

## ■ Analyse psychométrique d'outils d'évaluation utilisés auprès des enfants francophones

## ■ Psychometric Analysis of Assessment Tools Used with Francophone Children

Marie-Eve Gaul Bouchard

Elizabeth M. Fitzpatrick

Janet Olds

### Abrégé

Le processus évaluatif constitue une partie intégrante du travail clinique dans le domaine de la santé et de l'éducation. Si un nombre important d'outils d'évaluation normalisés existe pour les patients/clients de langue anglaise, il en va bien autrement des outils adaptés aux patients/clients francophones qu'ils soient adultes ou enfants/adolescents. Les professionnels de l'évaluation du langage sont donc continuellement confrontés à la pénurie d'outils adaptés à leur clientèle. La présente étude poursuit deux objectifs : 1) faire le point sur la situation de l'évaluation du langage et de la parole des enfants francophones du Canada en produisant un répertoire des outils existants et 2) faire une étude critique de 31 d'entre eux. Les résultats obtenus démontrent que bien que plusieurs outils soient disponibles, peu d'entre eux répond toutefois aux standards psychométriques. Les résultats suggèrent donc d'appuyer toute démarche susceptible de mener au développement et à l'adaptation d'outils d'évaluation permettant de mieux desservir la population francophone canadienne.

### Abstract

The assessment process constitutes an integral part of clinical work in health and education. Although a large number of standardized assessment tools exist for English speaking patients/clients, the situation is quite different for tools tailored to francophone patients/clients, be they adults or children/adolescents. The current study has two goals: 1) to describe the current situation in speech and language assessment of francophone children in Canada by compiling an inventory of existing tools and 2) to complete a critical review of 31 of these tools. The findings demonstrate that although a large number of tools are available, few meet psychometric standards. The findings suggest that measures aimed at promoting the development and adaptation of assessment tools in French should be supported.

**Mots clés :** évaluation, francophone, langage, enfants, validité, fidélité

**L**e processus évaluatif constitue une partie intégrante du travail clinique dans le domaine de la santé et de l'éducation. Ses objectifs se trouvent généralement dans l'une des quatre catégories suivantes : 1) diagnostiquer la présence d'un trouble ou d'un déficit; 2) établir et décrire les forces et faiblesses d'un patient; 3) cibler des objectifs et des stratégies d'intervention; et finalement 4) mesurer la réponse face au traitement (Lahey, 1988, 1990; McCauley & Swisher, 1984a; Merrell & Plante, 1997; Paul, 2007).

Marie-Eve Gaul Bouchard  
Université du Québec à  
Montréal  
Montréal, Québec Canada

Elizabeth M. Fitzpatrick  
Faculté des sciences de la  
santé  
Université d'Ottawa  
Ottawa, Ontario Canada

Janet Olds  
Centre hospitalier pour  
enfants de l'est de l'Ontario  
Ottawa, Ontario Canada

Cet article porte sur les tests normalisés en français. Les tests standardisés ou normés présentent des données servant de système de référence puisque ces données sont recueillies auprès d'un échantillon représentatif de la population à laquelle on administre le test. Ils constituent les outils sur lesquels se base la majorité des décisions relatives aux deux premiers objectifs (Huang, Hopkins, & Nippold (1997) et sont toutefois moins utiles pour établir et mesurer les objectifs d'intervention. D'autres épreuves, telles que les tests à critères, sont plus appropriés (McCauley & Swisher, 1984b; Kerr, Guildford, & Kay-Raining Bird, 2003, Lefebvre & Trudeau, 2005). Contrairement aux tests standardisés, les tests à critères examinent les performances d'un sujet par rapport à des critères prédéterminés et non par rapport aux performances d'un groupe normatif, ceci permettant d'obtenir des informations cliniques tout aussi utiles.

### Limites des outils de langue anglaise traduits ou adaptés en français

Si un nombre important d'outils d'évaluation normés de la parole et du langage existent pour les clients de langue anglaise, il en va bien autrement des outils de langue française développés et adaptés aux clients francophones qu'ils soient adultes ou enfants/adolescents. En effet, les professionnels de l'évaluation<sup>1</sup> du langage sont continuellement confrontés à la pénurie d'outils adaptés à leur clientèle. Afin de pallier à cette difficulté, ils ont parfois tenté de créer des tests originaux, mais ils ont surtout emprunté des outils conçus en Europe et traduits ou adaptés des tests anglais en français (Garcia & Desrochers, 1997). Les résultats de ces deux dernières stratégies n'ont cependant pas toujours été des plus heureux. En effet, dans la majorité des cas, ces adaptations/traductions ont été effectuées pour répondre à un besoin urgent de combler le manque d'outils comparables en français. Par conséquent, dans l'utilisation de ces versions « empruntées », les spécialistes se retrouvent souvent devant trois problèmes: 1) ces tests reposent rarement sur des normes culturelles et linguistiques appropriées pour la population ciblée; 2) ils ne sont plus équivalents, sur le plan linguistique, aux tests initiaux de langue anglaise; et 3) ils ne reposent pas sur des normes appropriées satisfaisant à des critères psychométriques pour la population visée. Les conséquences reliées à ces difficultés sont susceptibles de limiter la validité des informations pouvant être tirées de l'évaluation du langage, tout comme les décisions qui en découlent.

### Limites linguistiques et culturelles

Au plan linguistique et culturel, par exemple, quels renseignements utiles à l'évaluation d'un patient canadien-français peuvent être tirés d'un test de vocabulaire dont les images illustrent des variétés de plantes typiques de la France et peu connues au Canada ou dans lesquelles

<sup>1</sup>Par l'appellation *professionnels de l'évaluation*, nous entendons les orthophonistes, les audiologistes, les psychologues, les enseignants, ou tout autre professionnels qui, dans leur travail, ont à évaluer le langage.

figurent des personnages célèbres d'un pays étranger (Garcia & Desrochers, 1997). De la même manière, on peut se demander comment le clinicien interprétera-t-il la production d'un anglicisme ou d'une expression régionale, tous deux absents du manuel de correction, mais pourtant bien présents dans la réalité linguistique du patient évalué. L'ensemble des réponses admissibles dans chaque test doit donc prendre en compte ces comportements linguistiques régionaux, puisque toute déviation à ce principe peut conduire à une sous-estimation des capacités linguistiques réelles du client.

### Limites relatives aux critères linguistiques

Au plan linguistique, des problèmes en lien avec la traduction et l'adaptation même de l'outil peuvent fausser l'interprétation et le diagnostic clinique qui en ressort. Par exemple, dans la construction des tests, les auteurs contrôlent généralement avec minutie la longueur des paragraphes ou des textes qui font partie de l'évaluation. Or, il n'est pas rare de voir que les textes traduits, bien qu'équivalents en ce qui a trait au sens, ne le sont pas quant à la longueur du texte. Or, un texte racontant la même histoire qu'un autre récit mais en moins ou plus de mots, pourrait induire une variation dans le niveau de difficulté, ce qui se reflète inévitablement dans la performance obtenue au test. La traduction des items peut également modifier le niveau de difficulté des diverses structures grammaticales ou la difficulté relative de diverses structures grammaticales composant les items. Par exemple, le pronom pluriel d'un test anglais est marqué par « they » alors qu'en français, il est marqué par « ils » ou « elles ». Des items qui ne sont pas comparables à la version originale pour ce qui est du niveau de difficulté laissent le clinicien dans le doute à propos de la validité réelle des habiletés évaluées par le test. Ces quelques exemples démontrent bien qu'adapter un texte ne s'en tient pas uniquement à la traduction formelle des items mais implique trois procédures: 1) adapter le matériel d'emprunt de manière à assurer sa pertinence culturelle pour la population ciblée, 2) prendre connaissance des variables (p.ex., longueur des mots, fréquence d'occurrence dans la langue) qui ont été considérés dans le choix des stimuli de manière à effectuer les mêmes contrôles dans le choix des stimuli français correspondants, et finalement 3) bien comprendre les objectifs de l'outil utilisé afin de préserver la comparabilité du nouvel outil avec la version originale.

