spécialisé en développement de la parole et du langage.	Anglais	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Test of Preschool Early Literacy (TOPEL) (2007)	Lonigan, C. J. Wagner, R. K., Torgesen, J. K. and Rashotte, C. A./PRO-ED	Communication, Alphabétisation.	3,0 à 5,11 ans	Éducateur de la petite enfance, éducateur spécialisé, psychologue, professionnels qui établissent un diagnostic ou autres professionnels. Avoir des connaissances et une expérience dans l'administration et l'interprétation d'évaluations.	Anglais	Évaluation directe individuelle
Vineland Social-Emotional Early Childhood Scales (SEEC) (1998)	Sparrow, S.S., Cicchetti, D.V., Balla, D.V. /American Guidance Services, Inc	Socioémotionnelle.	0 à 5 ans	Diplôme universitaire en développement de la petite enfance avec une formation dans les techniques d'entrevue. Expérience dans l'administration, la notation et l'interprétation du <i>Vineland SEEC Scales</i> .	Anglais	Questionnaire du parent complété avec l'aide de l'intervieweur
Young Children's Achievement Test (YCAT) (2000)	Hresko, W., Peak, P., Herron, S., Bridges, D. /PRO-ED	Communication, Alphabétisation, Concept des nombres (math),	4 à 7 ans	Diplôme de bachelier dans un domaine lié aux sphères ou une formation en évaluation.	Anglais, espagnol	Évaluation directe individuelle

Annexe E. Description des outils d'évaluation (suite)

3. Les outils d'évaluation ayant une double fonction

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Achenbach System of empirically based assessment (ASEBA) (1999)	Achenbach T. M. and Rescorla, L. A./ASEBA	Communication, Socio-émotionnelle. Comportements	1,5 à 5 ans	Posséder une scolarité équivalente à une 5 ^e année. Interprétation : formation universitaire.	Disponibilité dans 75 langues disponibles en français	Complété par le parent ou le soignant (caregiver)
Adaptive Behavior Assessment System (ABAS-II) (2003)	Harrison, P., Oakland, T./Pearson	Motricité Communication Cognition Comportement adaptatif Autonomie Ressource communautaire Responsabilité individuelle	0 à 89 ans	Liste de contrôle : Aucune formation requise Interprétation : niveau d'expérience élevé en évaluation.	Anglais, espagnol	L'enseignant, l'intervenant ou le parent rapporte les fréquences de comportement de l'enfant
Arizona Articulation Proficiency Scale, 3rd Edition (2001)	Barker Fudala, J. / PRO-ED Inc	Communication.	1,6 à 18 ans	Diplôme universitaire en orthophonie.	Anglais	Évaluation directe individuelle
Battelle Developmental Inventory (BDI) (1984)	Newborg, J., Stock, J.R., Wnek, J., Guidubaldi, J and Svinicki, J.S. /Riverside Publishing Co.	Motricité, communication, Cognition, dév. émotionnel et social, Comportement adaptatif.	0 à 8 ans	Enseignant préscolaire, éducateur spécialisé, psychologue, professionnel de la santé, un assistant, etc. Être familiarisé avec les procédures psychométriques.	Anglais, espagnol	Évaluation directe individuelle Observation, entrevue du parent
Behavior Assessment System for Children (BASC-2) (2004)	Reynolds, C.R. and Kamphaus, R. W. /Pearson	Socioémotionnelle Communication Comportement	2 à 25 ans	Administration : parents ou enseignants. Interprétation : formation universitaire.	Anglais, espagnol	Observation, Questionnaire complété par le parent ou l'enseignant, auto-évaluation

Annexe E. Description des outils d'évaluation (suite)

3. Les outils d'évaluation ayant une double fonction (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Brigance Comprehensive Inventory of Basic Skills-Revised (CIBS-R) — 0-7 ans (1999)	Brigance, A.H. /Curriculum Associates, Inc.	Habilités motrices et comportements, motricité, autonomie, langage, connaissances générales et maîtrise des concepts, développement socio-affectif, pré lecture, lecture, écriture, calcul.	0 à 7 ans	Aucune formation requise.	Anglais, espagnol	Évaluation directe individuelle
Brigance Diagnostic Inventory of Early Development II (IED-II) (2004)	Brigance, A.H. /Curriculum Associates	Motricité, Communication, Cognition, Socioémotionnelle, Alphabétisation, Concept des nombres (math)	0 à 7 ans	Enseignant, assistant, thérapeute, infirmier, médecin. Aucune formation spécifique requise.	Anglais, espagnol	Observation, entrevue du parent Questionnaire complété par le parent ou l'enseignant, évaluation directe individuelle
Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edition (BOT-2) (2006)	Bruininks, R.H., Bruininks, B.D./Pearson	Motricité.	4 à 21 ans	Physiothérapeute, ergothérapeute, enseignant en éducation physique adaptée, professionnel en éducation spécialisée.	Anglais	Évaluation directe individuelle
Comprehensive Test of	Wagner, R., <u>Torgesen</u> , J.,	Alphabétisation.	5 à 24,11	Professionnel ayant une	Anglais	Évaluation

Phonological Processing (CTOPP) (1999)	Rashotte, C., Pearson, N.A. / Ann Arbor Publishers		ans	formation en évaluation.		directe individuelle
--	--	--	-----	--------------------------	--	----------------------

Annexe E. Description des outils d'évaluation (suite)

