

Faculté des sciences de la motricité

Analyse des propriétés psychométriques des outils d'évaluation du contrôle moteur chez les enfants au développement typique

Une scoping review avec synthèse quantitative

Auteur : PERROT François
Promotrice : SCHEPENS Bénédicte
Année académique 2025-2026
Master en Kinésithérapie et Réadaptation [60.0] - KINE2M

UCLouvain FSM : Déclaration d'utilisation des aides et outils Intelligence Artificielle (IA)

J'ai utilisé l'IA pour ce projet :	☒ Oui ☐ Non
------------------------------------	---

Si vous avez utilisé l'IA dans ce travail, veuillez préciser comment ci-dessous :

Texte (par exemple, vérification orthographique, génération de texte, production d'idées, ...)

Le cas échéant, veuillez résumer (500 caractères maximum):

L'IA a été utilisée pour clarifier certaines idées, reformuler des passages longs ou peu clairs, améliorer le style, corriger l'orthographe et la syntaxe, proposer des formulations alternatives et limiter les redondances, et à traduire certains termes anglais en français.

Images

Le cas échéant, veuillez résumer (500 caractères maximum):

/

Code, aide à la programmation, ou algorithmes

Le cas échéant, veuillez résumer (500 caractères maximum):

L'IA a été un support précieux pour optimiser l'utilisation d'Excel, la programmation et à la vérification du script R utilisé pour les synthèses quantitatives et faciliter l'usage des divers outils et bases de données.

Autres utilisations

Le cas échéant, veuillez résumer (500 caractères maximum):

L'IA a été utilisée comme soutien à l'extraction et à l'analyse des données afin d'identifier les éléments clés. Les informations obtenues ont ensuite été comparées à mes propres analyses et vérifiées dans les sources originales. Toute divergence faisait l'objet d'une vérification personnelle.

Je connais les règles de de la Faculté des sciences de la motricité de l'UCLouvain en matière d'intelligence artificielle. Je déclare que toute utilisation d'aides ou d'outils d'intelligence artificielle est explicitement mentionnée dans le présent formulaire de déclaration

Nom	Perrot	Prénom	François
-----	--------	--------	----------

Remerciements

J'aimerais remercier toutes les personnes qui m'ont aidé, guidé et soutenu durant la réalisation de ce travail.

Je tiens avant tout à adresser toute ma gratitude à ma promotrice, Madame Bénédicte Schepens, pour ses conseils judicieux, sa patience ainsi que sa grande disponibilité qui ont su guider mon travail.

Je souhaite également exprimer ma reconnaissance à ma famille, mes amis et mes proches pour leur soutien indéfectible et leurs relectures attentives de ce travail.

INTRODUCTION.....	1
1. DEFINITION DU CONTROLE MOTEUR.....	2
2. LE DÉVELOPPEMENT MOTEUR CHEZ L'ENFANT	2
3. LES INSTRUMENTS D'ÉVALUATION MOTRICE CHEZ L'ENFANT	3
4. OBJECTIF DU MÉMOIRE.....	4
MÉTHODOLOGIE	6
1. PROTOCOLE DE RECHERCHE	6
2. STRATÉGIE DE RECHERCHE	6
3. CRITÈRES D'ÉLIGIBILITÉ.....	7
4. SÉLECTION DES ÉTUDES	8
5. EXTRACTION DES DONNÉES.....	9
5.1. Données extraites pour la synthèse descriptive	9
5.2. Données extraites pour la synthèse quantitative	10
5.3. Processus d'extraction et vérification des données	11
6. ÉVALUATION DE LA QUALITE METHODOLOGIQUE ET ANALYSE STATISTIQUE	11
6.1. Évaluation méthodologique selon COSMIN.....	11
6.2. Synthèse descriptive, qualitative et quantitative	12
RÉSULTATS.....	15
1. ÉTUDES INCLUSES	15
2. QUALITE DES ETUDES.....	16
3. CARACTERISTIQUES GENERALES DES ETUDES	18
4. CARTOGRAPHIE DES OUTILS D'ÉVALUATION IDENTIFIÉS.....	19
4.1. Vue d'ensemble des principaux outils.....	19
4.2. Batteries standardisées à passation individuelle.....	19
4.3. Tests de terrain et outils de screening	19
4.4. Outils de mobilité fonctionnelle	19
4.5. Questionnaires et outils déclaratifs	20
5. PROPRIÉTÉS PSYCHOMÉTRIQUES : SYNTHÈSE QUALITATIVE.....	20
5.1. BOT-2.....	20
5.2. KTK.....	21
5.3. Autres outils	22
6. PROPRIÉTÉS PSYCHOMÉTRIQUES : SYNTHÈSE QUANTITATIVE	22
6.1. TGMD-3 : fidélité inter-évaluateur	23
6.2. TGMD-3 : validité structurelle (CFA).....	23
6.3. TGMD-3 : cohérence interne et structure dimensionnelle	25
6.4. TGMD-3 : fidélité test-retest.....	27
6.5. MABC-2 AB1 : synthèse exploratoire de la reproductibilité par item	28
6.6. Analyses de sous-groupes	32
6.7. Analyses de sensibilité	32
6.8. Biais de publication	32

DISCUSSION :	33
1. CONTRASTE ENTRE LES RESULTATS DE FIDELITE ET DE VALIDITE STRUCTURELLE.....	33
2. STRUCTURE DIMENSIONNELLE DU TGMD-3 : CONVERGENCE DES RÉSULTATS	35
3. L’EFFET DE L’AGE ET DE LA VARIANCE D’ECHANTILLON SUR LES INDICES DE FIDELITE	37
4. TRADUCTION, ADAPTATION CULTURELLE ET HETEROGENEITE	38
5. VARIABILITÉ DE LA REPRODUCTIBILITÉ DES ITEMS DU MABC-2 AB1 : AU-DELÀ DE DRAWING TRAIL	40
6. VALIDITÉ STRUCTURELLE DU MABC-2 : CONVERGENCE ET DIVERGENCE ENTRE BANDES D’ÂGE.....	41
7. LA DISTINCTION ENTRE SOUS-DOMAINES MOTEURS : UNE FRAGILITÉ TRANSVERSALE AU CORPUS	42
8. QUEL OUTIL UTILISER ET A QUEL MOMENT ?.....	43
9. LIMITES ET PERSPECTIVES	44
10. PERSPECTIVES	46
CONCLUSION	47
BIBLIOGRAPHIE	50
ANNEXES :	54
ANNEXE 1 : LISTE DES ABRÉVIATIONS.....	54
ANNEXE 2 : GUIDELINES PRISMA-SCR	56
ANNEXE 3 : ÉQUATIONS DE RECHERCHE EN FONCTION DES BASES DE DONNEES	57
ANNEXE 4 : GRILLE A.....	57
ANNEXE 5 : ÉVALUATION DÉTAILLÉE DES PROPRIÉTÉS PSYCHOMÉTRIQUES ET DU RISQUE DE BIAIS COSMIN DES ÉTUDES PORTANT SUR LE TGMD	68
ANNEXE 6 : ÉVALUATION DÉTAILLÉE DES PROPRIÉTÉS PSYCHOMÉTRIQUES ET DU RISQUE DE BIAIS COSMIN DES ÉTUDES PORTANT SUR LE MABC-2	70
ANNEXE 7 : TABLEAU DES INDICATEURS POUR LES SYNTHÈSES QUANTITATIVES ET SCRIPT R ...	72

Introduction

Chez l'enfant, les compétences motrices constituent une composante essentielle de son développement et de ses trajectoires de santé (Robinson et al., 2020). Le développement moteur est étroitement associé au développement cognitif et perceptif ainsi qu'à la participation à la vie familiale, scolaire et communautaire (Griffiths et al., 2018 ; Mancini et al., 2016). Son évaluation rigoureuse revêt une importance clinique et scientifique majeure (Scarfò et al., 2024). Si les déficits moteurs sévères sont généralement diagnostiqués avant l'âge de deux ans, les déficits plus légers ne deviennent souvent apparents qu'en contexte préscolaire ou scolaire, lorsque l'enfant est confronté à des tâches motrices plus complexes et comparé à ses pairs ; leur identification précoce constitue alors une étape déterminante vers une prise en charge adaptée (Griffiths et al., 2018).

Au-delà du dépistage, ces outils permettent également d'objectiver l'effet des interventions thérapeutiques et de soutenir la recherche fondamentale sur les mécanismes de l'acquisition motrice chez l'enfant (Adolph et Robinson, 2015 ; Griffiths et al., 2018). Face à cet enjeu, il n'existe cependant aucun étalon-or de l'évaluation motrice chez l'enfant : une multitude d'outils et de méthodes ont été développés au fil des décennies, variant tant par leur facilité d'utilisation que par leur mode de référencement, normé ou critérié (Griffiths et al., 2018). Cette diversité méthodologique impose aux professionnels de santé et aux chercheurs de connaître les propriétés psychométriques de chaque instrument afin d'en guider le choix et d'interpréter les résultats de manière appropriée (Griffiths et al., 2018 ; Mookink et al., 2018).

Dans ce contexte, ce mémoire propose une cartographie des principaux instruments standardisés pour évaluer les manifestations fonctionnelles du contrôle et du développement moteur chez les enfants de 1 à 12 ans au développement typique, complétée par une analyse de leurs propriétés psychométriques.

Le choix de cibler une population au développement typique répond à un objectif méthodologique précis. L'étude de ces enfants permet d'examiner le fonctionnement de l'instrument en dehors de l'influence d'une pathologie et de constituer des valeurs de référence nécessaires à l'interprétation des performances individuelles (Scarfò et al., 2024). Ces données permettent notamment de situer les résultats d'enfants présentant un trouble du développement par rapport aux

performances attendues selon l'âge. Elles ne remplacent toutefois pas la validation spécifique de l'instrument dans les populations cliniques auxquelles il est destiné. Elles contribuent également à documenter les propriétés de mesure de l'instrument dans des conditions standardisées d'utilisation et à établir les références nécessaires à l'interprétation des performances individuelles (Scarfò et al., 2024).

1. Définition du contrôle moteur

L'exécution des activités motrices, qu'elles impliquent la locomotion, la manipulation d'objets ou le maintien d'une posture statique, nécessite des stratégies de contrôle postural visant à maintenir la stabilité tout en permettant un mouvement efficient (Shumway-Cook et Woollacott, 2014). Le contrôle moteur désigne l'ensemble des mécanismes sensoriels, neurophysiologiques et biomécaniques permettant de produire des mouvements coordonnés, adaptés et efficaces. Il repose sur l'intégration d'informations issues des systèmes visuel, proprioceptif et vestibulaire, ainsi que sur des processus de planification, d'exécution et d'ajustement des mouvements impliquant notamment les structures cérébelleuses, corticales et sous-corticales (Latash, 2012 ; Levin et Piscitelli, 2022).

2. Le développement moteur chez l'enfant

Chez l'enfant, le contrôle moteur est associé à l'acquisition des compétences locomotrices, à l'exploration de l'environnement et à la participation aux activités sociales (Adolph et Franchak, 2017). Le contrôle postural peut être envisagé selon des composantes statiques et dynamiques, en fonction des exigences de stabilité et de mouvement (Campolettano et al., 2018). Le maintien de l'équilibre et l'adaptation de la posture reposent sur des mécanismes de contrôle postural, définis comme l'ensemble des processus permettant de stabiliser le corps et d'ajuster la posture en fonction des contraintes de la tâche et de l'environnement (Shumway-Cook et Woollacott, 2014).

Les travaux d'Assaiante (1998) mettent en évidence le caractère évolutif du contrôle postural au cours du développement, en particulier à travers la progression du contrôle du tronc et de la coordination intersegmentaire. Ces processus sont associés à la maturation du système nerveux central et à l'expérience motrice.

Entre 1 et 12 ans, le développement moteur suit une trajectoire progressive associée à l'âge, tout en présentant d'importantes variations interindividuelles liées aux expériences motrices et aux contextes environnementaux (Adolph et Robinson,

2015). Dans une perspective dynamique, ces évolutions résultent de l'interaction entre la maturation neurologique, l'expérience motrice et les contraintes propres à la tâche et à l'environnement (Newell, 1986 ; Thelen, 1995).

Entre 1 et 3 ans, l'enfant acquiert la marche indépendante et développe des capacités de locomotion et d'adaptation aux contraintes environnementales.

Entre 4 et 6 ans, les habiletés locomotrices et de coordination, telles que le saut, la coordination bilatérale et le maintien de l'équilibre dynamique, se diversifient et se structurent progressivement à travers des tâches motrices variées (Gallahue et al., 2021).

Entre 7 et 9 ans, les caractéristiques de la coordination motrice évoluent vers une organisation plus stable des mouvements. Une diminution des mouvements syncinétiques involontaires et une structuration progressive des actions motrices sont décrites dans la littérature, associées à une réduction du recours au contrôle cognitif conscient (Haywood et Getchell, 2009).

Entre 10 et 12 ans, l'organisation motrice continue de se préciser parallèlement à la maturation du système nerveux central et à l'expérience accumulée au cours des activités motrices. Les performances observées à cet âge restent également influencées par les opportunités de pratique et les facteurs environnementaux (Case-Smith et O'Brien, 2015 ; Stodden et al., 2008).

3. Les instruments d'évaluation motrice chez l'enfant

Face à la diversité des manifestations du contrôle et du développement moteur, plusieurs batteries de tests ont été élaborées pour évaluer les performances, les compétences et la coordination motrices chez l'enfant.

Parmi les outils les plus fréquemment utilisés en recherche et en pratique clinique pédiatrique figurent : le Test of Gross Motor Development (TGMD), le Movement Assessment Battery for Children-2 (MABC-2), le Bruininks-Oseretsky Test of Motor Proficiency-2 (BOT-2), les Peabody Developmental Motor Scales-2 (PDMS-2) et le Körperkoordinationstest für Kinder (KTK). Ces instruments n'évaluent pas directement l'ensemble des mécanismes neurophysiologiques du contrôle moteur, mais leurs manifestations fonctionnelles à travers des tâches standardisées de locomotion, d'équilibre, de coordination, de manipulation d'objets ou de motricité fine.

L'utilisation de ces tests repose sur l'hypothèse que leurs propriétés psychométriques permettent une interprétation suffisamment fiable et valide des

scores obtenus. La qualité de cette interprétation dépend notamment de la fidélité, c'est-à-dire de la reproductibilité des mesures, de la validité des interprétations fondées sur les scores, ainsi que de l'erreur de mesure associée à l'instrument utilisé. L'absence de données robustes concernant ces propriétés peut limiter l'interprétation des résultats, notamment dans le cadre du suivi longitudinal ou de l'évaluation d'interventions thérapeutiques (Aadland et al., 2022 ; Messick, 1995 ; Terwee et al., 2007).

L'analyse systématique de ces propriétés est importante pour orienter le choix des instruments selon l'âge, le domaine évalué et le contexte d'utilisation, et pour interpréter les résultats de manière rigoureuse et reproductible.

Au-delà de ces tests standardisés largement référencés dans la littérature internationale, d'autres instruments sont également utilisés : tests de terrain conçus pour une administration collective en milieu scolaire (MOBAK, CAMSA), batteries préscolaires (KiMo), outils de mobilité fonctionnelle (TUG) et questionnaires parentaux (LDCDQ, DCDQ). Leurs propriétés psychométriques sont documentées de manière inégale et leur comparaison avec les principaux tests standardisés reste limitée.

4. Objectif du mémoire

Plusieurs limites persistent dans la littérature consacrée à ces outils d'évaluation motrice chez l'enfant. Les propriétés psychométriques ne sont pas systématiquement documentées pour toutes les tranches d'âge et tous les instruments. Les données relatives à l'erreur de mesure et à la sensibilité au changement sont également moins fréquemment rapportées, malgré leur importance dans le suivi de l'évolution des performances. Enfin, l'hétérogénéité des populations, de protocoles et des méthodes statistiques complique la comparaison entre études. Cette scoping review a pour objectif principal de cartographier et de synthétiser les principaux outils standardisés disponibles pour évaluer les manifestations fonctionnelles du contrôle et du développement moteur chez les enfants de 1 à 12 ans au développement typique, en analysant les propriétés psychométriques rapportées dans la littérature, notamment la fidélité, la validité, la sensibilité au changement et l'erreur de mesure. Pour les deux familles d'outils les mieux documentées dans le corpus, le TGMD et le MABC-2, cette cartographie a été complétée par des synthèses quantitatives ciblées lorsque plusieurs études rapportaient des données jugées suffisamment probantes.

Ce travail vise également à proposer des indications d'utilisation en fonction de l'âge de l'enfant, du domaine moteur évalué et de l'objectif de l'évaluation. Il vise enfin à identifier les propriétés et les populations encore insuffisamment documentées afin de dégager des perspectives pour les recherches futures et de soutenir une sélection raisonnée des instruments en pratique clinique et en recherche.

Méthodologie

1. Protocole de recherche

Cette scoping review a été conduite selon les recommandations méthodologiques du JBI pour les scoping reviews et rapportée conformément à l'extension PRISMA-ScR (Peters et al., 2020 ; Tricco et al., 2018). La checklist PRISMA-ScR, utilisée pour guider la rédaction et assurer la transparence du processus, est présentée en Annexe 2.

2. Stratégie de recherche

Le cadre PCC (liste des abréviations en Annexe 1), recommandé pour les scoping reviews (Peters et al., 2020), a été utilisé pour structurer la question de recherche et identifier les principaux concepts de la stratégie documentaire. La question de recherche était la suivante : quels outils standardisés sont utilisés pour évaluer le contrôle et le développement moteur chez les enfants de 1 à 12 ans au développement typique, quelles propriétés psychométriques sont rapportées pour ces outils et quelles lacunes persistent dans la littérature ? Les différents éléments du cadre PCC sont détaillés dans le tableau 1.

Tableau 1. Cadre PCC de la stratégie de recherche

PCC	
Population (P)	Enfants âgés de 1 à 12 ans au développement typique
Concept (C)	Outils standardisés d'évaluation du contrôle et du développement moteur ; Propriétés de mesure rapportées : fidélité, validité, erreur de mesure, sensibilité au changement et données normatives.
Contexte (C)	Contextes de recherche et de pratique clinique pédiatrique, sans restriction géographique ni de date de publication

À partir de ces éléments, une équation de recherche a ensuite été élaborée afin de cibler de manière systématique les articles pertinents dans les bases de données PubMed, Embase et Scopus. La stratégie de recherche combinait quatre blocs de termes relatifs à la population pédiatrique, au contrôle et au développement moteur, aux instruments d'évaluation et aux propriétés psychométriques. Un bloc d'exclusion ciblait les populations présentant un développement atypique. La syntaxe, les champs interrogés et le niveau de spécificité ont été adaptés aux caractéristiques propres à PubMed, Embase et Scopus. Les équations complètes sont présentées en Annexe 3.

Les termes de recherche ont été formulés en anglais dans les trois bases de données. Aucune restriction de date de publication n'a été appliquée. La stratégie a permis d'identifier 375 références dans PubMed, 241 dans Scopus et 126 dans Embase. La recherche documentaire et le processus de sélection se sont déroulés entre le 1^{er} novembre 2025 et le 30 juin 2026. En complément de la recherche électronique, les listes de références des articles examinés en texte intégral ont été vérifiées. Ce chaînage des références bibliographiques a permis d'identifier huit études supplémentaires non repérées par l'équation initiale. Cette démarche est conforme aux recommandations de PRISMA-ScR concernant la description des différentes sources d'information utilisées (Tricco et al., 2018).

3. Critères d'éligibilité

Les études primaires étaient incluses lorsqu'elles répondaient aux critères suivants :

(1) il s'agissait d'études quantitatives évaluant au moins une propriété psychométrique d'un outil de mesure du contrôle ou du développement moteur, notamment des études de validation, des études transversales ou longitudinales, des études de fidélité test-retest, intra-évaluateur ou inter-évaluateur, ainsi que des essais contrôlés randomisés lorsqu'ils rapportaient des données psychométriques exploitables ; (2) elles portaient sur des enfants âgés de 1 à 12 ans au développement typique ; (3) elles utilisaient un instrument reposant sur une procédure standardisée de passation et de cotation, qu'il s'agisse d'une batterie de performance motrice, d'une échelle normée, d'un test de terrain ou d'un questionnaire validé ; (4) elles étaient publiées en anglais ou en français.

Les études dont la tranche d'âge dépassait partiellement 12 ans pouvaient être retenues lorsque les données relatives aux enfants âgés de 1 à 12 ans étaient présentées séparément, lorsque la majorité de l'échantillon se situait dans la tranche d'âge cible ou lorsque l'étude portait sur une bande d'âge standardisée de l'instrument incluant cette population.

Les revues systématiques portant spécifiquement sur les propriétés psychométriques ou l'utilisation d'un instrument éligible pouvaient également être retenues à des fins de cartographie et de contextualisation. Elles étaient analysées séparément des études primaires et n'étaient pas intégrées aux synthèses quantitatives. Les revues narratives ou critiques ne constituaient pas des sources de données primaires ; lorsqu'elles apportaient des informations pertinentes pour la description ou l'interprétation d'un instrument, elles pouvaient être conservées

comme références contextuelles sans être intégrées aux analyses quantitatives. Les études étaient exclues lorsqu'elles portaient exclusivement sur des enfants présentant un développement atypique ou un trouble moteur identifié, lorsqu'elles n'utilisaient pas d'instrument d'évaluation du contrôle ou du développement moteur, ou lorsqu'elles ne rapportaient aucune information pertinente concernant ses propriétés psychométriques. Les études qualitatives, les études de cas ainsi que les publications dont la méthodologie ne permettait pas d'extraire ou d'interpréter des données psychométriques pertinentes étaient également exclues.

4. Sélection des études

À partir des recherches documentaires effectuées, l'ensemble des références identifiées dans PubMed, Embase et Scopus a été importé dans Rayyan, une plateforme permettant la gestion des références, la détection des doublons et la sélection des études dans le cadre des revues systématiques et des scoping reviews (Ouzzani et al., 2016).

Après suppression des doublons, les titres et résumés ont été examinés au regard des critères d'éligibilité prédéfinis. Les références potentiellement pertinentes ont ensuite été évaluées en texte intégral. La sélection a été réalisée par un seul évaluateur. En cas d'incertitude, l'article était réexaminé à partir du texte intégral et des critères définis préalablement.

Parmi les 742 références identifiées, 300 doublons ont été supprimés. Les titres et résumés des 442 références restantes ont ensuite été examinés, conduisant à l'exclusion de 3666 références. Parmi les 76 rapports recherchés en texte intégral, trois n'ont pas pu être obtenus. Les 73 rapports disponibles ont été évalués selon les critères d'éligibilité et sept ont été exclus. Au total, 66 études issues des bases de données ont ainsi été retenues.

Huit références supplémentaires ont été identifiées par chaînage des références bibliographiques et évaluées selon les mêmes critères d'éligibilité. Elles ont toutes été retenues, portant le corpus final à 74 publications. Le processus de sélection est présenté dans le diagramme de flux PRISMA-ScR (figure 1). Les références bibliographiques incluses ont ensuite été organisées à l'aide du logiciel EndNote.

5. Extraction des données

Dans le cadre de cette scoping review, l'extraction des données a été organisée selon deux niveaux complémentaires. La Grille A, présentée en Annexe 4, a été appliquée à l'ensemble des 74 publications incluses afin de réaliser une cartographie descriptive des instruments identifiés, des populations étudiées et des propriétés psychométriques rapportées.

La Grille B correspondait à une extraction approfondie des propriétés psychométriques et de la qualité méthodologique, fondée sur le cadre COSMIN. Elle a été appliquée uniquement aux études portant sur les instruments faisant l'objet d'une synthèse quantitative, soit le TGMD, avec 19 études présentées en Annexe 5, et le MABC-2, avec 17 études présentées en Annexe 6.

5.1. Données extraites pour la synthèse descriptive

Pour chaque étude, les éléments suivants ont été systématiquement recueillis :

Informations générales sur l'étude

- Référence bibliographique (auteur, année, titre, revue)
- Pays et contexte de réalisation

Caractéristiques de la population étudiée

- Taille de l'échantillon
- Âge moyen et plage d'âge
- Répartition par sexe
- Statut développemental de la population et présence éventuelle d'un sous-groupe clinique
- Critères d'inclusion et d'exclusion

Description du test ou outil d'évaluation

- Nom du test ou de l'instrument utilisé
- Domaine évalué : locomotion, motricité globale, motricité fine, coordination, contrôle postural
- Format du test, matériel requis et mode de cotation
- Conditions d'administration et durée

Propriétés psychométriques rapportées

Conformément aux catégories de propriétés de mesure retenues dans cette revue, les éléments suivants ont été extraits lorsqu'ils étaient disponibles :

- Fidélité : test-retest, inter-évaluateur et intra-évaluateur ;

- Cohérence interne, notamment l'alpha de Cronbach ou l'oméga de McDonald ;
- Validité : validité de contenu, de construit et de critère, notamment concurrente ou prédictive ;
- Sensibilité au changement ;
- Erreur de mesure, notamment le SEM et le MDC (abréviations en Annexe 1) ;
- Normes ou valeurs de référence selon l'âge

Conception de l'étude

- Type de plan d'étude : étude de validation, étude transversale, longitudinale, normative ou autre plan pertinent ;
- Nombre de sessions de mesure et intervalle entre les évaluations lorsqu'une fidélité temporelle était étudiée ;
- Modalités de cotation et nombre d'évaluateurs lorsque pertinent ;
- Autres caractéristiques méthodologiques utiles à l'interprétation des propriétés psychométriques.

5.2. Données extraites pour la synthèse quantitative

Seules les propriétés psychométriques rapportées dans au moins deux études et jugées suffisamment comparables en termes d'instrument, de version, de niveau de score, de population, de bande d'âge et de méthode de mesure ont été retenues pour la synthèse quantitative.

Les données quantitatives extraites étaient :

- Les coefficients de corrélation intraclasse (ICC) ;
- Les coefficients de cohérence interne (l'alpha de Cronbach) ;
- Les coefficients de corrélation utilisés pour l'analyse de la validité ou de la structure dimensionnelle ;
- Les indices d'ajustement des modèles factoriels, notamment le CFI et le RMSEA ;
- Les intervalles de confiance associés ;
- Les effectifs des échantillons ;
- Les données nécessaires à l'estimation de l'erreur standard lorsque celle-ci n'était pas directement rapportée ;
- Les indices d'erreur de mesure, notamment le SEM et le MDC, lorsqu'ils étaient disponibles.

5.3. Processus d'extraction et vérification des données

L'extraction et la vérification des données ont été réalisées par un seul évaluateur à l'aide de formulaires standardisés. Les informations extraites ont été vérifiées à partir des articles en texte intégral avant leur intégration dans les tableaux descriptifs et les synthèses quantitatives. Lorsqu'une information était absente, imprécise ou non directement exploitable, cette situation a été signalée dans les grilles d'extraction.

Le recours à un évaluateur unique pour la sélection des études, l'extraction des données et l'évaluation méthodologique constitue une limite de la présente scoping review, abordée dans la section consacrée aux limites de l'étude.

6. Évaluation de la qualité méthodologique et analyse statistique

Dans le cadre de cette scoping review, l'évaluation de la qualité méthodologique n'a pas conduit à l'exclusion d'études. L'ensemble des études répondant aux critères d'éligibilité a été conservé, tandis que leur qualité méthodologique a été rapportée de manière transparente afin de contextualiser l'interprétation des résultats (Peters et al., 2020).

6.1.Évaluation méthodologique selon COSMIN

L'évaluation COSMIN Risk of Bias a été appliquée aux études portant sur le TGMD et le MABC-2, soit les instruments faisant l'objet de la synthèse quantitative approfondie. Pour chaque étude, la qualité méthodologique a été évaluée séparément pour chaque propriété psychométrique analysée, conformément aux recommandations de Mokkink et al. (2018).