### Limites relatives aux normes

Finalement, en ce qui concerne les normes d'un test, elles permettent de situer un individu par rapport à son groupe de référence. En d'autres mots, un score faible au test se traduit par une performance faible dans l'échantillon de normalisation et, par inférence, un score faible dans la population. Toutefois, pour être en mesure de faire une telle inférence, l'outil doit d'abord répondre aux critères psychométriques que sont: 1) la validité, c'est-à-dire que le test mesure véritablement ce qu'il prétend mesurer (McCauley & Swisher, 1984b) et 2) la fidélité, c'est-à-dire

la capacité de reproduire les résultats obtenus auprès d'un individu dans un intervalle de temps donné (McCauley & Swisher, 1984b). Ainsi, les scores dérivés d'un outil ne satisfaisant pas à ces qualités psychométriques sont très difficiles à interpréter et les décisions cliniques impossibles à justifier sur la base de la performance au test uniquement. Il faut donc considérer la standardisation comme une police d'assurance; elle ne prévient pas tous les cas aberrants, mais elle fournit un cadre de référence permettant de vérifier efficacement les hypothèses cliniques et de confirmer la présence de certains déficits.

Les auteurs de ces outils adaptés/traduits ainsi que les cliniciens qui les utilisent sont bien conscients des imperfections des outils qu'ils ont élaborés et qui sont décrites précédemment. Toutefois, malgré ces lacunes, les professionnels considèrent qu'à choisir entre une absence totale de ressources et des outils imparfaits, il est plus constructif d'adopter la seconde stratégie (Garcia & Desrochers, 1997).

### Les outils de langue française

La réalité du manque d'outil et l'insatisfaction des cliniciens quant à la disponibilité des tests d'évaluation en français sont entre autres discutées par Garcia, Paradis, Sénécal & Laroche (2006) et Kerr, Guildford, & Kay-Raining Bird (2003). Malgré que certains travaux se soient intéressés à compiler les outils de langue française les plus fréquemment utilisés par les audiologistes et orthophonistes canadiens (p.ex. Garcia et al., 2006; Garcia & Desrochers, 1997), aucun article n'a tenté de dresser un répertoire des outils d'évaluation existants. Pourtant un tel travail s'avère des plus pertinent puisqu'il semble que la sélection des tests utilisés par une majorité de professionnels, peu importe la langue de pratique, est guidée principalement par des critères subjectifs et non objectifs. L'exemple qui suit le démontre bien. Record et Tomblin (1994) ont examiné la pratique de 27 orthophonistes-cliniciens, membres de l'*American Speech-Language-Hearing Association* (ASHA). Ils rapportent que l'opinion concernant les qualités psychométriques de l'outil, ainsi que la familiarité avec le test représentaient, pour ces professionnels, les critères les plus décisifs dans la sélection des tests à utiliser pour évaluer le langage d'enfants. De plus dans un sondage sur les habitudes d'évaluation des audiologistes et orthophonistes canadiens, Kerr et al. (2003) rapportent que, dans leur travail auprès des enfants, les répondants du sondage utilisaient généralement les cinq mêmes tests pour exécuter les tâches liées à l'évaluation et à la réadaptation. Il est justifié de s'interroger à savoir dans quelle mesure une telle pratique serait guidée, du moins en partie, par un manque de connaissances de l'existence d'autres outils.

### Étude des qualités psychométriques des outils d'évaluation

Un autre aspect important sur lequel les professionnels canadiens rapportent de l'insatisfaction concerne la qualité psychométrique des mesures auxquelles ils ont accès (Garcia et al., 2006; Garcia & Desrochers, 1997).

Aucune étude ne s'est intéressée à évaluer la qualité psychométrique des tests de langue française utilisés pour évaluer les jeunes canadiens francophones. Connaître la qualité psychométrique des outils actuellement utilisés s'avère crucial puisque l'utilisation de tests standardisés qui répondent aux critères de psychométrie est nécessaire afin de répondre à la préoccupation grandissante d'une pratique fondée sur des données probantes (*evidence-based practice*; Shipley & McAfee, 2004). De plus, l'Ordre des audiologistes et orthophonistes de l'Ontario (2000), ainsi que l'Ordre des orthophonistes et audiologistes du Québec (2003) préviennent leurs membres que le fait de tirer des conclusions basées sur des tests non standardisés ou qui ne possèdent pas des niveaux de qualités psychométriques appropriés va à l'encontre du code déontologique de leur ordre professionnel.

Il est possible de retrouver dans la littérature diverses études de qualités psychométriques menées sur les tests américains de langage et d'articulation utilisés auprès des enfants. Une première est celle de McCauley et Swisher (1984a) dans laquelle les auteurs ont examiné 30 tests de langage et d'articulation pour les enfants d'âge préscolaire, en utilisant 10 critères psychométriques portant principalement sur la validité, la fidélité, et les caractéristiques de l'échantillon de normalisation. Les résultats révèlent que la plupart des informations nécessaires pour évaluer la validité et la fidélité des tests étaient absentes du manuel de l'utilisateur. En effet, seulement 20% des tests remplissaient 50% des critères.

Reprenant les critères de McCauley et Swisher (1984a) et en ajoutant d'autres pour considérer, entre autres, les effets de plancher et de plafond potentiels dans certains tests, Plante et Vance (1994) ont mené, 10 ans plus tard, une analyse de 20 tests destinés à évaluer le langage d'enfants d'âge préscolaire. Leurs résultats démontrent que seulement 38% des tests répondaient à la moitié des critères. Ils en concluent que malgré l'évolution des standards de pratique, cette plus grande sensibilité ne se transpose toutefois pas en gains substantiels concernant la qualité des informations fournies aux praticiens par les éditeurs de tests.

Finalement, Hutchinson (1996) suggère que les professionnels devraient se baser sur 20 critères pour évaluer les qualités psychométriques d'un test à utiliser dans une démarche clinique. En plus des critères reliés à la validité, la fidélité et à la standardisation, l'auteur aborde la question des fondements théoriques du test et soulève des questions relatives aux objectifs, au développement et à l'interprétation du test.

### Objectifs de l'étude

Le présent travail a donc deux principaux objectifs : 1) générer un répertoire des outils d'évaluation disponibles en français pour les enfants francophones nord-américains et 2) faire une revue des qualités psychométriques de certains outils standardisés faisant partie du répertoire généré. Dans cet article, nous présenterons succinctement les principaux résultats obtenus dans le développement

du répertoire des tests et nous détaillerons les résultats du deuxième objectif.

Par cette recherche, nous voulons faire le point sur les qualités psychométriques de certains outils standardisés servant à l'évaluation du langage des enfants francophones au Canada. Nous désirons ainsi encourager la discussion à ce sujet et fournir aux professionnels davantage d'informations susceptibles de les aider à choisir des instruments. Nous souhaitons également les informer sur les répercussions relatives à l'utilisation des tests sur le diagnostic et la prise en charge subséquente des clients/patients, lorsque les qualités psychométriques de ces tests sont potentiellement défaillantes. Finalement, nous espérons que la présente étude stimulera la poursuite de travaux visant le développement d'outils spécifiquement destinés aux jeunes enfants canadiens dont la langue maternelle est le français « québécois », l'adaptation de tests déjà existants (en anglais ou en français « européen ») ainsi qu'une meilleure standardisation des outils actuellement en usage. Afin de répondre aux deux objectifs, l'étude a été divisée en deux parties. La première partie décrit la démarche menant à la création d'un répertoire des outils d'évaluation du langage utilisés en français tandis que la deuxième partie se veut une critique de certains des outils répertoriés.