3. Les outils d'évaluation ayant une double fonction (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Devereux Early Childhood Assessment for infants and toddlers (DECA — I/T) (2007)	Mackrain, M., LeBuffe, P. and Powell, G. / Kaplan Early Learning Company	Socioémotionnelle.	1 à 36 mois	Posséder une scolarité équivalant à une 6 ^e année. Administration : toute personne qui et qui a observé les comportements de l'enfant un minimum de deux heures par jour, deux jours/semaine et pendant une période de plus de quatre semaines. Un membre du programme formé note et interprète l'évaluation standardisée et le profil.	Anglais	Observation et questionnaire complété par le parent ou le soignant (caregiver)
Expressive One-Word Picture Vocabulary Test (EOWPVT) (2000)	Martin, N and Brownell, R. / Academic Therapy Publications	Communication.	2 à 80 ans	Administration : personne formée à utiliser l'outil. Interprétation : formation universitaire et de la pratique.	Anglais Espagnol	Évaluation directe individuelle
Learning Accomplishment Profile-Diagnostic (LAP-D), 3rd Edition (2005)	Hardin, B.J., Peisner — Feinberg, E., Weeks, S.W., LeMay, D.W., Griffin, P.M., Sanford, A.R./Kaplan Press	Motricité, Communication, Cognition, Comportement adaptatif, Concept des nombres (math)	2,6 à 6,0 ans	Enseignant qualifié, assistant, clinicien, éducateur spécialisé, orthophoniste et autres personnes familiarisées avec le développement des enfants.	Anglais, espagnol	Évaluation directe individuelle
MacArthur-Bates Communicative Development Inventory (2nd Edition), (2007) Les Inventaires MacArthur du développement de la Communication	Fenson, L., Marchman, V.A., Dale, P.S., Reznick, J.S., Thal, D. and Bates, E./Paul H. Brookes Publishing Company	Communication.	8 à 37 mois	Parents. Aucune formation particulière.	Anglais, français, espagnol, et autres langues (selon le WSOSPI).	Questionnaire complété par le parent

Annexe E. Description des outils d'évaluation (suite)

3. Les outils d'évaluation ayant une double fonction (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Merrill-Palmer-Revised Scales of Development (M-P-R) (2004)	Roid, G.H. and Sampers, J.L. / Western Psychological Services	Motricité, Communication, Cognition, Socioémotionnelle, Comportement adaptatif.	0 à 6,6 ans	Diplôme en psychologie ou en relation d'aide ou autres domaines connexes avec un cours en évaluation.	Anglais, espagnol	Évaluation directe individuelle par l'évaluateur et questionnaire complété par le parent le soignant (caregiver)
OWLS: Listening Comprehension (LC) Scale & Oral Expression (OE) Scale (1995)	Carrow-Woolfolk, E. / Ann Arbois publishers	Communication.	3 à 21,11 ans	Psychologue scolaire, orthophoniste, personne formée pour établir le niveau scolaire de l'enfant (educational diagnosticians).	Anglais	Évaluation directe individuelle
OWLS: Written Expression (WE) Scale (1996)	Carrow-Woolfolk, E. / Ann Arbois publishers	Expression écrite, Communication.	5,0 – 21,11 ans	Psychologue scolaire, orthophoniste, personne formée pour établir le niveau scolaire de l'enfant (educational diagnosticians).	Anglais	Individuellement ou en petits groupes
Peabody Developmental Motor Scales, 2nd Edition (PDMS-2) * (2000)	Folio, M.R., Fewell, R.R. / Pro-ED	Motricité fine et globale.	0 à 6,0 ans	Ergothérapeute, physiothérapeute, personne qui établissent des diagnostics, intervenant en intervention précoce, enseignant en éducation physique adaptée, psychologue, etc.	Anglais	Évaluation directe individuelle
Peabody Picture Vocabulary Test, 4th edition (PPVT-IV) (2007)	Dunn, L.M, and Dunn, D.M. / Pearson Assessments	Communication.	2,6 à 90 ans	Diplôme de bachelier incluant des cours en mesure et évaluation et une formation en développement de la parole et du langage.	Anglais, espagnol	Évaluation directe individuelle
Test of Early Mathematics Ability, 3rd Edition (TEMA-	Ginsburg, H.P., Baroody, A.J. / Pro-ED	Concept des nombres (math).	3,0 à 8,11 ans	Diplôme en psychologie ou en relation d'aide ou autres domaines connexes	Anglais	Évaluation directe

3) (2003)					avec un cours en évaluation.		individuelle
-----------	--	--	--	--	------------------------------	--	--------------

Annexe E. Description des outils d'évaluation (suite)

3. Les outils d'évaluation ayant une double fonction (suite)

Nom de l'outil	Auteurs/éditeurs	Sphères	Étendue d'âge	Qualifications requises selon le concepteur de l'outil	Langue	Type d'évaluation
Test of Early Reading Ability, 3rd Edition (TERA-3) (2001)	Reid, D.K., Hresko, W.P., Hammill, D.D. /PRO-ED	Alphabétisation.	3,6 à 8,6 ans	Diplôme en psychologie ou en relation d'aide ou autres domaines connexes avec un cours en évaluation.	Anglais	Évaluation directe individuelle
Vineland Adaptive Behavior Scales (VABS) II Edition (2005)	Sparrow, S.S., Cicchetti, D.V., Balla, D.A. /NCS Pearson, Inc.	Motricité, Communication, Socioémotionnelle, Comportement adapté.	0 à 90 ans	Diplôme en psychologie ou dans un domaine connexe à la relation d'aide, ainsi qu'une expérience en évaluation et interprétation.	Anglais, espagnol	Questionnaire complété par le parent ou le soignant (caregiver)

Annexe F. Objectifs des outils (par ordre alphabétique)