La qualité méthodologique a été classée comme « très bonne », « adéquate », « douteuse » ou « inadéquate » selon le principe du worst score counts. La cotation méthodologique attribuée à chaque propriété correspondait ainsi au niveau le plus faible obtenu parmi les critères applicables.

Dans un second temps, les résultats de chaque propriété psychométrique ont été comparés aux critères de suffisance des propriétés de mesure proposées par COSMIN et classés comme suffisants (+), insuffisants (-) ou indéterminés (?). La qualité méthodologique de l'étude a ainsi été distinguée de la suffisance du résultat obtenu. Une propriété pouvait donc présenter une valeur jugée suffisante tout en reposant sur une étude de qualité méthodologique limitée.

Lorsque plusieurs études examinaient une même propriété pour un même instrument, les résultats ont été interprétés en tenant compte de leur cohérence, de leur qualité méthodologique, de leur précision et de leur applicabilité à la population cible. La confiance dans les résultats a été appréciée qualitativement à partir du risque de biais, de l'incohérence entre les études, de l'imprécision et de l'applicabilité des données. Cette appréciation n'a pas constitué une cotation GRADE formelle

6.2.Synthèse descriptive, qualitative et quantitative

Les analyses ont été organisées en trois étapes complémentaires. Une cartographie descriptive a d'abord été réalisée afin de présenter les outils identifiés, les populations étudiées et les propriétés psychométriques rapportées. Une synthèse qualitative narrative a ensuite été menée pour les outils ne disposant pas d'un nombre suffisant d'études comparables, notamment le BOT-2, le PDMS-2, le KTK et les autres instruments moins documentés.

Une synthèse quantitative ciblée a été réalisée pour le TGMD et le MABC-2. Avant tout regroupement, la comparabilité clinique, méthodologique et statistique des données a été examinée. Les coefficients de corrélation intraclasse ont été combinés lorsqu'ils portaient sur une même version de l'instrument, un même niveau de score, une propriété de mesure comparable et une population suffisamment similaire.

Les différents types de fidélité ont été analysés séparément lorsque le nombre d'études disponibles le permettait. Les versions TGMD-2 et TGMD-3 ont été analysées séparément. Pour le MABC-2, les différentes bandes d'âge n'ont pas été regroupées entre elles.

Pour le MABC-2 Bande d'âge 1, le faible nombre d'études disponibles ne permettait pas de constituer des regroupements distincts pour chaque type de fidélité. Certains regroupements par item ont donc inclus, à titre exploratoire, des estimations issues de protocoles de reproductibilité différents : fidélité test-retest, fidélité intra-évaluateur et, pour certains items, fidélité inter-évaluateur à partir d'enregistrements vidéo. Ces regroupements ont été interprétés comme une synthèse générale de la reproductibilité des scores par item, et non comme l'estimation homogène d'un type unique de fidélité.

Les résultats fondés sur des statistiques différentes, comme les corrélations de Pearson ou les coefficients kappa et les ICC, n'ont pas été combinés dans une même synthèse quantitative.

Pour les données jugées suffisamment comparables, une transformation adaptée à la nature de chaque indicateur a été appliquée avant le regroupement. Les ICC et les coefficients de corrélation inter-facteurs ont été transformés sur l'échelle z de Fisher. Les coefficients alpha ont également été transformés sur cette échelle dans le cadre d'une analyse exploratoire ; ce choix ne tenait toutefois pas explicitement compte du nombre d'items composant les scores et a donc été considéré comme une limite méthodologique. Le RMSEA a été analysé après transformation logarithmique. Les synthèses relatives au CFI ont été réalisées à titre exploratoire à partir d'une estimation approximative de sa précision lorsque les intervalles de confiance n'étaient pas rapportés. Les procédures appliquées à chaque indicateur sont détaillées en Annexe 7.

Les analyses ont été réalisées avec R version 4.6.1, à l'aide du package *meta* (version 8.5-0), selon un modèle à effets aléatoires estimé par la méthode du maximum de vraisemblance restreint (*Restricted Maximum Likelihood*, REML). Les estimations regroupées ont ensuite été retransformées dans leur échelle originale afin d'en faciliter l'interprétation.

En l'absence d'intervalles de confiance rapportés pour le CFI, une approximation de la précision a été utilisée dans le cadre d'une synthèse strictement exploratoire. Cette approximation ne permettait pas de restituer l'ensemble des sources d'incertitude propres aux modèles factoriels, notamment les degrés de liberté, la complexité du modèle et la méthode d'estimation. Les résultats correspondants ont donc été interprétés comme des indicateurs descriptifs de convergence plutôt que comme des estimations méta-analytiques standard.

L'absence d'un intervalle de confiance ou d'une erreur standard limitait l'estimation de la variance nécessaire à la pondération statistique. Lorsqu'une estimation suffisamment fiable de la précision ne pouvait pas être obtenue, l'étude n'était pas intégrée au regroupement quantitatif concerné, mais restait incluse dans la synthèse qualitative.

L'hétérogénéité entre les études a été appréciée à l'aide de la statistique Q de Cochran, de l'indice I^2 et de l'estimation de la variance τ^2 inter-études. Lorsque le

nombre d'études était faible, les estimations regroupées et les mesures d'hétérogénéité ont été interprétées avec prudence.

Lorsque l'influence potentielle d'une étude sur les résultats pouvait être anticipée en raison d'une caractéristique méthodologique particulière, une analyse de sensibilité a été réalisée en répétant le regroupement après son exclusion.

Les synthèses quantitatives portant sur les indices d'ajustement factoriel, notamment le CFI et le RMSEA, ont été considérées comme exploratoires en raison des différences entre les modèles statistiques, les méthodes d'estimation et les caractéristiques des échantillons.

L'évaluation du biais de publication par funnel plot ou test d'Egger n'était envisagée que pour les regroupements comprenant un nombre suffisant d'études. Aucun regroupement n'atteignant le seuil de dix études généralement retenu pour permettre une interprétation suffisamment informative de ces méthodes, ces analyses n'ont pas été réalisées.

RÉSULTATS

1. Études incluses

La recherche dans les bases de données PubMed, Embase et Scopus a généré 742 références (375 PubMed, 241 Scopus, 126 Embase). Après importation dans Rayyan et suppression de 300 doublons, 442 références ont été examinées sur la base du titre et du résumé. À ce stade, 366 références ont été exclues et 76 rapports ont été recherchés en texte intégral. Trois rapports n'ont pas pu être obtenus. Parmi les 73 rapports évalués pour l'éligibilité, sept ont été exclus : quatre ne répondaient pas au cadre PCC, deux portaient sur une population non conforme au développement typique et un correspondait à un doublon d'échantillon. Au total, 66 études issues des bases de données ont été retenues. Huit études supplémentaires identifiées par chaînage de citations ont été ajoutées (Bardid et al., 2019 ; Barnett et al., 2014 ; Farrokhi et al., 2014 ; Gharaei et al., 2019 ; Holm et al., 2013 ; Novak et al., 2016 ; Vandorpe et al., 2011 ; Zuvela et al., 2011), portant le corpus final à 74 publications. Le processus de sélection est présenté dans le diagramme de flux PRISMA-ScR en figure 1.

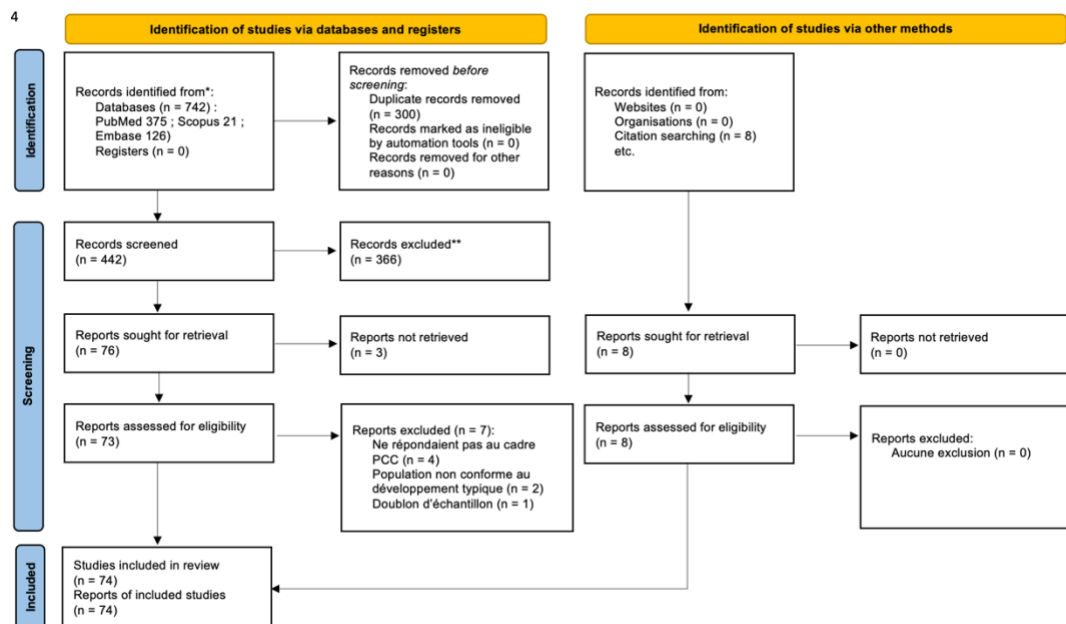


Fig. 1 : Diagramme de flux PRISMA-ScR pour la sélection des articles inclus.

Tableau 2. Répartition des articles inclus par outil et type d'analyse

Outil	N articles	Type d'analyse
TGMD	19	Synthèses quantitatives ciblées et synthèse qualitative
MABC-2	17	Synthèses quantitatives ciblées et synthèse qualitative
PDMS-2	8	Synthèse qualitative
KTK	7	Synthèse qualitative
BOT-2	7	Synthèse qualitative
Autres outils	16	Cartographie narrative
Total	74	

2. Qualité des études

Conformément au protocole, aucune étude n'a été exclue sur la seule base de sa qualité méthodologique. L'ensemble des 74 publications répondant aux critères d'inclusion définis pour la cartographie a été conservé, la qualité méthodologique étant rapportée de manière transparente afin de contextualiser l'interprétation des résultats (Peters et al., 2020). L'évaluation approfondie selon la checklist COSMIN Risk of Bias a été appliquée aux 36 articles portant sur le TGMD et le MABC-2, ces deux instruments ayant été retenus pour l'analyse psychométrique approfondie. L'évaluation méthodologique selon COSMIN a mis en évidence une qualité variable parmi les études portant sur ces deux instruments. Les principales limites concernaient la faible taille de certains échantillons, l'absence d'intervalles de confiance à 95 %, le manque de précision concernant le modèle d'ICC utilisé ainsi que l'hétérogénéité des procédures de passation et de cotation.

L'absence d'intervalles de confiance à 95 % associés aux coefficients de corrélation intraclasse constituait l'une des limites les plus fréquemment observées. Elle empêchait, dans certaines études, d'estimer avec suffisamment de précision la variance nécessaire au regroupement quantitatif. Les études concernées ont néanmoins été conservées dans la synthèse qualitative. Les autres motifs de non-intégration aux synthèses quantitatives comprenaient l'utilisation de statistiques non comparables aux ICC, l'évaluation d'une version différente de l'instrument, l'analyse d'une propriété psychométrique distincte ou l'absence d'une seconde étude suffisamment comparable. Les résultats détaillés par étude et par propriété psychométrique sont présentés en Annexe 5 pour le TGMD et en Annexe 6 pour le MABC-2.

Tableau 3. Articles pour la synthèse quantitative par outil (Grille B)

Outils	Études évaluées	Appréciation méthodologique synthétique	Articles avec au moins un ICC potentiellement intégrable à une synthèse quantitative	Principales limites
TGMD	19	Variable, de très bonne à douteuse	9 (Rintala 2017, Maïano 2022, Marinšek 2023, Estevan 2017, Wagner 2017, Barnett 2014, Farrokhi 2014, Maeng 2017, Magistro 2020)	petits échantillons, IC95 % absents, modèles ICC parfois mal précisés
MABC-2	17	Majoritairement douteuse à adéquate	7 (Jaikaew & Satiansukpong 2019, Arbi 2025, Serbetar 2019, Scarfò 2024, Ellinoudis 2011, Holm 2013, Niño-Cruz 2019)	faibles effectifs, effets plafond, propriétés inégalement rapportées

L'évaluation COSMIN ayant été réalisée séparément pour chaque propriété psychométrique, la colonne « Appréciation méthodologique synthétique » fournit uniquement une vue d'ensemble de la qualité méthodologique observée. Les cotations détaillées par étude et par propriété sont présentées dans les Annexes 5 et 6.

Les études comportant des données potentiellement regroupables dans le tableau 3 ne sont pas toutes représentées dans les forest plots de la section 6.

Pour le TGMD, Barnett et al. (2014) et Farrokhi et al. (2014) portent sur le TGMD-2 et n'ont pas été regroupées avec les études TGMD-3. De plus, les deux études rapportaient des types de fidélité différents : une fidélité inter-évaluateur pour Barnett et al. (2014) et une fidélité test-retest et intra-évaluateur pour Farrokhi et al. (2014). En l'absence d'une seconde étude suffisamment comparable pour chacune de ces propriétés, aucun regroupement quantitatif propre au TGMD-2 n'a été réalisé.

Pour le MABC-2, Jaikaew et Satiansukpong (2019) et Holm et al. (2013) portent sur la bande d'âge 2 (AB2), tandis que les figures 10a-f ne regroupent que la bande d'âge 1 (AB1) de 3 études (Arbi et al., 2025 ; Ellinoudis et al., 2011 ; Scarfò et al., 2024), conformément à la règle de non-mélange des bandes d'âge. Aucune de ces deux études AB2 ne dispose d'une seconde étude AB2 comparable pour constituer un regroupement propre à cette bande d'âge 2. Holm et al. (2013) rapportent en outre des résultats fondés sur des scores bruts non standardisés, qui ne sont pas directement assimilables à un ICC inter-évaluateur classique. Enfin, Serbetar et al. (2019), bien que portant sur la bande d'âge 1, rapportent des ICC pour le score total plutôt que des ICC distincts pour chacun des six items. Cette étude n'a donc pas été intégrée aux analyses item par item des figures 10a à 10f, lesquelles exigeaient une mesure comparable pour chaque item dans les études retenues.

3. Caractéristiques générales des études

Les 74 publications incluses couvrent une diversité de contextes géographiques et démographiques. La répartition par pays, par famille d'outil, est présentée dans le tableau suivant (chaque article étant compté une fois, pour le pays principal de réalisation).

Tableau 4. Répartition géographique des études incluses par famille d'outil

Famille	Pays représentés
TGMD (19)	Brésil (3), USA (3), Iran (2), Myanmar, Finlande, Canada-Québec, Slovénie, Espagne, Portugal, Allemagne, Canada, Indonésie, Australie, Italie (1 chacun)
MABC-2 (17)	Australie (2), Tchéquie (3), Thaïlande (2), Pays-Bas, Europe/Afrique (multicentrique), Tunisie, Brésil, Croatie, Italie, Grèce, Norvège, Colombie, Japon (1 chacun)
PDMS-2 (8)	USA, Brésil (2), Australie, Taïwan, Portugal, Chine, Brésil+Taïwan/USA
KTK (7)	Pays-Bas (2), Belgique-Flandre (2), Brésil, Belgique, Iran (1 chacun)
BOT-2 (7)	Irlande, Hong Kong, Australie, Corée du Sud, Iran, Belgique-Flandre (1 chacun)
Autres outils (16)	Portugal (3), Belgique (2), Brésil (2), Suisse, USA, Pays-Bas, Espagne, Pologne, Danemark, Indonésie, Irlande, Croatie (1 chacun)

La population cible de la revue était constituée d'enfants âgés de 1 à 12 ans au développement typique. Certaines publications incluaient toutefois une tranche d'âge dépassant partiellement cette limite. Rizkyanto (2024) portait sur des enfants âgés de 11 à 13 ans, Valentini et al. (2014) sur des enfants âgés de 3 à 13 ans et Scarfò et al. (2024) incluait les trois bandes d'âge du MABC-2, jusqu'à 16 ans. Kakebeeke et al. (2018) présentait des normes couvrant une population âgée de 3 à 18 ans. Enfin, la revue de Christopher et al. (2021) concernait une population couvrant l'ensemble du cycle de vie, avec un sous-ensemble pédiatrique pertinent pour la présente revue. Ces publications ont été conservées conformément aux critères d'éligibilité définis dans la Méthodologie. Deux études portant sur le PDMS-2 ont en revanche été exclues en raison d'une population non conforme au critère de développement typique.

4. Cartographie des outils d'évaluation identifiés

4.1. Vue d'ensemble des principaux outils

Tableau 5. Vue d'ensemble des cinq outils cibles du mémoire

Outil	Versions identifiées	Domaine moteur	N articles
TGMD	TGMD-2, TGMD-3 (dont TGMD-3-SF)	Locomoteur + contrôle d'objet	19
MABC-2	MABC-2 AB1/AB2/AB3, MABC-2 Checklist, MABC-3 (1 publication, non intégrée aux synthèses quantitatives)	Dextérité manuelle + visée-réception + équilibre	17
PDMS-2	PDMS original, PDMS-2	Motricité globale + fine (6 sous-échelles)	8
KTK	KTK original (4 sous-tests), KTK3 (forme courte), KTK3+	Équilibre dynamique, locomotion latérale, saut unipodal, saut latéral	7
BOT-2	BOTMP (ancien), BOT-2 (forme complète), BOT-2 Short Form	Motricité globale + fine	7

4.2. Batteries standardisées à passation individuelle

Les cinq outils ciblés par le mémoire (TGMD, MABC-2, BOT-2, PDMS-2, KTK) relèvent de ce format de passation individuelle structurée en sous-tests. S'y ajoutent, dans la famille « Autres outils », deux batteries standardisées de structure comparable : le Motor Competence Assessment (MCA) (Rodrigues et al., 2022 ; Soares et al., 2020) et le Motor Development Scale (MDS) (Okuda et al., 2020).

4.3. Tests de terrain et outils de screening

Plusieurs outils de terrain, conçus pour une utilisation en contexte scolaire ou de dépistage, complètent la cartographie : le MOBAC 1 et le MOBAC 3-4 (Quitério et al., 2018), le CAMSA (Menescardi et al., 2022), le FUS test (Makaruk et al., 2023), le FMS2 (Gavigan et al., 2025), et l'AST (Hoeboer et al., 2016 ; Hoeboer et al., 2018), ce dernier utilisant le KTK comme critère externe de validité concourante plutôt que comme objet d'étude propre. Le test POLYGON (Zuvela et al., 2011) appartient également à cette catégorie ; il a été validé par comparaison concourante au TGMD-2 (ρ de Spearman = -0,82).

4.4. Outils de mobilité fonctionnelle

Le Timed Up and Go (TUG) constitue le principal outil de cette catégorie. La revue de Hallemans et al. (2019) regroupe six protocoles distincts issus de six études primaires portant au total sur 2626 enfants. Christopher et al. (2021) propose une revue couvrant plusieurs populations au cours de la vie dont un sous-ensemble d'études concerne spécifiquement des enfants au développement typique. Le Supine-to-Stand Task, étudié dans la revue systématique de Cattuzzo et al. (2020), appartient également à cette catégorie.

4.5. Questionnaires et outils déclaratifs

Trois questionnaires complètent la cartographie, reposant sur un format déclaratif plutôt que sur une passation chronométrée directe : le MABC-2 Checklist, complété par l'enseignant (Banátová et al., 2023), le LDCDQ-BE, version belge francophone du Little Developmental Coordination Disorder Questionnaire (De Roubaix et Van Waelvelde, 2022), et le DCDQ dans son adaptation danoise (Topor et al., 2024).

5. Propriétés psychométriques : synthèse qualitative

Cette section présente les instruments pour lesquels le nombre ou la comparabilité des études disponibles ne permettaient pas de réaliser une synthèse quantitative pertinente : le PDMS-2, le BOT-2, le KTK et les autres instruments identifiés.

Tableau 6. Disponibilité des propriétés psychométriques par famille d'outil

Famille	ICC rapporté (Oui / Non)	Alpha ou oméga rapporté (Oui / Non)
PDMS-2 (8)	4 / 4	3 / 5
BOT-2 (7)	2 / 5	1 / 6
KTK (7)	0 / 7	2 / 5
Autres outils (16)	6 / 10	0 / 16

Le corpus PDMS-2 final comprenait huit publications. Trois publications supplémentaires initialement recensées n'ont pas été retenues faute d'accès au texte intégral, et deux autres ont été exclues en raison d'une population non conforme au critère de développement typique. Zanella et al. (2021) et Valentini et Zanella (2022) reposaient sur le même échantillon brésilien de 637 enfants, mais analysaient des propriétés différentes : la première étude rapportait notamment des coefficients de fidélité, tandis que la seconde proposait une analyse de Rasch. Elles ont donc été considérées comme deux analyses complémentaires d'un même échantillon et non comme deux échantillons indépendants. Seamon et al. (2025) réutilisait des calibrations Rasch déjà publiées pour construire une unité de mesure trans-instruments. Zhu et al. (2024) constituait une revue COSMIN du PDMS-2 portant sur 22 études et a été utilisée comme référence contextuelle.

5.1. BOT-2

Les sept publications portant sur le BOT-2 couvraient le BOTMP original (Lam, 2011 ; MacCobb et al., 2005) ainsi que le BOT-2 dans sa forme complète et ses formes abrégées (Bardid et al., 2019 ; Brown, 2019a ; Brown, 2019b ; Gharaei et al., 2019 ; Yoon et al., 2024). Gharaei et al. (2019) rapportaient un ICC test-retest

de 0,809 [IC95 % 0,685–0,888] pour le BOT-2 Brief Form dans un échantillon de 50 enfants. Yoon et al. (2024) rapportaient une cohérence interne de 0,90 pour la forme complète et de 0,69 pour la forme courte.

Bardid et al. (2019), sur un échantillon de 2538 enfants, ont analysé la structure du BOT-2 Short Form à l'aide d'un modèle de Rasch et ont retenu 12 des 14 items après recalibrage. Brown (2019a) a étudié la validité structurelle des sous-échelles et scores composites de la forme complète, tandis que Brown (2019b) a appliqué une approche similaire à la forme brève. Dans cette dernière étude, 9 des 14 items de la version standard présentaient un mésajustement au modèle de Rasch et une version révisée à cinq items a été proposée.

5.2.KTK

Aucun des sept articles KTK ne rapporte d'ICC au sens strict. Deux études (Hoeboer et al., 2016 ; Hoeboer et al., 2018) utilisent le KTK uniquement comme critère externe de validité concourante pour l'Athletic Skills Track, sans rapporter de propriété psychométrique propre à l'outil. Moreira et al. (2019) et Salami et al. (2023) rapportent une cohérence interne ($\alpha = 0,80$ et $0,86$ respectivement) ainsi qu'une fidélité inter/intra-évaluateur estimée par kappa de Cohen plutôt que par ICC. Coppens et al. (2021) cite des coefficients de fidélité test-retest issus d'études antérieures, non recalculés sur son propre échantillon et exprimés en r de Pearson. Vandorpe et al. (2011) est une étude de valeurs de référence qui ne recalcule pas de coefficient de fidélité sur son échantillon, mais rappelle ceux du manuel allemand original (Kiphard & Schilling, 1974/2007) avec un coefficient de 0,97 pour le score total et des coefficients compris entre 0,80 et 0,96 pour les sous-tests. Novak et al. (2016) compare une forme courte KTK3 à la forme complète KTK4 par corrélation de Pearson ($r = 0,97$) et accord de classification ($\kappa = 0,72$), constituant une étude de concordance inter-formes plutôt qu'une étude de fidélité ou de validité classique. Les échantillons de Vandorpe et al. (2011) et de Novak et al. (2016) provenaient d'un contexte géographique et d'équipes de recherche proches ; les informations disponibles ne permettaient pas d'établir avec certitude si leurs échantillons étaient totalement indépendants.

5.3. Autres outils

Parmi les 16 articles de cette catégorie, plusieurs rapportent des données de fidélité chiffrées. Kakebeeke et al. (2018) rapporte une fidélité intra-observateur ($ICC = 1,00$), inter-observateur ($ICC = 0,92-1,00$) et test-retest ($ICC = 0,67-0,84$), sans IC95 %. La revue sur le TUG de Hallemans et al. (2019), regroupant six études, rapporte des ICC allant de 0,76 à 0,999 selon le protocole et l'étude source, avec des IC95 % disponibles uniquement pour le sous-groupe Williams et al. (2005). Une divergence était observée entre Hallemans et al. (2019) et Christopher et al. (2021) concernant l'ICC test-retest attribué au protocole de Williams et al. (2005) pour les enfants âgés de 5 à 9 ans : 0,76 dans la première revue et 0,71 dans la seconde, alors qu'un même IC95 % [0,61–0,85] était rapporté. En l'absence de résolution de cette discordance entre les deux sources secondaires, cette estimation n'a pas été intégrée à une synthèse quantitative. Zuvela et al. (2011) rapportaient des ICC intra-évaluateur compris entre 0,83 et 0,98 selon les 21 tâches du test POLYGON. Les questionnaires LDCDQ-BE et DCDQ ainsi que les outils MOBAK, CAMSA et FUS ne rapportaient pas d'ICC ni de coefficient alpha dans les données extraites, leurs propriétés étant documentées par d'autres indicateurs de validité.

6. Propriétés psychométriques : synthèse quantitative

Une synthèse quantitative a été réalisée pour le TGMD-3 et le MABC-2 à l'aide du package *meta* dans R, selon un modèle à effets aléatoires estimé par la méthode REML. Les méthodes de transformation et d'estimation de la précision propres à chaque indicateur sont détaillées dans la Méthodologie, section 6.2. Les valeurs présentées dans le texte et les tableaux ont été retransformées dans leur échelle originale. Les synthèses quantitatives réalisées sont présentées sous forme de forest plots dans les figures 2 à 10f. Conformément à la règle méthodologique du mémoire, le TGMD-2 et le TGMD-3 n'ont jamais été regroupés ensemble, de même que les bandes d'âge AB1/AB2/AB3 du MABC-2 entre elles.

Compte tenu du faible nombre d'études disponibles dans la majorité des regroupements et de l'hétérogénéité méthodologique entre les études, les estimations regroupées sont présentées comme des résultats exploratoires et doivent être interprétées avec prudence. Cette réserve concerne particulièrement les synthèses portant sur les indices d'ajustement factoriel CFI et RMSEA.

6.1. TGMD-3 : fidélité inter-évaluateur

Cinq études portant sur la fidélité inter-évaluateur du score total du TGMD-3 ont été regroupées (figure 2) : Rintala et al. (2017), Maeng et al. (2017), Magistro et al. (2020), Maïano et al. (2022) et Marinšek et al. (2023). L'ICC inter-évaluateur regroupé, obtenu à l'aide d'un modèle à effets aléatoires REML après transformation en z de Fisher, était de 0,960 [IC95 % 0,786–0,993], avec une variance inter-études $\tau^2 = 1,016$ et une hétérogénéité $I^2 = 99,8$ %.

Après exclusion de Maeng et al. (2017), dont l'échantillon était limité à dix participants, l'estimation regroupée demeurerait identique à 0,960. Son intervalle de confiance s'élargissait toutefois à [0,670–0,996] et la variance inter-études augmentait de $\tau^2 = 1,016$ à $\tau^2 = 1,339$.

Malgré une estimation regroupée élevée, l'hétérogénéité extrême observée limite fortement l'interprétation d'une valeur moyenne commune aux cinq études. Les ICC individuels variaient de 0,620 [0,589–0,649] chez Rintala et al. (2017) à 0,998 [0,997–0,998] chez Maïano et al. (2022). Les différences méthodologiques et les caractéristiques des échantillons susceptibles de contribuer à cette dispersion sont examinées dans la Discussion.