## Partie 1 : Répertoire des outils d'évaluation du langage en français

### Méthodologie

La méthodologie utilisée pour mener à la création du répertoire des outils d'évaluation du langage compte plusieurs étapes. En tout premier lieu, une revue de littérature a été effectuée sur *PubMed* et aussi en consultant tous les exemplaires de la *Revue d'orthophonie et d'audiologie (Journal of Speech-Language Pathology and Audiology)* depuis 1985. Dans une seconde étape, une recherche élargie dans Internet a permis d'identifier des sites de regroupements professionnels européens dans les domaines de la logopédie et de la neuropédiatrie. Ces sites se sont avérés très utiles afin d'identifier des outils développés pour les enfants francophones d'origine européenne (p.ex. : française, belge, etc.).

Dix chercheurs ont par la suite été contactés afin de vérifier dans quelles mesures ils connaissaient les outils trouvés. Ces chercheurs occupaient des postes de recherches et d'enseignement dans des facultés telles que celles de l'éducation, de la psychologie, de l'audiologie et de l'orthophonie et étaient spécialisés en psychométrie ou dans l'évaluation des enfants. Ces chercheurs occupaient des postes en recherche et en enseignement dans des universités canadiennes. Certains d'entre eux travaillaient entre autres à l'adaptation en français québécois d'outils initialement conçus en anglais ou à la création de nouveaux tests destinés à la population pédiatrique québécoise. Leur contact a mené à l'obtention de documents généraux de référence portant sur l'évaluation du langage des enfants mais aussi sur les facteurs importants devant être considérés dans

l'évaluation de cette population, comme par exemple les œuvres de Rondal (1997) et Chevrier-Muller & Narbona (1999). Toutefois, peu d'entre eux connaissaient les tests européens qui leur ont été présentés. Les tests « adaptés » ou les versions « maison » d'outils en français initialement conçus en anglais leur étaient davantage familiers.

Des rencontres ont par la suite été effectuées avec sept praticiens (orthophonistes, psychologues, spécialistes en déficience auditive, et audiologistes) travaillant dans divers milieux incluant commissions scolaires, centres hospitaliers et cliniques privées dans le but d'en apprendre davantage sur les outils qu'ils connaissaient et sur ceux qu'ils utilisaient les plus fréquemment dans le cadre de leurs évaluations. Nous les avons également interrogés à propos de leurs besoins en matière d'évaluation du langage pour les enfants nord-américains de langue française et sur les facteurs qui motivaient généralement la sélection des tests qu'ils utilisaient dans leur travail quotidien. Parmi les outils les plus fréquemment nommés, se retrouvaient quelques tests développés pour les enfants européens de langue française, mais surtout des versions « maisons » adaptés en français, de tests anglophones pour lesquels les normes américaines servaient de critères de référence.

Finalement, la visite de diverses testothèques, matériathèques, médiathèques et docutothèques des facultés/programmes de psychologie, d'orthophonie et d'éducation a permis de mettre en lumière l'existence d'une multitude de versions « maison » de divers tests fréquemment utilisés en langue anglaise ayant été adaptés en français. Ces visites ont également permis d'identifier et de se familiariser avec d'autres épreuves ayant été développées pour les enfants francophones. Bien que les outils d'évaluation de la parole et du langage sont surtout administrés par les orthophonistes ou autres thérapeutes, nous avons également inclus les tests d'évaluation cognitive de l'enfant comprenant une composante de l'évaluation du langage.

### Résultats

À la suite de discussions avec des professionnels de l'évaluation dans les domaines de la santé et de l'éducation, les tests publiés à partir de 1948 ont été inclus dans le répertoire. Les tests publiés avant cette année n'étant plus ou que rarement utilisés par les professionnels. Neuf domaines généralement reconnus, tant par les professionnels que par les maisons d'édition de tests, pour être liés au langage et à son développement ont été considérés : 1) la reconnaissance/perception de la parole, 2) la production de la parole/articulation, 3) la mémoire auditive verbale, 4) la conscience phonologique, 5) la compréhension du langage, 6) la production du langage, 7) la lecture, 8) l'écriture/l'orthographe, et 9) la cognition/intelligence. Pour chaque test, les informations suivantes ont été répertoriées : le nom complet du test, le/les auteurs, la date de parution (la date initiale, ainsi que les parutions ultérieures), l'âge (ou les caractéristiques spéciales) de la population visée, les caractéristiques de l'échantillon (nombre d'enfants, origine, spécification de l'échantillon),

et finalement le type d'adaptation. À ce sujet, l'appellation *pour enfants francophones* sous-entend un test ayant été développé spécifiquement pour les enfants de langue française; celle *adaptation maison* signifie une traduction libre effectuée par des cliniciens sans étude de normalisation et de standardisation. Par conséquent, l'échantillon du test original sert de critère de référence pour comparer les résultats obtenus par un patient. Finalement, l'appellation *adaptation et validation* signifie qu'une traduction officielle en français du test originalement développé dans la langue anglaise a été effectuée et qu'une étude de normalisation a été menée auprès d'un groupe d'enfants francophones (et dans certains cas francophones canadiens) pour guider les décisions cliniques.

Le Tableau 1 présente les caractéristiques d'adaptation des outils du répertoire des tests. Chaque test ne représente pas qu'un seul domaine puisque certains outils particuliers (p. ex., Batterie d'évaluation du langage écrit et de ses troubles, BELEC) ciblent plusieurs domaines identifiés au Tableau 1. Le répertoire compte 110 outils. La majorité des tests ont été développés pour les enfants francophones, puis des adaptations/validations et finalement des adaptations « maison ». Il est toutefois possible que le nombre d'outils dans la catégorie « validation maison » soit sous-estimé puisque durant nos rencontres avec les professionnels, certains étaient hésitants à montrer ces outils « maison » ou à discuter de leur existence. Le Tableau 1 présente donc les outils « maison » auxquels nous avons eu accès.

L'analyse du répertoire permet de faire plusieurs observations. Tout d'abord, les outils évaluant le langage, que ce soit dans son volet expressif ou réceptif, constituent la catégorie possédant le plus grand nombre de tests. On note que peu d'outils (10) existent pour la reconnaissance/perception de la parole comparativement au nombre pour la production de la parole/articulation (13). On observe également que certains domaines, tels que la lecture et l'orthographe sont presque exclusivement composés

de tests développés ou adaptés/validés pour les enfants francophones comparativement à ceux évaluant le langage réceptif ou expressif qui comptent en plus de nombreuses adaptations « maison ». En effet, sur les 51 outils répertoriés pour les domaines reliés à l'alphabétisation (lecture/orthographe), seulement deux représentent des adaptations de tests anglophones et aucune adaptation « maison » n'a été répertoriée. À l'opposé, pour les domaines du langage réceptif et expressif, 10 tests sont des adaptations « maison ». De telles disparités entre les domaines peuvent être liées aux besoins en évaluation qui s'avèrent plus grands et sans doute plus diversifiés dans les domaines du langage réceptif ou expressif que ceux pour l'évaluation de la lecture et l'écriture. Finalement, les tests développés pour les enfants francophones d'origine européenne sont nombreux et présents dans tous les domaines étudiés.

