

Analyse critique des caractéristiques psychométriques des batteries de performances évaluant
la cognition mathématique

J. Cattini ^{1,2}

A. Lafay ^{3,4}

¹ Orthophoniste, Luxembourg-ville, Luxembourg

² Collaboratrice scientifique, CPLU, RUCHE, Université de Liège, Liège, Belgique

³ Maitresse de conférences universitaire, Département de psychologie, Université Savoie
Mont Blanc, Chambéry, France

⁴ Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

Correspondance : Julie Cattini > juliecattini@hotmail.com

Conflit d'intérêt :

La première autrice (Julie Cattini) déclare être membre du conseil d'administration de Tool2care mentionné dans l'article. Elle n'en tire toutefois aucun bénéfice financier.

La seconde autrice (Anne Lafay) est co-autrice des outils d'évaluation intitulés Examath 8-15 et Examath 5-8 : batteries informatisées d'examen des habiletés mathématiques qui ont été inclus dans la recension des outils d'évaluation mathématique et qui ont fait l'objet de l'analyse psychométrique. Elle déclare avoir des liens financiers avec la maison d'édition de l'outil d'évaluation (HappyNeuron).

Résumé

L'identification et la prise en charge du Trouble Spécifique des Apprentissages en Mathématiques sont indispensables pour encourager la réussite scolaire et l'insertion professionnelle. L'étude identifie et analyse les caractéristiques psychométriques de treize batteries de performances en cognition mathématique pour les enfants et les adolescents. Les trois batteries possédant la meilleure qualité psychométrique sont : Examath 5-8, Examath 8-15 et Tedi-Math Grands. Des recommandations sont proposées au regard des données disponibles.

Mots-clefs : Psychométrie, Orthophonie, Cognition mathématique, Enfant

Abstract:

The identification and intervention of Mathematics Learning Disabilities are essential to encourage academic success and professional integration. This study identifies and analyzes the psychometric characteristics of thirteen batteries of mathematical cognition performance for children and adolescents. The three batteries with the best psychometric quality are: Examath 5-8, Examath 8-15, and Tedi-Math Grands. Based on the available data, recommendations are proposed.

Keywords: Psychometry, Speech-and-Language Therapy, Mathematical cognition, Child

Resumen:

La identificación y el tratamiento de los problemas de aprendizaje en matemáticas son esenciales para favorecer el éxito escolar y la integración en el mundo laboral. El estudio identifica y analiza las características psicométricas de trece baterías de rendimiento en cognición matemática para niños y adolescentes. Las tres baterías con mejor calidad psicométrica son : Examath 5-8, Examath 8-15 y Tedi-Math Grands. Se proponen recomendaciones a la luz de los datos disponibles.

Palabras clave: Psicometría, Logopedia, Cognición matemática, Niño

Remerciements

Nous tenons à remercier Marion Balla et Jonathan Burnay pour leur aide lors de l'extraction des données.

Introduction

L'évaluation des compétences mathématiques revêt une importance capitale dans notre société quotidienne. Chaque aspect de nos vies, des simples achats aux activités de loisirs, est imprégné de nombres et de calculs. Les performances mathématiques des individus sont associées à leur succès futur, que ce soit sur le plan professionnel, personnel ou académique (Crawford & Cribb, 2013). En effet, rencontrer des obstacles dans l'apprentissage des mathématiques affecte non seulement le parcours scolaire d'un enfant, mais aussi sa formation professionnelle à l'âge adulte, ainsi que sa qualité de vie tout au long de son existence. C'est pourquoi évaluer de manière précise les compétences mathématiques revêt une importance cruciale pour identifier et adresser les difficultés d'apprentissage dans ce domaine.

Évaluation chez l'enfant et l'adolescent

L'évaluation peut répondre à quatre objectifs principaux : détecter, diagnostiquer, comprendre le profil de l'enfant et évaluer son évolution (Burnay et al., 2024 ; Youngstrom et al., 2017).

La détection des difficultés d'apprentissage en mathématiques d'un enfant est généralement réalisée à partir des observations informelles réalisées par l'entourage. Celles-ci sont obtenues lors de l'anamnèse et/ou de l'entretien avec l'enseignant qui rapportent des comportements les alertant par rapport à leurs connaissances du développement de l'enfant (par exemple, « mon enfant a 9 ans et il compte encore sur ses doigts »). Toutefois, certaines épreuves de première ligne peuvent être proposées par le médecin de famille, l'équipe pédagogique, le psychologue ou encore l'orthophoniste avant de réaliser un bilan approfondi.

Lorsqu'un enfant a été identifié comme rencontrant des difficultés d'apprentissage en mathématiques, la question du diagnostic se pose rapidement. Il pourrait par exemple

présenter un trouble spécifique des apprentissages en mathématiques¹ (TSAM). Selon le manuel Diagnostique et Statistique des troubles Mentaux (DSM-5 ; APA, 2016), plusieurs critères ont été définis et permettent de poser un diagnostic de trouble spécifique des apprentissages avec « déficit du calcul » et/ou avec « difficultés du raisonnement mathématique ». En particulier, l'enfant présente des performances qui sont largement sous le niveau attendu à son âge depuis plus de six mois, et ce de manière quantifiable. Pour valider ces critères, il s'agit donc de proposer une évaluation individuelle avec une interprétation normée² ou critériée³. Bien entendu, ces données doivent être complétées par des données cliniques complémentaires (par exemple, histoire scolaire, relevé de notes, évaluation informelle) afin de documenter le début des difficultés d'apprentissage ou encore leur impact fonctionnel.

Une fois, les difficultés de l'enfant objectivées, celui-ci acquiert, selon les pays, le droit de bénéficier d'une intervention remboursée par les organismes de santé et de la mise en place d'aménagements pédagogiques. A ce stade, l'évaluation se poursuit afin de comprendre le profil du patient, autrement dit, ses forces et faiblesses. Les évaluations dynamiques s'avèrent les plus appropriées pour déterminer les cibles et les moyens thérapeutiques les plus pertinents ou encore pour fournir des recommandations à l'équipe pédagogique (Duboisdindien & Cattini, 2024). La modalité d'évaluation dynamique s'oppose à la modalité statique qui est la plus couramment utilisée. L'évaluation dynamique comprend des procédures conçues pour évaluer les performances d'un patient dans des contextes variés

¹ « Les compétences scolaires ou universitaires perturbées sont nettement au-dessous du niveau escompté pour l'âge chronologique du sujet, et ce de manière quantifiable. Cela interfère de façon significative avec les performances scolaires, universitaires ou professionnelles, ou avec les activités de la vie courante, comme le confirment des tests de niveau standardisés administrés individuellement ainsi qu'une évaluation clinique complète. Pour les individus âgés de 17 ans et plus, des antécédents avérés de difficultés d'apprentissages perturbantes peuvent se substituer à une évaluation standardisée ».