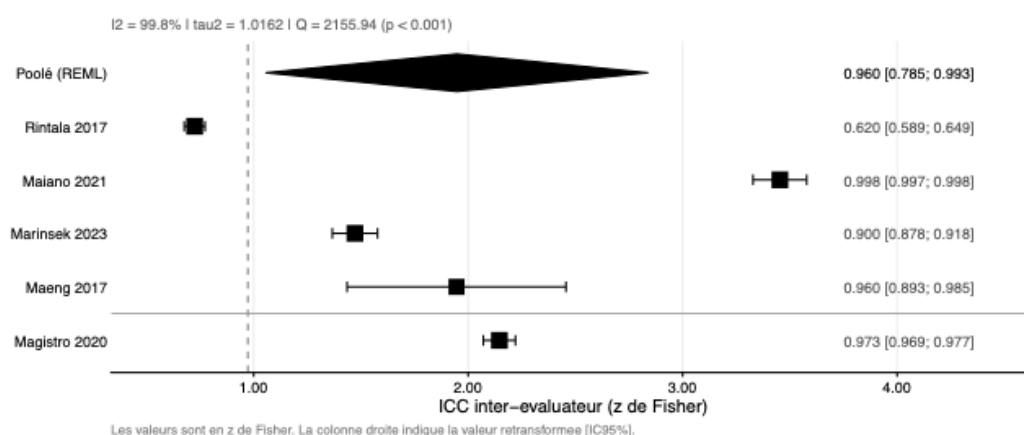


Fig. 2 : Forest Plot de l'ICC inter-évaluateur du TGMD-3

6.2. TGMD-3 : validité structurelle (CFA)

Les synthèses quantitatives du CFI et du RMSEA ont été réalisées à titre exploratoire. Les études combinaient des modèles factoriels, des méthodes d'estimation et des caractéristiques d'échantillon partiellement différents ; les valeurs regroupées décrivent donc la convergence des indices rapportés, sans constituer à elles seules une validation formelle d'un modèle factoriel commun.

Pour le modèle à deux facteurs, quatre études ont été regroupées (figure 3). Le CFI regroupé était de 0,956 [IC95 % 0,934–0,970]. Une hétérogénéité faible à modérée

était observée ($I^2 = 49,9 \%$, $p = 0,112$). L'absence de significativité statistique du test d'hétérogénéité ne permet toutefois pas de conclure à l'absence de différences entre les études, notamment en raison de leur faible nombre.

Pour le modèle bifacteur, trois études ont été regroupées (figure 4). Le CFI regroupé était de 0,986 [IC95 % 0,978–0,991] et le RMSEA (figure 5) de 0,042 [IC95 % 0,035–0,051]. L'indice I^2 était estimé à 0 % pour les deux analyses ($p = 0,960$ pour le CFI et $p = 0,964$ pour le RMSEA). Compte tenu du faible nombre d'études, cette absence d'hétérogénéité statistiquement détectée doit être interprétée avec prudence.

Dans les études disponibles, les indices d'ajustement rapportés étaient donc globalement compatibles avec un ajustement satisfaisant des modèles à deux facteurs et bifactoriels. Les valeurs plus favorables observées pour les modèles bifactoriels ne permettent cependant pas d'établir leur supériorité formelle, les synthèses ne reposant pas nécessairement sur des ensembles d'études et des procédures statistiques strictement identiques.

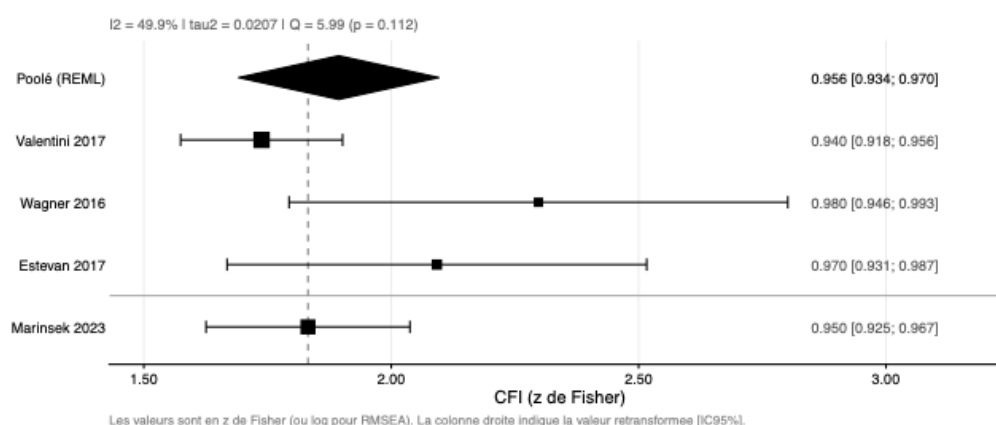


Fig. 3 : Forest Plot du CFI du TGMD-3, modèle à deux facteurs

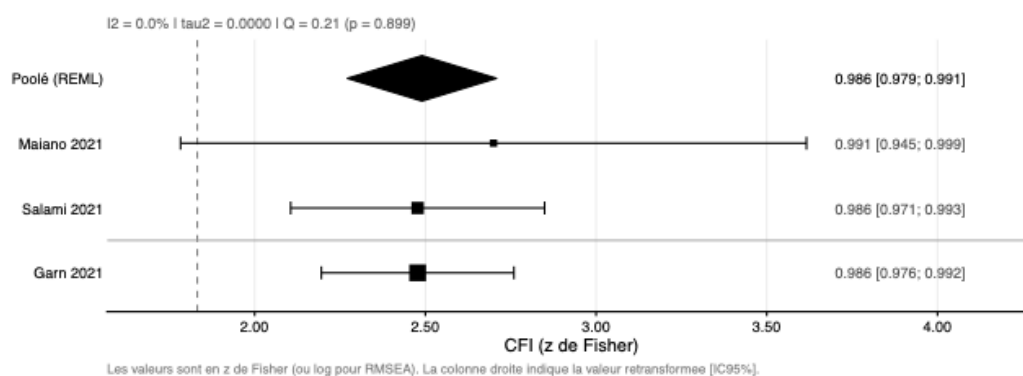


Fig. 4 : Forest Plot du CFI du TGMD-3, modèle bifacteur

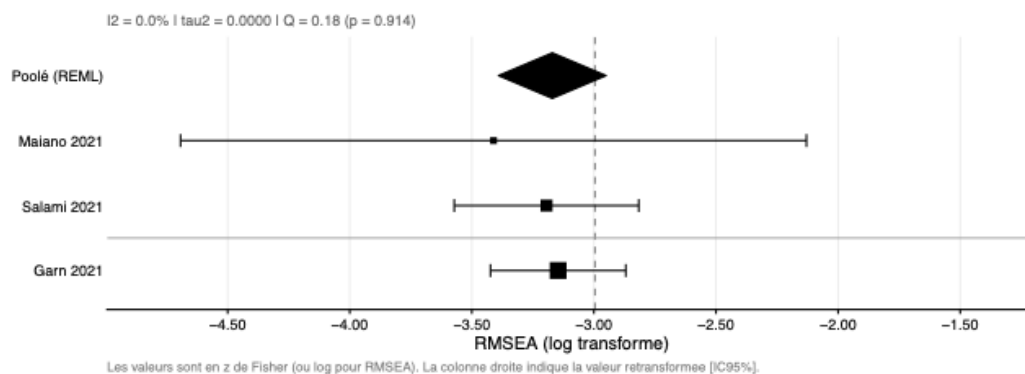


Fig. 5 : Forest Plot du RMSEA du TGMD-3, modèle bifacteur

Tableau 7. Indices d'ajustement regroupés du TGMD-3 selon le modèle testé

Modèle	k	CFI [IC95 %]	regroupé I ² (CFI)	RMSEA [IC95 %]	regroupé I ² (RMSEA)
Deux facteurs	4	0,956 [0,934–0,970]	49,9 % (p = 0,112)	non modélisé pour ce modèle	—
Bifacteur	3	0,986 [0,978–0,991]	0,0 % (p = 0,960)	0,042 [0,035–0,051]	0,0 % (p = 0,964)

Maïano et al. (2022) (IC90 % [0,000 ; 0,071]) a été conservée dans ce pooling plutôt qu'exclue, malgré l'imprécision de son intervalle par rapport à Salami et al. (2022) ([0,027 ; 0,053]) et Garn et Webster (2021) ([0,033 ; 0,053]). Après transformation logarithmique, cette imprécision se traduisait par une variance plus élevée et, par conséquent, par une pondération plus faible dans le modèle à effets aléatoires. Aucun seuil d'exclusion fondé uniquement sur la largeur de l'intervalle de confiance n'a été appliqué.

6.3. TGMD-3 : cohérence interne et structure dimensionnelle

La synthèse quantitative des coefficients alpha de Cronbach a été réalisée à titre exploratoire (figure 6), compte tenu des différences entre les méthodes de cotation et les structures de score utilisées dans les études. L'alpha regroupé du score total du TGMD-3 était de 0,863 [IC95 % 0,785–0,915], avec une variance inter-études $\tau^2 = 0,0629$, un test d'hétérogénéité $Q = 220,13$ ($p < 0,001$) et une hétérogénéité élevée ($I^2 = 98,6$ %). Une analyse de sensibilité a été réalisée en excluant Valentini et al. (2017), seule étude utilisant une cotation de type skills-to-test et rapportant un coefficient alpha de 0,740 [0,716–0,762]. Après cette exclusion, l'alpha regroupé était de 0,891 [IC95 % 0,864–0,913], avec $\tau^2 = 0,0092$ et $I^2 = 89,1$ %. Cette analyse indiquait une légère augmentation de l'estimation regroupée et une diminution de l'hétérogénéité, qui demeurait néanmoins élevée.

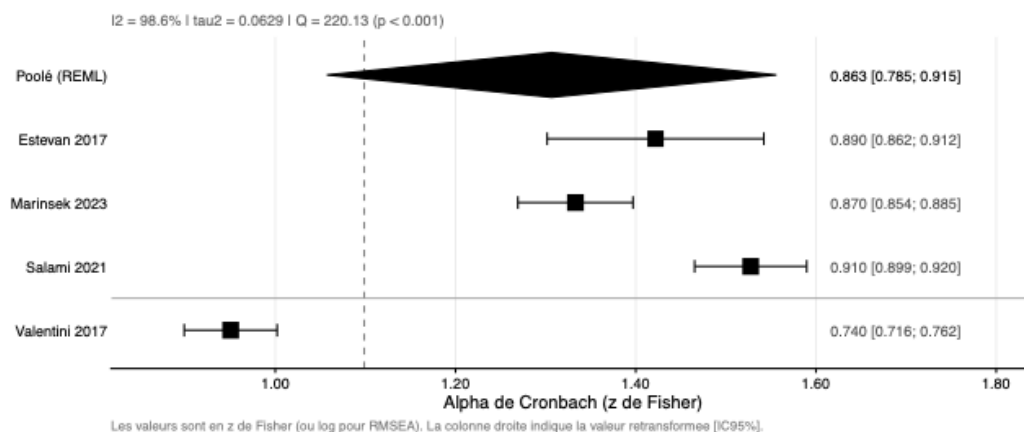


Fig. 6 : Forest Plot de l'alpha de Cronbach TGMD-3, score total

La corrélation regroupée entre les facteurs Locomoteur et Ball Skills était de $r = 0,871$ [IC95 % 0,827–0,904], avec $\tau^2 = 0,0230$, $Q = 28,40$ ($p < 0,001$) et $I^2 = 89,4$ %. Les quatre études incluses rapportaient des corrélations individuelles comprises entre 0,820 et 0,910 (figure 7). La valeur la plus faible était rapportée par Marinšek et al. (2023) avec $r = 0,820$, tandis que Valentini et al. (2017) rapportaient une corrélation de $r = 0,890$.

Malgré une corrélation moyenne élevée entre les deux facteurs, l'hétérogénéité importante indiquait une variabilité de l'intensité de cette association entre les études.

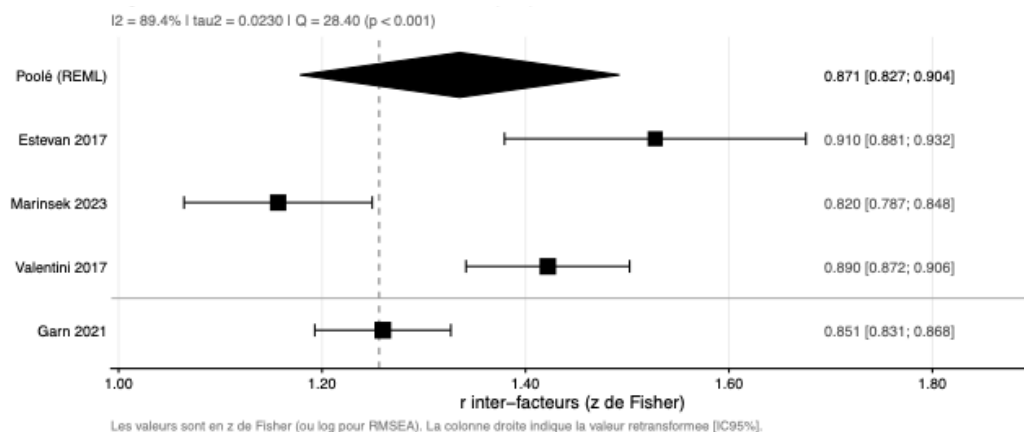


Fig. 7 : Forest Plot de la corrélation entre les facteurs Locomoteur et Ball Skills du TGMD-3

Les ICC individuels rapportés pour le TGMD-2 et le TGMD-3 sont présentés séparément dans la figure 8. Cette représentation est descriptive et ne correspond pas à une estimation regroupée commune, les deux versions de l'instrument n'ayant pas été combinées quantitativement.

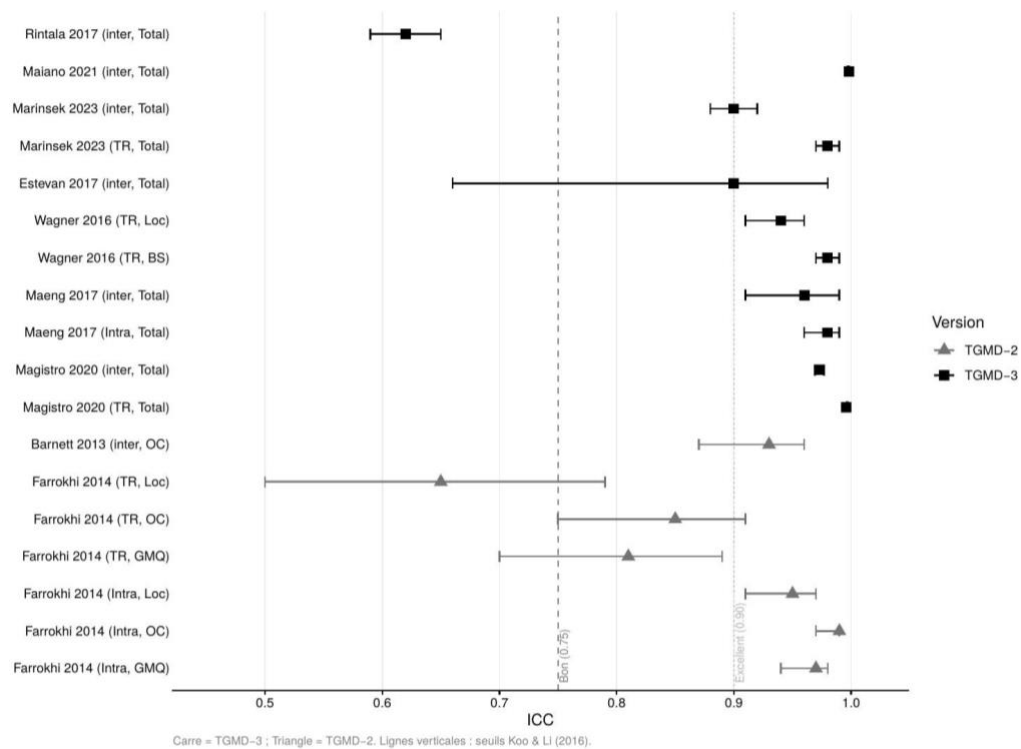


Fig. 8 : Forest plot des ICC individuels du TGMD avec IC95 %.

Carré = TGMD-3 ; triangle = TGMD-2.

6.4.TGMD-3 : fidélité test-retest

Trois études portant sur la fidélité test-retest du TGMD-3 ont été regroupées : Magistro et al. (2020), Marinšek et al. (2023), Wagner et al. (2017). Pour la sous-échelle Locomoteur (figure 9a), l'ICC regroupé, obtenu à l'aide d'un modèle à effets aléatoires REML après transformation en z de Fisher, était de 0,98 [IC95 % 0,89–1,00], avec une variance inter-études $\tau^2 = 0,542$ et une hétérogénéité $I^2 = 99,2$ %.

Pour la sous-échelle Ball Skills (figure 9b), l'ICC regroupé était de 0,99 [IC95 % 0,95–1,00], avec $\tau^2 = 0,381$ et $I^2 = 98,9$ %.

Aucun regroupement n'a été réalisé pour le score total, Wagner et al. (2017) ne rapportant les ICC test-retest que pour les sous-échelles. Malgré des estimations regroupées élevées, l'hétérogénéité extrême observée dans les deux analyses limite fortement l'interprétation d'une valeur moyenne commune aux études. Les facteurs susceptibles de contribuer à cette dispersion sont examinés dans la Discussion.

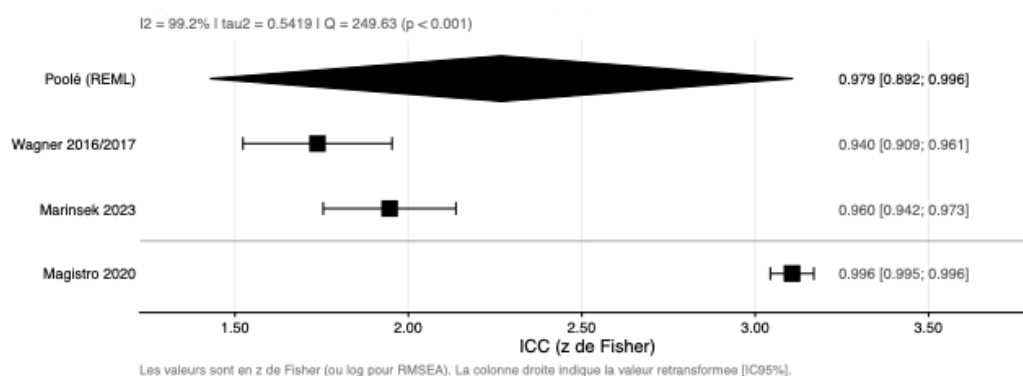


Fig. 9a : Forest Plot de l'ICC test-retest du TGMD-3, Locomoteur

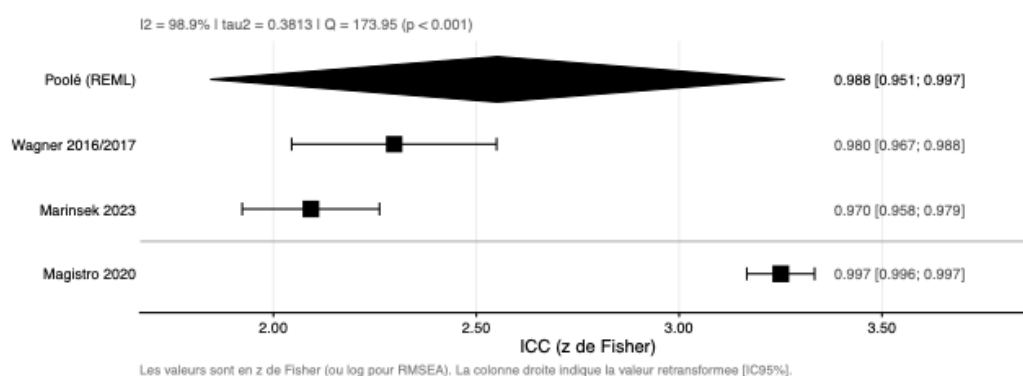


Fig. 9b : Forest Plot de l'ICC test-retest du TGMD-3, Ball Skills

6.5.MABC-2 AB1 : synthèse exploratoire de la reproductibilité par item

Six items du MABC-2 Bande d'âge 1 disposaient de données ICC suffisamment comparables pour permettre une synthèse quantitative : Posting Coins avec la main préférée (MD1p), Threading Beads (MD2), Drawing Trail (MD3), Catching Beanbag (CA1), Aiming (CA2) et Walking Heels Raised (B2).

Les six analyses reposaient sur un noyau commun de trois études : Arbi et al. (2025), Ellinoudis et al. (2011), Scarfò et al. (2024). L'étude de Niño-Cruz et al. (2019), menée auprès de 29 enfants en Colombie, a également été intégrée aux analyses de MD1p et de CA2, portant le nombre d'études incluses à quatre pour ces deux items. Les protocoles de reproductibilité n'étaient pas strictement identiques entre les études. Arbi et al. (2025) et Ellinoudis et al. (2011) rapportaient une fidélité test-retest, Scarfò et al. (2024) une fidélité intra-évaluateur, tandis que Niño-Cruz et al. (2019) rapportait une fidélité inter-évaluateur à partir d'enregistrements vidéo pour MD1p et CA2. Les analyses suivantes sont donc présentées comme des synthèses exploratoires de la reproductibilité par item.

Pour les quatre autres items, les intervalles de confiance rapportés par Niño-Cruz et al. (2019) présentaient des bornes identiques après arrondi. Cette présentation ne permettait pas de reconstruire une erreur standard suffisamment fiable pour la transformation en z de Fisher. Ces résultats n'ont donc pas été intégrés aux regroupements correspondants, qui restaient composés de trois études.

Pour MD1p (figure 10a), l'ICC regroupé était de 0,960 [IC95 % 0,861–0,989], avec $\tau^2 = 0,4171$ et $I^2 = 95,9$ %. Pour MD2 (figure 10b), l'ICC regroupé était de 0,973 [IC95 % 0,912–0,992], avec $\tau^2 = 0,2639$ et $I^2 = 93,9$ %. Pour MD3 (figure 10c), l'estimation regroupée était de 0,893 [IC95 % 0,661–0,969], avec $\tau^2 = 0,2987$ et $I^2 = 94,1$ %.

Pour CA1 (figure 10d), l'ICC regroupé était de 0,943 [IC95 % 0,815–0,983], avec $\tau^2 = 0,2875$ et $I^2 = 94,3$ %. Pour CA2 (figure 10e), l'ICC regroupé était de 0,908 [IC95 % 0,829–0,952], avec $\tau^2 = 0,0815$ et $I^2 = 74,9$ %. Enfin, pour B2 (figure 10f), l'ICC regroupé était de 0,971 [IC95 % 0,921–0,990], avec $\tau^2 = 0,1992$ et $I^2 = 92,7$ %.

Les estimations regroupées étaient élevées pour les six items, mais l'hétérogénéité restait importante à extrême dans l'ensemble des analyses. Compte tenu du faible nombre d'études et de cette hétérogénéité, les résultats sont considérés comme exploratoires et ne peuvent pas être généralisés comme des estimations universelles de la reproductibilité de chaque item.

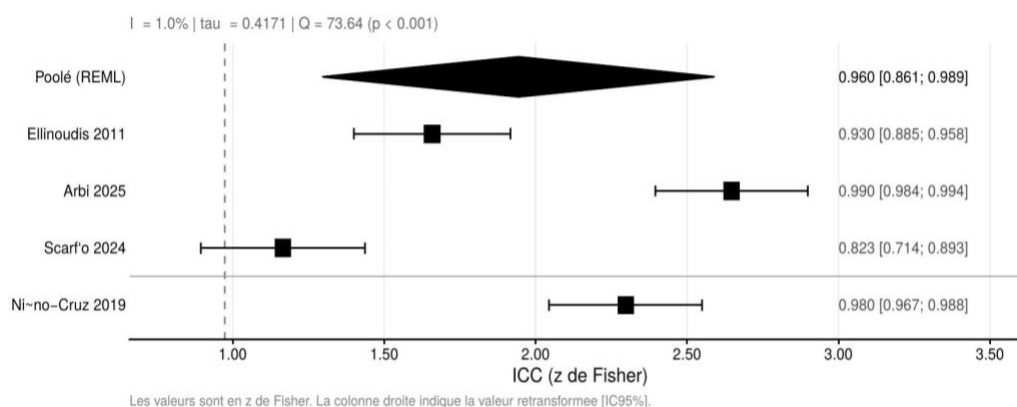


Fig. 10a : Forest Plot du MD1p : Posting coins, main préférée (k = 4).

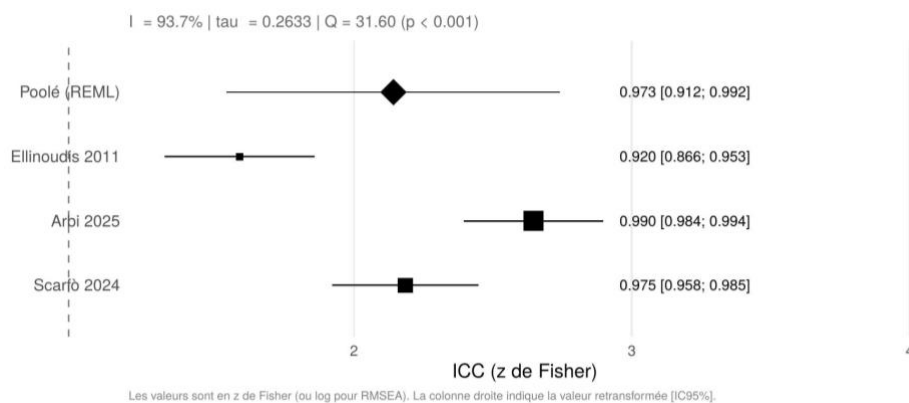


Fig. 10b : Forest plot du MD2 : Threading beads (k = 3).

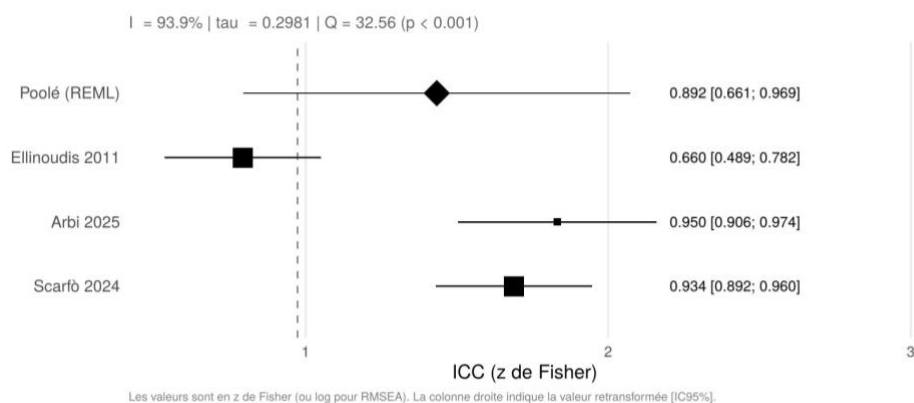


Fig. 10c : Forest plot du MD3 : Drawing Trail (k = 3).

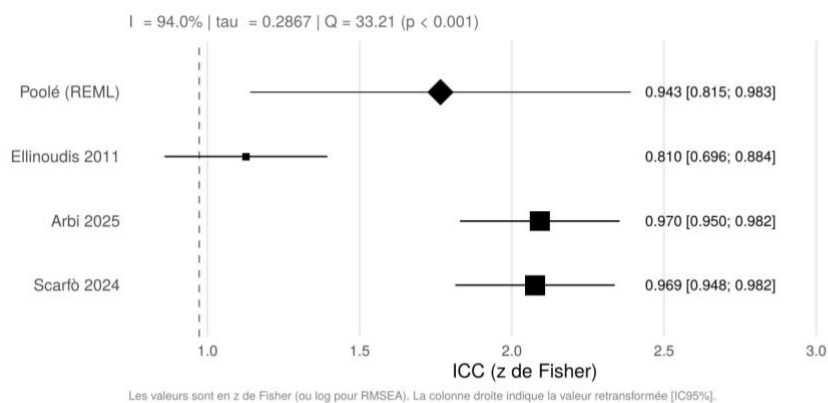


Fig.10d : Forest plot du CA1 : Catching beanbag (k = 3).