## Partie 2 : Critique de tests d'évaluation du langage en français

### Méthodologie

Pour les 110 outils répertoriés, des critères d'exclusion ont été appliqués pour l'analyse psychométrique. Tout d'abord, de ce nombre, les 13 tests ayant eu une *adaptation « maison »* n'ont pas été considérés pour l'analyse. De plus, 66 outils furent retirés de la liste de tests puisque le test ou le manuel de l'examineur n'était pas disponible pour l'évaluation. Ainsi, un total de 31 tests ou outils furent soumis à l'analyse psychométrique (Tableau 2).

La procédure utilisée pour l'analyse psychométrique des outils rejoint celle d'abord proposée par McCauley et Swisher (1984a) puis réactualisée par Plante et Vance (1994). Des critères ont été ajoutés de manière à représenter plus fidèlement de nouvelles tendances observées aujourd'hui par les maisons d'édition d'outils psychométriques. Ces six critères additionnels proviennent principalement d'écrits scientifiques dont Hutchison (1996) et Beran (2003). Au total, 16 critères représentent des caractéristiques de base

devant être obligatoirement considérées par le clinicien pour utiliser un test dans le but de poser un diagnostic ou d'émettre une décision clinique à propos d'une performance d'un patient/client au test. Puisque ces tests sont utilisés par des professionnels de différents domaines, ces critères sont importants car ils assurent d'obtenir des informations plus objectives. Conséquemment, les décisions cliniques qui en découlent risquent moins d'être influencées par la manière dont les praticiens conceptualisent et interprètent les construits évalués. Ces critères sont présentés et expliqués en Annexe A.

**Tableau 1**

*Caractéristiques d'adaptation des outils répertoriés*

Domaine	Développé pour les enfants francophones		Adaptation/Validation		Adaptation maison
	Canadiens	Non-canadiens	Canadiens	Non-canadiens	
Reconnaissance de la parole		8	1		1
Production de la parole	1	11			1
Conscience phonologique	1	14			1
Mémoire auditive		12		1	
Langage réceptif		28	3	4	5
Langage expressif		37	1	5	5
Lecture	1	26		2	0
Écriture/Orthographe		22	0		0
Cognition/Rendement intellectuel	2	9	3	7	0

L'analyse psychométrique a été effectuée par une doctorante en psychologie. Au cours du processus, à la suite des discussions, l'évaluatrice et deux autres chercheurs en sont venus à un consensus sur les critères s'appliquant à chaque test. Le nombre de critères de chaque test a été additionné, chaque critère possédant une importance relative équivalente (i.e. 1 critère = 1 point). Un test a été considéré comme satisfaisant un critère si le manuel présentait, dans son entier, suffisamment d'information en lien avec le critère en question pour en permettre l'évaluation.

### Résultats

Le Tableau 2 présente le nombre de tests satisfaisants chaque critère. En moyenne, 14,18 tests (écart-type = 11,18) ont satisfait chaque critère. Les critères les plus fréquemment satisfaits sont ceux en lien avec (1) les objectifs du test, (2) la définition des construits, (12) la taille de l'échantillon de normalisation, (15) l'administration et les qualifications du clinicien, et finalement (16) l'année de publication du test et de l'étude de standardisation. À l'opposé, les critères les moins fréquemment satisfaits sont ceux qui abordent la question de (3) la validité de construit, (6) la validité prédictive, (9) de la fidélité inter-juges, (13) des effets plancher/plafonds, et finalement (14) des positions extrêmes. Aucun des tests analysés n'a satisfait les critères 9 et 14.

Le Tableau 3 présente les tests satisfaisants à chaque critère. Malgré l'apparition plus fréquente de certains outils, le patron observé à la Figure 1 n'est toutefois pas dysharmonieux (par exemple, quelques tests qui rencontrent remplissent tous les critères, alors que d'autres n'en remplissent que très peu). En moyenne, chaque test a rempli 6,81 critères (écart-type : 2,91). Quatorze critères est le nombre le plus élevé de caractéristiques ayant été satisfaites, et un seul test, l'Échelle d'intelligence de Weschler pour enfants, 4<sup>ième</sup> édition, (WISC-IV), y parvient. Deux tests, l'Échelle d'intelligence de Weschler pour la période préscolaire et primaire, 3<sup>ième</sup> édition (WPPSI-III) et les Inventaires MacArthur du développement de la communication (IMDC), une adaptation et validation d'un test anglais avec des enfants franco-québécois, satisfont 12 critères. L'Échelle de vocabulaire en images Peabody (ÉVIP), qui est une adaptation et validation du *Peabody Picture Vocabulary Test* avec des individus d'origine canadienne française a répondu à 10 critères. À l'opposé, le Test de lecture California (California Lecture Test) ne satisfait qu'à deux critères. En somme, 16 tests évalués répondent à moins de sept critères d'analyse.

### Discussion

Cette étude visait en tout premier lieu à répertorier des outils existants pour l'évaluation du langage des enfants francophones canadiens. Les résultats ont montré que parmi les 110 outils retrouvés, très peu d'entre eux ont été spécifiquement développés pour les enfants francophones nord-américains. Premièrement, les résultats révèlent que les professionnels doivent

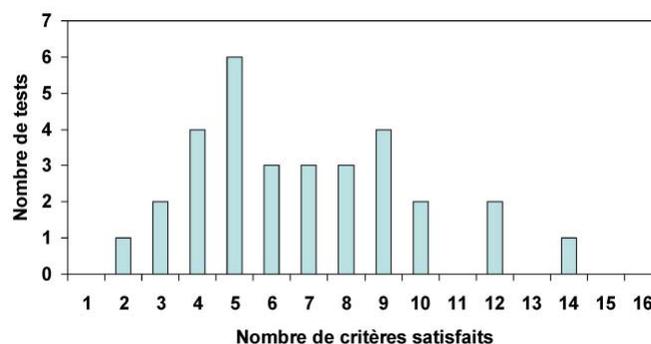


Figure 1. Distribution des tests par nombre de critères psychométriques satisfaits

avoir recours à plusieurs outils d'évaluation de conception « maison » ou des outils d'abord conçus pour une population américaine anglophone. L'utilisation de tels outils peut compromettre la validité et la fiabilité de l'évaluation si ces derniers ne sont pas développés selon des règles précises (voir Vallerand, 1989). Conséquemment, la marge d'erreur possible dans les résultats issus du processus d'évaluation pourrait être considérablement élevée. La démarche de vérification d'hypothèses cliniques se trouve potentiellement compromise et la mesure de l'interprétation de la sévérité du trouble, erronée (Garcia et al., 2006; Garcia & Desrochers, 1997).

Les résultats montrent également que plusieurs outils répertoriés sont conçus et validés en Europe. Or, la réalité des enfants européens se distingue de celle de la population pédiatrique canadienne française à plusieurs égards. Par exemple, comme l'utilisation et le contenu du vocabulaire sont intimement liés à la culture, le français européen et le français canadien présentent certainement des différences marquées; leurs systèmes d'éducation, tant dans l'organisation de sa structure que dans le contenu académique de ses programmes, diffèrent et ne correspondent pas à la réalité du système d'éducation canadien français comme le Québec (p.ex. un enfant français de 10 ans est en 2<sup>ième</sup> année de son primaire moyen dans le cycle III, alors qu'un enfant allemand débute sa première année de secondaire; l'enfant suisse fait sa première année de transition et finalement l'enfant québécois est en 5<sup>ième</sup> année du 3<sup>ième</sup> cycle du primaire). Les cliniciens utilisant les outils doivent donc être tout particulièrement vigilants et bien connaître le système scolaire d'où proviennent les enfants formant l'échantillon de normalisation.