² Comparaison quantitative des résultats d'un patient à ceux d'un échantillon de pairs appariés qui ont accompli la même tâche et dont les données ont été traitées pour fournir des mesures de tendance centrale (Denman et al., 2019).

³ Comparaison des performances d'un patient par rapport à un niveau ou à un critère de référence préétabli telles que les compétences attendues au regard du niveau scolaire ou du programme d'étude (Denman et al., 2019).

et/ou avec des aides thérapeutiques différentes grâce à une relation interactive entre le patient et l'évaluateur (Denman et al., 2019).

Le dernier objectif est l'évaluation de l'évolution du patient avec ou sans intervention. Dans le cas de la mise en place d'une intervention, l'orthophoniste pourra, par exemple, proposer une évaluation critériée ou descriptive⁴ réalisée en mesure répétée pendant l'intervention (ou routine d'évaluation) ou encore une évaluation pré-post avant et après l'intervention (Martinez Perez et al., 2021 ; Olswang & Bain, 1994).

Caractéristiques psychométriques des outils francophones d'évaluation mathématique

En 2018, Lafay et Cattini ont recensé les outils disponibles en français pour l'évaluation des compétences en mathématique et ont répertorié les caractéristiques psychométriques de 22 outils. Les auteurs ont considéré 21 critères établis comme étant des caractéristiques de base avant d'utiliser un test dans le but de poser un diagnostic ou d'émettre une décision clinique à propos d'une performance d'un enfant à un test. Parmi les outils recensés, seulement trois outils évaluant spécifiquement les habiletés mathématiques remplissaient les conditions attendues pour plus de la moitié des caractéristiques psychométriques (Examath 8-15, Lafay & Helloin, 2016 ; Tedi-Math Grands, Noël & Grégoire, 2015 ; WIAT-II, Wechsler, 2005). Certains critères étaient très bien considérés (par exemple, la standardisation) alors que d'autres ne l'étaient peu ou pas (par exemple, le pouvoir discriminant). La limite principale de ce travail était l'absence de pondération des différentes caractéristiques psychométriques selon l'objectif d'évaluation. Les résultats offraient une analyse complète des différentes caractéristiques psychométriques des outils sans fournir des clés d'interprétation aux cliniciens. De ce fait, deux tests pouvaient afficher le même résultat en pourcentage, tout en ayant des caractéristiques psychométriques distinctes, dont certaines peuvent revêtir une plus grande importance que d'autres lorsqu'il s'agit de

⁴ L'évaluation permet de fournir des données descriptives ou qualitatives sur les capacités du patient (Denman et al., 2019).

choisir un test selon son objectif d'évaluation (AERA et al., 2014 ; Daub et al., 2021). De plus, depuis cette étude, d'autres outils francophones d'évaluation des habiletés mathématiques pour l'enfant ont été édités et sont utilisés par les cliniciens.

Une mise à jour de la recension ainsi qu'une prise en compte des limites de l'étude précédente sont nécessaires. La grille élaborée par l'équipe de Tool2Care (Burnay et al., 2024) qui est utilisée de manière transversale pour les outils d'évaluation orthophonique et (neuro)psychologique (voir par exemple Cattini & Maillart, 2024) répond aux besoins méthodologiques de la présente étude.

Objectifs

La présente étude vise à mettre à jour la recension des batteries de performances disponibles en français pour l'évaluation mathématique et à analyser de manière critique leurs caractéristiques psychométriques afin de fournir des recommandations cliniques tenant compte des objectifs de l'évaluation (c'est-à-dire, détecter, diagnostiquer, comprendre et évaluer l'évolution).

La section suivante présente la méthodologie utilisée, à savoir les étapes de recherche des batteries, d'extraction des informations et d'analyse de la qualité méthodologique des batteries.

Méthodologie

Recherche des batteries de performances francophones évaluant les habiletés mathématiques chez les enfants et les adolescents

Une recherche a été menée afin de repérer les outils d'évaluation répondant aux critères suivants : (1) être une batterie de performances (ou de tests approfondis)⁵ en mathématiques ou comprenant des épreuves évaluant des habiletés mathématiques ; (2) être destiné à des enfants francophones de 3 à 18 ans ; (3) être disponible auprès d'une maison

⁵ Afin de faciliter la lecture du présent article, le terme « batterie de performances » sera utilisé tant pour les tests de performances que les tests approfondis.

d'édition en septembre 2023 ; (4) proposer un étalonnage ou des critères d'interprétation ; et (5) disposer d'un manuel.

L'éligibilité des outils répertoriés dans Lafay et Cattini (2018) a été analysée au regard des nouveaux critères d'inclusion et une recherche a été réalisée sur les sites web des éditeurs (c'est-à-dire, Cogilud, ECPA, Gerip, HappyNeuron, Hogrefe, Mot-à-mot, Ortho Edition, Orthopatic) en septembre 2023 par la seconde autrice.

Extraction des informations et des caractéristiques concernant la population d'étalonnage

Les deux autrices ont extrait les données descriptives des outils d'évaluation concernant la population d'étalonnage et les habiletés mathématiques évaluées par les batteries de performances. Ces grilles d'extraction ont été reprises de Lafay et Cattini (2018).