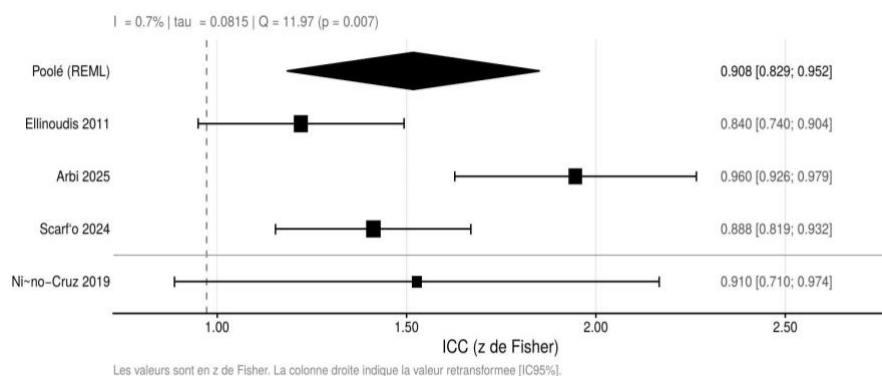


Fig. 10e : Forest Plot du CA2 : Aiming (k = 4).

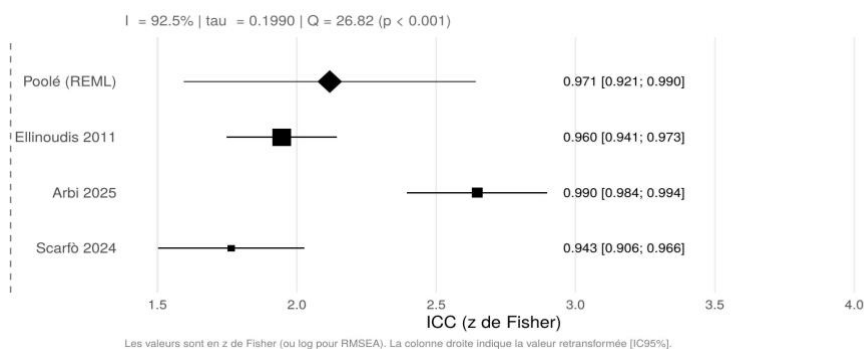


Fig. 10f : Forest plot du B2 : Walking heels raised (k = 3).

Tableau 8. Synthèse des six items du MABC-2 AB1 poolés (figures 10a-f)

Item	ICC brut (min-max)	ICC poolé [IC95 %]	τ^2	Q (p)	I²
MD1p (Posting coins)	0,823–0,990 (+ Niño-Cruz 0,980)	0,960 [0,861–0,989]	0,4171	73,64 (p < 0,001)	95,9 %
MD2 (Threading beads)	0,920–0,990	0,973 [0,912–0,992]	0,2639	32,91 (p < 0,001)	93,9 %
MD3 (Drawing Trail)	0,660–0,950	0,893 [0,661–0,969]	0,2987	34,12 (p < 0,001)	94,1 %
CA1 (Catching beanbag)	0,810–0,970	0,943 [0,815–0,983]	0,2875	34,94 (p < 0,001)	94,3 %
CA2 (Aiming)	0,840–0,960 (+ Niño-Cruz 0,910)	0,908 [0,829–0,952]	0,0815	11,97 (p = 0,007)	74,9 %
B2 (Walking heels raised)	0,943–0,990	0,971 [0,921–0,990]	0,1992	27,40 (p < 0,001)	92,7 %

Note : Les analyses ont été réalisées à l'aide d'un modèle à effets aléatoires REML après transformation en z de Fisher. Les résultats doivent être interprétés avec prudence en raison du faible nombre d'études et de l'hétérogénéité élevée.

6.6. Analyses de sous-groupes

Les analyses de sous-groupes prévues selon la tranche d'âge, la version de l'instrument et le contexte culturel n'ont pas été réalisées, la majorité des sous-groupes potentiels ne comprenant qu'une ou deux études. Ce nombre limité ne permettait pas d'obtenir des comparaisons quantitatives suffisamment informatives. Les différences observées entre les groupes sont donc décrites qualitativement dans la Discussion.

6.7. Analyses de sensibilité

Une analyse de sensibilité a été réalisée pour la fidélité inter-évaluateur du TGMD-3 en excluant Maeng et al. (2017), dont l'échantillon était limité à dix participants. L'estimation regroupée demeurerait identique, mais l'intervalle de confiance s'élargissait et la variance inter-études augmentait (voir section 6.1)

Une seconde analyse de sensibilité a été réalisée pour le coefficient alpha du TGMD-3 en excluant Valentini et al. (2017), seule étude utilisant une cotation de type *skills-to-test*. Les résultats de cette analyse sont présentés dans la section 6.3.

Une analyse de sensibilité exploratoire a également été réalisée pour les items MD1p et CA2 du MABC-2 AB1 en excluant Scarfò et al. (2024), dont le protocole de reproductibilité différait de celui des autres études. Après exclusion de Scarfò et al. (2024), l'ICC regroupé était de 0,976 [IC95 % 0,927–0,992] pour MD1p et de 0,916 [IC95 % 0,798–0,966] pour CA2.

Le chevauchement d'échantillon entre Garn et Webster (2021) et Webster et Ulrich (2017) n'a pas nécessité d'analyse de sensibilité supplémentaire. Ces deux articles reposaient sur la même base normative américaine, mais ils n'ont pas été inclus simultanément dans un même regroupement quantitatif portant sur une propriété identique. Aucun double comptage n'était présent dans les analyses concernées.

6.8. Biais de publication

Aucun des regroupements quantitatifs ne comportait au moins dix études. Le nombre maximal d'études était de cinq pour les analyses du TGMD-3 et de quatre pour celles du MABC-2 AB1. En conséquence, le biais de publication n'a pas été évalué à l'aide d'un funnel plot ou du test d'Egger, ces méthodes étant peu informatives lorsque le nombre d'études est limité. Cette absence d'évaluation formelle du biais de publication est prise en compte dans les limites méthodologiques de la revue.

DISCUSSION :

Cette scoping review enrichie de synthèses quantitatives ciblées avait pour objectif de cartographier les propriétés psychométriques des principaux instruments utilisés pour évaluer les manifestations fonctionnelles du contrôle et du développement moteur chez les enfants âgés de 1 à 12 ans au développement typique, et d’approfondir quantitativement certaines propriétés du TGMD et du MABC-2 lorsque les données disponibles étaient suffisamment comparables. Les résultats permettent notamment de discuter la reproductibilité des propriétés psychométriques entre différents contextes, la structure dimensionnelle des instruments et les implications de leur utilisation selon le contexte d’évaluation. Les analyses ont été réalisées sur les échelles transformées décrites dans la Méthodologie, tandis que les estimations présentées dans le texte et les tableaux sont rapportées après retransformation dans leur échelle originale.

1. Contraste entre les résultats de fidélité et de validité structurelle

L’un des principaux résultats de la synthèse quantitative est le contraste entre l’hétérogénéité très élevée observée pour les coefficients de fidélité du TGMD-3 et l’hétérogénéité plus faible observée pour les indices de validité structurelle. Ce contraste apparaît notamment à la comparaison des figures 2, 3 et 4.

La figure 2 illustre le pooling de l’ICC inter-évaluateur TGMD-3 ($k = 5$: Maeng et al. (2017), Magistro et al. (2020), Maïano et al. (2022), Marinšek et al. (2023) et Rintala et al. (2017)). L’ICC regroupé retransformé est de 0,960 [IC95 % 0,786–0,993], et l’ $I^2 = 99,8$ %. Malgré une estimation moyenne élevée, l’intervalle de confiance relativement large et surtout l’hétérogénéité extrême limitent fortement l’interprétation d’une valeur commune aux cinq études. Ce résultat suggère que la fidélité observée variait fortement selon les études, sans permettre d’identifier à partir de cette synthèse une cause unique à cette variabilité.

Un constat comparable est observé pour la fidélité test-retest (figures 9a et 9b) : l’ICC regroupé est de 0,98 [IC95 % 0,89–1,00] pour la sous-échelle Locomoteur ($I^2 = 99,2$ %) et de 0,99 [IC95 % 0,95–1,00] pour Ball Skills ($I^2 = 98,9$ %). Ces niveaux d’hétérogénéité indiquent que des coefficients moyens élevés peuvent masquer d’importantes variations entre études. Plusieurs facteurs pourraient y contribuer, notamment les caractéristiques des échantillons, les conditions de passation, la formation des évaluateurs et les méthodes statistiques utilisées. Les

données disponibles ne permettent toutefois pas d'estimer la contribution respective de chacun de ces facteurs.

Les figures 3 et 4 présentent les regroupements des indices CFI pour les deux modèles structurels du TGMD-3. Pour les indices de validité structurelle, quatre études portant sur le modèle à deux facteurs (figure 3) conduisaient à un CFI regroupé de 0,956 [IC95 % 0,934–0,970] avec un $I^2 = 49,9$ %, non significatif ($p = 0,112$). Pour le modèle bifacteur (figure 4), les trois études disponibles conduisaient à un CFI regroupé de 0,986 [IC95 % 0,978–0,991], avec I^2 estimé à 0 %. Les résultats disponibles montrent ainsi une convergence des valeurs de CFI rapportées dans ces études, en particulier pour le modèle bifacteur. Toutefois, le faible nombre d'études ne permet pas de conclure à une homogénéité générale de la structure factorielle du TGMD-3 entre les populations.

Ce contraste pourrait être lié à la nature différente des propriétés évaluées. Les coefficients de fidélité décrivent la stabilité ou l'accord des scores entre évaluateurs ou entre passations. Ils dépendent donc à la fois des caractéristiques de l'instrument et de celles de l'échantillon, des conditions de passation, de la formation des évaluateurs et du modèle statistique utilisé.

Les indices issus des analyses factorielles confirmatoires portent davantage sur la structure latente des scores que sur l'accord entre évaluateurs. Ils peuvent être moins directement influencés par certaines conditions locales de cotation, mais restent dépendants des caractéristiques de l'échantillon, de la taille de celui-ci, du modèle testé et de la méthode d'estimation. La moindre hétérogénéité observée pour certains indices CFA ne peut donc pas être interprétée comme la démonstration d'une propriété structurelle universelle du TGMD-3.

Les résultats disponibles montrent une convergence de certaines structures factorielles du TGMD-3 dans plusieurs contextes d'étude, sans permettre de conclure à leur invariance entre populations. En pratique, ces résultats invitent à distinguer la structure théorique de l'instrument de la fidélité effectivement obtenue dans un contexte donné : l'existence d'une structure factorielle similaire dans plusieurs populations ne garantit pas que les coefficients de fidélité observés dans une population puissent être directement transposés à une autre.

Une étude portant sur le TGMD-3, identifiée après la clôture du processus de sélection, apporte toutefois un élément externe de mise en perspective. Duncan et al. (2022), auprès de 1 608 enfants irlandais âgés de 5 à 11 ans, rapportaient un

ajustement insuffisant du modèle standard à deux facteurs et treize items ($CFI = 0,800$; $TLI = 0,756$; $RMSEA = 0,074$ [$0,069-0,080$]). Ces résultats étaient moins favorables que ceux observés dans les études incluses dans la synthèse quantitative du modèle à deux facteurs.

Cette étude n'a pas été intégrée au corpus ni aux analyses principales, puisqu'elle a été identifiée après leur clôture. Elle est donc utilisée uniquement comme élément externe de contextualisation et ne constitue pas une analyse de sensibilité.

Dans cette étude, les auteurs ont proposé une forme courte à sept items présentant de meilleurs indices d'ajustement ($CFI = 0,951$; $RMSEA = 0,050$ [$0,041-0,060$]). Cette version reposant sur une structure différente de celle du TGMD-3 complet, ses indices ne peuvent toutefois pas être directement comparés à ceux des modèles à treize items inclus dans la présente synthèse. Cette étude externe montre que les indices d'ajustement favorables observés dans les études incluses ne sont pas nécessairement retrouvés dans tous les échantillons. Elle renforce ainsi la prudence nécessaire dans l'interprétation de la convergence observée et souligne que les indices de validité structurelle restent dépendants des caractéristiques de l'échantillon et des choix de modélisation.

2. Structure dimensionnelle du TGMD-3 : convergence des résultats

Les indices disponibles pour le modèle bifacteur étaient particulièrement favorables, avec un $RMSEA$ regroupé de $0,042$ [$IC95 \% 0,035-0,051$] (figure 5). Cette estimation se situe dans la zone généralement interprétée comme indiquant un bon ajustement selon les repères proposés par Browne et Cudeck (1992), un $RMSEA$ inférieur à $0,05$ étant habituellement considéré comme favorable et des valeurs comprises entre $0,05$ et $0,08$ comme acceptables. Ces seuils restent néanmoins des repères d'interprétation et ne suffisent pas, à eux seuls, à établir la validité d'un modèle.

La figure 6 présente le regroupement de la cohérence interne du score total du TGMD-3 à partir de quatre études : Estevan et al. (2017), Marinšek et al. (2023), Salami et al. (2022) et Valentini et al. (2017). L' α regroupé est de $0,863$ [$IC95 \% 0,785 - 0,915$], valeur compatible avec les critères de suffisance habituellement retenus pour la cohérence interne. Cette estimation était toutefois associée à une hétérogénéité très élevée ($I^2 = 98,6 \%$).

Valentini et al. (2017) rapportaient la valeur la plus faible ($\alpha = 0,740$ [$IC95 \% 0,716-0,762$]) et se distinguaient des trois autres études, dont les coefficients étaient

compris entre 0,870 et 0,910. Une différence méthodologique pourrait contribuer à cet écart : Valentini et al. (2017) ont calculé l'alpha à partir de scores agrégés par habileté selon une approche « skills-to-test », tandis que les autres études utilisaient un niveau de cotation comprenant davantage d'unités. Comme l'alpha de Cronbach dépend notamment du nombre d'items et de leurs corrélations moyennes, cette différence dans l'unité de cotation pourrait contribuer à la valeur plus faible observée (Tavakol et Dennick, 2011). Elle ne permet toutefois pas d'attribuer à elle seule l'hétérogénéité observée.

L'unité d'analyse plus fine retenue par les trois autres études avec davantage d'items produirait ainsi un alpha plus élevé, sans nécessairement refléter une différence réelle de fidélité.

Malgré l'hétérogénéité statistique élevée observée sur certains indices regroupés (alpha, $I^2 = 98,6 \%$; r inter-facteurs, $I^2 = 89,4 \%$), plusieurs résultats convergent vers l'existence d'une dimension générale importante dans le score du TGMD-3. Ces résultats invitent à interpréter avec prudence la contribution spécifique des sous-scores Locomoteur et Ball Skills après prise en compte du facteur général. La figure 7 présente le résultat conceptuellement le plus fort de cette synthèse : la corrélation regroupée entre les facteurs Locomoteur et Ball Skills était élevée : $r = 0,871$ [IC95 % 0,827 – 0,904] (Estevan et al., 2017 ; Garn et Webster 2021 ; Marinšek et al., 2023, Valentini et al., 2017). Malgré une hétérogénéité élevée ($I^2 = 89,4 \%$), toutes les études rapportaient des corrélations élevées, comprises entre 0,820 chez Marinšek et al. (2023) et 0,910 chez Estevan et al. (2017). Garn et Webster (2021) rapportaient une valeur de 0,851 et Valentini et al. (2017) de 0,890. La valeur regroupée dépasse ainsi le seuil de 0,85 parfois utilisé pour questionner la validité discriminante entre deux facteurs fortement corrélés. Ce seuil doit toutefois être interprété comme un repère et non comme une frontière absolue permettant, à lui seul, de conclure à l'absence de distinction entre les deux dimensions.

Cette interprétation est également soutenue par plusieurs indicateurs issus des modèles bifactoriels. Garn et Webster (2021) rapportaient un Explained Common Variance (ECV) de 0,769 pour le facteur général ainsi que des coefficients omega-hiérarchiques spécifiques faibles pour les sous-échelles Locomoteur ($\omega_{HS} = 0,168$) et Ball Skills ($\omega_{HS} = 0,216$). Ces résultats indiquaient que, après prise en compte du facteur général, une part limitée de variance spécifique subsistait dans les deux

sous-scores. Salami et al. (2022) rapportaient également un ECV élevé pour le facteur général (0,840), ainsi que de faibles coefficients omega spécifiques pour les sous-échelles. Ces résultats renforcent l'intérêt du score global, tout en invitant à la prudence dans l'interprétation indépendante des sous-scores.

3. L'effet de l'âge et de la variance d'échantillon sur les indices de fidélité

L'hétérogénéité des coefficients de fidélité peut également être examinée à la lumière des caractéristiques des échantillons. La figure 8 présente les ICC individuels du TGMD et leurs intervalles de confiance, dans une représentation descriptive. Cette dispersion rappelle qu'un coefficient de fidélité ne dépend pas uniquement de l'instrument, mais également de la variabilité des performances dans la population étudiée et de la précision avec laquelle l'ICC est estimé.

L'ICC dépend du rapport entre la variance inter-individuelle et la variance d'erreur : à erreur de mesure comparable, un échantillon présentant une faible dispersion des performances peut conduire à un ICC plus faible qu'un échantillon plus hétérogène (Portney et Watkins, 2009). Cette propriété pourrait contribuer à certaines différences entre études. Rintala et al. (2017) ont évalué la fidélité sur trois sous-groupes de 20 enfants constitués selon le niveau de compétence, rapportant un ICC inter-évaluateur du score total de 0,62 [IC95 % 0,59–0,65]. À l'inverse, Marinšek et al. (2023), sur un échantillon de 452 enfants couvrant une large plage d'âge, rapportaient un ICC inter-évaluateur de 0,90 [IC95 % 0,88–0,92]. Cette comparaison ne permet pas d'attribuer la différence observée à la seule composition des échantillons, mais elle illustre l'influence potentielle de la dispersion des performances sur la valeur de l'ICC.

La taille de l'échantillon influence quant à elle principalement la précision de l'estimation. Magistro et al. (2020), avec un échantillon de 5 210 participants, rapportaient un ICC inter-évaluateur de 0,973 [IC95 % 0,969–0,977], associé à un intervalle de confiance très étroit. À l'inverse, Maeng et al. (2017), avec seulement dix participants, rapportaient un ICC également élevé (0,96) mais associé à un intervalle de confiance plus large [0,91–0,99]. Ces deux situations montrent qu'une valeur ponctuelle d'ICC doit être interprétée conjointement avec son intervalle de confiance et les caractéristiques de l'échantillon.

Ces constats renforcent l'importance de rapporter précisément la distribution des âges et des performances dans les échantillons de fidélité, ainsi que les intervalles de confiance associés aux ICC. Lorsque les effectifs le permettent, des analyses

stratifiées selon l'âge peuvent compléter l'analyse globale afin d'examiner d'éventuelles différences développementales. Les ICC devraient également être accompagnés d'indicateurs absolus d'erreur de mesure, tels que le SEM et le MDC, qui apportent une information complémentaire moins directement dépendante de la variance interindividuelle que l'ICC et sont particulièrement utiles pour l'interprétation clinique.

La tranche d'âge couverte par chaque instrument conditionne également son utilisation. Le TGMD-2 et le TGMD-3 ciblent principalement les enfants d'âge préscolaire et scolaire, période durant laquelle les performances motrices évoluent avec l'âge. Plusieurs études du corpus rapportaient ainsi des associations positives entre l'âge et les scores, notamment Farrokhi et al. (2014) et Estevan et al. (2017). Le MABC-2 adopte une organisation différente, avec trois bandes d'âge disposant de tâches adaptées au niveau développemental attendu. Cette organisation ne supprime toutefois pas tout risque d'effet plafond. Des effets plafond ont notamment été rapportés pour plusieurs items de la Bande d'âge 2 dans une population thaïlandaise, suggérant que la capacité discriminative de certaines tâches peut varier selon l'âge et le contexte de validation.

Le PDMS-2 couvre spécifiquement la période allant de la naissance à 72 mois et apparaît particulièrement pertinent pour la petite enfance et la période préscolaire précoce. La revue COSMIN de Zhu et al. (2024) rapporte des résultats globalement favorables pour plusieurs propriétés psychométriques, tout en soulignant des données encore limitées concernant notamment la validité transculturelle, l'erreur de mesure et la sensibilité au changement.

4. Traduction, adaptation culturelle et hétérogénéité

La qualité des procédures de traduction et d'adaptation culturelle constitue une source potentielle de variabilité entre les études. Une traduction imprécise des critères de cotation pourrait modifier leur interprétation par les évaluateurs et ainsi contribuer à des différences de fidélité inter-évaluateur. Cette hypothèse ne peut toutefois pas être testée directement à partir des données disponibles et doit être considérée comme une explication possible parmi d'autres.

Les études du corpus différaient dans le niveau de détail et la rigueur des procédures de traduction rapportées. Marinšek et al. (2023) ont décrit une procédure particulièrement structurée pour l'adaptation slovène du TGMD-3 : trois traducteurs bilingues, un expert indépendant en motricité, deux relecteurs

anglophones natifs et un coordinateur de projet ont participé aux différentes étapes de traduction, rétrotraduction et harmonisation. Les divergences ont ensuite été discutées collectivement avant l'établissement de la version finale. Cette procédure était également complétée par une évaluation de la validité de contenu et par une formation spécifique des évaluateurs.

Cette rigueur méthodologique pourrait avoir contribué à la qualité des coefficients de fidélité observés dans l'étude slovène, mais elle ne permet pas d'expliquer à elle seule leur niveau élevé, compte tenu notamment de la taille et de l'hétérogénéité de l'échantillon ainsi que des procédures de cotation. Il est donc préférable de considérer la traduction et l'adaptation culturelle comme l'un des facteurs susceptibles d'influencer les résultats plutôt que comme une cause démontrée de l'hétérogénéité observée.

Estevan et al. (2017) ont également accordé une attention particulière à l'adaptation de la version espagnole du TGMD-3. Leur estimation de fidélité inter-évaluateur était toutefois fondée sur un très faible sous-échantillon, ce qui limitait fortement sa précision et expliquait son exclusion du regroupement principal indépendamment de la qualité de la traduction.

Au-delà de la traduction linguistique, certaines tâches peuvent présenter une pertinence culturelle différente selon les populations. Dans le TGMD-3, certains gestes issus de pratiques sportives peuvent être plus ou moins familiers selon le contexte. Dans le MABC-2, plusieurs études du corpus ont également identifié des difficultés récurrentes pour certains items, notamment Drawing Trail, qui présentait une fidélité plus faible, des effets plafond ou une contribution structurelle limitée selon les populations étudiées. Verbecque et al. (2025) ont notamment retiré cet item de leur structure factorielle optimisée, tandis que Jaikaew et Satiansukpong (2019) ont rapporté des effets plafond sur plusieurs tâches de la bande d'âge 2.

Ces résultats montrent que la validation d'une version adaptée ne devrait pas se limiter à une traduction littérale. Elle devrait idéalement distinguer la traduction linguistique, l'adaptation culturelle des tâches et des critères de cotation, puis la vérification locale des propriétés psychométriques.

5. Variabilité de la reproductibilité des items du MABC-2 AB1 : au-delà de Drawing Trail

Parmi les six items, MD1p (Posting coins) présente l'indice I^2 le plus élevé du groupe (95,9 %) et une dispersion brute quasi identique à celle de CA1 (0,167 contre 0,160 point d'ICC). MD3 (Drawing Trail), pour sa part, reste l'item avec la plus grande dispersion brute entre études (écart de 0,29 point d'ICC entre Ellinoudis et al. (2011) et Arbi et al. (2025)) et la borne inférieure de l'IC95 % regroupé la plus basse du groupe (0,661). CA1 (Catching beanbag, écart de 0,16) présente une dispersion comparable à celle de MD1p. CA2 (Aiming) affiche paradoxalement l' I^2 le plus bas du groupe (74,9 %) malgré un écart brut de 0,12. Toutefois, la significativité statistique de son hétérogénéité reste confirmée ($p = 0,007$). À l'inverse, MD2 (Threading beads) et B2 (Walking heels raised) sont les deux items les plus stables, avec une dispersion brute limitée (0,05–0,07 point).

Arbi et al. (2025) rapportent pour MD3 une erreur standard de mesure exprimée en pourcentage (SEM%) de 34,97 % à 84,74 % selon le groupe d'âge, et un MDC% pouvant atteindre 141,79 % dans le groupe 3 à 4 ans. Une telle valeur indique qu'un changement très important par rapport au niveau initial serait nécessaire pour pouvoir être considéré comme supérieur à l'erreur de mesure avec le niveau de confiance retenu. Ces résultats suggèrent une faible précision de mesure de MD3 dans cet échantillon.

Plusieurs travaux ont questionné la contribution de MD3 à la cohérence ou à la structure du MABC-2 AB1, et certains modèles ont conduit à son retrait. Ce constat ne signifie toutefois pas que l'item soit dépourvu d'intérêt clinique : Drawing Trail sollicite des compétences graphomotrices fines potentiellement pertinentes pour l'identification de difficultés spécifiques. Son retrait pourrait améliorer certains indicateurs psychométriques sans nécessairement améliorer la couverture du construit évalué. Les résultats de la présente synthèse montrent par ailleurs que l'hétérogénéité ne se limite pas à MD3, puisque MD1p et CA1 présentent également une variabilité importante entre études.

L'intégration de Niño-Cruz et al. (2019) a été possible pour MD1p et CA2 uniquement, les quatre autres items présentant dans l'article source des intervalles de confiance dont les bornes devenaient identiques après arrondi, ce qui empêchait d'estimer de manière suffisamment fiable l'erreur standard nécessaire à la transformation en z de Fisher. Pour MD1p, l'ICC regroupé passait de

0,949 [IC95 % 0,749–0,991] à 0,960 [IC95 % 0,861–0,989], tandis que I^2 diminuait légèrement de 96,9 % à 95,9 %. Pour CA2, l'estimation regroupée restait à 0,908, avec une diminution de I^2 de 83,3 % à 74,9 %. L'ajout de l'étude colombienne ne modifiait donc pas substantiellement les estimations centrales, malgré un protocole de reproductibilité différent et un échantillon limité à 29 enfants.

Une analyse de sensibilité excluant Scarfò et al. (2024) conduisait à des estimations de 0,976 [IC95 % 0,927–0,992] pour MD1p et de 0,916 [IC95 % 0,798–0,966] pour CA2. Ces résultats montrent que les estimations centrales restaient élevées lorsque cette étude était retirée. Ils suggèrent également que les différences de protocole entre études contribuent à l'hétérogénéité observée, sans permettre d'attribuer celle-ci à une étude particulière de manière certaine.

6. Validité structurelle du MABC-2 : convergence et divergence entre bandes d'âge

Au-delà de la fidélité par item (figures 10a-f), deux études du corpus documentent la validité structurelle du MABC-2 par analyse factorielle confirmatoire, sans toutefois rapporter de données de fidélité exploitables. Psotta et Brom (2016), le plus grand échantillon MABC-2 AB1 du corpus ($N = 399$, 202 garçons et 197 filles, 3-6 ans), confirment une structure à trois facteurs spécifiques correspondant à la Dextérité manuelle, la Visée-réception, l'Équilibre plutôt qu'un modèle reposant sur un facteur général unique, et ce séparément chez les enfants de 3-4 ans et ceux de 5-6 ans. Ce résultat contraste avec celui de Schulz et al. (2011), qui rapportent un facteur général indifférencié pour l'AB1 dans leur échantillon. Les auteurs relèvent une saturation secondaire de l'item MD2 (Threading beads) sur le facteur Équilibre chez les enfants de 5-6 ans (charge = 0,22), qu'ils mettent en relation avec la composante bimanuelle de la tâche. Le modèle à trois facteurs chez les plus âgés présente un RMSEA de 0,048, et les corrélations inter-facteurs expliquent entre 20 et 49 % de variance commune selon les paires de facteurs chez les 5-6 ans (31 à 62 % chez les 3-4 ans).