Outils d'évaluation	Sphères	Objectifs
Achenbach System of empirically based assessment (ASEBA) (1999)	Communication, socioémotionnelle.	<ul style="list-style-type: none"> - Dresser le profil comportemental de l'enfant. - Identifier les domaines pour l'intervention. - Évaluer les effets de l'intervention.
Adaptive Behavior Assessment System (ABAS-II)/Système d'évaluation du comportement adapté II (formulaires pour l'enseignant, le gardien, parent ou le tuteur principal d'enfants âgés 0-5 ans) (2003)	Motricité Communication Cognition Comportement adapté Autonomie Ressource communautaire Responsabilité individuelle	<ul style="list-style-type: none"> - Déterminer comment une personne répond aux demandes quotidiennes. - Développer des objectifs d'intervention. - Déterminer l'admissibilité aux services et prestations de sécurité sociale. - Évaluer les personnes ayant une déficience intellectuelle, des difficultés d'apprentissage, le TDA/H ou d'autres déficiences. - Évaluer la capacité des adultes à vivre de façon autonome. - Documenter le profil adaptatif
Arizona Articulation Proficiency Scale, 3rd Edition (2001)	Communication.	<ul style="list-style-type: none"> - Identifier les prononciations incorrectes (<i>misarticulations</i>) et la capacité articulatoire globale. - Identifier les enfants qui ont besoin d'une évaluation plus complète quant à leurs compétences langagières ou leurs capacités d'apprentissage. - Identifier et sélectionner les enfants pour un service d'orthophonie. - Déterminer les progrès réalisés grâce à l'orthophonie. - Discuter du développement de la parole, des besoins d'intervention, et des progrès.
Assessment, Evaluation and Programming System (AEPS) ou Évaluation, intervention, suivi (EIS) : 0-3 ans (2002)	Motricité, Communication, Cognition, Socioémotionnelle, Comportement adapté.	<ul style="list-style-type: none"> - Créer un programme individuel d'évaluation et de programmation. - Développer un curriculum d'intervention. - Mesurer les progrès développement de l'enfant au fil du temps.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Battelle Developmental Inventory (BDI) (1984)	Motricité, communication, Cognition, dév. émotionnel et social, Comportement adaptatif.	<ul style="list-style-type: none"> - Évaluer et identifier des enfants présentant des retards de développement. - Évaluer la maturité scolaire des enfants ayant un développement typique. - Planifier et dispenser l'enseignement. - Évaluer un groupe d'enfants ayant des retards de développement.
Bayley Scales Of Infant and Toddler Development (Third Edition) (2005)	Motricité, communication, Cognition, Socio-émotionnelle, Comportement adaptatif.	<ul style="list-style-type: none"> - Identifier les domaines faisant l'objet d'une déficience ou d'un retard. - Développer des curriculums pour les programmes. - Évaluer les effets des interventions. - Indiquer les domaines qui requerraient une évaluation plus approfondie. - Utiliser pour évaluer retard global de développement.
Behavior Assessment System for Children (BASC-2) (2004)	Socioémotionnelle, communication.	<ul style="list-style-type: none"> - Diagnostiquer les troubles émotionnels et comportementaux. - Identifier les retards aux fins de placement dans les programmes éducationnels. - Mener des recherches sur les troubles émotionnels et comportementaux des enfants. - Évaluer les interventions et les programmes.
Boehm Test of Basic Concepts, 3 rd Edition* (2000)	Cognition, Communication, Alphabétisation, Concept des nombres (math).	<ul style="list-style-type: none"> - Identifier les élèves ayant potentiellement des difficultés d'apprentissage et qui pourraient avoir besoin d'évaluations supplémentaires. - Identifier les élèves dans la classe qui connaissent la plupart des concepts, mais qui ont de la difficulté à comprendre des concepts clés. - Déterminer les progrès suite à une intervention. - Évaluer les compétences de l'enfant dans la compréhension du langage.
Boehm Test of Basic Concepts, Preschool, 3 rd Edition (2001)	Cognition, Communication, Alphabétisation, Concept des nombres (math).	<ul style="list-style-type: none"> - Identifier les enfants ne possédant pas les prérequis académiques afin de leur offrir une intervention rapidement.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Bracken Basic Concept Scale – Revised (BBCS-R) * (1998)	Cognition, Communication, Concept des nombres (math).	<ul style="list-style-type: none"> - Évaluer les Prérequis académiques chez l'enfant. - Mesurer la compréhension de concepts éducatifs fondamentaux et fonctionnels.
Brigance Comprehensive Inventory of Basic Skills-Revised (CIBS-R) (1999)	Habiletés motrices et comportements, motricité, autonomie, langage, connaissances générales et maîtrise des concepts, développement socio-affectifs, prélecture, lecture, écriture, calcul.	<ul style="list-style-type: none"> - Déterminer la performance scolaire. - Planifier l'enseignement. - Suivre les progrès de l'enfant.
Brigance Diagnostic Inventory of Early Development II (IED-II) (2004)	Motricité, Communication, Cognition, Socioémotionnelle, Alphabétisation, Concept des nombres (math)	<ul style="list-style-type: none"> - Identifier les forces et les faiblesses de l'enfant. - Diagnostiquer et évaluer l'enfant en classe. - Suivre les progrès. - Évaluer les programmes.
Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edition (BOT-2) (2006)	Motricité.	<ul style="list-style-type: none"> - Évaluer l'aptitude motrice. - Développer et évaluer des programmes ciblant la sphère motrice.

Clinical Evaluation of Language Fundamentals, 4th Edition (CELF-4, 2003)*	Communication.	- Identifier et diagnostiquer les troubles du langage.
---	----------------	--

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Clinical Evaluation of Language Fundamentals – Preschool, 2nd Edition (CELF-P 2) (2004)	Communication, Alphabétisation.	- Évaluer les compétences communicationnelles nécessaires pour faire la transition du préscolaire au primaire.
Comprehensive Test of Phonological Processing (CTOPP) (1999)	Alphabétisation.	- Identifier les personnes ayant des capacités phonologiques nettement inférieures à leurs pairs. - Déterminer les forces et les faiblesses en habileté de traitement phonologique. - Documenter le progrès dans le traitement phonologique à la suite d'une intervention. - Recueillir des données sur le traitement phonologique pour la recherche.
Developmental Assessment of Young Children (DAYC) (1998)	Motricité, Communication, Cognition, Socioémotionnelle, Capacité d'adaptation.	- Identifier les enfants ayant des retards possibles dans les sphères énumérées.
Developmental Profile 3 (DP-3) (2007)	Motricité, Cognition, Communication, socioémotionnelle.	- Effectuer le dépistage. - Mesurer les progrès de l'enfant. - Recueillir des données pour la recherche.
Devereux Early Childhood Assessment for infants and toddlers (DECA – I/T) (2007)	socioémotionnelle.	- Identifier les enfants qui peuvent éprouver des difficultés sociales et émotionnelles. - Comparer la santé émotionnelle et sociale de l'enfant au cours du temps. - Développer des programmes pour favoriser la croissance sociale et émotionnelle de l'enfant.