Les informations descriptives extraites indépendamment par les deux autrices étaient les suivantes : (1) auteurs de l'outil ; (2) date de publication ; (3) informatisation de la passation ; (4) caractéristiques de l'échantillon de référence (c'est-à-dire, âge, classe, pays, niveau socioéconomique) ; (5) effectif de l'échantillon de référence (c'est-à-dire, nombre total, nombre de groupes et nombre par groupe) ; et (6) moment de la réalisation de l'étalonnage. Le pourcentage d'accord interjuge entre les autrices de la présente étude était de 85.4%. Tous les points de désaccord ont été vérifiés et traités pour arriver à un consensus. La seconde autrice (AL) a ensuite répertorié les domaines mathématiques couverts par les différentes batteries de performances et les résultats ont été vérifiés par la première autrice (JC).

Evaluation des caractéristiques psychométriques des outils

La seconde phase de l'étude a consisté à analyser de manière critique les caractéristiques psychométriques des outils. Dans ce cadre, a été utilisée la grille d'analyse critique de Tool2Care (Burnay et al., 2024) prenant en compte les caractéristiques

psychométriques essentielles, importantes ou secondaires selon l'objectif de l'évaluation (voir tableau 1). Les critères d'évaluation sont divisés en trois catégories principales : (1) les preuves de validité (c'est-à-dire, la capacité d'un outil d'évaluation à mesurer le construit pour lequel il a été conçu), (2) les preuves de fidélité (c'est-à-dire, la capacité d'un test à fournir des résultats constants ou fiables quel que soit le nombre d'administrations, les correcteurs impliqués et le contenu spécifique du test) et (3) les informations concernant les mesures de tendance centrale (c'est-à-dire, la distribution des scores obtenus par un échantillon de personnes représentatif d'une population définie). Concernant la pondération vis-à-vis de la fidélité, le score dépend de la qualité des indices (par exemple, un coefficient de consistance interne égal ou supérieur à 0.90 reçoit le score maximal, tandis qu'un coefficient strictement inférieur à 0.70 obtient un score de 0). En revanche, en ce qui concerne la validité, le score est déterminé par la présence ou l'absence de qualité (c'est-à-dire, 0 ou 1). La pondération appliquée est basée sur une échelle à trois niveaux : les caractéristiques essentielles sont pondérées deux fois (x2), celles jugées importantes reçoivent une pondération simple (x1), tandis que les caractéristiques considérées comme secondaires ont une pondération nulle (x0). La note finale est sur un total de 10 points. Pour davantage d'informations concernant la grille d'évaluation, nous renvoyons le lecteur à l'article de Burnay et al. (2024) dans ce même numéro.

Quelques adaptations mineures au protocole décrit ci-dessus ont toutefois été réalisées dans le cadre de cette étude. Nous avons jugé pertinent de différencier l'absence ou la présence de validité théorique sur deux niveaux : la présence de données/modèles théoriques justifiant le choix des épreuves (c'est-à-dire, 0.5 ou 0) et la sélection des items (c'est-à-dire, 0.5 ou 0).

Tableau 1

Récapitulatif des caractéristiques psychométriques essentielles, importantes et secondaires selon l'objectif de l'évaluation

Caractéristiques psychométriques	Objectifs de l'évaluation			
	Détecter	Diagnostiquer	Comprendre	Evaluer l'évolution
Validité de la structure interne	-	+	+	-
Validité théorique et de contenu	-	++	++	-
Validité concourante *	++	++	+	-
Validité prédictive	++	+	+	++
Sensibilité	++	++	-	-
Spécificité	-	++	-	-
Sensibilité au changement	-	-	-	++
Fidélité test-retest	+	++	+	++
Fidélité interjuge	+	++	+	++
Consistance interne	+	++	+	++
Distribution des données	++	++	-	-
Erreur standard de mesure	+	++	+	++

Notes. * ou validité convergente ou validité divergente.

Légende. ++ = essentiel ; + = important ; - = secondaire.

L'extraction et l'encodage des données avec la grille d'analyse critique de Tool2Care ont été réalisées de manière indépendante par un chercheur spécialisé en psychométrie et la première autrice (JC). La seconde autrice (AL) n'a pas été impliquée dans l'analyse critique des outils au regard de ses conflits d'intérêt. L'accord interjuge a été calculé à l'aide d'un kappa de Cohen et était de 0.69. Tous les points de désaccord ont été traités afin d'arriver à un consensus.

Résultats

Répertoire des batteries de performances et des domaines mathématiques évalués

La recherche des outils a permis de répertorier sept batteries de performances spécifiques à la cognition mathématique et six batteries de performances générales comprenant des épreuves en mathématique. Quatre outils supplémentaires par rapport à l'étude de Lafay et Cattini (2018) ont été identifiés : BMT-i (Billard et al., 2019), EPOCY 2-3 (Zanga, 2011), Examath 5-8 (Helloin & Lafay, 2021) et UDN-3 (Meljac et al., 2021).

Le tableau 2 présente les caractéristiques générales des batteries de performances : le titre, les auteurs, la date de publication, la modalité de présentation (informatisée ou papier),

ainsi que les caractéristiques de l'échantillon (c'est-à-dire, âge, classe, effectif, pays et niveau socioéconomique) et du moment d'étalonnage. Les batteries permettent d'évaluer les enfants scolarisés de la Petite Section de Maternelle à la 3^{ème} au collège. Huit batteries ont été élaborées pour la population française, trois pour la population française et belge francophone et une pour la population franco-québécoise. L'effectif par groupe est généralement inférieur à 100 avec une étendue de 17 à 371. Les batteries de performances évaluent un panel de domaines mathématiques variant d'un outil à l'autre. Nous renvoyons le lecteur au tableau 3 pour le détail des domaines mathématiques couverts par chaque outil.

Caractéristiques psychométriques des outils

Le relevé synthétique des données de validité et de fidélité des treize batteries de performances analysées est disponible dans le tableau 4.