Kita et al. (2016) apportent également des données en faveur de la validité factorielle du MABC-2 dans la bande d'âge 2, à partir d'un échantillon japonais de 132 enfants âgés de 7 à 10,8 ans. Les auteurs jugent l'ajustement du modèle satisfaisant selon leurs critères a priori, bien que les valeurs exactes du CFI et du RMSEA ne soient pas disponibles dans les données extraites de la présente revue. Cette étude documente en revanche des différences significatives entre les normes

japonaises et britanniques sur les domaines Dextérité manuelle et Équilibre, ainsi qu'un effet du sexe (les filles obtiennent de meilleurs scores en dextérité manuelle et en équilibre). Ces écarts suggèrent que l'application directe des normes britanniques à la population japonaise mérite d'être considérée avec prudence et renforcent l'intérêt de données normatives locales. Ni Kita et al. (2016) ni Psotta et Brom (2016) ne rapportent d'ICC permettant d'évaluer la fidélité dans ces mêmes échantillons.

7. La distinction entre sous-domaines moteurs : une fragilité transversale au corpus

Au-delà des propriétés spécifiques à chaque outil, plusieurs résultats suggèrent que la séparation entre sous-domaines moteurs n'est pas toujours empiriquement nette. Cette difficulté concerne des distinctions différentes selon les instruments : Locomoteur/Ball Skills pour le TGMD-3, motricité fine et motricité globale pour le PDMS-2 et certains sous-scores ou composites du BOT-2.

Pour le TGMD-3, la corrélation inter-facteurs regroupée entre Locomoteur et Ball Skills atteint $r = 0,871$ [IC95 % 0,827–0,904]. Cette valeur dépasse le seuil de 0,85 parfois utilisé comme critère de remise en question de la validité discriminante entre deux facteurs. Comme discuté précédemment, ce seuil ne constitue toutefois pas à lui seul une démonstration de l'absence de distinction entre les deux dimensions.

Le BOT-2 et son prédécesseur, le BOTMP, apportent également plusieurs éléments suggérant que certains sous-domaines ou scores composites ne présentent pas toujours une structure psychométrique aussi nette que celle postulée par l'organisation théorique du test. MacCobb et al. (2005) rapportaient des difficultés à retrouver une structure factorielle cohérente entre certaines composantes du BOTMP. Les analyses de Rasch de Brown (2019a) identifiaient des difficultés d'ajustement pour certains composites du BOT-2 complet, tandis que Bardid et al. (2019) retenaient 12 des 14 items du BOT-2 Short Form après recalibrage. Yoon et al. (2024) rapportaient également des différences de validité concurrente entre certains composites et le score global. Ces résultats ne portent pas tous sur la même propriété psychométrique, mais convergent vers la nécessité d'interpréter avec prudence certains sous-scores du BOT-2.

Le PDMS-2 fournit un exemple particulièrement net d'une difficulté de validité discriminante. Rebelo et al. (2021), dans un échantillon portugais de 392 enfants, rapportaient une corrélation très élevée entre les facteurs de motricité globale et de

motricité fine ($r^2 = 0,96$, soit $r \approx 0,98$). La variance partagée entre les deux facteurs dépassait les valeurs de variance moyenne extraite rapportées pour chacun d’eux, ce que les auteurs interprétaient comme un problème de validité discriminante.

Pris ensemble, ces résultats montrent que la question de la distinction entre sous-domaines ne se limite pas à un seul instrument. Leur hétérogénéité méthodologique ne permet toutefois pas de conclure à un mécanisme structurel commun aux différentes batteries. Ils invitent plutôt à vérifier empiriquement la pertinence des sous-scores avant de les interpréter comme des dimensions motrices indépendantes. Le MABC-2 nuance toutefois ce constat. Kokštejn et al. (2018) et Kita et al. (2016) rapportent, dans des populations indépendantes de Tchéquie et du Japon, de faibles corrélations entre les facteurs Dextérité manuelle et Visée-réception ($r = 0,01$ et $r = 0,15$ respectivement). Ces résultats apportent des éléments en faveur d’une distinction entre ces deux dimensions dans les populations étudiées. Ils ne permettent toutefois pas de conclure à une indépendance générale des trois domaines du MABC-2, d’autant que d’autres résultats du corpus mettent en évidence des variations structurelles selon l’âge et des fragilités au niveau de certains items (figures 10a-f).

8. Quel outil utiliser et à quel moment ?

Les résultats de cette scoping review permettent de proposer un cadre de sélection instrumentale fondé sur les propriétés psychométriques disponibles, les tranches d’âge couvertes, le domaine moteur ciblé et le contexte d’utilisation.

Tableau 9. Cadre de sélection instrumentale selon le contexte d'utilisation

Contexte	Âge	Instrument à considérer	Justification	Limites
Dépistage des habiletés motrices fondamentales et évaluation qualitative du mouvement	3–10 ans	TGMD-3	Indices d’ajustement favorables pour le modèle bifacteur dans les études disponibles ; CFI regroupé = 0,986 ; RMSEA regroupé = 0,042 ; associations positives entre l’âge et les scores rapportées dans plusieurs études	Fidélité fortement variable entre études ($I^2 = 99,8\%$) ; forte corrélation entre Locomoteur et Ball Skills ($r = 0,871$), invitant à la prudence dans l’interprétation indépendante des sous-scores
Évaluation normative des performances motrices et contribution à l’identification de difficultés	3–16 ans	MABC-2	Normes organisées par bandes d’âge ; système de classification par percentiles ; utilisation clinique largement documentée	Fidélité par item hétérogène ; plusieurs items, notamment MD1p et MD3, présentent une variabilité importante selon les études

Contexte	Âge	Instrument à considérer	Justification	Limites
compatibles avec un TDC				
Évaluation globale de la motricité fine et globale du nourrisson et de l'enfant préscolaire	0–6 ans	PDMS-2	Couverture dès la naissance ; six sous-échelles ; plusieurs propriétés psychométriques globalement favorables	Validité transculturelle encore insuffisamment documentée ; données plus limitées pour l'erreur de mesure et la sensibilité au changement
Évaluation de la coordination motrice globale en milieu scolaire	5–14 ans	KTK	Normes flamandes disponibles chez 2 470 enfants ; plusieurs versions abrégées ou adaptées étudiées	Normes historiques du KTK original fondées sur des données anciennes ; peu d'études du corpus rapportent des indices de fidélité directement recalculés avec IC95 %
Profil moteur détaillé en contexte clinique ou de recherche	4–21 ans	BOT-2	Large couverture d'âge et de domaines moteurs ; plusieurs études soutiennent la fidélité ou la validité de formes complètes ou abrégées	Certaines sous-échelles ou formes abrégées présentent des résultats structurels ou de cohérence interne moins favorables ; peu d'études avec ICC et IC95 %
Évaluation de terrain avec contraintes de temps ou de matériel	Selon l'outil	CAMS A / MOBA K / FUS	Administration généralement plus courte et matériel limité selon l'instrument	Propriétés psychométriques moins documentées dans le corpus ; objectifs et domaines évalués variables selon l'outil

Il est important de souligner que ce cadre ne constitue pas une hiérarchie de qualité entre les outils, mais un guide de sélection adapté au contexte. Le TGMD-3 et le MABC-2 répondent à des questions cliniques différentes : le premier évalue la qualité du processus moteur (comment l'enfant réalise le mouvement), le second situe la performance de l'enfant relativement à une norme populationnelle. Ces deux perspectives sont complémentaires et peuvent être combinées dans une évaluation clinique complète.

9. Limites et perspectives

Plusieurs limites méthodologiques doivent être prises en compte dans l'interprétation des résultats.

Premièrement, l'utilisation d'un bloc d'exclusion ciblant les populations atypiques a amélioré la spécificité de la stratégie de recherche, mais a pu conduire à l'exclusion de certaines études incluant un sous-groupe d'enfants au développement typique lorsqu'une population clinique était également mentionnée dans le titre ou le résumé. La vérification des listes de références des articles inclus a permis

d'identifier plusieurs publications supplémentaires, mais ne permet pas d'exclure totalement ce risque de perte de sensibilité.

Deuxièmement, le nombre d'études disponibles pour la majorité des regroupements était faible, généralement compris entre trois et quatre études, avec un maximum de cinq pour l'analyse de fidélité inter-évaluateur du TGMD-3. Dans ces conditions, le test Q de Cochran présente une faible puissance et l'estimation de l'hétérogénéité par I^2 est elle-même incertaine. Les valeurs de I^2 doivent donc être interprétées avec prudence, en particulier lorsqu'elles sont fondées sur seulement trois études.

Troisièmement, plusieurs synthèses quantitatives ont nécessité une reconstruction ou une approximation de la précision des estimations lorsque les études sources ne rapportaient pas directement les erreurs standards ou les intervalles de confiance nécessaires. Pour certains ICC, l'erreur standard sur l'échelle transformée a été obtenue par approximation selon la méthode delta. Cette procédure peut être moins précise lorsque le coefficient est très proche de 1. Une transformation directe des bornes de l'intervalle de confiance aurait alors constitué une approche plus précise. La synthèse des coefficients alpha doit également être considérée comme exploratoire. Leur transformation sur l'échelle de Fisher ne tient pas explicitement compte du nombre d'items composant les scores, alors que celui-ci influence directement la valeur de l'alpha.

Enfin, les synthèses du CFI et du RMSEA reposaient sur des approximations de leur précision qui ne tenaient pas compte de l'ensemble des caractéristiques du modèle d'équations structurelles, notamment les degrés de liberté, la méthode d'estimation et la complexité du modèle. Ces regroupements doivent donc être interprétés comme des analyses descriptives exploratoires de la convergence des indices plutôt que comme des méta-analyses structurelles formelles.

Quatrièmement, le pooling par item du MABC-2 (figures 10a-f) mêle des designs non strictement identiques : test-retest à une semaine (Arbi et al., 2025 ; Ellinoudis et al., 2011), et intra-évaluateur à 48 heures (Scarfò et al., 2024), et, pour MD1p et CA2 uniquement, inter-évaluateur vidéo (Niño-Cruz, 2019). La durée entre les mesures est susceptible d'influencer les coefficients de reproductibilité, des intervalles très courts limitant notamment la possibilité d'un changement réel entre les deux évaluations. Cette hétérogénéité de protocole pourrait contribuer aux valeurs élevées de I^2 observées, parallèlement aux différences de populations et de contextes culturels entre les études.

Cinquièmement, aucun regroupement ne comportait au moins dix études, seuil généralement recommandé pour permettre une interprétation suffisamment informative d'un funnel plot ou d'un test d'Egger. Le biais de publication n'a donc pas été évalué formellement.

Sixièmement, trois publications initialement identifiées pour le PDMS-2 n'ont pas pu être intégrées à l'analyse faute d'accès au texte intégral (Stokes, 1990 ; Gebhard, 1994 ; Darrah, 2007). Leur absence réduit notamment la couverture des travaux historiques consacrés à cet instrument.

Enfin, l'ensemble du processus de sélection, d'extraction et d'évaluation méthodologique a été réalisé par un seul évaluateur. L'absence de sélection et d'extraction indépendantes par un second évaluateur augmente le risque d'erreurs de sélection, d'extraction ou de cotation méthodologique. L'utilisation de formulaires standardisés et la vérification des données à partir des textes intégraux ont permis de structurer le processus, sans supprimer ce risque.

10. Perspectives

Plusieurs pistes de recherche se dégagent de cette synthèse. En priorité, la réalisation de validations locales systématiques dans des contextes culturels sous-représentés dans la littérature actuelle permettrait d'enrichir les bases de données nécessaires à des synthèses quantitatives plus robustes. La publication systématique des IC95 % des ICC, des IC90 % du RMSEA et des IC95 % de l'alpha de Cronbach dans les futures études de validation faciliterait les méta-analyses sans recours à l'estimation depuis N.

Sur le plan méthodologique, lorsque plusieurs études rapportent des matrices de corrélation comparables, une approche de type Two-Stage Meta-Analytic Structural Equation Modeling (TSSEM) permettrait de synthétiser directement les relations entre items ou dimensions avant d'estimer un modèle structurel commun. Cette approche serait plus appropriée que le regroupement d'indices d'ajustement isolés pour examiner la reproductibilité d'une structure factorielle entre études.

Enfin, la variabilité inter-études observée pour plusieurs items du MABC-2 AB1, notamment MD1p et MD3, justifie leur réévaluation dans de futurs travaux transculturels. Ces études pourraient examiner plus spécifiquement l'influence de l'âge, du contexte culturel, des expériences graphomotrices et des modalités de passation sur la difficulté et la reproductibilité de ces tâches, afin de déterminer si une adaptation des items ou de leurs normes améliore leurs propriétés de mesure.

CONCLUSION

Sur un corpus final de 74 publications, cette scoping review enrichie de synthèses quantitatives ciblées dresse un état des lieux des propriétés psychométriques des principaux instruments standardisés utilisés pour évaluer les manifestations fonctionnelles du contrôle et du développement moteur chez les enfants au développement typique. Le TGMD et le MABC-2 ont fait l'objet d'un approfondissement méthodologique et quantitatif, tandis que le PDMS-2, le BOT-2 et le KTK ont été examinés principalement dans le cadre de la synthèse qualitative. Ce travail s'inscrit dans la continuité des synthèses antérieures tout en actualisant les données disponibles. Griffiths et al. (2018) avaient réalisé une revue systématique des propriétés psychométriques de plusieurs instruments d'évaluation motrice pédiatrique, tandis que Rey et al. (2020) avaient spécifiquement synthétisé les données de fidélité des différentes versions du TGMD sans réaliser de méta-analyse. La présente revue complète ces travaux en intégrant des études plus récentes, notamment consacrées au TGMD-3, et en proposant, lorsque les données étaient jugées suffisamment comparables, des synthèses quantitatives ciblées permettant d'estimer des coefficients regroupés et de quantifier explicitement l'hétérogénéité entre études.

Le principal enseignement de cette revue est que des coefficients psychométriques globalement élevés ne garantissent pas à eux seuls la stabilité des propriétés d'un instrument entre études et contextes. Pour la fidélité inter-évaluateur du score total du TGMD-3, l'ICC regroupé atteignait 0,960 [IC95 % 0,786–0,993], mais l'hétérogénéité était extrême ($I^2 = 99,8 \%$), ce qui nuance l'interprétation globalement favorable de la fidélité rapportée dans les synthèses antérieures. Les caractéristiques des échantillons ainsi que les procédures de passation, de cotation, de formation des évaluateurs, de traduction et d'adaptation culturelle pourraient contribuer à ces différences, sans que leur influence puisse être directement établie à partir des données disponibles. À l'inverse, la validité structurelle du TGMD-3 présentait une convergence plus importante : le modèle bifactoriel rapportait un CFI regroupé de 0,986 et un RMSEA regroupé de 0,042, sur un faible nombre d'études et selon des procédures exploratoires. Cette convergence, associée à la corrélation élevée entre les facteurs Locomoteur et Ball Skills ($r = 0,871$ [IC95 % 0,827–0,904]) et aux valeurs d'Explained Common Variance rapportées (0,769 à 0,840),

soutient l'intérêt du score global du TGMD-3 et invite à interpréter avec prudence ses sous-scores de manière indépendante.

Pour le MABC-2, les synthèses par item de la bande d'âge 1 ont mis en évidence des estimations généralement élevées mais associées à une variabilité importante. Drawing Trail (MD3), déjà identifié comme présentant certaines fragilités psychométriques, montrait une dispersion importante des coefficients ; cette variabilité n'était toutefois pas limitée à cet item, Posting Coins (MD1p) présentant l'hétérogénéité statistique la plus élevée parmi les six items regroupés. Plus largement, plusieurs résultats du corpus montrent que les sous-domaines théoriquement distingués par les batteries motrices ne sont pas toujours nettement séparés sur le plan empirique, une observation qui concerne le TGMD-3, le PDMS-2 et certains composites du BOT-2/BOTMP, le MABC-2 présentant des résultats plus nuancés selon les études. Ces observations soulignent l'intérêt d'examiner la validité d'un instrument à plusieurs niveaux : score global, sous-domaines et items individuels.

Ces résultats doivent être replacés dans les limites méthodologiques de la revue : le nombre d'études disponibles pour chaque synthèse était généralement faible, plusieurs erreurs standards ont dû être estimées ou reconstruites en l'absence de données directement rapportées, et aucun regroupement n'atteignait le seuil recommandé pour évaluer le biais de publication. La sélection, l'extraction et l'évaluation méthodologique ont par ailleurs été réalisées par un seul évaluateur.

Sur le plan pratique, cette synthèse ne permet pas d'établir une hiérarchie univoque entre les instruments ; leur intérêt dépend de l'âge de l'enfant, du domaine évalué et de l'objectif de l'évaluation. Le TGMD-3 paraît particulièrement pertinent pour l'observation qualitative des habiletés motrices fondamentales chez les enfants de 3 à 10 ans. Le MABC-2 permet de situer les performances par rapport à des normes d'âge entre 3 et 16 ans et peut contribuer à l'identification de difficultés compatibles avec un trouble développemental de la coordination. Le PDMS-2 apparaît pertinent pour la petite enfance, le BOT-2 permet un profil moteur détaillé en contexte clinique ou de recherche, et le KTK constitue un outil de terrain centré sur la coordination globale en milieu scolaire. Ces instruments répondent à des objectifs différents et doivent être considérés comme complémentaires plutôt que concurrents.

Enfin, les futures études de validation devraient systématiquement rapporter les intervalles de confiance associés aux coefficients de fidélité et de cohérence interne, décrire précisément les caractéristiques des échantillons et les modèles statistiques employés, et documenter de manière standardisée les procédures de traduction et d'adaptation culturelle, accompagnées d'une vérification locale des propriétés psychométriques. Ces recommandations rejoignent les observations de Rey et al. (2020) sur la nécessité de protocoles standardisés de formation des évaluateurs pour l'utilisation du TGMD. Une meilleure standardisation des méthodes de validation et du rapportage des résultats permettrait de réaliser des synthèses quantitatives plus robustes et de soutenir une sélection raisonnée des instruments en fonction de la population, de l'objectif clinique et du contexte d'évaluation.

Bibliographie

- Aadland, K. N., Nilsen, A. K. O., Lervåg, A. O., & Aadland, E. (2022). Structural validity of a test battery for assessment of fundamental movement skills in Norwegian 3-6-year-old children. *Journal of sports sciences*, 40(15), 1688-1699. <https://doi.org/10.1080/02640414.2022.2100622>
- Adolph, K. E., & Robinson, S. R. (2015). Motor development. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (Vol. 2, pp. 113-157). <https://doi.org/10.1002/9781118963418.childpsy204>
- Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *WIREs Cognitive Science*, 8(1-2), e1430. <https://doi.org/10.1002/wcs.1430>
- Arbi, Z. E., Mnejja, K., Negra, Y., Albaret, J. M., Rebai, H., & Sahli, S. (2025). Movement Assessment Battery for Children—second edition (MABC-2) : aspects of reliability for Tunisian children (age band 1). *Sport Sciences for Health*, 21(2), 1197-1207. <https://doi.org/10.1007/s11332-025-01359-0>
- Assaiante, C. (1998). Development of locomotor balance control in healthy children. *Neurosci Biobehav Rev*, 22(4), 527-32. [https://doi.org/10.1016/s0149-7634\(97\)00040-7](https://doi.org/10.1016/s0149-7634(97)00040-7)
- Banátová, K., Valtr, L., & Cuberek, R. (2023). Cultural Adaptation of the MABC-2 Checklist as a Screening Tool for Developmental Coordination Disorder in Schools. *Occupational therapy in health care*, 37(4), 496-512. <https://doi.org/10.1080/07380577.2022.2056778>
- Bardid, F., Utesch, T., & Lenoir, M. (2019). Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form : An item response theory perspective. *Scandinavian journal of medicine & science in sports*, 29(12), 1980-1987. <https://doi.org/10.1111/sms.13527>
- Barnett, L. M., Minto, C., Lander, N., & Hardy, L. L. (2014). Interrater reliability assessment using the Test of Gross Motor Development-2. *Journal of science and medicine in sport*, 17(6), 667-70. <https://doi.org/10.1016/j.jsams.2013.09.013>
- Brown, T. (2019). Structural Validity of the Bruininks-Oseretsky Test of Motor Proficiency – Second Edition (BOT-2) Subscales and Composite Scales. *Journal of Occupational Therapy, Schools, & Early Intervention*, 12(3), 323-353. <https://doi.org/10.1080/19411243.2019.1590755>
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Campolettano, E. T., Gellner, R. A., & Rowson, S. (2018). Assessing static and dynamic postural control in a healthy population. *Biomedical Sciences Instrumentation*, 54(1), 24-31.
- Cattuzzo, M. T., de Santana, F. S., Safons, M. P., Ré, A. H. N., Nesbitt, D. R., Santos, A. B. D., Feitoza, A. H. P., & Stodden, D. F. (2020). Assessment in the Supine-To-Stand Task and Functional Health from Youth to Old Age : A Systematic Review. *International Journal of Environmental Research and Public Health*, 17(16), 5794.
- Chien, C.-W. (2007). *Using the Rasch model to validate the Peabody Developmental Motor Scales-second edition in infants and pre-school children*. Master's (Research) thesis, James Cook University. DOI : 10.25903/pmf0-7n42.
- Christopher, A., Kraft, E., Olenick, H., Kiesling, R., & Doty, A. (2021). The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan : a systematic review. *Disability and Rehabilitation*, 43(13), 1799-1813. <https://doi.org/10.1080/09638288.2019.1682066>
- Coppens, E., Laureys, F., Mostaert, M., D'Hondt, E., Deconinck, F. J. A., & Lenoir, M. (2021). Validation of a Motor Competence Assessment Tool for Children and Adolescents (KTK3+) With Normative Values for 6- to 19-Year-Olds. *Frontiers in Physiology, Volume 12 - 2021*. <https://doi.org/10.3389/fphys.2021.652952>
- Duncan, M. J., Martins, C., Ribeiro Bandeira, P. F., Issartel, J., Peers, C., Belton, S., O'Connor, N. E., & Behan, S. (2022). TGMD-3 short version : Evidence of validity and associations with sex in Irish children. *Journal of sports sciences*, 40(2), 138-145. <https://doi.org/10.1080/02640414.2021.1978161>
- Ellinoudis, T., Evaggelidou, C., Kourtessis, T., Konstantinidou, Z., Venetsanou, F., & Kambas, A. (2011). Reliability and validity of age band 1 of the Movement Assessment Battery for Children--second edition. *Research in developmental disabilities*, 32(3), 1046-51. <https://doi.org/10.1016/j.ridd.2011.01.035>
- Estevan, I., Molina-García, J., Queralt, A., Álvarez, O., Castillo, I., & Barnett, L. (2017). Validity and Reliability of the Spanish Version of the Test of Gross Motor Development-3. *Journal of Motor Learning and Development*, 5(1), 69-81. Retrieved Aug 7, 2026, from <https://doi.org/10.1123/jmld.2016-0045>
- Farrokhi, A., Zareh Zadeh, M., Karimi Alvar, L., Kazemnejad, A., & Ilbeigi, S. (2014). Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City. *Journal of Physical Education and Sport Management*, 5, 18-28. <https://doi.org/10.5897/JPEM12.003>
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, 11(1), 93-103. <https://doi.org/10.1177/014662168701100107>
- Garn, A. C., & Webster, E. K. (2021). Bifactor structure and model reliability of the Test of Gross Motor Development - 3rd edition. *Journal of science and medicine in sport*, 24(1), 67-73. <https://doi.org/10.1016/j.jsams.2020.08.009>

- Gavigan, N., Goss, H., & Belton, S. (2025). Inter-rater reliability of the FMS² assessment tool among generalist primary-school teachers. *Journal of science and medicine in sport*, 28(12), 1019–1025. <https://doi.org/10.1016/j.jsams.2025.07.008>
- Gharaei, E., Shojaei, M., & Daneshfar, A. (2019). The validity and reliability of the Bruininks–Oseretsky Test of Motor Proficiency, 2nd edition brief form, in preschool children. *Annals of Applied Sport Science*, 7(2), 3–12. <https://doi.org/10.29252/aassjournal.7.2.3>
- Gallahue, D. L., Goodway, J., & Ozmun, J. C. (2021). Understanding motor development : infants, children, adolescents, adults (Eighth edition.). Jones & Bartlett Learning.
- Griffiths, A., Toovey, R., Morgan, P. E., & Spittle, A. J. (2018). Psychometric properties of gross motor assessment tools for children : a systematic review. *BMJ Open*, 8(10), e021734. <https://doi.org/10.1136/bmjopen-2018-021734>
- Hallems, A., Klingels, K., Schepens, B., Thér , J., Schepens, K., & Verbecque, E. (2019). The Timed Up and Go Test in Children : Does Protocol Choice Matter? A Systematic Review. *Pediatric Physical Therapy*, 31(1), 22–31. <https://doi.org/10.1097/PEP.0000000000000558>
- Haywood, K., & Getchell, N. (2009). *Lifespan motor development* (5th ed.). Human Kinetics.
- Hoeboer, J., De Vries, S., Krijger-Hombergen, M., Wormhoudt, R., Drent, A., Krabben, K., & Savelsbergh, G. (2016). Validity of an Athletic Skills Track among 6- to 12-year-old children. *Journal of Sports Sciences*, 34(21), 2095–2105. <https://doi.org/10.1080/02640414.2016.1151920>
- Hoeboer, J., Krijger-Hombergen, M., Savelsbergh, G., & De Vries, S. (2018). Reliability and concurrent validity of a motor skill competence test among 4- to 12-year old children. *Journal of Sports Sciences*, 36(14), 1607–1613. <https://doi.org/10.1080/02640414.2017.1406296>
- Holm, I., Tveter, A. T., Aulie, V. S., & Stuge, B. (2013). High intra- and inter-rater chance variation of the movement assessment battery for children 2, ageband 2. *Research in developmental disabilities*, 34(2), 795–800. <https://doi.org/10.1016/j.ridd.2012.11.002>
- Hua, J., Gu, G., Meng, W., & Wu, Z. (2013). Age band 1 of the Movement Assessment Battery for Children-Second Edition : exploring its usefulness in mainland China. *Research in developmental disabilities*, 34(2), 801–8. <https://doi.org/10.1016/j.ridd.2012.10.012>
- Jaikaew, R., & Satiansukpong, N. (2019). Movement Assessment Battery for Children-Second Edition (MABC2) : Cross-Cultural Validity, Content Validity, and Interrater Reliability in Thai Children. *Occupational Therapy International*, 2019(1), 4086594. <https://doi.org/10.1155/2019/4086594>
- Kakebeeke, T. H., Knaier, E., Chaouch, A., Caflisch, J., Rousson, V., Largo, R. H., & Jenni, O. G. (2018). Neuromotor development in children. Part 4 : new norms from 3 to 18 years. *Developmental Medicine & Child Neurology*, 60(8), 810–819. <https://doi.org/10.1111/dmcn.13793>
- Kita, Y., Suzuki, K., Hirata, S., Sakihara, K., Inagaki, M., & Nakai, A. (2016). Applicability of the Movement Assessment Battery for Children-Second Edition to Japanese children : A study of the Age band 2. *Brain & development*, 38(8), 706–13. <https://doi.org/10.1016/j.braindev.2016.02.012>
- Lam, H. M. Y. (2011). Assessment of preschoolers' gross motor proficiency : revisiting Bruininks–Oseretsky Test of Motor Proficiency. *Early Child Development and Care*, 181(2), 189–201. <https://doi.org/10.1080/03004430.2011.536640>
- Latash, M. L. (2012). Fundamentals of Motor Control. Elsevier Inc. <https://doi.org/10.1016/C2011-0-05693-4>
- Levin, M. F., & Piscitelli, D. (2022). Motor Control : A Conceptual Framework for Rehabilitation. *Motor Control*, 26(4), 497–517. <https://doi.org/10.1123/mc.2022-0026>
- MacCobb, S., Greene, S., Nugent, K., & O'Mahony, P. (2005). Measurement and prediction of motor proficiency in children using Bayley infant scales and the Bruininks–Oseretsky test. *Physical & occupational therapy in pediatrics*, 25(1), 59–79.
- Maeng, H., Webster, E. K., Pitchford, E. A., & Ulrich, D. A. (2017). Inter- and Intrarater Reliabilities of the Test of Gross Motor Development-Third Edition Among Experienced TGMD-2 Raters. *Adapted physical activity quarterly : APAQ*, 34(4), 442–455. <https://doi.org/10.1123/apaq.2016-0026>
- Magistro, D., Piumatti, G., Carlevaro, F., Sherar, L. B., Esliger, D. W., Bardaglio, G., Magno, F., Zecca, M., & Musella, G. (2020). Psychometric properties of the Test of Gross Motor Development-Third Edition in a large sample of Italian children. *Journal of science and medicine in sport*, 23(9), 860–865. <https://doi.org/10.1016/j.jsams.2020.02.014>
- Makaruk, H., Porter, J. M., Webster, E. K., Makaruk, B., Bodasińska, A., Zieliński, J., Tomaszewski, P., Nogal, M., Szyszka, P., Starzak, M., Śliwa, M., Banaś, M., Biegajło, M., Chaliburda, A., Gierczuk, D., Suchecki, B., Molik, B., & Sadowski, J. (2023). The fus test : a promising tool for evaluating fundamental motor skills in children and adolescents. *BMC Public Health*, 23(1), 1912. <https://doi.org/10.1186/s12889-023-16843-w>
- Maïano, C., Morin, A. J. S., April, J., Webster, E. K., Hue, O., Dugas, C., & Ulrich, D. (2022). Psychometric Properties of a French-Canadian Version of the Test of Gross Motor Development – Third Edition (TGMD-3) : A Bifactor Structural Equation Modeling Approach. *Measurement in Physical Education and Exercise Science*, 26(1), 51–62. <https://doi.org/10.1080/1091367X.2021.1946541>
- Menescardi, C., Villarrasa-Sapiña, I., Lander, N., & Estevan, I. (2022). Canadian Agility Movement Skill Assessment (CAMSA) in a Spanish Context : Evidences of Reliability and Validity. *Measurement in Physical Education and Exercise Science*, 26(3), 245–255. <https://doi.org/10.1080/1091367X.2021.2020794>