Early Learning Accomplishment Profile (E-Lap) (1995)	Motricité, Communication, Cognition, socioémotionnelle.	- Fournir des orientations aux programmes d'intervention précoce dans l'évaluation et la programmation pour les tout-petits, jeunes enfants et enfants ayant des retards de développement.
--	--	--

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Expressive One-Word Picture Vocabulary Test (EOWPVT) (2000)	Communication.	<ul style="list-style-type: none"> - Mesurer le vocabulaire. - Dépister un retard du langage chez le jeune enfant. - Identifier les difficultés de lecture et des mots exprimés. - Évaluer le vocabulaire anglais d'un nouvel apprenant. - Mesure l'aphasie (lorsqu'utilisé avec le Receptive One-Word Vocabulary Test). - Mesurer les capacités cognitives. - Évaluer un programme d'intervention élaboré pour améliorer le vocabulaire ou la compréhension de lecture.
Expressive Vocabulary Test, 2nd Edition (EVT-2) (2007)	Communication.	<ul style="list-style-type: none"> - Évaluer rapidement le vocabulaire expressif. - Évaluer l'anglais d'un nouvel apprenant. - Comparer avec le vocabulaire réceptif afin d'identifier les forces et les faiblesses. - Évaluer l'expression orale comme fondation aux habiletés d'écriture. - Évaluer le progrès.
Functional Emotional Assessment Scale (FEAS) (2001)	Socioémotionnelle.	<ul style="list-style-type: none"> - Dépister des problèmes chez l'enfant qui connaît régulièrement des troubles émotionnels, mais sans pouvoir diagnostiquer un trouble spécifique. - Aider les cliniciens à identifier les sphères critiques qui requièrent une évaluation plus approfondie. - Décrire le profil socioémotionnel de l'enfant.
High/Scope Preschool Child Observation Record (COR) (1999)	Cognition, langage, communication, socioémotionnelle	<ul style="list-style-type: none"> - Évaluer les connaissances et les capacités de l'enfant. - Mesurer les progrès de l'enfant dans les programmes d'intervention précoce.

<p>High/scope Child Observation Record for Infants and Toddlers (COR-IT). (2002)</p>	<p>Communication, Cognition, socioémotionnelle.</p>	<ul style="list-style-type: none"> - Évaluer le développement de l'enfant et ses capacités afin d'aider les programmes à planifier des activités pour l'enfant. - Suivre les effets de ces activités et le progrès de l'enfant. - Évaluer l'efficacité d'un programme et son curriculum. - Donner un cadre pour communiquer avec les parents de l'enfant.
--	---	---

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
<p>Individual Growth Development Indicators (IGDIs) for Infants and Toddlers (2003)</p>	<p>Motricité, Communication, Cognition, socioémotionnelle.</p>	<ul style="list-style-type: none"> - Estimer et surveiller la croissance et les progrès du nourrisson et du jeune enfant (conçu pour être utilisé souvent).
<p>Infant-Toddler Developmental Assessment (IDA) (1995)</p>	<p>Motricité, Communication, Cognition, socioémotionnelle.</p>	<ul style="list-style-type: none"> - Déterminer la nécessité d'un suivi, d'une consultation, d'une intervention ou d'autres services pour l'enfant et la famille. - Développer un plan de service individualisé pour la famille.
<p>Infant/Toddler Sensory Profile : user's manual (2002)</p>	<p>Évalue les difficultés sensorielles et vise à comprendre comment le processus sensoriel affecte l'enfant dans sa performance fonctionnelle journalière.</p>	<ul style="list-style-type: none"> - Examiner les tendances chez les tout-petits qui sont à risque ou qui ont une incapacité spécifique liée aux capacités de traitement sensoriel.
<p>Infant Toddler Social and Emotional Assessment (ITSEA) (2005)</p>	<p>Socioémotionnel.</p>	<ul style="list-style-type: none"> - Identifier les jeunes enfants qui peuvent avoir des comportements anormaux, une psychopathologie ou des lacunes pour certaines compétences. - Établir l'admissibilité à un service (en combinaison avec des observations directes, une évaluation additionnelle ou des discussions avec les parents).