Une dernière réalité mise en lumière par cette recherche est celle des outils francophones développés au Québec et publiés dans les années '60 et '70. De tels tests, bien qu'intéressants sont désuets. Ils ne sont plus adaptés à la réalité des enfants d'aujourd'hui tant au plan de leur contenu que des théories sur lesquels ils s'appuient. Ces dernières ne correspondent plus aux

**Tableau 2***Tests remplissant les critères de qualités psychométriques*

Critère	Nombre de tests	Tests remplissant les conditions du critère
1	31	BELEC, ELO, N-EEL, L2MA, IMDC, EVIP, WPPSI-III, LMC-R, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, BTBC, BTBC-PS, DORLEC, ALOUETTE-R, BQAL, Jeannot et Georges, EIHM-E, ECOSSE, EC-SL-SE, BPAL, VL-Kh, TV-BP, WISC-IV, EIHM-10/24, ITPA, LAC, TDD
2	28	BELEC, ELO, N-EEL, L2MA, IMDC, EVIP, WPPSI-III, LMC-R, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, BTBC, BTBS-PS, DORLEC, ALOUETTE-R, BQAL, Jeannot et Georges, EIHM-E, ECOSSE, EC-SL-SE, VL-Kh, TV-BP, WISC-IV, EIHM-10/24, TDD
3	5	EVIP, WPPSI-III, BTBC, WISC-IV, TDD
4	9	IMDC, WPPSI-III, EPFL, EIHM-E, EC-SL-SE, BPAL, WISC-IV, EIHM-10/24, TDD
5	13	IMDC, EVIP, WPPSI-III, Cube de Kohs, TNO, TIMÉ-2, BTBC-PS, BQAL, EC-SL-SE, BPAL, WISC-IV, TDD
6	3	IMDC, BTBC, WISC-IV
7	14	IMDC, EVIP, WPPSI-III, LMC-R, TNO, TIMÉ-2, BTBC, BTBC-PS, EIHM-E, EC-SL-SE, BPAL, WISC-IV, EIHM-10/24,
8	12	N-EEL, IMDC, EVIP, WPPSI-III, TNO, TIMÉ-2, EPFL, BTBC, BTBC-PS, EIHM-E, WISC-IV, EIHM-10/24
9	0	
10	6	IMDC, EVIP, WPPSI-III, BTBC, BTBC-PS, WISC-IV
11	13	N-EEL, L2MA, IMDC, WPPSI-III, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, ECOSSE, EC-SL-SE, WISC-IV, EIHM-10/24
12	24	BELEC, L2MA, IMDC, EVIP, WPPSI-III, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, BTBC, BTBC-PS, ALOUETTE-R, CRT, Jeannot et Georges, EIHM-E, ECOSSE, EC-SL-SE, BPAL, VL-Kh, TV-BP, WISC-IV, EIHM-10/24
13	2	ELO, WISC-IV
14	0	
15	30	BELEC, ELO, N-EEL, L2MA, IMDC, EVIP, WPPSI-III, LMC-R, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, BTBC, BTBC-PS, ALOUETTE-R, BQAL, CRT, DORLEC, Jeannot et Georges, EIHM-E, EC-SL-SE, BPAL, VL-Kh, TV-BF, WISC-IV, EIHM-10/24, ITPA, LAC, TDD
16	28	BELEC, ELO, N-EEL, L2MA, IMDC, EVIP, WPPSI-III, LMC-R, Cube de Kohs, TNO, TIMÉ-2, ODEDYS, EPFL, BTBC, BTBC-PS, ALOUETTE-R, BQAL, CRT, DORLEC, EIHM-E, EC-SL-SE, BPAL, VL-Kh, WISC-IV, EIHM-10/24, ITPA, LAC, TDD

**Liste des tests évalués**

- ALOUETTE-R : *L'Alouette*. Lefavrais, P. (2006).
- BELEC : *Batterie d'évaluation du langage écrit et de ses troubles*. Mousty, P., Leybaert, J., Alegria, J., Content, A., & Morais, J. (1994).
- BPAL : *Le temps d'apprendre à lire : Batterie prédictive de lecture-écriture d'Inizan*. Inizan, A. (2000).
- BQAL : *Bilan qualitatif des apprentissages de la lecture*. Campeau-Fillion, F., & Gauthier, G. (1990).
- BTBC : *Test des concepts de base de Boehm (TCBB)*. Chevrier, J. M. (1974).
- BTBC-PS : *Test des concepts de base de Boehm (TCBB-P)*. Chevrier, J. M. (1974).
- CRT : *Test de lecture California (CRT)*. Claes, M., Dehant, A., & Lamy, J. (1977).
- Cube de Kohs : *Test des cubes de Kohs*. Kohs, C. (1972).
- DORLEC : *Disposition Orthographe Lecture*. Lobrot, M. (1980).
- EC-SL-SE : *Échelle composite du savoir lire et du savoir écrire au cours préparatoire*. Inizan, A., Inizan, A., & Bartout, D. (2002).
- ECOSSE : *Épreuve de Compréhension Syntaxico-Sémantique*. Lecocq, P. (1996).
- EIHM-10/24 : *Épreuve individuelle d'habileté mentale (EIHM-Adultes)*. Chevrier, J.-M. (1989).
- EIHM-E : *Épreuve individuelle d'habileté mentale*. Chevrier, J.-M. (1989).
- ELO : *Évaluation du langage oral*. Khomsi, A. (2001).
- EPFL : *Épreuve de performance fonctionnelle en lecture*. Ruel, P.-H. (1976).
- EVIP : *Échelle de vocabulaire en images Peabody (EVIP)*. Dunn, L. M., Theriault-Whalen, C. M., & Dunn, L. M. (1991).
- N-EEL : *Nouvelles épreuves pour l'examen du langage*. Chevrié-Muller, C., & Plaza, M. (2001).
- IMDC : *Inventaires MacArthur du développement de la communication (CDI)*. Fenson, L., Dale, P. S., Reznick, J. S., & Thal, D. J. (1993).  
Adaptation par Trudeau, N., Frank I., & Poulin-Dubois, D. (1999).
- ITPA : *Illinois Test of Psycholinguistic Abilities*. Kirk & McCarthy. (1961). Adaptation par Salomon, A. (1975).
- Jeannot et Georges : *Épreuve de Jeannot et Georges*. Hermabessier, G. & Sax, H. (1972).
- LAC : *Lindamood auditory conceptualization test*. Lindamood, H.C. & Lindamood, P.C. (1979). Adaptation par Silencieux, S. (1999).
- L2MA : *Langage oral, langage écrit, mémoire & attention*. Chevrié-Muller, C., Simon, A. M., & Fournier, S. (1997).
- LMC-R : *Épreuve d'évaluation de la compétence en lecture*. Khomsi, A. (1990).
- ODEDYS : *Outil de Dépistage des Dyslexies*. Valdois, S., Jacquier-Roux, M., & Zorman, M. (2002).
- TDD : *Test de dyslexie*. Griffin, J. R., & Howard, W. (1998).
- TNO : *Test du niveau d'orthographe*. Doutriaux, F., & Lepez, R. (1980).
- TIMÉ-2 : *Test d'identification du monde écrit de 6 à 8 ans*. Ecalle, J. (2003).
- TV-BP : *Test de vocabulaire de Binois et Pichot*. (1958).
- VL-Kh : *Vitesse en Lecture*. Khomsi, A., Pasquet, F., Parbeau-Guéno, & Nanty, I. (2005).
- WISC-IV : *Échelle d'intelligence de Weschler pour enfants*, 4<sup>ème</sup> édition. (2005).
- WPPSI-III : *Échelle d'intelligence de Weschler pour la période préscolaire et primaire 3<sup>ème</sup> édition (WPPSI-III)*. Weschler, D. (2002).