- Messick, S. (1995). Validity of psychological assessment : Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, 27(5), 1171-1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Moreira, J. P. A., Lopes, M. C., Miranda-Júnior, M. V., Valentini, N. C., Lage, G. M., & Albuquerque, M. R. (2019). Körperkoordinationstest Für Kinder (KTK) for Brazilian Children and Adolescents : Factor Analysis, Invariance and Factor Score. *Frontiers in Psychology, Volume 10 - 2019*. <https://doi.org/10.3389/fpsyg.2019.02524>
- Newell, K. M. (1986). Motor Development in Children : Aspects of Coordination and Control. *Motor Development in Children : Aspects of Coordination and Control*. https://doi.org/10.1007/978-94-009-4460-2_19
- Niño-Cruz, G. I., Camargo-Lemos, D. M., Velásquez-Escobar, L. I., Rodríguez-Ortiz, J. K., & Patiño-Segura, M. S. (2019). Age band 1 of the movement assessment battery for children -2. Reliability of the spanish version. *Revista chilena de pediatría*, 90(5), 522-532. <https://doi.org/10.32641/rchped.v90i5.881>
- Novak, A. R., Bennett, K. J. M., Beavan, A., Pion, J., Spiteri, T., Fransen, J., & Lenoir, M. (2016). The applicability of a short form of the KörperKoordinationsTest für Kinder for measuring motor competence in children aged 6-11 years. *Journal of Motor Learning and Development*, 5(1), 1-15. <https://doi.org/10.1123/jmld.2016-0028>
- Okuda, P. M. M., Félix, E., Cogo-Moreira, H., Liu, T., Surkan, P. J., Martins, S. S., & Caetano, S. C. (2020). Construct validity of the Motor Development Scale (MDS). *Trends Psychiatry Psychother*, 42(1), 82-85. <https://doi.org/10.1590/2237-6089-2018-0114>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11 : Scoping Reviews. *JBIM Manual for Evidence Synthesis*. <https://doi.org/10.46658/JBIMES-20-12>
- Portney, L. G., & Watkins, M. P. (2009). Foundations of Clinical Research : Applications to Practice. 3rd edition.
- Psotta, R., & Brom, O. (2016). Factorial Structure of the Movement Assessment Battery for Children Test-Second Edition in Preschool Children. *Perceptual and motor skills*, 123(3), 702-716. <https://doi.org/10.1177/0031512516666072>
- Quitério, A., Martins, J., Onofre, M., Costa, J., Mota Rodrigues, J., Gerlach, E., Scheur, C., & Herrmann, C. (2018). MOBAC 1 Assessment in Primary Physical Education : Exploring Basic Motor Competences of Portuguese 6-Year-Olds. *Perceptual and Motor Skills*, 125(6), 1055-1069. <https://doi.org/10.1177/0031512518804358>
- Rebelo, M., Serrano, J., Duarte-Mendes, P., Monteiro, D., Paulo, R., & Marinho, D. A. (2021). Evaluation of the Psychometric Properties of the Portuguese Peabody Developmental Motor Scales-2 Edition : A Study with Children Aged 12 to 48 Months. *Children*, 8(11), 1049.
- Rey, E., Carballo-Fazanes, A., Varela-Casal, C., Abelairas-Gómez, C., & ALFA-MOV Project collaborators (2020). Reliability of the test of gross motor development: A systematic review. *PloS one*, 15(7), e0236070. <https://doi.org/10.1371/journal.pone.0236070>
- Rintala, P., Sääkslahti, A., & Iivonen, S. (2017). Reliability Assessment of Scores From Video-Recorded TGMD-3 Performances. *Journal of Motor Learning and Development*, 5, 1-17. <https://doi.org/10.1123/jmld.2016-0007>
- Rizkyanto, W. I., Gani, I., Iswanto, A., Yudhistira, D., & Shahril, M. I. (2024). Validity and reliability of the Gross Motor Development III Test for Indonesian children. *Fizjoterapia Polska*.
- Robinson, L. E., Stodden, D. F., Barnett, L. M., Lopes, V. P., Logan, S. W., Rodrigues, L. P., & D'Hondt, E. (2015). Motor competence and its effect on positive developmental trajectories of health. *Sports Medicine*, 45(9), 1273-1284.
- Rodrigues, L. P., Luz, C., Cordovil, R., Pombo, A., & Lopes, V. P. (2022). Motor Competence Assessment (MCA) Scoring Method. *Children (Basel)*, 9(11). <https://doi.org/10.3390/children9111769>
- Salami, S., Bandeira, P. F. R., Gomes, C. M. A., & Dehkordi, P. S. (2022). The Test of Gross Motor Development—Third Edition : A Bifactor Model, Dimensionality, and Measurement Invariance. *Journal of Motor Learning and Development*, 10(1), 116-131. <https://doi.org/10.1123/jmld.2020-0069>
- Salami, S., Ribeiro Bandeira, P. F., Martins, C., Hardy, L. L., Shams, A., & Dehkordi, P. S. (2023). The Körperkoordinations Test Für Kinder for Iranian Youth : Factor Structure, Measurement Invariance, and Covariates. *Journal of Motor Learning and Development*, 11(1), 165-187. <https://doi.org/10.1123/jmld.2021-0060>
- Scarfò, E., Ranucci, M., Berardi, A., Simeon, R., & Galeoto, G. (2024). Evaluation of the Psychometric Properties of the Movement Assessment Battery Second Edition (M-ABC2) : A Cross-Sectional Study. *Children*, 11(5), 555. <https://doi.org/10.3390/children11050555>
- Seamon, B. A., Sears, C. L., Anderson, E., & Velozo, C. A. (2025). Validating a Universal Measurement Scale and Standardized, Reproducible Unit for Gross Motor Development. *Archives of Physical Medicine and Rehabilitation*, 106(7), 1021-1027. <https://doi.org/10.1016/j.apmr.2024.12.008>

- Serbetar, I., Loftesnes, J. M., & Mamen, A. (2019). Reliability and Structural Validity of the Movement Assessment Battery for Children-2 in Croatian Preschool Children. *Sports*, 7(12), 248. <https://doi.org/10.3390/sports7120248>
- Shumway-Cook, A., & Woollacott, M. (2014). *Motor control : Translating research into clinical practice* (4th ed.). Wolters Kluwer Health.
- Soares, Â. A. A., Martins, C. M. D. L., Nobre, G. C., & Cattuzzo, M. T. (2020). Evidências de validade de construto, critério e fidedignidade da Motor Competence Assessment em pré-escolares. *Journal of Physical Education (Maringá)*, 31(1), e-3176. <https://doi.org/10.4025/jphyseduc.v31i1.3176>
- Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Robertson, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity : An Emergent Relationship. *Quest*, 60(2), 290-306. <https://doi.org/10.1080/00336297.2008.10483582>
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Thelen, E. (1995). Motor development. A new synthesis. *Am Psychol*, 50(2), 79-95. <https://doi.org/10.1037//0003-066x.50.2.79>
- Topor, M., Henriksen, A. H., Ahmt Rohde, R., Bugge, A., Bundgård Fals, E., Jacobsen, I., Andersen, K., & Wienecke, J. (2024). Developmental coordination disorder questionnaire – translation and adaptation into Danish. *Scandinavian Journal of Occupational Therapy*, 31(1), 2423698. <https://doi.org/10.1080/11038128.2024.2423698>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garrity, C., Lewin, S., Godfrey, C. M., Macdonald, M. T., Langlois, E. V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., & Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR) : Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467-473. <https://doi.org/10.7326/M18-0850>
- Valentini, N. C., Ramalho, M. H., & Oliveira, M. A. (2014). Movement assessment battery for children-2 : translation, reliability, and validity for Brazilian children. *Research in developmental disabilities*, 35(3), 733-40. <https://doi.org/10.1016/j.ridd.2013.10.028>
- Valentini, N. C., Zanella, L. W., & Webster, E. K. (2017). Test of Gross Motor Development—Third Edition : Establishing Content and Construct Validity for Brazilian Children. *Journal of Motor Learning and Development*, 5(1), 15-28. <https://doi.org/10.1123/jmld.2016-0002>
- Valentini, N. C., Duarte, M. G., Zanella, L. W., & Nobre, G. C. (2022). Test of Gross Motor Development-3 : Item Difficulty and Item Differential Functioning by Gender and Age with Rasch Analysis. *International journal of environmental research and public health*, 19(14). <https://doi.org/10.3390/ijerph19148667>
- Vandorpe, B., Vandendriessche, J., Lefevre, J., Pion, J., Vaeyens, R., Matthys, S., Philippaerts, R., & Lenoir, M. (2011). The KörperkoordinationsTest für Kinder : reference values and suitability for 6-12-year-old children in Flanders. *Scandinavian journal of medicine & science in sports*, 21(3), 378–388. <https://doi.org/10.1111/j.1600-0838.2009.01067.x>
- Wagner, M. O., Webster, E. K., & Ulrich, D. A. (2017). Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation) : Results of a Pilot Study. *Journal of Motor Learning and Development*, 5(1), 29-44. <https://doi.org/10.1123/jmld.2016-0006>
- Williams, E. N., Carroll, S. G., Reddihough, D. S., Phillips, B. A., & Galea, M. P. (2005). Investigation of the timed'Up & Go'test in children. *Developmental Medicine & Child Neurology*, 47(8), 518-524. <https://doi.org/10.1111/j.1469-8749.2005.tb01185.x>
- Yoon, D., Choi, D., Kim, M., Ji, S., Joung, Y. S., & Kim, E. Y. (2024). Validity of the BOT-2 Short Form for Korean School-Age Children : A Preliminary Study. *Children*, 11(6), 724.
- Zanella, L. W., Valentini, N. C., Copetti, F., & Nobre, G. C. (2021). Peabody Developmental Motor Scales - Second Edition (PDMS-2) : Reliability, content and construct validity evidence for Brazilian children. *Res Dev Disabil*, 111, 103871. <https://doi.org/10.1016/j.ridd.2021.103871>
- Zhu, Y., Hu, J., Ye, W., Korivi, M., & Qian, Y. (2024). Assessment of the measurement properties of the Peabody Developmental Motor Scales-2 by applying the COSMIN methodology. *Italian Journal of Pediatrics*, 50. <https://doi.org/10.1186/s13052-024-01645-6>
- Zuvela, F., Bozanic, A., & Miletic, D. (2011). POLYGON - A New Fundamental Movement Skills Test for 8 Year Old Children : Construction and Validation. *Journal of sports science & medicine*, 10(1), 157–163.

ANNEXES :

Annexe 1 : Liste des abréviations

Abréviation	Signification complète
AB1 ; AB2 ; AB3	Bande d'âge (bande d'âge) du MABC-2 : AB1 = 3 à 6 ans ; AB2 = 7 à 10 ans ; AB 3 = 11 à 16 ans
AST	Athletic Skills Track
B2	Walking Heels Raised — item d'équilibre du MABC-2 AB1
BOT-2	Bruininks–Oseretsky Test of Motor Proficiency – Second Edition
BOT-2-BF	Bruininks–Oseretsky Test of Motor Proficiency – Second Edition, Brief Form
BOTMP	Bruininks–Oseretsky Test of Motor Proficiency — version originale
CA1	Catching Beanbag — premier item « visée et réception » du MABC-2 AB1
CA2	Aiming — deuxième item « visée et réception » du MABC-2 AB1
CAMSA	Canadian Agility and Movement Skill Assessment
CFA	Confirmatory Factor Analysis — analyse factorielle confirmatoire
CFI	Comparative Fit Index — indice comparatif d'ajustement
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DCD	Developmental Coordination Disorder — trouble développemental de la coordination
DCDQ	Developmental Coordination Disorder Questionnaire
ECV	Explained Common Variance — variance commune expliquée
FMS	Fundamental Movement Skills — habiletés motrices fondamentales
FMS²	Fundamental Movement Skills Square assessment tool
FUS	Fundamental, Universal and Specific Skills Test
ICC	Intraclass Correlation Coefficient — coefficient de corrélation intra classe
IC ; IC90 % ; IC95 %	Intervalle de confiance ; intervalle de confiance à 90 % ; Intervalle de confiance à 95 %
KiMo	Kinder-Motoriktest — batterie d'évaluation motrice destinée aux enfants
KTK	Körperkoordinationstest für Kinder
KTK3 ; KTK3+	Version courte du KTK avec trois sous-tests ; Version du KTK3 avec une tâche supplémentaire
KTK4	Version complète du KTK comprenant quatre sous-tests
LDCDQ	Little Developmental Coordination Disorder Questionnaire
LDCDQ-BE	Version belge francophone du Little Developmental Coordination Disorder Questionnaire
MABC ; MABC-2 ; MABC-3	Movement Assessment Battery for Children ; Second Edition ; Third Edition
MCA	Motor Competence Assessment
MDC	Minimal Detectable Change — changement minimal détectable
MDC%	Changement minimal détectable exprimé en pourcentage
MD1p	Posting Coins – Preferred Hand — insertion de pièces avec la main préférée
MD2	Threading Beads — enfilage de perles
MD3	Drawing Trail — tracé graphique
MDS	Motor Development Scale
MOBAK	Motorische Basiskompetenzen — compétences motrices de base
N	Effectif total d'un échantillon
n	Effectif d'un groupe ou d'un sous-échantillon
PCC	Population, Concept, Contexte
PDMS ; PDMS-2	Peabody Developmental Motor Scales ; Peabody Developmental Motor Scales – Second Edition
PERF-FIT	Performance and Fitness Test Battery
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR	Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
Q	Statistique Q de Cochran utilisée pour tester l'hétérogénéité
r	Coefficient de corrélation
r²	Coefficient de détermination
RCT	Randomized Controlled Trial — essai contrôlé randomisé
REML	Restricted Maximum Likelihood — maximum de vraisemblance restreint

Abréviation	Signification complète
RMSEA	Root Mean Square Error of Approximation — racine de l'erreur quadratique moyenne d'approximation
SD	Standard Deviation — écart-type
SE	Standard Error — erreur standard
SEM	Standard Error of Measurement — erreur standard de mesure
SEM%	Erreur standard de mesure exprimée en pourcentage
SEz	Erreur standard calculée sur l'échelle transformée en z de Fisher
SMC-Kids	Simple Motor Competence-check for Kids
TGMD ; TGMD-2 ; TGMD-3	Test of Gross Motor Development ; Second Edition ; Third Edition
TGMD-3-SF	Test of Gross Motor Development – Third Edition, Short Form
TLI	Tucker–Lewis Index — indice de Tucker-Lewis
TSSEM	Two-Stage Meta-Analytic Structural Equation Modeling — modélisation par équations structurelles méta-analytique en deux étapes
TUG ; TUG-Child	Timed Up and Go ; Timed Up and Go — version pédiatrique
α	Alpha de Cronbach — indice de cohérence interne
ω	Oméga de McDonald — indice de cohérence interne
ω_H	Oméga hiérarchique
I^2	I-squared — proportion de la variabilité totale attribuable à l'hétérogénéité entre les études
τ^2	Tau-squared — estimation de la variance réelle entre les études
κ / kappa	Coefficient kappa d'accord
ρ / rho	Coefficient de corrélation de Spearman
k	Nombre d'études incluses dans une synthèse quantitative
p	Valeur de probabilité associée à un test statistique
$\bar{x}$	Moyenne arithmétique
z	Valeur obtenue après transformation en z de Fisher

Annexe 2 : Guidelines PRISMA-SCR

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	P. de garde
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	4 ^{ème} de couverture
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	1
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	4
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	N/A
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	7
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	6
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	6 – 7 (Annexe 3 P. 57)
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	8
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	9
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	10
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	11
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	12

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	15
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	18 : Annexe 4 p. 56 – 68
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	16 : Annexe 5-6 p. 69 – 72
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	18 – 32 : Annexe 4 – 6 : p. 56 – 72
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	19
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	33
Limitations	20	Discuss the limitations of the scoping review process.	44
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	47
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	N/A

Annexe 3 : Équations de recherche en fonction des bases de données

Équation de recherche complète

	(child*[Title/Abstract] OR toddler*[Title/Abstract] OR preschool*[Title/Abstract] OR "children"[MeSH Terms] OR "Child"[MeSH] OR "Preschool Child"[MeSH] OR "Infant"[MeSH]) AND ("motor control"[Title/Abstract] OR "motor skills"[Title/Abstract] OR "postural control"[Title/Abstract] OR "balance"[Title/Abstract] OR "coordination"[Title/Abstract] OR locomotion[Title/Abstract] OR "gross motor"[Title/Abstract] OR "fine motor"[Title/Abstract] OR "Motor Skills"[MeSH] OR "Postural Balance"[MeSH] OR "Locomotion"[MeSH] OR "Psychomotor Performance"[MeSH]) AND ("TGMD"[Title/Abstract] OR "Test of Gross Motor Development"[Title/Abstract] OR "MABC"[Title/Abstract] OR "Movement Assessment Battery for Children"[Title/Abstract] OR "Bruininks-Oseretsky"[Title/Abstract] OR "BOT-2"[Title/Abstract] OR "PDMS"[Title/Abstract] OR "Peabody Developmental Motor"[Title/Abstract] OR "KTK"[Title/Abstract] OR "Körperkoordinationstest"[Title/Abstract] OR "AIMS"[Title/Abstract] OR "Alberta Infant Motor Scale"[Title/Abstract]) AND (
PubMed	psychometric*[Title/Abstract] OR "test-retest"[Title/Abstract] OR "inter-rater reliability"[Title/Abstract] OR "intra-rater reliability"[Title/Abstract] OR "intraclass correlation"[Title/Abstract] OR "Cronbach alpha"[Title/Abstract] OR "measurement error"[Title/Abstract] OR "minimal detectable change"[Title/Abstract] OR "standard error of measurement"[Title/Abstract] OR "construct validity"[Title/Abstract] OR "criterion validity"[Title/Abstract] OR "floor effect"[Title/Abstract] OR "ceiling effect"[Title/Abstract] OR "Psychometrics"[MeSH] OR "Reproducibility of Results"[MeSH]) AND NOT ("autism"[Title/Abstract] OR "cerebral palsy"[Title/Abstract] OR "developmental coordination disorder"[Title/Abstract] OR "DCD"[Title/Abstract] OR "ADHD"[Title/Abstract] OR "intellectual disability"[Title/Abstract] OR "Down syndrome"[Title/Abstract] OR "neurological disorder"[Title/Abstract] OR "developmental delay"[Title/Abstract] OR "disability"[Title/Abstract] OR "impairment"[Title/Abstract]) AND ("humans"[Filter] AND ("english"[Filter] OR "french"[Filter]) AND ("child"[Filter] OR "preschoolchild"[Filter])).
Embase	('child'/exp OR child* OR toddler* OR preschool* OR "school age") AND ('motor control'/exp OR "postural control" OR locomotion OR coordination OR "gross motor" OR "fine motor" OR "motor skill") AND ('psychometry'/exp OR "psychometric properties" OR reliability OR validity OR sensitivity OR "measurement error" OR "minimal detectable change" OR norm*) AND (test* OR assessment* OR evaluation* OR scale* OR measure*)
Scopus	(TITLE-ABS-KEY(child* OR toddler* OR preschool* OR "school-age")) AND (TITLE-ABS-KEY("motor control" OR "postural control" OR locomotion OR coordination OR "gross motor" OR "fine motor" OR "motor skills")) AND (TITLE-ABS-KEY("psychometric properties" OR reliability OR validity OR "internal consistency" OR "test-retest" OR "inter-rater" OR sensitivity OR specificity OR "measurement error" OR "minimal detectable change" OR norms)) AND (TITLE-ABS-KEY(test OR assessment OR evaluation OR scale OR tool OR measure*))