Kaufman Survey of Early Academic and Language Skills (K-SEALS) (1993)	Communication, Alphabétisation, Concept des nombres.	<ul style="list-style-type: none"> - Évaluer la préparation à l'école. - Identifier les enfants doués. - Évaluer l'efficacité d'un programme. - Effectuer des recherches ciblant le développement des jeunes enfants.
Kaufman Test of Educational Achievement, 2nd Edition (KTEA-II) * (2004)	Communication, Alphabétisation, Concept des nombres (math).	<ul style="list-style-type: none"> - Évaluer les compétences académiques clés. - Dépister et effectuer une pré-orientation. - Mesurer le progrès de l'enfant suite à une intervention.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Learning Accomplishment Profile-Diagnostic (LAP-D), 3rd Edition (2005)	Motricité, Communication, Cognition, Comportement adaptatif, Concept des nombres (math)	<ul style="list-style-type: none"> - Évaluer le développement des compétences individuelles. - Évaluer les progrès de l'enfant afin d'individualiser et de développer des activités appropriées.
Learning Accomplishment Profile — (LAP-3) (2005)	Motricité, Communication, Cognition.	<ul style="list-style-type: none"> - Fournir des orientations aux programmes de la petite enfance pour l'évaluation et la programmation des tout-petits, jeunes enfants, et enfants ayant des retards de développement.
Leiter International Performance Scale-Revised (LEITER-R) (1997)	Cognition, socioémotionnelle.	<ul style="list-style-type: none"> - Identifier les enfants qui ont une déficience cognitive. - Suivre les petites améliorations des capacités cognitives. - Développer des stratégies d'intervention qui répondent aux incapacités identifiées.
MacArthur-Bates Communicative Development Inventory (2nd Edition), (2007) Les Inventaires MacArthur du développement de la Communication	Communication.	<ul style="list-style-type: none"> - Dépister les retards dans le développement du langage. - Identifier les compétences problématiques. - Élaborer des stratégies d'intervention et en évaluer les effets.

Merrill-Palmer-Revised Scales of Development (M-P-R) (2004)	Motricité, Communication, Cognition, Socioémotionnelle, Comportement adaptatif.	<ul style="list-style-type: none"> - Évaluer les enfants nés prématurément. - Identifier les retards de développement et les problèmes d'apprentissage. - Évaluer le développement cognitif général chez les enfants qui parlent anglais et espagnol. - Dépister ceux qui ont été référés pour une évaluation pour un possible retard de développement ou pour une déficience. - Réévaluer les personnes préalablement identifiées comme ayant un retard de développement.
---	--	---

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Metropolitan Readiness Test (MRT6), 6 th Edition (1994)	Alphabétisation, Concept des nombres (math)	<ul style="list-style-type: none"> - Identifier les compétences des enfants en lecture et en mathématiques.
Mullen Scales of Early Learning (1995)	Motricité, Communication, Cognition.	<ul style="list-style-type: none"> - Identifier les enfants ayant des besoins éducatifs spéciaux pour une évaluation plus approfondie. - Identifier les forces et faiblesses qui sous-tendent le style d'apprentissage de l'enfant afin d'élaborer des plans individuels d'intervention.
The Ounce Scale (2003)	Motricité, Communication, Cognition, socioémotionnelle.	<ul style="list-style-type: none"> - Fournir des lignes directrices et des normes pour l'observation et l'interprétation de la croissance et le comportement. - Utiliser les résultats pour la reddition de comptes. - Planifier les programmes.
OWLS: Listening Comprehension (LC) Scale & Oral Expression (OE) Scale (1995)	Communication.	<ul style="list-style-type: none"> - Identifier les étudiants qui ont des troubles d'apprentissage, des troubles du langage ou des difficultés connexes. - Concevoir des interventions ciblées répondant aux difficultés académiques potentielles. - Suivre le progrès. - Recueillir des données longitudinales pour la recherche.

<p>OWLS: Written Expression (WE) Scale (1996)</p>	<p>Expression écrite, Communication.</p>	<ul style="list-style-type: none"> - Identifier les étudiants qui ont des troubles d'apprentissage, des troubles du langage ou des difficultés connexes. - Concevoir des interventions ciblées répondant aux difficultés académiques potentielles. - Suivre le progrès. - Recueillir des données longitudinales pour la recherche.
---	--	--