Données sur la validité

La validité théorique est la mieux représentée dans le manuel des outils avec la mention de modèles théoriques liés au contenu de l'évaluation pour neuf des treize batteries et des preuves d'un contrôle des items pour six d'entre elles. Les batteries de performances spécifiques à la cognition mathématique démontrent une meilleure validité théorique que les batteries plus générales, pour lesquelles il n'est pas rare que les auteurs fournissent des références plus globales ou en lien avec l'objectif principal de l'outil d'évaluation (par exemple, les compétences langagières) sans préciser les fondements théoriques pour les sous-épreuves mathématiques. Six des treize outils évaluent les preuves de relation avec d'autres variables (c'est-à-dire, validité concourante ou convergente) pour l'ensemble de la batterie de performances ou certaines sous-épreuves. Les autres indices de validité sont moins fréquemment présents. La structure interne est évaluée pour trois batteries de performance à l'aide d'analyse factorielle exploratoire (c'est-à-dire, BMT-i ; Tedi-Math Grands ; Zareki-R – Dellatolas & Von Aster, 2006). La validité prédictive a été évaluée pour trois batteries

évaluant spécifiquement les compétences en cognition mathématique : Examath 5-8, Examath 8-15 et Tedi-Math Grands. La validité discriminante, caractéristique essentielle pour diagnostiquer les difficultés d'un patient, a été évaluée pour seulement deux batteries de performances : Examath 5-8 et Examath 8-15. Ces deux batteries montrent une bonne spécificité (c'est-à-dire, pouvoir qu'un test possède pour repérer une personne saine comme étant effectivement saine) avec des scores généralement supérieurs à 80% tandis que la sensibilité (c'est-à-dire, pouvoir qu'un test possède pour repérer un enfant en difficulté comme étant effectivement en difficulté) est plus faible avec des scores généralement inférieurs à 80% sur la base des diagnostics posés par les orthophonistes des enfants recrutés.

Tableau 2*Relevé des informations générales et des caractéristiques concernant la population d'étalonnage des batteries d'évaluation analysées*

Nom de l'outil *	Auteurs	Date	Présentation informatisée	Échantillon					Moment de l'étalonnage		
				Âge	Classe**	Nbre total	Nbre de groupes	Nbre par groupe		Pays	Niveau socioéconomique
Batteries de performances en mathématiques											
ERLA	Legeay et al.	2009	Non	/	/	/	/	/	/	/	/
Examath 5-8	Helloin et Lafay	2021	Oui	5-8	GSM-CE1	339	3	97 à 124	France	Répartition proche des indices INSEE	Mars-juin
Examath 8-15	Lafay et Helloin	2016	Oui	8-15	CE2-3 ^{ème}	443	5	74 à 127	France et Belgique	Répartition proche des indices INSEE	Mars-mai
Tedi-Math	Van Nieuwenhoven et al.	2001	Non	/	MSM-CE2	583	8	67 à 76	France et Belgique	/	Novembre et mai
Tedi-Math Grands	Noël et Grégoire	2015	Les deux	/	CE2-5 ^{ème}	254	5	46 à 56	France	Répartition proche des indices INSEE	Mars-juin
UDN-3	Meljac et al.	2020	Les deux	4-16	PS-1 ^{ère}	543	14	17 à 58	France	Répartition proche des indices INSEE	Mai-décembre
Zareki-R	Dellatolas et Von Aster	2006	Non	6-11;5	CP-CM2	249	5	43 à 59	France	50% ZEP, 28% ne parlent pas français à la maison	Janvier à mars
Batteries de performances comprenant des épreuves en mathématiques											
BMT-i	Billard et al.	2019	Oui	4-13	MSM-5 ^{ème}	1074	9	105 à 140	France	Répartition proche des indices INSEE	Septembre-juin
EPOCY 2-3	Zanga	2011	Non	/	CP-CM1	1908	8	197 à 371	France	Répartition proche des indices INSEE	Janvier-février et juin
Exalang 3-6	Thibault et Helloin	2006	Oui	3-5;5	MSM-GSM	468	6	59 à 96	France et Belgique	Répartition proche des indices INSEE	/
Exalang 8-11	Thibault et al.	2012	Oui	8-11	CE2-CM2	461	3	93 à 150	France	Répartition proche des indices INSEE	Février-avril
Exalang 11-15	Thibault et al.	2009	Oui	11-15	6 ^{ème} -3 ^{ème}	322	4	85 à 97	France	Répartition proche des indices INSEE	Janvier-mai
WIAT-II	Wechsler	2005	Non	6-18	CP-1 ^{ère}	304	12	18 à 32	Canada (Québec)	Niveau d'études des parents	Février à décembre

Notes. *Les tests sont classés en ordre alphabétique selon leur titre. ** Les classes sont indiquées en référence au système scolaire en France.

Légendes. CE1 = Cours élémentaire 1; CE2 = Cours élémentaire 2; CM1 = Cours moyen 1; CM2 = Cours moyen 2; CP = Cours préparatoire; BMT-i = Batterie modulable de tests informatisée ; EPOCY 2-3 = Echelle de Positionnement en cycles 2 et 3 ; ERLA = Batterie d'Exploration du Raisonnement et du Langage Associé ; Exalang 3-6 = batterie informatisée pour l'EXAMen du LANGage et des compétences transversales chez l'enfant de 3 à 6 ans ; Exalang 8-11 = batterie informatisée pour l'EXAMen du LANGage et des compétences transversales chez l'enfant de 8 à 11 ans ; Exalang 11-15 = batterie informatisée pour l'EXAMen du LANGage oral, du langage écrit et des compétences transversales chez le collégien ; Examath 5-8 = batterie informatisée d'EXAMen des habiletés MATHématiques pour l'enfant de 5 à 8 ans ; Examath 8-15 = batterie informatisée d'EXAMen des habiletés MATHématiques pour l'enfant de 8 à 15 ans ; GSM = Grande section de maternelle ; INSEE = Institut national de la statistique et des études économiques ; MSM = Moyenne section de maternelle ; PSM = Petite Section de Maternelle ; UDN-3 = Utilisation Du Nombre, 3^e édition ; WIAT-II = Wechsler Individual Achievement Test, 2^{ème} édition ; ZAREKI = Die Neuropsychologische Testbatterie für ZAHlenarbeit und REchnen bei KIndern ; ZEP = Zone d'éducation prioritaire.