**Tableau 3***Nombres de critères remplis par chaque test*

Nom du test	Nombre de critères	Critères*
ALOUETTE-R	5	1, 2, 12, 15, 16
BELEC	5	1, 2, 12, 15, 16
BPAL	7	1, 4, 5, 7, 12, 15, 16
BQAL	5	1, 2, 5, 15, 16
BTBC	10	1, 2, 3, 6, 7, 8, 10, 12, 15, 16
BTBC-PS	9	1, 2, 5, 7, 8, 10, 12, 15, 16
CRT	2	12, 15
Cube de Kohs	7	1, 2, 5, 11, 12, 15, 16
DORLEC	4	1, 2, 15, 16
EC-SL-SE	9	1, 2, 4, 5, 7, 11, 12, 15, 16
ECOSSE	4	1, 2, 11, 12
EIHM-10/24	9	1, 2, 4, 7, 8, 11, 12, 15, 16
EIHM-E	8	1, 2, 4, 7, 8, 12, 15, 16
ELO	5	1, 2, 13, 15, 16
EPFL	8	1, 2, 4, 8, 11, 12, 15, 16
EVIP	10	1, 2, 3, 5, 7, 8, 10, 12, 15, 16
IMDC	12	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 15, 16
ITPA	3	1, 15, 16
Jeannot et Georges	4	1, 2, 12, 15
L2MA	6	1, 2, 11, 12, 15, 16
LAC	3	1, 15, 16
LMC-R	5	1, 2, 7, 15, 16
N-EEL	6	1, 2, 8, 11, 15, 16
ODEDYS	6	1, 2, 11, 12, 15, 16
TDD	7	1, 2, 3, 4, 5, 15, 16
TIMÉ-2	9	1, 2, 5, 7, 8, 11, 12, 15, 16
TNO	8	1, 2, 5, 7, 8, 11, 12, 15
TV-BP	4	1, 2, 12, 15
VL-Kh	5	1, 2, 12, 15, 16
WISC-IV	14	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16
WPPSI-III	12	1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 15, 16

\*Synthèse des critères d'analyse :

Critère 1 : Objectifs du test clairement identifiés

Critère 2 : Définition du ou des construits mesurés par le test

Critère 3 : Information en ce qui a trait à la validité de contenu

Critère 4 : Information en ce qui a trait à la validité de construit

Critère 5 : Information en ce qui a trait à la validité concordante

Critère 6 : Information en ce qui a trait à la validité prédictive

Critère 7 : Information en ce qui a trait à la cohérence interne

Critère 8 : Information en ce qui a trait à la fidélité test-retest

Critère 9 : Information en ce qui a trait à la fidélité inter-juges

Critère 10 : Information en ce qui a trait à l'erreur standard de mesure (ESM)

Critère 11 : Information en ce qui a trait à l'échantillon de normalisation

Critère 12 : Information en ce qui a trait à la taille des échantillons

Critère 13 : Information en ce qui a trait aux effets planchers et plafonds

Critère 14 : Information en ce qui a trait à la manière dont sont représentés les enfants dans les extrêmes

Critère 15 : Information en ce qui a trait à la passation/administration du test

Critère 16 : Information sur l'année de publication et de standardisation de l'échantillon de normalisation

connaissances actuelles dans le domaine du développement de l'enfant et de ses troubles. Par exemple, les découvertes récentes dans les sciences neurologiques et cognitives ont modifié les façons de conceptualiser certains troubles neurologique du développement et par le fait même la manière de les évaluer.

Les résultats obtenus à la suite de la revue des qualités psychométriques de 31 des 110 tests répertoriés suggèrent que la majorité ne répond pas à plusieurs des critères définissant un bon test standardisé. Ainsi, plus de la moitié des outils ne parviennent pas à fournir les preuves empiriques de validité et de fidélité nécessaires pour diagnostiquer avec confiance de la présence ou non d'une déficience du langage ou des autres fonctions associées. On observe que les caractéristiques les plus souvent satisfaites sont celles qui exigent généralement le moins d'investissement financier et de temps, telles que les critères à propos des objectifs du test (critère 1), de la description de l'administration et des qualifications de l'examineur (critère 15). À l'inverse, les critères les moins souvent remplis sont ceux qui exigent divers investissements – de connaissances psychométriques, de temps et d'argent – tels que les critères reliés aux démonstrations empiriques de validité (critères 3, 4, 5, et 6) et de fidélité (critères 7, 8, et 9). Ces résultats sont reliés principalement à la sévérité des critères qui ont été appliqués lors de l'analyse. Toutefois, même une diminution des exigences n'aurait sans doute pas modifié les résultats obtenus puisque la plupart des échecs observés dans les outils analysés est reliée à une absence de l'information plutôt qu'à une mauvaise performance des tests à un ou plusieurs critères.

Les résultats de la présente recherche sont alarmants puisque la majorité des professionnels œuvrant auprès des enfants font des tests standardisés, l'outil privilégié pour guider les diverses décisions cliniques qu'ils doivent prendre (voir Kerr, Guildford, & Kay-Raining Bird, 2003). Par exemple, ils doivent 1) identifier et diagnostiquer des troubles de langage, 2) établir des objectifs de traitement et finalement 3) mesurer les progrès en cours de traitement. Non seulement les tests standardisés ne sont pas conçus pour répondre efficacement à de tels motifs (spécialement les objectifs 2 et 3; voir McCauley & Swisher, 1984a, 1984b), mais ils deviennent encore moins efficaces lorsque la qualité psychométrique des outils ne peut être établie comme c'est le cas actuellement

avec de nombreux outils. En contexte francophone au Canada, grâce au travail de chercheurs, plusieurs professionnels de l'évaluation du langage sont déjà sensibilisés au fait que les normes d'outils standardisés en anglais ne sont pas directement transférables au français (Lefebvre & Trudeau, 2005; Garcia et al., 2006).

Les conséquences liées à la mauvaise utilisation de tests standardisés sont nombreuses. Non seulement elle compromet la rigueur du processus d'évaluation, la validité de son résultat mais également l'efficacité de l'intervention qui en découle. Le client peut se voir refuser des services dont il a besoin et auxquels il aurait droit ou se faire imposer des services dont il n'a pas besoin. Les coûts sociaux de cette réalité ne sont également pas à négliger. Ceux-ci sont difficiles à quantifier, mais ils sont néanmoins bien réels. Avec les mesures de coupure et de rationalisation toujours plus importantes dans les secteurs médicaux, scolaires et de réadaptation, des traitements et des objectifs thérapeutiques qui ne sont pas nécessaires représentent des pertes économiques énormes de temps, d'argent et de ressources puisque tant le clinicien que l'enfant se voient occupés à travailler sur des objectifs qui ne s'avèrent pas des besoins réels du patient.

Ainsi, rehausser la qualité des outils d'évaluation des troubles du langage accessibles aux professionnels qui oeuvrent auprès de la population pédiatrique canadienne francophone se doit de devenir une priorité puisque le niveau de qualité des services dispensés en dépend et que les conséquences qui découlent d'un tel écart entre la situation d'évaluation actuelle et celle souhaitée sont nombreuses. Malheureusement, les maisons d'édition de tests sont peu sensibles à cette réalité et peu intéressées à ce secteur d'activité entre autres parce que le marché français au Canada est restreint et les profits qui pourraient découler d'un tel investissement sont probablement trop modestes (Garcia & Desrochers, 1997). Par conséquent, les solutions aux problèmes de l'évaluation du langage en français doivent inévitablement passer par une volonté collective des professionnels du Canada d'améliorer la qualité des outils disponibles et d'optimiser leur accessibilité. Pour ce faire, nous croyons qu'il faut orienter nos actions dans trois directions.