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Peabody Developmental Motor Scales, 2nd Edition (PDMS-2) * (2000)	Motricité fine et globale.	<ul style="list-style-type: none"> - Évaluer les habiletés motrices des enfants. - Planifier des interventions individualisées et identifier les objectifs et les activités à y inclure. - Estimer les compétences motrices globales de l'enfant comparées à ses pairs ou évaluer ses propres habiletés motrices fines versus globales.
Peabody Picture Vocabulary Test, 4 th edition (PPVT-IV) (2007)	Communication.	<ul style="list-style-type: none"> - Mesurer l'amélioration du vocabulaire ou la réponse à l'enseignement. - Diagnostiquer les difficultés en lecture. - Mesurer le potentiel linguistique et, chez un non-lecteur, de mesurer son développement ou ses retards, ses difficultés d'écriture ou de langage expressif, ou d'autres incapacités telles que l'aphasie; - Évaluer le vocabulaire d'une personne apprenant l'anglais.
Phonological Awareness and Literacy Screenings-PreKindergarten (PALS-PreK) (2004)	Alphabétisation.	<ul style="list-style-type: none"> - Évalue les connaissances de base en littératie d'enfants d'âge préscolaire. - Offrir des repères pour les enseignants afin qu'ils adaptent leur enseignement aux besoins des enfants.
Pictorial Test of Intelligence, 2nd Edition (PTI-2) (2001)	Communication, Cognition, Concept des nombres (math).	<ul style="list-style-type: none"> - Estimer les capacités cognitives afin de prendre une décision sur l'admissibilité de l'enfant pour un placement dans un programme pour une déficience intellectuelle. - Prédire le degré de succès de l'enfant qui est promu, saute une année scolaire, ou est placé dans un programme enrichi. - Identifier les objectifs à court et à long terme (avec d'autres informations). - Effectuer de la recherche.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Preschool Language Assessment Instrument, 2nd Edition (PLAI-2) (2003)	Communication.	<ul style="list-style-type: none"> - Identifier les enfants ayant des difficultés de langage et de communication pouvant entraver la performance scolaire. - Structurer l'enseignement ou l'intervention selon le niveau de fonctionnement de l'enfant. - Comparer les compétences réceptives et expressives du discours. - Délimiter les comportements pragmatiques qui interfèrent avec une communication efficace. - Évaluer les capacités de l'enfant à s'exprimer lorsqu'il a peu ou pas de langage expressif. - Évaluer l'efficacité de l'intervention.
Preschool Language Scales (PLS-4, 2002)*	Communication.	<ul style="list-style-type: none"> - Évaluer le développement du langage. - Déterminer si un enfant à un trouble du langage et connaître la source du problème (audition, expression orale ou les deux).
Qualls Early Learning Inventory (QELI) (2002)	Communication, Cognition, Alphabétisation, Concept des nombres (math), Approche visant l'apprentissage (<i>approaches to learning</i>)	<ul style="list-style-type: none"> - , établir un diagnostic sur six domaines clés de l'apprentissage qui sont important pour la réussite scolaire.
Ready to Learn: A Dyslexia Screener (2004)	Motricité, Communication, Cognition, Alphabétisation, Concept des nombres (math)	<ul style="list-style-type: none"> - Dépister. - Suivre les progrès.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Receptive One-Word Picture Vocabulary Test (ROWPVT) (2000)	Communication. (Measures receptive vocabulary only).	<ul style="list-style-type: none"> - Documenter le développement du vocabulaire. - Évaluer les habiletés de lecture. - Dépister les retards de langage. - Examiner les mots et concepts récupérés par les personnes aphasiques. - Obtenir une indication indirecte de certaines compétences cognitives. - Évaluer un programme d'intervention.
Scales of Independent Behavior-Revised (SIB-R) (1996)	Motricité, Communication, Socioémotionnelle, Capacité d'adaptation.	<ul style="list-style-type: none"> - Évaluer le comportement (adaptatif — 14 domaines — et inadapté — 8 domaines).
Social Competence and Behavior Evaluation (SCBE-Preschool Edition) (1995)	Socioémotionnel.	<ul style="list-style-type: none"> - Évaluer l'efficacité d'une intervention. - Évaluer l'adaptation à l'école. - Décrire l'adaptation et le fonctionnement de l'enfant dans sa classe afin d'aider l'enseignant à planifier une intervention.
Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition [Early SB5]/ (2005) <i>Échelle d'intelligence Stanford-Binet</i>	Cognitif.	<ul style="list-style-type: none"> - Évaluation clinique et neuropsychologique. - Évaluation psychoéducative pour le référencement en éducation spécialisée. - Offrir de l'information aux intervenants. - Effectuer des recherches sur les capacités et aptitudes.
Test of Early Language Development, 3rd Edition (TELD-3) (1999)	Communication.	<ul style="list-style-type: none"> - Dépister les enfants pour l'intervention précoce. - Identifier les enfants qui nécessitent une plus vaste évaluation clinique. - Aider les cliniciens à développer des plans d'intervention. - Suivre les progrès.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Test of Early Mathematics Ability, 3 rd Edition (TEMA-3) (2003)	Concept des nombres (math).	<ul style="list-style-type: none"> - Déterminer les forces et les faiblesses. - Identifier les enfants doués. - Mesurer les progrès. - Évaluer les programmes. - Évaluer la préparation à l'école. - Découvrir les causes de mauvaises performances scolaires en mathématiques. - Guider le choix des stratégies d'enseignement et des mesures correctives.
Test of Early Reading Ability, 3 rd Edition (TERA-3) (2001)	Alphabétisation.	<ul style="list-style-type: none"> - Identifier les enfants dont les compétences de lectures se situent de manière importante sous celles de leurs pairs. - Identifier les forces et les faiblesses des enfants. - Documenter les progrès à la suite d'une intervention. - Effectuer des recherches.
Test of Preschool Early Literacy (TOPEL) (2007)	Communication, Alphabétisation.	<ul style="list-style-type: none"> - Mesurer et quantifier les capacités liées à l'alphabétisation et les changements. - Identifier les enfants qui ont des problèmes liés à l'alphabétisation ou qui sont à risque d'en développer. - Suivre le progrès de l'enfant.
Vineland Adaptive Behavior Scales (VABS) II Edition (2005)	Motricité, Communication, Socioémotionnelle, Comportement adaptatif.	<ul style="list-style-type: none"> - Évaluer les retards de développement. - Effectuer un suivi des progrès. - Planifier un programme. - Déterminer les effets des différentes interventions sur les habiletés fonctionnelles des enfants (projets de recherche).
Vineland Socio-Emotional Early Childhood Scales (SEEC) (1998)	Socioémotionnelle.	<ul style="list-style-type: none"> - Planifier un programme et sélectionner des activités adaptées aux besoins des enfants. - Suivre le progrès et évaluer les réussites de l'enfant après avoir complété le programme.
Wechsler Individual Achievement Test, 2 nd Edition (WIAT-II) (2001)	Communication, Alphabétisation, Concept des nombres (math).	<ul style="list-style-type: none"> - Identifier les forces et les faiblesses académiques. - Diagnostiquer un trouble d'apprentissage spécifique. - Éclairer des décisions concernant l'admissibilité à des services éducatifs. - Concevoir des objectifs pédagogiques et des plans d'intervention.

Annexe F. Objectifs des outils (par ordre alphabétique) (suite)

Outils d'évaluation	Sphères	Objectifs
Wechsler Preschool and Primary Scale of Intelligence, 3rd Edition (WPPSI-III) (2002)	Cognition Verbale	<ul style="list-style-type: none"> - Identifier et qualifier les enfants ayant des retards cognitifs pour des services. - Évaluer les enfants pour des retards cognitifs, des incapacités intellectuelles, pour l'autisme et la douance. - Déterminer l'impact d'une lésion cérébrale traumatique sur la fonction cognitive des enfants.
Young Children's Achievement Test (YCAT) (2000)	Communication, Alphabétisation, Concept des nombres (math),	<ul style="list-style-type: none"> - Mesurer les compétences et les aptitudes qui déterminent le succès à l'école.