Tableau 3

Domaines mathématiques évalués par les batteries de performances analysées

Nom de l'outil *	Traitement du nombre	Comptine numérique	Dénombrement	Numération	Transcodage	Calcul	Résolution de problèmes	Raisonnement	Vocabulaire mathématique
Batteries de performances en mathématiques									
ERLA		+	+		+		+	+	
Examath 5-8	+	+	+	+	+	+	+	+	+
Examath 8-15	+		+	+	+	+	+	+	+
Tedi-Math	+	+	+	+	+	+	+	+	
Tedi-Math Grands	+			+	+	+	+	+	+
UDN-3	+	+	+	+	+		+	+	+
Zareki-R	+	+	+		+	+	+		
Batteries de performances comprenant des épreuves en mathématiques									
BMT-i	+	+	+	+	+	+	+		
EPOCY 2-3				+		+	+		
Exalang 3-6			+						+
Exalang 8-11							+	+	+
Exalang 11-15								+	
WIAT-II						+	+	+	

Notes. *Les tests sont classés en ordre alphabétique selon leur titre.

Légendes. Voir légendes du tableau 2

Données sur la fidélité

Dix des treize batteries de performances fournissent des données de fidélité : six outils fournissent des estimations de la consistance interne (Examath 5-8 ; Tedi-Math, Van Nieuwenhoven et al., 2001 ; Tedi-Math Grands ; UDN-3 ; BMT-i ; WIAT-II), cinq outils fournissent des données de fidélité test-retest (Examath 5-8 ; Examath 8-15 ; Exalang 3-6, Thibault & Helloin, 2006 ; Exalang 8-11, Thibault et al., 2012 ; Exalang 11-15, Thibault et al., 2009) et un outil indique des informations pour la fidélité interjuge (Exalang 8-11). Trois outils ne fournissent aucune information concernant le degré de précision et de fidélité des scores : EPOCY 2-3, ERLA (Legeay et al., 2009) et Zareki-R.

Sensibilités aux changements

Deux outils fournissent des informations concernant la sensibilité aux changements afin de fournir des preuves essentielles pour mieux caractériser l'évolution d'un patient : EPOCY 2-3 et UDN-3. L'EPOCY 2-3 fournit les données pour pouvoir calculer la taille d'effet (η^2) de sensibilité au changement développemental pour le score global en mathématiques tandis que l'UDN-3 fournit les données pour chaque sous-épreuve.

Distribution des scores

La majorité des outils indiquent si les données sont normales ou proposent des données corrigées, excepté EPOCY 2-3, ERLA et UDN-3. Cependant, ces trois batteries de performances proposent une interprétation critériée (et non normée) des scores. Cela rend cette caractéristique psychométrique secondaire dans le cas où l'objectif de l'évaluation est de mieux comprendre le profil du patient.

L'erreur standard de mesure est également calculable pour la majorité des outils fournissant des données corrigées. En revanche, cette donnée n'est généralement pas fournie de manière transparente dans le manuel et doit être calculée par l'utilisateur.

Tableau 4

Relevé synthétique des preuves de validité et de fidélité des batteries de performances

Nom de l'outil *	Validité						Fidélité			Score		
	Théorique		Structure interne	Relations avec d'autres variables	Prédictive	Sensibilité et spécificité	Sensibilité au changement	Temporelle	Interjuge	Consistance interne	Distribution	Erreur standard de mesure
	Epreuves	Items										
Batteries de performances en cognition mathématique												
ERLA	+	-	-	-	-	-	-	-	-	-	-	-
Examath 5-8	+	+	-	+	+	+	-	≥ .90	-	< .70	+	+
Examath 8-15	+	+	-	+	+	+	-	≥ .70	-	-	+	+
Tedi-Math	+	+	-	-	-	-	-	-	-	≥ .80	+	+
Tedi-Math Grands	+	+	+	+	+	-	-	-	-	≥ .70	+	+
UDN-3	+	-	-	-	-	-	+	-	-	< .70	-	-
Zareki-R	+	+	+	+	-	-	-	-	-	-	+	-
Batteries de performances comprenant des épreuves en cognition mathématique												
BMT-i	+	+	+	-	-	-	-	-	NA	< .70	+	+
EPOCY 2-3	+	-	-	+	-	-	+	-	-	-	-	-
Exalang 3-6	-	-	-	-	-	-	-	≥ .90	-	-	+	+
Exalang 8-11	-	-	-	-	-	-	-	≥ .70	≥ .80	-	+	+
Exalang 11-15	-	-	-	-	-	-	-	≥ .80	-	-	+	+
WIAT-II	-	-	-	+	-	-	-	-	-	≥ .80	+	+
Score	9/13	6/13	3/13	6/13	3/13	2/13	2/13	5/13	1/12	6/13	10/13	9/13
Pourcentage	69 %	46 %	23 %	46 %	23 %	15 %	15 %	39 %	8%	46 %	77 %	69 %

Notes. *Les tests sont classés en ordre alphabétique selon leur titre.

Légendes. + = présente ; - = absente ; NA = Non applicable

Voir légendes du tableau 2.

Analyse critique des batteries selon l'objectif de l'évaluation

Les scores obtenus à l'aide de la grille d'analyse critique Tool2Care sont très variables avec une moyenne de 4.11 et un écart-type de 2.1 sur un score total de 10 points. Les résultats de l'analyse critique selon l'objectif de l'évaluation sont disponibles dans le tableau 5. Ce tableau comprend également le type d'interprétation des données à réaliser (critériée, descriptive ou normée) ainsi que la modalité de l'interaction entre l'administrateur et le patient (statique ou dynamique). La majorité des batteries de performances (10/13) ciblent plus d'un objectif d'évaluation. Neuf batteries de performances sur treize obtiennent un résultat faible (≤ 3) pour un des objectifs d'évaluation fixés par les auteurs.

Pour les huit outils visant à détecter les difficultés du patient, le score maximal est de 8 sur 10 avec les batteries Examath 5-8 et 8-15. Les autres fournissent des preuves pour la totalité des caractéristiques psychométriques essentielles, c'est-à-dire la validité concurrente, la validité prédictive, la sensibilité et la distribution des données. Les six autres outils obtiennent un score égal ou inférieur à 3 sur 10. Aucun d'entre eux ne fournit des preuves de validité prédictive ou encore de sensibilité qui sont des caractéristiques essentielles afin de s'assurer que l'outil a la capacité de détecter les personnes à risque de rencontrer des difficultés d'apprentissage en mathématique.