Tout d'abord, tel que suggéré par McCauley et Swisher (1984a), il est impératif de continuer à sensibiliser les cliniciens aux conséquences et aux risques associés à une utilisation d'outils aux qualités psychométriques défaillantes. Il faut également s'assurer qu'ils possèdent une connaissance solide de la psychométrie théorique et appliquée, ce qui n'est pas le cas actuellement. En effet, dans un sondage effectué par Kerr et ses collègues (2003), auprès des orthophonistes et audiologistes membre de l'ACOA, seulement 17% des répondants ont affirmé se sentir pleinement confiants en ce qui a trait à leurs connaissances psychométriques et à leur capacité à évaluer la qualité des tests. Les auteurs rapportent également que le niveau de confiance exprimé n'était pas significativement lié à la capacité réelle des participants à identifier les problèmes

liés à l'utilisation de tests. La justesse de leur perception personnelle n'était donc que très peu valable. Nul doute qu'il y a place à de l'amélioration en ce qui concerne les connaissances liées à l'utilisation des tests psychométriques. Bien que les manuels accompagnants les outils d'évaluation fournissent des informations non négligeables, les livres, les articles ainsi que la possibilité de suivre des ateliers de formation sur l'évaluation des tests et l'interprétation qui en découle pourraient s'avérer des ressources intéressantes. Ces ressources devraient être facilement accessibles aux professionnels, idéalement durant leurs heures de travail. La structure de certains programmes de formation universitaire devrait également veiller à offrir aux étudiants la possibilité de s'initier aux principes et à la pratique liée à l'élaboration des outils d'évaluation.

Le développement des outils constitue la seconde direction d'action. Tout d'abord, il faudrait améliorer la communication entre les professionnels du Canada français de manière à éviter le dédoublement d'efforts menant actuellement au développement de plusieurs versions françaises des mêmes tests américains. Le travail effectué par le Groupe de recherche sur l'évaluation des troubles de la communication (GRETCOM) a certainement grandement contribué à une meilleure connaissance de l'état actuel des ressources en français au Canada (Paradis, Desrochers, & Garcia, 2002). Il faudrait également songer à établir des partenariats entre chercheurs qui oeuvrent en milieu universitaire et cliniciens. Durant trop longtemps, ces derniers ont dû effectuer la double tâche d'évaluer et de produire des outils. Or, dans les contextes actuels, leurs responsabilités professionnelles, trop importantes et lourdes, les empêchent de faire évoluer aussi rapidement que nécessaire la situation de l'élaboration des outils d'évaluation en français. Finalement, avant d'entreprendre le long processus de traduction/adaptation d'un outil de mesure, il est impératif de se demander dans quelle mesure ce travail en vaut vraiment la peine. En effet, tel que le rappellent Garcia et ses coll. (2006) ainsi que George (1997), il ne faut jamais oublier que même si un outil est valide et justifié pour une population, il ne l'est pas nécessairement pour une autre. Il s'avère donc raisonnable de s'interroger non seulement sur la qualité psychométrique de l'outil d'intérêt mais également sur sa spécificité et sa pertinence culturelle pour une population d'intérêt. C'est pourquoi Sperber, Devellis, et Boehlecke (1994) proposent que les outils jugés pertinents pour diverses populations linguistiques soient développés de façon simultanée dans les deux langues au lieu de traduire à partir d'un outil déjà standardisé dans une des deux langues. En ce qui concerne le développement de nouveaux outils, il est primordial de s'assurer que ceux-ci répondent aux qualités psychométriques. Ils doivent présenter, entre autres, une bonne fidélité et une bonne validité et avoir un bon étalonnage des normes.

Finalement, la troisième et dernière suggestion touche la sensibilisation des gestionnaires et autres partenaires financiers à l'importance et à la nécessité d'investir

davantage dans des fonds destinés à la promotion et au développement d'outils adaptés au Canadien français. La plus grande conscientisation des professionnels demeurera vaine si elle ne se voit pas accompagnée de bourses ou de subventions leur permettant de créer les outils répondant plus adéquatement à leurs besoins.

## Conclusion

La présente étude avait comme principal objectif de faire le point sur la situation de l'évaluation du langage et de la parole des enfants francophones du Canada en produisant d'abord un répertoire de 110 outils et en faisant une étude critique de 31 de ces outils. Les résultats obtenus démontrent qu'il existe plusieurs outils à la disposition des professionnels, bon nombre d'entre eux ayant été spécifiquement conçus pour les enfants francophones européens, mais très peu pour les enfants francophones nord-américains. Nos observations mettent également en lumière l'existence d'un écart considérable entre la qualité des outils d'évaluation actuels en français et celle souhaitée selon les standards psychométriques. Nous suggérons ainsi d'appuyer toute démarche susceptible de mener au développement et à l'adaptation d'outils d'évaluation en français. Lorsqu'il y aura une concertation et une collaboration entre plusieurs chercheurs et cliniciens et lorsque les organismes subventionnaires en feront une priorité, il sera alors possible d'espérer pouvoir développer des outils de qualité nécessaires et utiles à l'évaluation de la parole et du langage chez les enfants francophones du Canada.

## Références

- Beran, T. N. (2003). The role of validity in psychological measurement for school psychology applications. *Canadian Journal of School Psychology, 18*(1-2), 223-243.
- Chevrier-Muller, C., & Narbona, S. (1999). *Le langage de l'enfant : aspects normaux et pathologiques*. Paris, Masson.
- Garcia, L. J., & Desrochers, A. (1997). L'évaluation des troubles du langage et de la parole chez l'adulte francophone. *Revue d'orthophonie et d'audiologie, 21*(4), 271-284.
- Garcia, L. J., Paradis, J., Sénécal, I., & Laroche, C. (2006). Utilisation et satisfaction à l'égard des outils en français évaluant les troubles de la communication. *Revue d'orthophonie et d'audiologie, 30*(4), 239-249.
- George, L. K. (1997). Choosing among established assessment tools: Scientific demands and practical constraints. *Generations, 21*(1), 32-36.
- Huang, R., Hopkins, J., & Nippold, M. A. (1997). Satisfaction with standardized language testing: A survey of speech language pathologists. *Language, Speech, and Hearing Services in Schools, 28*, 12-29.
- Hutchinson, T. A. (1996). What to look for in the technical manual: Twenty questions for users. *Language, Speech and Hearing Services in Schools, 27*, 109-121.
- Kerr, M. A., Guildford, S., & Kay-Raining Bird, E. (2003). Standardized language test use: A Canadian survey. *Journal of Speech-Language Pathology and Audiology, 27*(1), 10-27.
- Lahey, M. (1990). Who shall be called language disordered? Some reflections and one perspective. *Journal of Speech and Hearing Disorders, 33*, 612-660.
- Lahey, M. (1988). *Language disorders and language development*. New York: American Macmillan.
- Lefebvre, P., & Trudeau, N. (2005). L'orthophoniste et les tests normalisés. *Fréquences, 17*(2), 17-20.
- McCaughey, R. J., & Swisher, L. (1984a). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*, 34-42.
- McCaughey, R. J., & Swisher, L. (1984b). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders, 49*, 338-348.
- Merrell, A. W., & Plante, E. (1997). Norm-referenced test interpretation in the diagnostic process. *Language, Speech, and Hearing Services in Schools, 28*, 50-58.
- Paradis, J., Desrochers, A., & Garcia, L. (2002). Le Groupe de recherche sur l'évaluation des troubles de la communication : historique, objectifs et réalisations. *Reflète, 8*(1), 61-72.
- Paul, R. (2007). *Language disorders from infancy to adolescence* (3rd ed.). St. Louis, MO: Mosby.