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus

1. Les outils d'évaluation diagnostiques

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Clinical Evaluation of Language Fundamentals, 4th Edition (CELF-4) * (2003)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		
Developmental Assessment of Young Children (DAYC) (1998)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		
Functional Emotional Assessment Scale (FEAS) (2001)	Concomitante : sous le seuil (< 0,5); Prédictive : sous le seuil (< 0,5)				0,65 ou plus			
Infant Toddler Social and Emotional Assessment (ITSEA) (2005)	Contenu : oui Conceptuelle : intercorrélations avec les sous-échelles sous le seuil pour deux domaines (< 0,50) et acceptable pour deux domaines (> 0,50) Concomitante : sous le seuil (< 0,5)	Concomitante : Acceptable (0,50 – 0,69)			0,65 ou plus	Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

1. Les outils d'évaluation diagnostiques (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Mullen Scales of Early Learning (1995)	<p>Concomitante : Acceptable (> 0,5);</p>	<p>Concomitante : Élevée (> 0,70)</p>		<p>Contenu : aucune information du concepteur; Conceptuelle : acceptable Concomitante : Acceptable (> 0,50) Prédictive : oui pour les enfants ayant des besoins spéciaux</p>	0,65 ou plus	Élevé (0,80 ou plus)		<p>Accord inter-juges : acceptable</p> <p>Fidélité test-retest : acceptable</p> <p>Consistance interne : acceptable</p>
(The) Ounce Scale (2003)	<p>Contenu : oui Concomitante : sous le seuil (< 0,5)</p>				Moins de 0,65			
Preschool Language Scale (PLS-4) (2002)	<p>Contenu : oui Conceptuelle : corrélation de 0,66 entre <i>Auditory Comprehension</i> et <i>Expressive Communication</i> Concomitante : acceptable (> 0,5)</p>				0,65 ou plus	Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

1. Les outils d'évaluation diagnostiques (suite)

Nom de l'outil	Données de validité rapportées par :			Données de fidélité rapportées par :				
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Pictorial Test of Intelligence, 2nd Edition (PTI-2) (2001)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		
Scales of Independent Behavior-Revised (SIB-R) (1996)		Concomitante : Sous le seuil (< 0,5), mais un peu plus haut pour les individus en situation de handicap.				Élevé (0,80 ou plus)		
Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition [Early SB5]/ (2005) <i>Échelle d'intelligence Stanford-Binet</i>		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

1. Les outils d'évaluation diagnostiques (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Wechsler Individual Achievement Test, 2nd Edition (WIAT-II) (2001)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Wechsler Preschool and Primary Scale of Intelligence, 3rd Edition (WPPSI-III) (2002)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Assessment, Evaluation and Programming System (AEPS) ou Évaluation, intervention, suivi (EIS) : 0-3 ans (2002)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Bayley Scales Of Infant and Toddler Development (Third Edition) (2005)	Contenu : oui Conceptuelle : oui Concomitante : acceptable (> 0,5) Developmental Profile Prédictive : acceptable (> 0,4)				0,65 ou plus	Élevé (0,80 ou plus)		
Boehm Test of Basic Concepts, Preschool, 3rd Edition * (2001)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Boehm Test of Basic Concepts, 3 rd Edition* (2000)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Clinical Evaluation of Language Fundamentals – Preschool, 2nd Edition (CELF-P 2) * (2004)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Developmental Profile 3 (DP-3) (2007)	Contenu : analyse factorielle exploratoire et confirmatoire Conceptuelle : oui Concomitante : acceptable (> 0,5); Prédictive : acceptable (> 0,4)				0,65 ou plus			
Early Learning Accomplishment Profile (E-Lap) (1995)	Concomitante : acceptable (> 0,5); Prédictive : acceptable (> 0,4)				0,65 ou plus			

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Expressive Vocabulary Test, 2nd Edition (EVT-2) (2007)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		
High/scope Child Observation Record for Infants and Toddlers (COR-IT). (2002)	Concomitante : acceptable (> 0,5)	Concomitante : Élevée (> 0,70)			0,65 ou plus	Élevé (0,80 ou plus)		
Individual Growth Development Indicators (IGDIs) for Infants and Toddlers (2003)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus) pour la communication, mouvement, social et résolution de problème		
Infant-Toddler Developmental Assessment (IDA) (1995)	Concomitante : acceptable (> 0,5)				0,65 ou plus	Élevé (0,80 ou plus)		
Kaufman Survey of Early Academic and Language Skills (K-SEALS) (1993)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Kaufman Test of Educational Achievement, 2nd Edition (KTEA-II) * (2004)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Leiter International Performance Scale-Revised (LEITER-R) (1997)	Concomitante : acceptable (> 0,5)	Concomitante : Élevée (> 0,70)			0,65 ou plus	Élevé (0,80 ou plus)		
Metropolitan Readiness Test (MRT6), 6 th Edition (1994)		Niveau 2 : Concomitante : Acceptable (0,50 — 0,69)				Niveau 1 et 2 : Élevé (0,80 ou plus)		
Phonological Awareness and Literacy Screenings-PreKindergarten (PALS-PreK) (2004)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Preschool Language Assessment Instrument, 2nd Edition (PLAI-2) (2003)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		
Ready to Learn: A Dyslexia Screener (2004)		Concomitante : Acceptable (0,50 – 0,69)				Acceptable (0,50 – 0,69)		
Receptive One-Word Picture Vocabulary Test (ROWPVT) (2000)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Test of Early Language Development, 3rd Edition (TELD-3) (1999)		Concomitante : Élevée (> 0,70)				Élevé (0,80 ou plus)		
Test of Preschool Early Literacy (TOPEL) (2007)	Contenu : oui Concomitante : Acceptable (> 0,50)	Concomitante : Acceptable (0,50 – 0,69)			0,65 ou plus	Élevé (0,80 ou plus)		
Vineland Socio-Emotional Early Childhood Scales (SEEC) (1998)	Concomitante : Acceptable (> 0,50)				0,65 ou plus			

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus (suite)