Les batteries de performances avec le meilleur score pour le diagnostic sont également Examath 5-8 et Examath 8-15 avec un score de 8 et de 7 respectivement. Les autres fournissent des preuves pour la totalité des caractéristiques psychométriques essentielles (par exemple, le pouvoir discriminant et l'erreur standard de mesure). Les cinq outils supplémentaires (c'est-à-dire, Tedi-Math, Tedi-Math Grands, Exalang 3-6, Exalang 8-11 et Exalang 11-15) conçus par les auteurs pour le diagnostic obtiennent un score égal ou inférieur à 5 sur 10. Par exemple, aucun d'entre eux ne fournit de données concernant le pouvoir discriminant, c'est-à-dire la capacité de l'outil à identifier de manière précise les individus

présentant des difficultés spécifiques d'apprentissage en mathématiques par rapport à ceux qui n'en rencontrent pas.

Concernant la compréhension du profil du patient, Examath 5-8, Examath 8-15 et Tedi-Math Grands obtiennent les meilleurs résultats parmi les dix outils visant cet objectif, avec des scores de 7, 6 et 6 respectivement. Cependant, Tedi-Math Grands fournit des preuves de sa structure interne, contrairement aux deux autres batteries de performances. Cette caractéristique psychométrique, essentielle pour la compréhension du profil du patient, indique la présence de dimensions attendues et d'items représentant ensemble une dimension cohérente par rapport au modèle de référence. Nous pouvons noter que cette caractéristique est présente pour le Zareki-R (score total de 4 sur 10) mais, également, la BMT-i qui obtient un score global de 6 pour la compréhension du profil du patient bien que les auteurs n'annoncent pas cet objectif d'évaluation dans le manuel .

Enfin, seuls les auteurs d'Exalang 3-6 spécifient un objectif d'évaluation de l'évolution du patient, mais les caractéristiques psychométriques essentielles ne sont pas rencontrées. Par ailleurs, cette batterie de performances évalue plus particulièrement les compétences langagières et propose en complément des épreuves permettant d'évaluer le vocabulaire en mathématique et les capacités en dénombrement. EPOCY 2-3 et l'UDN 3 pourraient être considérés au vu des données disponibles au niveau de la sensibilité au changement, caractéristique essentielle pour cet objectif d'évaluation. Examath 5-8, Examath 8-15 et Tedi-Math Grands qui ont l'avantage de fournir des preuves de validité prédictive sont également des candidats possibles.

Tableau 5*Evaluation de la qualité psychométrique des batteries de performances*

Nom de l'outil *	Forme de l'évaluation			Score – Grille d'évaluation			
	Interprétation	Interaction avec l'administrateur	Objectif(s) de l'outil selon les auteurs	Détecter (/10)	Diagnostiquer (/10)	Comprendre (/10)	Evaluer l'évolution (/10)
Batteries de performances en cognition mathématique							
ERLA	Critériée	Statique et dynamique	Comprendre	0	1	1	0
Examath 5-8	Normée	Statique et dynamique	Détecter, diagnostiquer et comprendre	8	8	7	5
Examath 8-15	Normée	Statique	Détecter, diagnostiquer et comprendre	8	7	6	4
Tedi-Math	Normée	Statique	Diagnostiquer et comprendre	3	4	4	3
Tedi-Math Grands	Normée	Statique	Diagnostiquer et comprendre	6	5	6	4
UDN-3	Critériée	Statique et dynamique	Comprendre	0	1	1	2
Zareki-R	Normée	Statique	Détecter et comprendre	3	3	4	0
Batteries de performances comprenant des épreuves en cognition mathématique							
BMT-i	Normée	Statique	Détecter	3	5	6	3
EPOCY 2-3	Critériée	Statique	Détecter et comprendre	2	2	2	2
Exalang 3-6	Normée	Statique	Détecter, diagnostiquer, et évaluer l'évolution	3	3	2	3
Exalang 8-11	Normée	Statique	Détecter et diagnostiquer	3	3	2	3
Exalang 11-15	Normée	Statique	Diagnostiquer et comprendre	3	3	2	3
WIAT-II	Normée	Statique	Comprendre	3	3	5	4

Notes. *Les tests sont classés en ordre alphabétique selon leur titre. Les cases grisées correspondent aux objectifs ciblés par les concepteurs de l'outil.

Légendes. Voir légendes du tableau 2.

Discussion

L'objectif de la présente étude était de (1) recenser les batteries de performances disponibles en français pour l'évaluation mathématique et (2) d'analyser de manière critique leurs caractéristiques psychométriques.

Concernant le premier objectif, la recherche a mis en évidence un total de treize batteries de performances, dont sept sont spécifiquement mathématiques et six sont des batteries de performances comprenant des épreuves mathématiques. Les batteries de performances évaluent un panel de domaines mathématiques variant d'un outil à l'autre : des habiletés de base telles que le traitement du nombre, la comptine numérique ou le dénombrement, des habiletés plus élaborées dépendant d'un apprentissage scolaire telles que la numération, le transcodage, le calcul, la résolution de problèmes et des habiletés transversales (liées au langage notamment) telles que le raisonnement et le vocabulaire mathématique.

Concernant le deuxième objectif, l'analyse des treize batteries de performances a montré que la qualité psychométrique semble évoluer positivement avec le temps : en effet, les trois batteries possédant la meilleure qualité psychométrique (en fonction de l'objectif principal visé et indiqué par les auteurs) sont parmi les plus récentes : Examath 5-8 en 2021, Examath 8-15 en 2016 et Tedi-Math Grands en 2015. La BMT-i montre également une bien meilleure documentation des caractéristiques psychométriques que l'ancienne version publiée en 2012 (EDA, Billard & Touzin, 2012). Cependant, les caractéristiques psychométriques étudiées pour la BMT-i ne sont pas en adéquation avec l'objectif de l'évaluation : la détection. Elles correspondent davantage à l'objectif de comprendre le profil du patient avec des données concernant la validité de structure interne ou encore de consistance interne. Cependant, l'absence de données concernant la sensibilité, la validité prédictive ou encore la

validité concurrente ne permet de pas de prendre une décision clinique éclairée pour l'utilisation de cet outil à des fins de détection.