Annexe 4 : Grille A

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
1 TGMD	Aye, T. (2017) Reliability of the Test of Gross Motor Development (2nd edition) chez l'enfant du Myanmar	Myanmar N : 50 (inter) ; 25 test-retest ; 12 intra Âge : M = 5.4 ans M=5.4 ; SD non rapportée TD	Test of Gross Motor Development Version : TGMD-2 (Ulrich 2000) Domaine : Locomoteur + Object control (GMQ global) Durée : 10-15	ICC inter-rater ; ICC intra-rater ; ICC test-retest ; alpha inter-rater ICC : Oui α/ω : Oui Validité concurrente : Non
2 TGMD	Rintala, P. (2017) Reliability assessment of scores from video-recorded TGMD-3 performances	Finlande N : 20 Âge : 3-9 ans Non rapportée (page 3-9 ans uniquement) TD	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2013/2019) Domaine : Locomotor skills + Ball skills + Total score Durée : non rapporté	ICC intra-rater ; ICC inter-rater ; % accord ; ICC par item ICC : Oui α/ω : Non Validité concurrente : Non
3 TGMD	Maïano, C. (2022) Psychometric Properties of a French-Canadian Version of the TGMD-3 : Bifactor SEM	Canada (Québec) N : 127 Âge : 5-11 ans Non rapportée (page 5-11 ans uniquement) TD	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2019) Domaine : Locomoteur + Ball skills (facteur G + facteurs spécifiques) Durée : non rapporté	ICC inter-rater ; omega ; validité structurelle bifactor-ESEM ; DIF ICC : Oui α/ω : Oui (ω) Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
4 TGMD	Valentini, N.C. (2021) Test of Gross Motor Development-3 Validity and Reliability : A Screening Form	Brésil N : 1192 (TGMD complet 772 ; SF 420 ; test-retest 114 ; critère 300) Âge : 3-10 ans (M=7.1, SD=2.0) M=7.1 ; SD=2.0 TD	Test of Gross Motor Development-3 Screening Form Version : TGMD-3-SF (6 items, version brésilienne) Domaine : Locomoteur (gallop, hop, skip) + Ball skills (one-hand strike, kick, overhand throw) Durée : Non rapporté explicitement (forme courte SF, 6 items ; durée probablement réduite par rapport au TGMD-3 complet [15-20 min], mais non chiffrée dans l'article)	EFA/CFA ; invariance ; alpha ; validité concurrente vs forme longue ; ICC inter/intra/test-retest ICC : Oui α/ω : Oui Validité concurrente : Oui (vs forme longue)
5 TGMD	Marinšek, M. (2023) Psychometric Properties of the Slovenian Version of the TGMD-3	Slovénie N : 452 (inter/validité) ; 24 intra ; 68 test-retest Âge : 3-10 ans (M=7.32, SD=2.16) M=7.32 ; SD=2.16 TD	Test of Gross Motor Development Version : TGMD-3-SI (Ulrich 2019) Domaine : Locomoteur (6 items) + Ball skills (7 items) Durée : 15-20 (valeur standard du manuel TGMD-2/TGMD-3 ; non rapportée spécifiquement dans cette étude)	CVI ; ICC inter/intra/test-retest ; CFA ; invariance sexe ; alpha ICC : Oui α/ω : Oui Validité concurrente : Non
6 TGMD	Estevan, I. (2017) Validity and Reliability of the Spanish Version of the TGMD-3	Espagne N : 178 (fiabilité inter/intra sur 4 enfants) Âge : 3-11 ans (M=6.94, SD=1.89) M=6.94 ; SD=1.89 TD	Test of Gross Motor Development Version : TGMD-3 (version espagnole) Domaine : Locomoteur (6 items) + Ball skills (7 items) Durée : 15-20 (valeur standard du manuel TGMD-2/TGMD-3 ; non rapportée spécifiquement dans cette étude)	Alpha ; ICC inter/intra-rater ; CFA ; validité de construit liée à l'âge ; SEM ICC : Oui α/ω : Oui Validité concurrente : Non
7 TGMD	Lopes, V.P. (2016) Reliability and construct validity of the TGMD-2 in Portuguese children	Portugal N : 330 (164 filles) ; test-retest 22 Âge : 5-10 ans (M=7.9, SD=1.3) M=7.9 ; SD=1.3 TD	Test of Gross Motor Development Version : TGMD-2 (version portugaise, Ulrich 2000) Domaine : Locomoteur (6 items) + Object control (6 items) Durée : 15-20 (valeur standard du manuel TGMD-2/TGMD-3 ; non rapportée spécifiquement dans cette étude)	Alpha ; test-retest Bland-Altman ; inter-rater Kappa ; CFA ICC : Non (Kappa/ratio, pas ICC) α/ω : Oui Validité concurrente : Non
8 TGMD	Webster, E.K. (2017) Evaluation of the Psychometric Properties of the Test of Gross Motor Development - Third Edition	USA N : 807 ; test-retest 30 Âge : 3-10.9 ans (M=6.33, SD=2.09) M=6.33 ; SD=2.09 TD	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2016/2019) Domaine : Locomoteur (6 items) + Ball skills (7 items) Durée : 15-20 (valeur standard du manuel TGMD-2/TGMD-3 ; non rapportée spécifiquement dans cette étude)	Alpha ; ICC test-retest ; accord inter-évaluateur % ; EFA/CFA ; difficulté/discrimination items ICC : Oui (test-retest) α/ω : Oui Validité concurrente : Non
9 TGMD	Wagner, M.O. (2017) Psychometric Properties of the TGMD-3 (German Translation) - Results of a Pilot-Study	Allemagne N : 189 (inter/intra 30 ; test-retest 104) Âge : 3.17-10.67 ans (M=7.15, SD=2.02) M=7.15 ; SD=Non rapportée TD	Test of Gross Motor Development Version : TGMD-3 allemand, 12 items (stationary strike exclu) Domaine : Locomoteur (6 items) + Ball skills (6 items) Durée : 15-20 (12 items, item stationary strike exclu ; valeur standard manuel, non rapportée spécifiquement dans l'étude)	ICC test-retest ; ICC inter/intra ; alpha ; CFA ; invariance ; validité divergente/concurrente/prédictive ICC : Oui α/ω : Oui Validité concurrente : Oui (partielle)
10 TGMD	Garn, A.C. (2021) Bifactor structure and model reliability of the TGMD-3	USA N : 862 Âge : 3-10 ans (M=6.51, SD=2.23) M=6.51 ; SD=Non rapportée Mixte : base normative avec 11% handicap documenté	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2019) Domaine : FMS général + Locomoteur + Ball skills Durée : non rapporté	CFA/ESEM bifactor ; omega hiérarchique ; ECV ; RPB ; alpha ICC : Non α/ω : Oui (alpha/ ω) Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
11 TGMD	Temple, V.A. (2016) A Peek at the Developmental Validity of the TGMD-3	Canada N : 277 (129 garçons, 148 filles) Âge : 8-9 ans Non rapportée (plage 8-9 ans uniquement) Mixte : TD + 8 enfants avec trouble	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2013) Domaine : Locomoteur (6 items) + Ball skills (7 items) Durée : non rapporté	Validité développementale ; changement grade 3-4 ; POMP ; différences de sexe ICC : Non α/ω : Non Validité concurrente : Non
12 TGMD	Rizkyanto, W.I. (2024) Validity and reliability of the Gross Motor Development III Test for Indonesian children	Indonésie N : 290 (180 garçons, 110 filles) Âge : 11-13 ans (déborde cible 1-12) Non rapportée (plage 11-13 ans uniquement) TD	Gross Motor Development III Test Version : TGMD-3 / TGMD III (Ulrich 2013/2019) Domaine : Locomoteur + Ball skills (13 items) Durée : non rapporté	Validité contenu Aiken V ; validité empirique Pearson ; alpha Cronbach ICC : Non α/ω : Oui (alpha .99 aberrant) Validité concurrente : Non
13 TGMD	Valentini, N.C. (2017) Test of Gross Motor Development-Third Edition : Establishing content and construct validity for Brazilian children	Brésil N : 597 (295 G / 302 F) ; retest 128 ; intra 100 ; inter 50 Âge : 3-10 ans (G M=6.76 ; F M=6.58) M=G=6.76 ; F=6.58 ; SD=G=2.11 ; F=2.06 TD	Test of Gross Motor Development Version : TGMD-3-BR (Ulrich 2014, version brésilienne) Domaine : Locomoteur + Ball skills (2 facteurs) Durée : non rapporté	Validité contenu (CVI, kappa) ; alpha inter/intra ; test-retest Pearson ; alpha ; AFC ICC : Non (alpha/Pearson, pas ICC exploitable) α/ω : Oui Validité concurrente : Non
14 TGMD	Salami, S. (2022) The Test of Gross Motor Development-Third Edition : A Bifactor Model, Dimensionality, and Measurement Invariance	Iran N : 496 (53.8% filles) ; inter 47 ; intra 80 Âge : 3-10.9 ans (M=7.23, SD=2.03) M=7.23 ; SD=2.03 TD	Test of Gross Motor Development Version : TGMD-3 (Ulrich 2019) Domaine : Locomoteur + Ball skills (unifactoriel / 2 facteurs / bifactor) Durée : non rapporté	Validité structurelle ; alpha ; omega ; invariance sexe ; fidélité inter/intra non chiffrée ICC : Non (valeurs exactes non communiquées) α/ω : Oui Validité concurrente : Non
15 TGMD	Soares, I.A. (Aires Soares) (2020) Evidence of construct validity, criteria and reliability of the Motor Competence Assessment battery in preschoolers	Brésil N : 251 (validité) ; 53 (test-retest, pour la MCA) Âge : 3-5 ans (M=4.53, SD=0.77) M=4.53 ; SD=0.77 TD	Test of Gross Motor Development (utilisé comme outil CRITÈRE pour valider la Motor Competence Assessment) Version : TGMD-2 (Ulrich 2000) Domaine : Locomotion (6 habiletés) + Contrôle d'objet (5 habiletés) Durée : non rapporté	ICC intra-évaluateur (vidéo) ; validité critère (corrélation MCA vs TGMD-2) ICC : Oui (vidéo, intra-évaluateur) α/ω : Non Validité concurrente : Oui (TGMD-2 = critère pour la MCA)
16 TGMD	Barnett, L.M., Minto, C., Lander, N. & Hardy, L.L. (2014) Interrater reliability assessment using the Test of Gross Motor Development-2	Australie N : 37 Âge : 4-8 ans (M=6.2, DS=0.8) M=6.2 ; SD=0.8 TD	Test of Gross Motor Development Version : TGMD-2 (Ulrich, 2000) Domaine : Contrôle d'objet UNIQUEMENT (6 habiletés : dribble, strike, overhand throw, underhand roll, kick, catch) — pas de données locomotrices Durée : Non rapportée (6 visites sur 2 mois pour la collecte complète)	ICC inter-évaluateur (sous-test + 6 habiletés) ; kappa par composante de skill (24 composantes) ICC : Oui α/ω : Non Validité concurrente : Non
17 TGMD	Farrokhi, A., Zareh Zadeh, M., Karimi Alvar, L., Kazemnejad, A. & Ilbeigi, S. (2014) Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City	Iran N : 1438 (normatif) ; 63 (test-retest) ; 32 (intra-rater) Âge : 3 :0-10 :11 ans Non rapporté (normatif global ; tranches par âge fournies, pas de moyenne unique) TD	Test of Gross Motor Development Version : TGMD-2 (Ulrich, 2000) Domaine : Locomoteur + Object control (GMQ global) Durée : 15-20	Alpha (cohérence interne) ; ICC test-retest ; ICC intra-rater ; AFC (validité de construit) ; corrélation âge ; ANOVA âge×sexe ICC : Oui α/ω : Oui Validité concurrente : Oui (AFC + corrélation âge)

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
18 TGMD	Maeng, H., Webster, E.K., Pitchford, E.A., Ulrich, D.A. (2017) Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition Among Experienced TGMD-2 Raters	USA (Midwest) N : 10 (enfants) ; 5 évaluateurs Âge : 3 ans 7 mois – 10 ans 9 mois (M=6,57 ; SD=2,51) M=6.57 ; SD=2.51 TD	Test of Gross Motor Development Version : TGMD-3 (Ulrich, in press) Domaine : Locomoteur + Ball Skills Durée : 10-15 min/enfant (scoring vidéo)	ICC inter-rater ; ICC intra-rater ICC : Oui α/ω : Non Validité concurrente : Non
19 TGMD	Magistro, D., Piumatti, G., Carlevaro, F., Sherar, L.B., Eslinger, D.W., Bardaglio, G., Magno, F., Zecca, M., Musella, G. (2020) Psychometric properties of the Test of Gross Motor Development—Third Edition in a large sample of Italian children	Italie (Nord-Ouest, Turin) N : 5210 (analyse finale) / 5956 recrutés ; sous-échantillon test-retest n=450 Âge : 3-11 ans (M=8,38-8,40 ; SD≈1,97-1,98) M=8.38 ; SD=1.98 TD (10 exclus pour trouble physique/cognitif sévère)	Test of Gross Motor Development Version : TGMD-3 Domaine : Locomoteur + Ball Skills Durée : ~20 min/enfant	Validité structurelle (EFA/CFA/invariance) ; ICC inter-rater ; ICC test-retest ; kappa intra-rater ; IRT (GRM) ICC : Oui α/ω : Non (kappa seulement pour intra-rater) Validité concurrente : Non
20 MABC-2	Brown, T. (2009) The MABC-2 : a review and critique	Australie N : non applicable (revue critique narrative) Âge : 3–16 ans (plage complète de l'outil ; revue critique, pas d'échantillon propre) SD non applicable (revue critique,) TD + clinique	MABC-2 Version : MABC-2 Domaine : Dextérité + équilibre + visée Durée : 20–40	Reliability; validity; factor analysis ICC : Oui α/ω : Oui Validité concurrente : Non
21 MABC-2	Smits-Engelsman, B.C.M. (2011) Is the MABC-2 a reliable instrument for 3-year-olds?	Pays-Bas N : 50 Âge : 36-48 mois (3,0-4,0 ans) M=G=41 mois F=41,5 mois ; SD=Non rapporté TD	MABC-2 Version : MABC-2 AB1 Domaine : Dextérité + équilibre + visée Durée : 20–40	Test-retest reliability ICC : Oui α/ω : Non Validité concurrente : Non
22 MABC-2	Kokštein, J. (2018) Construct validity of the MABC-2 in preschool children	Tchéquie N : 510 Âge : 3-6 ans (M=4,9 ; DS=1,1) M=4,9 ans ; SD=1,1 TD	MABC-2 Version : MABC-2 AB1 Domaine : Dextérité + équilibre + visée Durée : 20–40	Construct validity; CFA ICC : Non α/ω : Non Validité concurrente : Non
23 MABC-2	Jaikaew, R. (2019) Cross-cultural validity, content validity, and interrater reliability in Thai children	Thaïlande N : 29 (fidélité ; 14G/15F) ; 5 enfants test préfinal ; 5 OT contenu Âge : 7-10 ans (M=9,08 ; DS=0,49) M=9,08 ans ; SD=0,49 TD	MABC-2 Version : MABC-2 AB2 Domaine : Dextérité + équilibre + visée Durée : 20–30	Validité transculturelle (protocole Beaton) ; validité de contenu (IOC) ; fidélité inter-évaluateur (ICC/item, scoring simultané) ICC : Oui (0.71–1.00) α/ω : Non Validité concurrente : Non
24 MABC-2	Jaikaew, R. (2021) Movement performance and movement difficulties in typical school-aged children	Thaïlande N : 360 recrutés / 335 retenus (171G/164F) après exclusion des valeurs aberrantes Âge : 7–10 ans M=Non rapporté ; SD=Non rapp TD	MABC-2 Version : MABC-2 AB2 Domaine : Dextérité + équilibre + visée Durée : 20–40	Validité de critère (cite ICC/CVI de Jaikaew 2019, pas mesuré ici) ; effet plafond ; données normatives de prévalence ICC : Non α/ω : Non Validité concurrente : Non
25 MABC-2	Banátová, K. (2022) Cross-cultural validation of the MABC-2 Checklist	Tchéquie N : Non rapporté Âge : 7–11 ans Moyenne d'âge et SD non rapportées TD	MABC-2 Checklist Version : MABC-2 Checklist Domaine : Questionnaire enseignant Durée : non applicable (questionnaire enseignant, pas de passation chronométrée)	Cross-cultural validity; reliability ICC : Oui α/ω : Oui Validité concurrente : Non
26 MABC-2	Verbecque, E. (2025) Structural validity of the MABC-2 : AB2 in European and African children	Europe/Afrique N : 3025 Âge : 7-10 ans M=Non rapporté ; SD=Non rapporté TD	MABC-2 Version : MABC-2 AB2 Domaine : Dextérité + équilibre + visée Durée : 20–40	Structural validity; CFA ICC : Non α/ω : Non Validité concurrente : Non
27 MABC-2	Wu, O.K. (2025) Test-Retest Reliability of the MABC-3 with Australian 3–6-Year-Olds	Australie N : 24 Âge : 3–6 ans M=4,33 ans ; SD=1,12 TD	MABC-3 Version : MABC-3 AB1 Domaine : Dextérité + équilibre + visée Durée : 20–40	Test-retest reliability; convergent validity ICC : Oui α/ω : Non Validité concurrente : Oui

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
28 MABC-2	Arbi, Z.E. (2025) MABC-2 : aspects of reliability for Tunisian children (AB1)	Tunisie N : 62 Âge : 3-6 ans (groupe 3-4 ans n=33, M=3,6±0,5 ; groupe 5-6 ans n=29, M=5,2±0,4) M=4,3 ans (total AB1) ; SD=0,9 (total AB1) ; sous-groupes : 3-4 ans SD=0,5 (n=33), 5-6 ans SD=0,4 (n=29) TD	MABC-2 Version : MABC-2 AB1 Domaine : Dextérité + équilibre + visée Durée : 20-40	Reliability; validity; adaptation ICC : Oui α/ω : Oui Validité concurrente : Non
29 MABC-2	Valentini, N.C., Ramalho, M.H. & Oliveira, M.A. (2014) Movement Assessment Battery for Children-2 : Translation, reliability, and validity for Brazilian children	Brésil N : 844 (440G/404F) ; sous-éch. test-retest n=168 ; longitudinal n=41 ; concourante n=43 Âge : 3-13 ans (M=8,31 ; DS=2,91) M=8,31 ans ; SD=2,91 TD (avec classification post-hoc DCD/à risque/TD)	Movement Assessment Battery for Children-2 Version : MABC-2 (Henderson, Sugden & Barnett, 2007) Domaine : Dextérité + équilibre + visée Durée : ~30 (moyenne rapportée, tous âges)	Validité de contenu (CVI, kappa) ; validité de façade ; fidélité inter-évaluateur (ICC) ; fidélité intra-évaluateur (ICC) ; fidélité test-retest (ICC, Pearson r) ; cohérence interne (alpha) ; validité discriminante ; validité prédictive ; validité concourante (vs TGMD-2) ICC : Oui (mais sans IC95% rapportés) α/ω : Oui Validité concurrente : Oui (vs TGMD-2, r=.30)
30 MABC-2	Serbetar, I., Loftesnes, J.M. & Mamen, A. (2019) Reliability and Structural Validity of the Movement Assessment Battery for Children-2 in Croatian Preschool Children	Croatie N : 683 (366G/317F) ; n=650 items complets ; test-retest analysé n=173 ; inter-évaluateur n=36 ; AFC n=606 Âge : 3-6 ans M=Réparti par année (3/4/5/6 ans) — pas de moyenne globale rapportée ; SD=Non rapporté TD (classification post-hoc feu tricolore : 9% impairment, 8% à risque, 83% TD)	Movement Assessment Battery for Children-2 Version : MABC-2 AB1 (Henderson, Sugden & Barnett, 2007) Domaine : Dextérité + équilibre + visée Durée : Non rapporté	Fidélité test-retest (ICC, SEM, SDC) ; fidélité inter-évaluateur (ICC, SEM, SDC) ; validité structurelle (AFC, modèle à 3 facteurs) ICC : Oui (IC95% complets, test-retest ET inter-évaluateur) α/ω : Non Validité concurrente : Non
31 MABC-2	Scarfò, E., Ranucci, M., Berardi, A., Simeon, R. & Galeoto, G. (2024) Evaluation of the Psychometric Properties of the Movement Assessment Battery Second Edition (M-ABC2) : A Cross-Sectional Study	Italie N : 108 (69G/39F) ; AB1 n=71, AB2 n=17, AB3 n=20 Âge : 3-16,11 ans (AB1/AB2/AB3) Non rapportée globalement (AB1 n=71 ; AB2 n=17 ; AB3 n=20) TD	Movement Assessment Battery for Children, Second Edition Version : MABC-2 (Henderson, Sugden & Barnett, 2007) Domaine : Dextérité + équilibre + visée Durée : 20-40	Fidélité test-retest intra-évaluateur (ICC/item/bande) ; fidélité test-retest inter-évaluateur (ICC/item/bande) ; normes italiennes (scores standards + percentiles) ICC : Oui (IC95% pour la quasi-totalité des items) α/ω : Non (non rapporté — limite reconnue par les auteurs) Validité concurrente : Non
32 MABC-2	Ellinoudis, T., Evaggelinou, C., Kourtessis, T., Konstantinidou, Z., Venetsanou, F. & Kambas, A. (2011) Reliability and validity of age band 1 of the Movement Assessment Battery for Children – Second Edition	Grèce N : 183 (sous-éch. test-retest n=60 ; AFC n=182) Âge : 3-5 ans (M=50 mois, DS=9) M=50 mois ; SD=9 mois TD	Movement Assessment Battery for Children – Second Edition Version : MABC-2 AB1 Domaine : Dextérité + équilibre + visée Durée : Non rapporté	Fidélité test-retest (ICC) ; cohérence interne (alpha) ; validité structurelle (AFC) ; corrélations item-total et composante-total ICC : Oui (IC95% complets) α/ω : Oui Validité concurrente : Non
33 MABC-2	Holm, I., Tveter, A.T., Aulie, V.S. & Stuge, B. (2013) High intra- and inter-rater chance variation of the movement assessment battery for children 2, ageband 2	Norvège N : 45 Âge : 7-9 ans (M=8.7, DS=0.7) M=8.7 ; SD=Non rapporté TD	Movement Assessment Battery for Children Version : MABC-2 AB2 (Henderson, Sugden & Barnett, 2007) Domaine : Dextérité manuelle + Visée-réception + Équilibre (8 items, ageband 2) Durée : 20-30	ICC intra-testeur (1-2 sem.) ; ICC inter-testeur (même jour) ; SEM ; SDC90% — calculés sur scores BRUTS ICC : Oui α/ω : Non Validité concurrente : Non
34 MABC-2	Niño-Cruz, G.I., Camargo-Lemos, D.M., Velásquez-Escobar, L.I., Rodríguez-Ortiz, J.K., Patiño-Segura, M.S. (2019) Age band 1 of the Movement Assessment Battery for Children-2. Reliability of the Spanish version	Colombie (Bucaramanga) N : 29 préscolaires (échantillonnage de convenance) Âge : 3-6 ans M=4,5 ans ; SD=0,9 TD	Movement Assessment Battery for Children Version : MABC-2 AB1 Domaine : Dextérité manuelle + Visée-réception + Équilibre Durée : Non rapporté	Fidélité vidéo inter-évaluateur (3 évaluateurs) ET intra-évaluateur (2 mesures) ; cohérence interne ; accord Bland-Altman ICC : Oui α/ω : Oui (sans IC95%) Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
35 MABC-2	Kita, Y., Suzuki, K., Hirata, S., Sakihara, K., Inagaki, M., Nakai, A. (2016) Applicability of the Movement Assessment Battery for Children-Second Edition to Japanese children : A study of the Age Band 2	Japon N : 132 (84 garçons, 48 filles) Âge : 7-10,8 ans (M=8,8 ; SD=1,2) M=8,8 ans ; SD=1,2 TD	Movement Assessment Battery for Children Version : MABC-2 AB2 Domaine : Dextérité manuelle + Visée-réception + Équilibre Durée : Non rapporté	Étude transversale : cohérence interne + validité factorielle (CFA/bootstrap Bollen-Stine) ; AUCUN design de fidélité ICC : Non α/ω : Oui (sans IC95%) Validité concurrente : Non
36 MABC-2	Psotta, R., Brom, O. (2016) Factorial Structure of the Movement Assessment Battery for Children Test—Second Edition in Preschool Children	Tchéquie N : 399 (202 garçons, 197 filles) Âge : 3 ans 0 mois à 6 ans 11 mois SD non rapportée TD	Movement Assessment Battery for Children Version : MABC-2 AB1 Domaine : Dextérité manuelle + Visée-réception + Équilibre Durée : Non rapporté	Étude transversale : validité structurelle uniquement (AFC, modèle 3 facteurs spécifiques vs facteur général), analysée séparément jeunes (3-4 ans) et grands (5-6 ans) préscolaires ; AUCUN design de fidélité ICC : Non α/ω : Non Validité concurrente : Non
37 PDMS-2	Tabatabainia, M.M. (1995) Construct validity of the Bruininks-Oseretsky Test of Motor Proficiency and the Peabody Developmental Motor Scales	Australie N : 40 (24G-16F) Âge : 4,5-5,5ans SD non rapportée (page 4,5-5,5 ans uniquement) TD	PDMS + BOTMP Version : PDMS original Domaine : Motricité globale + fine Durée : 90min	Construct validity; factor analysis ICC : Non α/ω : Non Validité concurrente : Non
38 PDMS-2	Tieman, B.L. (2005) Assessment of motor development in preschool children	USA N : non applicable (revue narrative, pas d'échantillon propre) Âge : 0–6 ans non applicable (revue narrative, aucun échantillon propre) TD	PDMS-2 Version : PDMS-2 Domaine : Motricité globale + fine Durée : 45–60	revue narrative ICC : Non α/ω : Non Validité concurrente : Non
39 PDMS-2	Sun, S.H. (2010) Development of the Preschooler Gross Motor Quality Scale vs PDMS	Taiwan N : 174 Âge : 3 - 6 ans par groupe d'âge (mois) : 3 ans SD=3,42 ; 4 ans SD=3,94 ; 5 ans SD=3,84 ; 6 ans SD=2,53 (pas de SD globale rapportée) TD	PGMQ et PDMS comparateur Version : PDMS-2 Domaine : Motricité globale Durée : non rapporté	Criterion validity vs PDMS ICC : Non α/ω : Non Validité concurrente : Oui
40 PDMS-2	Zanella, L.W. (2021) PDMS-2 : Reliability, content and construct validity — Brazilian children	Brésil N : 637 Âge : 0–71 mois (naissance à 5;11 ans) SD=18,6 mois TD	PDMS-2 Version : PDMS-2 Domaine : Motricité globale + fine Durée : 40–60	Reliability; validity COSMIN ICC : Oui α/ω : Oui Validité concurrente : Non
41 PDMS-2	Valentini, N.C. & Zanella, L.W. (2022) Peabody Developmental Motor Scales-2 : The Use of Rasch Analysis to Examine the Model Unidimensionality, Motor Function, and Item Difficulty	Brésil N : 637 (même cohorte que Zanella et al. 2021, ligne ci-dessus) Âge : 0–71 mois (M=21,7 ; SD=18,6) M=21,7 mois ; SD=18,6 TD (mêmes critères d'exclusion que Zanella et al. 2021 : troubles musculo-squelettiques, syndromes génétiques, malformations congénitales)	PDMS-2 — 6 sous-échelles (Reflexes, Stationary, Locomotor, Object Manipulation, Grasping, Visual-Motor Integration) Version : PDMS-2 (Folio & Fewell, 2000) Domaine : Motricité globale + fine (6 sous-échelles) Durée : 45–60	Analyse de Rasch (modèle de crédit partiel) : unidimensionnalité, fidélité de séparation items/personnes, hiérarchie de difficulté des items, infit/outfit, corrélations point-bisériales ICC : Oui (ICC intra->0,97 ; inter->0,92 — mêmes données que Zanella et al. 2021) α/ω : Non (fidélité de séparation Rasch rapportée, pas alpha/omega) Validité concurrente : Non
42 PDMS-2	Rebelo, M. (2021) Psychometric properties of the Portuguese PDMS-2 (12–48 months)	Portugal N : 392 Âge : 12–48 mois SD=8,79 mois TD	PDMS-2 Version : PDMS-2 Domaine : Motricité globale + fine Durée : 45–60	Psychometric properties; adaptation ICC : Oui α/ω : Oui Validité concurrente : Non
43 PDMS-2	Zhu, Y. (2024) Assessment of the measurement properties of PDMS-2 — COSMIN methodology	Chine N : 22 études Âge : 0–6 ans non applicable (revue systématique COSMIN, 22 études, aucun échantillon propre) TD	PDMS-2 Version : PDMS-2 Domaine : Motricité globale + fine Durée : 45–60	COSMIN review complet (grade A) ICC : Oui α/ω : Oui Validité concurrente : Oui