2. Les outils d'intervention et de programmation (suite)

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Young Children's Achievement Test (YCAT) (2000)		Concomitante : Acceptable (0,50 – 0,69)				Élevé (0,80 ou plus)		

3— Les outils d'évaluation ayant une double fonction

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Achenbach System of empirically based assessment (ASEBA) (1999)	Concomitante : acceptable [$> 0,5$] Prédictive : acceptable [$> 0,4$]	Concomitante : Acceptable [0,50 – 0,69]			0,65 ou plus	Élevé [0,80 ou plus]		
Adaptive Behavior Assessment System (ABAS-II) (2003)		Concomitante : Acceptable [0,50 – 0,69]				Élevé [0,80 ou plus]		
Arizona Articulation Proficiency Scale, 3rd Edition (2001)		Concomitante : Élevée [$> 0,70$]				Élevé [0,80 ou plus]		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus [suite]

3— Les outils d'évaluation ayant une double fonction [suite]

Nom de l'outil	Données de validité rapportées par :			Données de fidélité rapportées par :				
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Battelle Developmental Inventory (BDI) (1984)	Contenu : oui Conceptuelle : oui Concomitante : acceptable [$> 0,5$]	Concomitante : Élevée [$> 0,70$]		Contenu : oui	0,65 ou plus	Élevé [0,80 ou plus]		Consistance interne : acceptable
Behavior Assessment System for Children (BASC-2) (2004)	Contenu : analyse factorielle confirmatoire Conceptuelle : oui Concomitante : acceptable [$> 0,5$]	Concomitante : Élevée [$> 0,70$]			0,65 ou plus	Élevé [0,80 ou plus]		
Brigance Diagnostic Inventory of Early Development II (IED-II) (2004)	Contenu : oui Conceptuelle : oui Concomitante : acceptable [$> 0,5$] Prédictive : acceptable [$> 0,4$]	Concomitante : Élevée [$> 0,70$]			0,65 ou plus	Acceptable [0,50 – 0,69]		
Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edition (BOT-2) (2006)		Concomitante : Élevée [$> 0,70$]				Élevé [0,80 ou plus]		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus [suite]

3— Les outils d'évaluation ayant une double fonction [suite]

Nom de l'outil	Données de validité rapportées par :			Données de fidélité rapportées par :				
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Comprehensive Test of Phonological Processing (CTOPP) (1999)		Concomitante : Acceptable [0,50 – 0,69]				Élevé [0,80 ou plus]		
Devereux Early Childhood Assessment for infants and toddlers (DECA – I/T) [2007]	Contenu : oui Conceptuelle : oui Concomitante : acceptable [$> 0,5$]				0,65 ou plus			
Expressive One-Word Picture Vocabulary Test (EOWPVT) (2000)		Concomitante : Élevée [$> 0,70$]			0,65 ou plus	Élevé [0,80 ou plus]		
MacArthur-Bates Communicative Development Inventory (2nd Edition), (2007) Les Inventaires MacArthur du développement de la Communication	Contenu : oui Conceptuelle : intercorrélations $> 0,59$ Concomitante : acceptable [$> 0,5$]				0,65 ou plus			

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus [suite]

3— Les outils d'évaluation ayant une double fonction [suite]

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Merrill-Palmer-Revised Scales of Development [M-P-R] [2004]		Concomitante : Élevée [$> 0,70$]				Élevé [0,80 ou plus]		
OWLS: Listening Comprehension (LC) Scale & Oral Expression (OE) Scale (1995)		Concomitante : Élevée [$> 0,70$]				Élevé [0,80 ou plus]		
OWLS: Written Expression (WE) Scale (1996)		Concomitante : Élevée [$> 0,70$]				Élevé [0,80 ou plus]		
Peabody Developmental Motor Scales, 2nd Edition (PDMS-2) * (2000)		Concomitante : Acceptable [0,50 – 0,69]				Élevé [0,80 ou plus]		

Annexe G. Données de validité des outils sélectionnés selon les documents de référence retenus [suite]

3— Les outils d'évaluation ayant une double fonction [suite]

Nom de l'outil	Données de validité rapportées par :				Données de fidélité rapportées par :			
	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)	Kisker et coll. (2011)	WSOSPI, 2008	INSPQ (2010)	Halle et coll. (2011)
Peabody Picture Vocabulary Test, 4 th edition [PPVT-IV] [2007]	<p>surtout auprès d'enfants d'âge primaire.</p> <p>Contenu : justification du choix de mots, pas de révision d'experts externes</p> <p>Concomitante : acceptable [$> 0,5$]</p> <p>Prédictive : acceptable [$> 0,4$]</p>	<p>Concomitante : Acceptable [0,50 – 0,69]</p>			0,65 ou plus	Élevé [0,80 ou plus]		
Test of Early Mathematics Ability, 3rd Edition (TEMA-3) (2003)		<p>Concomitante : Acceptable [0,50 – 0,69]</p>				Élevé [0,80 ou plus]		
Test of Early Reading Ability, 3rd Edition [TERA-3] [2001]		<p>Concomitante : Acceptable [0,50 – 0,69]</p>				Élevé [0,80 ou plus]		
Vineland Adaptive Behavior Scales (VABS) II Edition (2005)	<p>Contenu : analyse factorielle confirmatoire = 0,96</p> <p>Conceptuelle : ICC de 0,56 à 0,65</p> <p>Concomitante : acceptable [$> 0,5$]</p>	<p>Concomitante : Acceptable [0,50 – 0,69]</p>			0,65 ou plus	Élevé [0,80 ou plus]		

Informations générales

3255, rue Foucher, Trois-Rivières (Québec) G8Z 1M6

Téléphone : 819 379-6868

Ligne sans frais : 1 888 379-7732

www.crditedmcq.qc.ca

Pour nous suivre sur les médias sociaux :

 #CRDITEDMCQIU

 CrditedmcqIU