L'analyse de l'ensemble des outils d'évaluation mathématique recensé est globalement meilleure que celle réalisée dans d'autres domaines, comme en attestent plusieurs articles ayant analysé les caractéristiques psychométriques des tests en langage oral (Leclercq & Veys, 2014 ; Macchi et al., sous presse), en langage écrit (Gaul-Bouchard et al., 2009) ou encore en parole (Cattini & Maillart, 2024 ; Meloni et al., 2022). L'analyse permet également d'identifier que la majorité des outils sont conçus pour le dépistage et le diagnostic mais que, pourtant, très peu fournissent de données sur leur pouvoir discriminant. Cela représente une limite importante à l'interprétation que les cliniciens peuvent faire de résultats obtenus grâce à l'utilisation de tels outils.

En ce qui concerne les formes de l'évaluation, nous avons évoqué précédemment l'intérêt de l'évaluation dynamique (critériée, normée ou descriptive) afin de comprendre le profil du patient ou encore déterminer les cibles et les moyens thérapeutiques les plus pertinents. Trois batteries de performances proposent des formes statiques et dynamiques à la fois : ERLA, Examath 5-8 et UDN-3. ERLA et UDN-3 s'inscrivent dans une approche dynamique de l'évaluation en proposant des étayages selon le résultat obtenu. Ces deux batteries de performances basées sur les théories piagésiennes se distinguent toutefois sur un point important : ERLA ne propose pas une standardisation des consignes de passation, contrairement à UDN-3 qui fournit une aide informatisée. Pour terminer, Examath 5-8 propose une modalité d'évaluation dynamique pour l'épreuve de résolution de problèmes arithmétiques verbaux en proposant différents aides à la résolution. Les formes d'évaluation semblent donc commencer à se diversifier.

De futurs outils devraient prendre le soin d'étudier – et de publier les données – relatives à la consistance interne, la sensibilité et spécificité, et la sensibilité au changement,

trois propriétés à ce jour négligées. De plus, alors que les orthophonistes ont trois outils, dont les propriétés psychométriques sont plutôt bonnes, pour une évaluation dont l'objectif est la détection ou le diagnostic du trouble des apprentissages en mathématiques, des efforts supplémentaires devront être réalisés pour le développement d'outils de qualité à des fins de compréhension du profil, mais surtout de mesure de l'évolution des patients.

Limites

Deux limites peuvent être soulevées. Premièrement, la présente étude n'a recensé que les batteries de performances éditées par des maisons d'édition mais n'a pas recherché l'existence – et analysé – d'outils publiés dans des revues scientifiques révisées par les pairs. L'étude de Cattini et Maillart (2024) a montré l'existence d'une diversité d'outils disponibles dans la littérature scientifique pour évaluer les processus de la parole. Cela permet de diversifier la pratique évaluative de l'orthophoniste (par exemple, répertoire des questionnaires permettant d'évaluer l'impact fonctionnel du trouble). Deuxièmement, il est important de rappeler qu'un outil obtenant une note faible ne signifie pas nécessairement que l'outil n'a aucune pertinence sur le plan clinique. Il s'agit, le plus souvent, d'un outil n'ayant pas encore fait l'objet d'études de validation et ne permettant pas à l'orthophoniste d'évaluer le patient de manière éclairée faute d'informations complémentaires. Par exemple, l'ERLA présente des scores très faibles quel que soit l'objectif d'évaluation. Il n'est pas pour autant juste d'en conclure sur base de cette analyse que l'outil ne remplit pas l'objectif d'évaluation énoncé par les auteurs, c'est-à-dire comprendre le profil du patient. Cependant, la validité et la fidélité de cette démarche doivent encore être démontrées en fournissant des preuves de validité théorique (au niveau du choix des items), de structure interne, de validité concurrente, de validité prédictive, de fidélité test-retest et/ou interjuge, ou encore de consistance interne.

Conclusion

Si l'utilisation de batteries de performances n'est pas suffisante pour proposer une évaluation complète et continue du patient, cela constitue généralement l'évaluation de première ligne utilisée par les orthophonistes (Daub et al., 2021 ; Kerr et al., 2003). Dans ce cadre, le choix de la ou des batteries de performances pour initier la prise d'informations cliniques est cruciale. De plus, plusieurs enquêtes ont montré que la sélection d'un outil est fortement affectée par la disponibilité du test sur le lieu de travail (Cattini & Willems, 2024 ; Ogiela & Montska, 2020). Il existe donc un enjeu majeur à renseigner les orthophonistes avant l'achat, souvent coûteux, d'un outil d'évaluation.

En considération de l'examen critique des outils d'évaluation effectué, ainsi que du raisonnement clinique des auteurs, et en attendant la disponibilité d'outils plus satisfaisants sur le plan psychométrique, les suggestions suivantes peuvent être avancées. Pour la détection, nous recommandons la passation de l'épreuve « Dépistage » d'Examath 5-8 pour les jeunes enfants et la BMT-i pour les enfants d'âge scolaire. Néanmoins, la prudence est de rigueur au regard des scores inférieurs à 80% pour le premier et l'absence de données concernant la sensibilité pour le second. Pour le diagnostic d'un TSAM, nous recommandons l'utilisation d'Examath 5-8 pour les jeunes enfants et d'Examath 8-15 pour les enfants d'âge scolaire et les adolescents. Pour comprendre les forces et les faiblesses du patient sur base de données théoriques valides, les batteries de performances précédemment citées sont recommandées ainsi que Tedi-Math Grands. Enfin, pour évaluer l'évolution du patient, nous recommandons d'utiliser un outil proposant *a minima* de calculer un intervalle de confiance⁶.

⁶ L'utilisation d'évaluations normées pour évaluer l'évolution d'un patient doivent être utilisées avec précaution en considérant les données de fidélité pour interpréter les résultats à l'aide de l'erreur standard de mesure (McCauley & Swisher, 1984).