Plante, E. & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*, 15-24.

Records, N. L., & Tomblin, J. B. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research, 37*, 144-156.

Rondal, J. A. (1997). *L'évaluation du langage*. Sprimond, Belgique: Mardaga.

ShIPLEY, K. G., & McAfee, I. G. (2004). *Assessment in speech-language pathology: A resource manual* (3rd ed.). Clifton Park, NY: Delmar learning.

Sperber, A. D., Devellis, R. F., & Boehlecke, B. (1994). Cross-cultural translation: Methodology and validation. *Journal of Cross-Cultural Psychology, 25*, 501-524.

Vallerand, R. J. (1989). Vers une méthodologie de validation transculturelle de questionnaires psychologiques : implications pour la recherche en langue française. *Psychologie canadienne, 30*(4), 662-689.

## Remerciements

Ce travail a bénéficié de l'appui financier du Réseau canadien de la recherche sur le langage et l'alphabétisation (RCRLA). Nos remerciements vont aussi à l'Institut de recherche du Centre hospitalier pour enfants de l'est de l'Ontario pour leur appui financier ainsi qu'au Masonic Foundation of Ontario.

## Note des auteurs

Adresser la correspondance à : Elizabeth Fitzpatrick, PhD, Faculté des sciences de la santé, Université d'Ottawa, 451 chemin Smyth (3071), Ottawa, ON K1H 8M5. Courriel : elizabeth.fitzpatrick@uottawa.ca

**Date soumis : le 30 mai 2008**

**Date accepté : le 11 mai 2009**



## Annexe A. Critères d'analyse des outils d'évaluation suivant ceux recommandés par McCauley & Schwiser (1984) ainsi que Hutchison (1994).

### Critère 1 : Objectifs du test clairement défini

Le manuel doit mentionner les objectifs du test de manière claire et simple.

### Critère 2 : Définition du construit

Le manuel doit définir le construit, i.e. la définition conceptuelle, théorique ou opérationnelle de ce qui est mesuré dans le test, de manière à ce que les utilisateurs puissent comprendre la manière dont les tâches et items s'articulent autour du construit, les rendant ainsi prédictibles et logiques.

### Critère 3 : Validité de contenu

Selon des experts du domaine, est-ce l'opérationnalisation du construit représente de manière suffisamment concrète la théorie ou le construit qui est avancé par le test? Est-ce tous les domaines sont présents et suffisamment représentés? Finalement, est-ce que les items représentent bien le domaine auquel ils appartiennent?

### Critère 4 : Validité de construit

Le manuel doit inclure une analyse empirique de la validité de construit et déterminer de manière claire la manière utilisée pour l'obtenir (corrélation inter-test, analyse factorielle). Ces analyses devraient pouvoir s'expliquer en lien avec la théorie avancée par le test. Ce n'est donc pas uniquement la valeur du coefficient qu'il importe de regarder mais également son lien avec la définition du construit qu'il sous-tend.

### Critère 5 : Validité concordante

Le manuel doit présenter la preuve d'une bonne validité concordante i.e. que le test corrèle bien avec d'autres mesures déjà validées pour mesurer le construit. La validité concordante permet de comparer le résultat du test avec ceux de tests normalisés.

### Critère 6: Validité prédictive

Le manuel doit présenter la preuve que le test peut prédire positivement les futures performances d'un enfant dans un test déjà accepté comme mesurant le construit.

### Critère 7 : Cohérence interne/homogénéité

Le manuel doit présenter une analyse empirique de la consistance interne et présenter de manière claire la manière utilisée pour l'obtenir (alpha de Cronbach, corrélation inter-items moyenne, corrélation item-total moyenne, « split-half »). La fidélité d'un instrument se détermine entre autres en évaluant dans quelle mesure les items censés mesurer un même construit mènent à des résultats similaires mais également dans quelle mesure ces résultats sont consistants.

### Critère 8 : Fidélité test-retest

Selon le construit mesuré, 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. Le manuel doit donc ainsi fournir des preuves empiriques d'un coefficient test-retest significatif (au moins .90 avec un alpha de 0.05 ou moins).

### Critère 9 : Fidélité inter-juge

Les utilisateurs d'un outil espèrent généralement que les résultats obtenus par un examiné soit le reflet de sa performance et non de la perception de l'examineur. Il importe donc que différents juges soient en mesure d'évaluer les performances de la même manière. Le manuel doit donc fournir des preuves empiriques d'un coefficient de fidélité inter-juge significatif (au moins .90 avec un alpha de 0.05 ou moins).

### Critère 10 : Erreur standard de mesure (*Standard Error of Measurement*)

Puisque le résultat obtenu à un test s'avère un reflet imparfait de la réelle performance en raison d'une multitude de sources d'erreurs possible (p. ex, le fait que le test lui-même ne représente pas parfaitement le construit qu'il mesure; l'état de l'examiné le jour de l'évaluation qui ne reflète pas son état habituel, etc...), les auteurs devraient toujours rapporter une marge d'erreur pour chacun des scores de tests.

### Critère 11 : Échantillon de standardisation

Il importe que le manuel présente une description précise de l'échantillon de standardisation, de manière à s'assurer que l'individu testé y est bien représenté. Trois informations s'avèrent nécessaires : (1) la localisation géographique/l'origine, (2) le statut socio-économique et finalement (3) la « normalité » des sujets et dans le cas contraire, les caractéristiques « spéciales » des sujets formant l'échantillon.

### Critère 12 : Taille de l'échantillon de normalisation

La loi de la limite inférieure exige un minimum de 100 personnes dans chaque sous-groupe (p.ex. par tranche d'âge) pour l'échantillon d'un test normalisé.

### Critère 13 : Présence d'effets planchers ou plafonds

Puisque les tests incluent généralement des items qui sont conformes avec l'âge, les niveaux scolaires et les niveaux de développement qu'ils tentent de couvrir, peu d'items permettent d'évaluer adéquatement les enfants situés dans les extrêmes. Il est ainsi possible de retrouver des effets planchers (les items, trop difficiles ne permettent pas de départager efficacement enfants qui se retrouvent dans la partie inférieure de la courbe) et des effets plafonds (les items, trop faciles, ne permettent pas de départager les enfants qui ont des performances supérieures et qui se retrouvent dans la partie extrême de la courbe). Il s'avère donc important de rapporter de tels effets ou d'augmenter le nombre d'enfants représentés dans les extrêmes, afin d'assurer leur bonne représentation.

### Critère 14 : Comment sont représentés les enfants dans les extrêmes

Puisqu'il y a généralement plus d'enfants dans le centre de la distribution, les extrêmes sont donc souvent moins bien représentés. Le manuel se doit d'expliquer la manière dont les concepteurs s'y sont pris pour représenter les enfants dans les extrêmes (« sur-échantillonnage » dans les extrêmes, extrapolation etc...). De telles techniques permettent ainsi une meilleure estimation des normes pour les extrêmes.

### Critère 15 : Administration du test

Le manuel doit fournir suffisamment d'information sur la passation de manière à offrir à un examinateur les moyens de répliquer la procédure d'administration et de cotation originale, qui a été utilisée durant la standardisation.

### Critère 16 : Année de publication et de standardisation de l'échantillon de normalisation

Puisque les populations ne constituent pas des entités perpétuellement stables dans le temps, les normes qui datent de plus de 5 ans s'avèrent inappropriées et désuètes pour la comparaison étant donné les changements dans les performances qui peuvent se produire durant cet intervalle de temps. Étant donné la difficulté de mettre à jour les normes dans un délai aussi court, les cliniciens sont appelés à se montrer prudent lorsque l'année de publication et d'adaptation des normes dépassent ce nombre d'années.