Bibliographie

- American Psychiatric Association. (2016). *DSM-5 : manuel diagnostique et statistique des troubles mentaux (5e éd.)*. Elsevier Masson.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Billard, C., Mirassou, A., & Touzin, M. (2018). *Batterie Modulaire de Tests – Informatisée*. Ortho Édition.
- Billard, C. & Touzin, M. (2012). *Evaluation Des fonctions cognitives et des Apprentissages de 4 à 11 ans*. Ortho Édition.
- Burnay, J., Grégoire, J., Monseur, C., & Willems, S. (2024). Lutter contre les freins à l’Evidence-Based Assessment : Création d’une grille pour examiner la qualité des outils d’évaluation en psychologie, neuropsychologie et orthophonie. *A.N.A.E.*, XX, xx-xx.
- Cattini, J., & Maillart, C. (2024). Analyse critique des caractéristiques psychométriques des outils francophones évaluant la parole chez les enfants d’âge préscolaire. *A.N.A.E.*, XX, xx-xx.
- Cattini, J., & Willems, S. (2024). Analyse des pratiques professionnelles et du sentiment de compétences des orthophonistes francophones pour la sélection raisonnée des tests normés. *A.N.A.E.*, XX, xx-xx.
- Crawford, C., & Cribb, J. (2013). *Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British Cohort Study*. Center for Analysis of Youth Transitions.
- Daub, O., Cunningham, B. J., Bagatto, M. P., Johnson, A. M., Kwok, E. Y., Smyth, R. E., & Cardy, J. O. (2021). Adopting a conceptual validity framework for testing in speech-

- language pathology. *American Journal of Speech-Language Pathology*, 30(4), 1894–1908. https://doi.org/10.1044/2021_AJSLP-20-00032
- Dellatolas, G., & Von Aster, M. (2006). *Batterie pour l'évaluation du traitement des nombres et du calcul chez l'enfant*. Édition du Centre de Psychologie Appliquée.
- Denman, D., Kim, J.-H., Munro, N., Speyer, R., & Cordier, R. (2019). Describing language assessments for school-aged children: A Delphi study. *International Journal of Speech-Language Pathology*, 21(6), 602-612. <https://doi.org/10.1080/17549507.2018.1552716>
- Duboisindien, G., & Cattini, J. (2024). L'expertise et le raisonnement cliniques de l'orthophoniste dans la pratique évaluative. In F. Brin-Henry (Ed.), *Orthophonie : Identité professionnelle et expertise*. De Boeck Supérieur.
- Gaul Bouchard, 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.
- Helloin, M.-C., & Lafay, A. (2016). *Examath 5-8 : batterie informatisée d'EXAmen des habiletés MATHématiques pour les enfants de 5 à 8 ans*. HappyNeuron.
- Kerr, M. A., Guildford, E., & Bird, E. K. (2003). Standardized language test use: a canadian survey. *Revue Canadienne d'Orthophonie et d'Audiologie*, 27(1), 10-28.
- Lafay, A., & Cattini, J. (2018). Analyse psychométrique des outils d'évaluation mathématique utilisés auprès des enfants francophones. *Revue Canadienne d'orthophonie et d'audiologie*, 42(2), 127–144.
- Lafay, A., & Helloin, M.-C. (2016). *Examath 8-15 : batterie informatisée d'EXAmen des habiletés MATHématiques pour les enfants de 8 à 15 ans*. HappyNeuron.
- Leclercq, A.-L., & Veys, E. (2014). Réflexions sur le choix de tests standardisés lors du diagnostic de dysphasie. *A.N.A.E.t*, 26(131), 374-382.
- Legeay, M. P., Morel, L., & Voye, M. (2009). *Mallette Bilan*. Cogilud.

- Macchi, L., Herman, F., Colli-Vaast, L., Merle, A., & Danchin, P. (sous presse). Propriétés psychométriques des tests francophones de langage oral chez l'enfant. *Études de Linguistique Appliquée : Revue de Didactologie Des Langues-Cultures*.
- Martinez Perez, T., Geurten, M., & Willems, S. (2021). Collecter des données sur l'évolution du patient pour augmenter son efficacité thérapeutique en orthophonie. *A.N.A.E.*, 33(173), 433-441.
- Meljac, C., Eynard, L-A., Bernardeau, C., Terriot, K., & Wierzbicki, C. (2020). *UDN-3 : évaluation dynamique de la pensée logique de 4 à 15 ans*. Editions Hogrefe France.
- Meloni, G., Loevenbruck, H., Vilain, A., Gillet-Perret, E., & MacLeod, A. (2022). Evaluation de la perception des sons de parole chez les populations pédiatriques : Réflexion sur les épreuves existantes. *GLOSSA*, 132, 1-27.
- Noël, M.-P., & Grégoire, J. (2015). *Test diagnostique des compétences de base en mathématiques pour les enfants du CE2 à la 5ème*. Édition du Centre de Psychologie Appliquée.
- Ogiela, D. A., & Montzka, J. L. (2021). Norm-referenced language test selection practices for elementary school children with suspected developmental language disorder. *Language, Speech, and Hearing Services in Schools*, 52(1), 288-303.
https://doi.org/10.1044/2020_LSHSS-19-00067
- Olswang, L. B., & Bain, B. (1994). Data collection: Monitoring children's treatment progress. *American Journal of Speech-Language Pathology*, 3(3), 55-66.
<https://doi.org/10.1044/1058-0360.0303.55>
- Thibault, M.-P., & Helloin, M.-C. (2006). *Exalang 3-6 : batterie informatisée pour l'EXAmen du LANGage et des compétences transversales chez l'enfant de 3 à 6 ans*. Orthomotus.

- Thibault, M.-P., Helloin, M.-C., & Lenfant, M. (2009). *Exalang 11-15 : batterie informatisée pour l'EXAmen du LANGage oral, du langage écrit et des compétences transversales chez le collégien*. Orthomotus.
- Thibault, M.-P., Lenfant, M., & Helloin, M.-C. (2012). *Exalang 8-11 : batterie informatisée pour l'EXAmen du LANGage et des compétences transversales chez l'enfant de 8 à 11 ans*. Orthomotus.
- Van Nieuwenhoven, C., Grégoire, J. et Noël, M.-P. (2001). *Test diagnostique des compétences de base en mathématiques*. Édition du Centre de Psychologie Appliquée.
- Wechsler, D. (2005). *Wechsler Individual Achievement Test (2e éd.)*. The Psychological Corporation.
- Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331-346. <https://doi.org/10.1111/cpsp.12207>
- Zanga, A. (2015). *EPOCY 2-3 Echelle de Positionnement en Cycles 2 et 3*. Édition du Centre de Psychologie Appliquée.