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
44 PDMS-2	Seamon, B.A. (2025) Validating a Universal Measurement Scale and Standardized, Reproducible Unit for Gross Motor Development	Brésil + Taïwan (données sources réanalysées) ; étude menée aux USA (Medical University of South Carolina) N : 4445 (réanalyse secondaire ; cumul de 3 échantillons : 637 + 424 + 3389) Âge : 0–77 mois (cumul des échantillons sources) non applicable (réanalyse combinant 3 échantillons hétérogènes : Zanella M=21,7/SD=18,6 ; Chien TD M=37,6/SD=21,2, handicap M=44,2/SD=16,5 ; Drachler et al. sans statistiques résumées) — pas de SD globale unique Majorité TD (637 Valentini & Zanella 2022 + 342/424 Chien 2007) + 655 avec développement atypique (retard dév. 49%, troubles sensorimoteurs 20%, PC 13%, autisme 7%, T21 5%, retard cognitif 3%, spina bifida 1,3% — % au sein du sous-groupe Chien)	GMU (nouvelle unité trans-instruments) ; PDMS-2 (sous-échelles Locomotor + Stationary) et Denver II réutilisés comme sources secondaires de calibration Version : PDMS-2 (Folio & Fewell, 2000) — calibrations réutilisées, pas de nouvelle administration Domaine : Motricité globale uniquement (locomotion + stationnaire ; pas de motricité fine) Durée : non applicable (réanalyse secondaire de calibrations Rasch déjà publiées ; aucune nouvelle passation)	Corrélations (Pearson/Spearman/désatténuées) entre difficulté d'item observée et difficulté théorique en GMU — PAS une étude de fidélité/validité COSMIN classique ICC : Non α/ω : Non Validité concurrente : Non — analyse de calibration Rasch, pas validité concurrente au sens classique
45 KTK	Hoeboer, J. (2016) Validity of an Athletic Skills Track among 6- to 12-year-old children	Pays-Bas N : 463 (252 garçons/211 filles, M=9±2 ans) — étude de validation principale ; étude pilote n=54 (28 garçons/26 filles), même école d'Amsterdam, KTK également administré Âge : 6–12 ans SD=2 (valeur arrondie rapportée, M=9±2 ans) TD	AST (Athletic Skills Track, AST-1 et AST-2) — KTK utilisé uniquement comme mesure de référence pour la validité concurrente Version : KTK protocole néerlandophone (Lenoir et al., 2014 ; Vandorpe et al., 2011), 4 sous-tests Domaine : Équilibre dynamique (marche arrière) ; déplacement latéral (plaques) ; saut unipodal en hauteur (pile de coussins) ; saut latéral (latte) Durée : ~25 min par enfant (explicitement rapporté : "It took ± 25 min to complete the KTK per child")	Aucune propriété psychométrique propre au KTK ; utilisé comme mesure de référence pour la validité concurrente de l'AST-1 et l'AST-2 ($r=-0,474$ à $-0,767$ selon l'âge et le track, sauf AST-2 chez les 7 ans $r=-0,290$) ; étude pilote (n=54) : $r=-0,645$, $p<0,001$ ICC : Non α/ω : Non Validité concurrente : Oui
46 KTK	Hoeboer, J. (2018) Reliability and concurrent validity of a motor skill competence test among 4- to 12-year old children	Pays-Bas N : 213 (104F/109G) — sous-étude validité concurrente AST-KTK ; étude 1 (fiabilité test-retest AST, n=717) n'implique pas le KTK Âge : 4–12 ans SD non rapportée (plage 4-12 ans uniquement, pour les deux sous-études) TD	AST (Athletic Skills Track) — KTK utilisé uniquement comme critère externe de validité concurrente Version : KTK protocole néerlandophone (Lenoir et al., 2014), 4 sous-tests Domaine : Locomotion latérale ; saut unipodal en hauteur ; équilibre dynamique Durée : Non rapporté pour le KTK dans cette étude	Aucune propriété psychométrique propre au KTK ; corrélations AST-KTK rapportées (validité concurrente, $r = -0,602$ à $-0,747$) ICC : Non α/ω : Non Validité concurrente : Oui
47 KTK	Moreira, J.P.A. (2019) Körperkoordinations test Für Kinder (KTK) for Brazilian Children and Adolescents : Factor Analysis, Invariance and Factor Score	Brésil N : 565 (282 garçons/283 filles, 49,9% garçons) Âge : 5–10 ans (M=7,93 ; DS=1,51) M=7,93 ans ; SD=1,51 TD	KTK Version : KTK original (4 sous-tests : WB, JS, MS, HH) Domaine : Équilibre dynamique (WB) ; locomotion latérale (JS, MS) ; saut unipodal en hauteur (HH) Durée : Non rapporté	Cohérence interne (alpha) ; validité structurelle (AFC, échantillon total/sexe/âge) ; invariance de mesure (sexe, âge) — NON confirmée au-delà du niveau métrique ; comparaison de 4 méthodes de calcul du score factoriel ICC : Non α/ω : Oui Validité concurrente : Non
48 KTK	Coppens, E. (2021) Validation of a Motor Competence Assessment Tool for Children and Adolescents (KTK3+) With Normative Values for 6- to 19-Year-Olds	Belgique N : 2271 (1112 garçons/1159 filles) ; sous-échantillon validation n=1248 (580G/668F) Âge : 6–19 ans SD non rapportée (normes par tranche d'âge annuelle 6-19 ans, pas de moyenne globale identifiée) TD	KTK3+ Version : KTK3+ (KTK3 : JS, MS, BB + tâche de coordination œil-main EHC, remplaçant le hopping for height) Domaine : Locomotion latérale (JS, MS) ; équilibre dynamique (BB) ; contrôle d'objet (EHC) Durée : Non rapporté individuellement (~45 min pour un groupe de ±25 enfants, anthropométrie + KTK3+)	Multicolinéarité (VIF) ; validité structurelle (MDS non métrique) ; corrélations inter-items ; validité de construit (MANOVA sexe×âge×sport) ; normes (percentiles, quotients moteurs) ; fiabilité test-retest citée (Faber 2014/Platvoet 2018, non recalculée) ICC : Non α/ω : Non Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
49 KTK	Salami, S. (2023) The Körperkoordinations Test Für Kinder for Iranian Youth : Factor Structure, Measurement Invariance, and Covariates	Iran N : 432 (262 garçons/170 filles, 61% garçons) Âge : 5–14 ans SD non rapportée (plage 5-14 ans rapportée) TD	KTK Version : KTK original (4 sous-tests : WB, HH, JS, MS) Domaine : Équilibre dynamique (WB) ; saut unipodal en hauteur (HH) ; locomotion latérale (JS, MS) Durée : 15–20 min par enfant	Cohérence interne ($\alpha=,86$) ; validité structurelle (AFC unifactorielle) ; invariance de mesure (sexe, âge) — partielle confirmée ; modèle MIMIC (covariables sexe/âge/IMC) ; fidélité inter/intra-évaluateur ($\kappa > 80\%$) ICC : Non α/ω : Oui Validité concurrente : Non
50 KTK	Vandorpe, B., Vandendriessche, J., Lefevre, J., Pion, J., Vaeyens, R., Matthys, S., Philippaerts, R. & Lenoir, M. (2011) The Körperkoordinations Test für Kinder : reference values and suitability for 6–12-year-old children in Flanders	Belgique (Flandre + Région de Bruxelles-Capitale) N : 2470 Âge : 6-12 ans SD non rapportée (plage 6-12 ans rapportée) TD (26 écoles primaires d'enseignement général)	Körperkoordinationstest für Kinder Version : KTK original, 4 sous-tests (Kiphard & Schilling, 1974/2007) Domaine : Équilibre dynamique (WB) ; locomotion latérale (MS) ; saut unipodal en hauteur (HH) ; saut latéral (JS) Durée : ~15	Valeurs de référence (normes âge/sexe, MQ) pour la population flamande anno 2008 ; comparaison scores bruts et MQ vs échantillon allemand de 1974 (test t à un échantillon) ; différences d'âge/sexe (ANOVA 2x6) ; distribution des scores et seuils de classification (chi carré) — pas de nouvelle mesure de fidélité/validité sur cet échantillon ICC : Non (pas d'ICC recalculé sur cet échantillon ; rappel des coefficients de fidélité test-retest originaux du manuel allemand de Kiphard & Schilling 1974/2007 : total=0,97 ; WB=0,80 ; MS=0,84 ; HH=0,96 ; JS=0,95 — non mesurés ici) α/ω : Non Validité concurrente : Non
51 KTK	Novak, A.R., Bennett, K.J.M., Beavan, A., Pion, J., Spiteri, T., Fransen, J. & Lenoir, M. (2016/2017) The Applicability of a Short Form of the KörperKoordinations Test für Kinder for Measuring Motor Competence in Children Aged 6-11 Years	Belgique (Flandre + Région de Bruxelles-Capitale) N : 2479 (1300 garçons/1179 filles) Âge : 6-11 ans SD non rapportée TD (26 écoles primaires d'enseignement général)	Körperkoordinationstest für Kinder — comparaison KTK3 (forme courte) vs KTK4 (forme complète) Version : KTK4 (complet, 4 sous-tests) vs KTK3 (forme courte, 3 sous-tests, HH exclu) Domaine : Équilibre dynamique (WB) ; locomotion latérale (MS) ; saut latéral (JS) ; saut unipodal en hauteur (HH, exclu dans KTK3) Durée : Non chiffré globalement (HH≈10 min/enfant ; WB+MS+JS≈10 min combinés, rapporté explicitement)	Corrélation de Pearson entre scores sommés (MQ) KTK3 et KTK4 ($r=0,97$ total, par âge et sexe) ; accord de classification par catégories de percentiles (χ^2 , Kappa de Cohen) — étude de concordance inter-formes du même outil, pas une étude de fidélité/validité classique vs un critère externe ICC : Non (corrélations de Pearson et Kappa de Cohen utilisés, pas d'ICC) α/ω : Non Validité concurrente : Non (concordance inter-formes KTK3/KTK4 du même outil, pas une validité concourante vis-à-vis d'un instrument externe)
52 BOT-2	MacCobb, S. (2005) Measurement and prediction of motor proficiency — Bayley + BOT	Irlande N : 76 (38 filles / 38 garçons) Âge : 8;2–10;1 ans SD non rapportée (âge ~9 ans, plage 8;2-10;1 ans uniquement) TD	BOTMP Version : BOTMP (ancien) Domaine : Motricité globale + fine Durée : 40–60	Construct validity; prédictif ICC : Non α/ω : Non Validité concurrente : Non
53 BOT-2	Lam, H.M.Y. (2011) Assessment of preschoolers' gross motor proficiency — BOT	Hong Kong N : 333 (171 garçons / 162 filles) Âge : 4;6–5;5 ans SD non rapportée (catégorie d'âge fixe "5 ans" du BOTMP, plage 4;6-5;5 ans) TD	BOTMP Version : BOTMP (ancien) Domaine : Motricité globale Durée : 40–60	Construct validity préscolaires ICC : Oui (ICC inter-évaluateur = 0,99 ; protocole de cotation non détaillé, sans IC95%) α/ω : Non Validité concurrente : Non
54 BOT-2	Brown, T. (2019) Structural Validity of the BOT-2 Subscales	Australie N : 117 (65 garçons / 52 filles) Âge : 8 à 12;2 ans (M=10;2, SD=1;4) M=10 ans 2 mois ; SD=1 an 4 mois TD	BOT-2 Version : BOT-2 BF Domaine : Motricité globale + fine Durée : 40–60	Structural validity; Rasch ICC : Non α/ω : Non Validité concurrente : Non
55 BOT-2	Yoon, D. (2024) Validity of the BOT-2 Short Form for Korean School-Age Children	Corée du Sud N : 283 (146 garçons / 137 filles) Âge : 7 à 12 ans (3 groupes : 7 ans ; 8–9 ans ; 10–12 ans ; M=9,1, SD=1,66) M=9,1 ans ; SD=1,66 TD (échantillon communautaire incluant quelques enfants avec TDAH, trouble du langage ou retard développemental — cf. limites)	BOT-2 Version : BOT-2 SF Domaine : Motricité globale + fine Durée : 15–20	Criterion validity; adaptation ICC : Non α/ω : Oui (α CF=0,90 ; α SF=0,69) Validité concurrente : Oui

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
56 BOT-2	Gharaei, E., Shojaei, M. & Daneshfar, A. (2019) The Validity and Reliability of the Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edition Brief Form, in Preschool Children	Iran N : 306 (164 filles/142 garçons) — sous-échantillon validité convergente n=30 ; fidélité test-retest n=50 ; sensibilité/spécificité n=60 (30+30) Âge : 4-7 ans SD non rapportée (plage 4-7 ans rapportée) TD (306 enfants sains, sélection aléatoire en grappes)	Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edition Version : BOT-2 BF (Brief Form) Domaine : Motricité globale + fine Durée : 15 (par test, BOT-2 et MABC administrés séparément)	Validité convergente (corrélation canonique BOT-2 vs MABC) ; fidélité test-retest (ICC, intervalle 2 semaines) ; sensibilité/spécificité (courbe ROC, point de coupure) ICC : Oui (ICC=0,809 ; IC95% [0,685–0,888], Table 4) α/ω : Non Validité concurrente : Oui (vs MABC, corrélation canonique $r=0,885$, $p<0,001$)
57 BOT-2	Bardid, F., Utesch, T. & Lenoir, M. (2019) Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form : An item response theory perspective	Belgique (Flandre + Région de Bruxelles-Capitale — projet Flemish Sports Compass) N : 2538 (6-8 ans : n=1243 ; 9-11 ans : n=1295) Âge : 6-11 ans SD non rapportée (plage 6-11 ans rapportée) TD (29 écoles primaires, échantillon représentatif ; données collectées à l'automne 2007)	Bruininks-Oseretsky Test of Motor Proficiency-2nd Edition Short Form Version : BOT-2 SF (14 items, 8 sous-tests) Domaine : Motricité globale + fine Durée : Non rapporté	Validité structurelle (modèle de Rasch — generalized partial credit model, GPCM) : unidimensionnalité, paramètres de seuil, statistiques d'ajustement χ^2 , fonctions d'information et de fiabilité par groupe d'âge (6-8 ans vs 9-11 ans) — pas une étude de fidélité/ICC classique ICC : Non (aucun ICC mesuré sur cet échantillon ; l'article cite, comme contexte issu du manuel BOT-2 SF, une fidélité test-retest $r\geq 0,80$ et inter-évaluateur $r=0,98$, non recalculées ici) α/ω : Non (fidélité de Rasch — fonctions d'information/fiabilité par niveau de compétence — rapportée à la place d'un alpha/omega classique) Validité concurrente : Non
58 BOT-2	Brown, T. (2019) Structural validity of the Bruininks-Oseretsky test of motor proficiency – Second edition brief form (BOT-2-BF)	Australie N : 123 Âge : 8-12 ans (M=10 ans 2 mois ; SD=1 an 4 mois) M=10 ans 2 mois ; SD=1 an 4 mois TD	Bruininks-Oseretsky Test of Motor Proficiency Version : BOT-2-BF (forme courte, 14 items) Domaine : 8 sous-tests représentés proportionnellement (FMP, FMI, MD, BC, B, RSA, ULC, S) Durée : 15-20 min	Validité structurelle (modèle de Rasch/RMM) : dimensionnalité, hiérarchie des items, DIF (genre), fiabilité de séparation items/personnes ICC : Non α/ω : Non (indices Rasch uniquement, pas d'alpha classique) Validité concurrente : Non
59 Autre	Ockerman, J., Velghe, S., Van Bladel, A. et al. (2024) Checks and balances : a meta-analysis on the known-groups validity of functional postural control tests in children	Belgique N : 1331 (pathologie) + 1889 (TD) — méta-analyse de 40 études cas-témoins Âge : M=8,68 ans (pathologie) / M=9,10 ans (TD) — critère d'inclusion : âge moyen <13 ans M=9,10 ans (sous-groupe TD, pertinent pour ce mémoire) ; M=8,68 ans (sous-groupe pathologie) ; SD=Non rapportée TD + pathologique (design cas-témoins, 7 groupes de pathologies)	Aucun des 5 outils du mémoire — méta-analyse de 22 tests posturaux (Pediatric Balance Scale, Berg Balance Scale, Kids-BESTest, reach tests, Timed Up and Go et variantes, One Leg Stance et variantes, perturbations sensorielles, marche en tandem, autres mesures chronométrées) Version : non applicable — revue de 22 outils différents, aucun TGMD/MABC-2/BOT-2/PDMS-2/KTK Domaine : Contrôle postural / équilibre fonctionnel Durée : non applicable (méta-analyse, pas une administration unique)	SMD (différences moyennes standardisées) ; niveau de preuve GRADE ; risque de biais SIGN — validité de groupes connus (known-groups validity), pas une étude de fidélité/validité classique d'un outil unique ICC : Non α/ω : Non Validité concurrente : Non (validité de groupes connus, pas concurrente)
60 Autre	Kakebeeke, T.H. (2018) Neuromotor development in children — new norms 3–18 years	Suisse N : 616 (304 garçons / 312 filles) Âge : 3–18 ans SD non rapportée classiquement (âge donné en médiane/IQR : M=7;9, IQR 6;7) TD	Normes neuromotrices Version : Normes générales Domaine : Motricité globale Durée : 20-30 (selon l'âge/niveau de l'enfant)	Normative data ; ICC intra-observateur ; ICC inter-observateur ; ICC test-retest (par composante + scores totaux) ICC : Oui α/ω : Non Validité concurrente : Non
61 Autre	Quitério, A. (2018) MOBAK 1 Assessment in Primary Physical Education (Portuguese 6-Year-Olds)	Portugal N : donnée précise non disponible dans les sources consultées Âge : 6 ans Données de dispersion (SD) non disponibles dans les sources consultées TD	MOBAK 1 Version : MOBAK 1 Domaine : Compétences motrices de base Durée : 45	Validity; reliability ICC : Non α/ω : Non Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
62 Autre	Verbecque, E. (2019) The Timed Up and Go Test in Children : Does Protocol Choice Matter? A Systematic Review	Belgique (revue multi-pays : données poolées Australie, Belgique, Brésil, Pakistan, USA) N : 2626 (6 études : Butz n=10 fidélité/160 normatif ; Habib n=180 ; Itzkowitz n=22 fidélité/1481 normatif ; Nicolini-Panisson & Donadio n=459/178 ; Verbecque n=172 ; Williams n=176/173/151) Âge : 3-18 ans (revue) ; par étude : Butz 5-12, Habib 5-13, Itzkowitz 5-17, Nicolini-Panisson & Donadio 3-18, Verbecque 3-5, Williams 3-9 SD non rapportée (revue de six études primaires, tranches d'âge hétérogènes selon l'étude) TD	Timed Up and Go (TUG) Version : Aucune version standardisée — 6 protocoles distincts (vitesse imposée/auto-sélectionnée ; motivation oui/non ; pieds nus/chaussures ; meilleur de 3 vs moyenne de 2-3 essais) Domaine : Équilibre dynamique fonctionnel (transfert assis-debout + marche 3m + demi-tour) Durée : Non rapporté explicitement dans la revue	ICC intra-évaluateur (within-session) ; ICC inter-évaluateur ; ICC test-retest ; SEM ; MDC95 ; données normatives par bande d'âge ; niveau de preuve Cochrane Back Review Group ICC : Oui α/ω : Non Validité concurrente : Non
63 Autre	Cattuzzo, M.T. (2020) Supine-To-Stand Task : Systematic Review	Brésil N : non applicable (revue systématique, pas d'échantillon propre) Âge : 3-22 ans environ (portée de la revue, de la jeunesse au grand âge) SD non applicable (revue systématique sans échantillon propre) TD	Supine-to-stand Version : Supine-to-stand Domaine : Motricité fonctionnelle Durée : non applicable (revue systématique)	Validity; functional health ICC : Non α/ω : Non Validité concurrente : Non
64 Autre	Okuda, P.M.M. (2020) Construct validity of the Motor Development Scale (MDS)	Brésil N : données précises de population non disponibles dans les sources consultées Âge : non disponible dans les sources consultées SD non disponible dans les sources consultées TD	MDS Version : MDS Domaine : Motricité globale Durée : 10–15	Construct validity ICC : Non α/ω : Non Validité concurrente : Non
65 Autre	Christopher, A. (2019) The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan : a systematic review	USA N : 77 articles inclus (revue multi-populations) ; sous-ensemble enfants TD pertinent = 4 références (Katz-Leurer 2008/2009, Nicolini-Panisson & Donadio, Williams et al. 2005) + Villamonte et al. 2010 (trisomie 21) Âge : 3-94 ans (ensemble des 77 études) ; sous-ensemble enfants TD : ~3-13 ans non applicable (revue systématique multi-populations, pas de M/SD unique pertinent pour le sous-ensemble enfants) TD + pathologie	Timed Up and Go (TUG) Version : Protocole original Podsiadlo & Richardson (1991) ; versions modifiées pour enfants incluses (adaptation des consignes uniquement, méthodologie conservée) Domaine : Mobilité fonctionnelle Durée : Non rapporté explicitement dans la revue	ICC test-retest/inter-rater/intra-rater par population ; validité concurrente (Pearson r / Spearman rho) ; validité prédictive (AUC, odds ratio, adultes typiques uniquement) ICC : Oui α/ω : Non Validité concurrente : Oui
66 Autre	Smits-Engelsman, B.C.M. (2021) Inter-rater and test-retest reliability of PERF-FIT for children	Pays-Bas N : n=29 (fidélité inter-juges) + n=72 (test-retest) — échantillons partiellement indépendants, 3 pays (Ghana/Afrique du Sud/Pays-Bas) Âge : 5–12 ans SD non rapportée (données N/âge confirmées) TD	P-MCS / PERF-FIT Version : PERF-FIT Domaine : Motricité liée à la forme physique Durée : non applicable (questionnaire parental, pas de passation chronométrée)	Reliability ICC : Oui α/ω : Non Validité concurrente : Non
67 Autre	De Roubaix, A. (2022) Psychometric properties of the Belgian LDCDQ (LDCDQ-BE)	Belgique N : 637 (parents) + 249 (enseignants) ; sous-échantillon comparé au MABC-2-NL : n=82 Âge : 3–5 ans (LDCDQ-BE et sous-échantillon de comparaison MABC-2) SD non rapportée TD	LDCDQ-BE Version : LDCDQ-BE Domaine : Questionnaire parental Durée : non applicable (questionnaire parental/enseignant, pas de passation chronométrée)	Psychometric; questionnaire ICC : Non α/ω : Non Validité concurrente : Non

N° / famille	Étude	Population	Instrument / domaine	Propriétés psychométriques rapportées
68 Autre	Rodrigues, L.P. (2022) Motor Competence Assessment (MCA) Scoring Method	Portugal N : 1000 (250/groupe d'âge, équilibré par sexe ; tiré d'une base de 2150) Âge : réparti en 4 groupes égaux (3-6, 7-10, 11-16, 17-22 ans), pas de moyenne globale unique (design par quotas) SD non applicable (design stratifié par bandes d'âge égales, pas de M/SD globale) TD	MCA Version : MCA Domaine : Motricité globale Durée : 45	Validity scoring ICC : Non α/ω : Non Validité concurrente : Non
69 Autre	Menescardi, C. (2022) CAMSA in a Spanish Context : Reliability and Validity	Espagne N : 749 (346 filles, 46,2 %) Âge : 8-14 ans (M=10,52, SD=1,68) SD=1,68 TD	CAMSA Version : CAMSA Domaine : Habiletés motrices combinées Durée : 10–15	Validity; reliability ICC : Non α/ω : Non Validité concurrente : Non
70 Autre	Makaruk, H. (2023) The FUS test : evaluating fundamental motor skills in children	Pologne N : 264 (139 filles, 125 garçons) Âge : 7–14 ans SD non rapportée TD	FUS test Version : FUS test Domaine : Habiletés motrices fondamentales Durée : $\approx$ 12-14 min/enfant (7-8 min administration + 5-6 min évaluation) ; $\approx$ 94-114 min pour un groupe de 13-15 élèves	Validity; reliability ICC : Non α/ω : Non Validité concurrente : Non
71 Autre	Topor, M. (2024) DCDQ — translation and adaptation into Danish	Danemark N : 36 (18 garçons, 18 filles) Âge : 5–7 ans (grade 0, M=6,08 ans) M=6,08 ans (grade 0) ; SD=Non rapportée TD	DCDQ Version : DCDQ Domaine : Questionnaire parental Durée : non applicable (questionnaire parental, pas de passation chronométrée)	Translation; validity ICC : Non α/ω : Non Validité concurrente : Non
72 Autre	Gavigan, N. (2025) Inter-rater reliability of the FMS2 among generalist teachers	Irlande N : 29 préscolaires (échantillonnage de convenance) Âge : 3-6 ans SD non rapportée TD	FMS2 Version : FMS2 Domaine : Habiletés motrices fondamentales Durée : Non rapporté	Inter-rater reliability ICC : Oui α/ω : Non Validité concurrente : Non
73 Autre	Mota, J. (2025) MOBAK 3–4 : Construct Validity and Score Reliability	Portugal N : 436 (233 garçons / 203 filles) Âge : 8–10 ans SD=0,6 TD	MOBAK 3-4 Version : MOBAK 3-4 Domaine : Compétences motrices de base Durée : 45	Construct validity; reliability ICC : Non α/ω : Non Validité concurrente : Non
74 Autre	Zuvela, F., Bozanic, A. & Miletic, D. (2011) POLYGON - a new fundamental movement skills test for 8 year old children : construction and validation	Croatie N : 95 (48 garçons / 47 filles) Âge : 8 ans ($8,1 \pm 0,3$ ans) M=8,1 ans ; SD=0,3 TD	FMS-POLYGON (TGMD-2 = comparateur de validité concurrente) Version : FMS-POLYGON (nouveau test norm-référencé, 4 tâches chronométrées) — comparateur : TGMD-2 (Ulrich, 2000) Domaine : Locomoteur + contrôle d'objet + franchissement d'obstacles + résistance (classification Mrakovic et al., 1993) Durée : NR (items chronométrés individuellement, 3,3-7,6 sec chacun ; durée totale du polygone complet non précisée)	ICC intra-rater (21 tâches + polygone complet) ; validité concurrente (rho de Spearman vs TGMD-2) ICC : Oui (intra-rater uniquement : 0,83-0,98 pour les 21 tâches ; 0,98 pour le polygone complet) α/ω : Non Validité concurrente : Oui (rho=-0,82, $p<0,05$, vs TGMD-2)

Annexe 5 : Évaluation détaillée des propriétés psychométriques et du risque de biais COSMIN des études portant sur le TGMD

Étude	Version	N	Propriétés analysées	Résultats principaux	Cotation du résultat	Risque de biais COSMIN	Utilisation dans la synthèse
Ave, T. (2017)	TGMD-2	50	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne	Inter : Loc= 95 ; OC= 88 ; GMD= 89 ; Test-retest : Loc= 82 ; OC= 79 ; GMD= 76 ; Intra : Loc= 98 ; OC= 95 ; GMD= 97 ; α/ω : Alpha inter-rater .98/.96/.96	Suffisant (+) pour les ICC rapportés	IC absents ; N test-retest/intra faibles	Non — aucun IC95% rapporté (inter-rater ni test-retest) ; exclu du pooling quantitatif, conservé en synthèse qualitative uniquement (décision du 25/06/2026, alignée sur le traltem...
Rintala, P. (2017)	TGMD-3	60 (3sous-échantillon de 20)	fidélité inter-évaluateur, fidélité intra-évaluateur	Inter : LS= 56 ; BS= 64 ; TS= 62 ; Intra : A: LS= 69 (.65-.72), BS= 77 (.75-.79), TS= 75 (.73-.76) ; B: LS/BS/TS= 73	Mixte : Inter-évaluateur insuffisant (-), intra-évaluateur partiellement suffisant (+)	N=20 ; intervalle intra trop long ; items faibles	Oui inter-rater avec IC, prudence
Maïano, C. (2021)	TGMD-3	127	fidélité inter-évaluateur, cohérence interne, validité de construit	Inter : Total= 998 ; Loc= 997 ; BS= 996 ; α/ω : ω G= 849 ; Loc S= 691 ; BS S= 620 ; Structure : CFI= 991, RMSEA= 033	Suffisant (+)	N>=100 ; IC rapportés ; méthode rigoureuse	Oui inter-rater ; non test-retest
Valentini, N.C. (2021)	TGMD-3-SF	1192	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : Total= 96 ; LOCS= 91 ; BS= 86 ; range .86-.99 ; Test-retest : Total= 86 ; LOCS= 87 ; BS= 78 ; Intra : Total= 92 ; LOCS= 92 ; BS= 82 ; range .60-.92 ; α/ω : α total= 75 ; LOCS= 60 ; BS= 71 ; Structure : CFI= 99, RMSEA= 02	Globalement suffisant (+), avec cohérence interne insuffisante pour LOCS	IC absents pour ICC ; forme courte uniquement	Non — aucun IC95% rapporté (inter-rater/test-retest/intra-rater) ; exclu du pooling quantitatif, conservé en synthèse qualitative uniquement (décision du 25/06/2026) ; s'ajoute au...
Marinšek, M. (2023)	TGMD-3	452	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : Total= 90 ; Loc= 82 ; BS= 92 ; items .71-.92 ; Test-retest : Total= 98 ; Loc= 96 ; BS= 97 ; Intra : Total= 99[.98-.1.00] ; Loco= 98[.98-.99] ; BS= 99[.98-.99] ; items .95[.88-.98... ; α/ω : α total= 87 ; Loc= 85 ; BS= 86 ; Structure : CFI= 95, RMSEA= 06 (.05-.08)	Suffisant (+)	IC rapportés ; N élevé ; test-retest 4 sem.	Oui inter/intra/test-retest avec IC
Estevan, I. (2017)	TGMD-3	178	fidélité inter-évaluateur, fidélité intra-évaluateur, cohérence interne, validité de construit, erreur de mesure	Inter : ICC= 90 ; Intra : ICC= 98 ; α/ω : .85-1.00 ; α total= 89 (.87-.92) ; Loc= 80 ; BS= 85 ; Structure : CFI= 970, RMSEA= .073	Mixte : fidélité suffisante (+), ajustement structurel contrasté	Fidélité sur 4 enfants seulement	Oui inter/intra avec réserve ; alpha poolable
Lopes, V.P. (2016)	TGMD-2	330	fidélité inter-évaluateur, fidélité test-retest, cohérence interne, validité de construit	Inter : Kappa : 1.0 plupart items ; .70 hop/underhand roll ; Test-retest : Ratio accord global= 96 (.09) [LoA 95% : 80-1.13] ; Loc= 1.0 (.08) [LoA 95% : ... ; α/ω : α 12 items= 69 ; Loc= 46 ; OC= 64 ; Structure : CFI= .956, RMSEA= .036 (.010-.054)	Mixte : structure acceptable, cohérence interne insuffisante pour certaines sous-échelles	Méthodes de fidélité non ICC ; N test-retest=22	Non pour ICC ; synthèse qualitative
Webster, E.K. (2017)	TGMD-3	807	fidélité inter-évaluateur, fidélité test-retest, cohérence interne, validité de construit	Inter : .98% accord, pas ICC ; Test-retest : Loc= 97 ; BS= 95 ; Total= 97 ; α/ω : α par âge : Loc 93-97 ; BS 89-97 ; Total .95-.98 ; Structure : CFI= 95, RMSEA= .10	Mixte : fidélité suffisante (+), RMSEA défavorable	Test-retest n=30 ; inter-rater non ICC ; doublon Garn	Non — aucun IC95% rapporté pour le test-retest ; exclu du pooling quantitatif, conservé en synthèse qualitative uniquement (décision du 25/06/2026)
Wagner, M.O. (2016/2017)	TGMD-3	189	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : Loc= 88 ; BS= 97 ; Test-retest : Loc= 94 ; BS= 98 ; Intra : Loc= 97 ; BS= 99 ; α/ω : α Loc= 76 (acceptable) ; BS= 89 (bon) [N.B. : IC95% intra-rater Loc= 94-.99 /... ; Structure : CFI= 98, RMSEA= .04 (.01-.07)	Suffisant (+)	Inter/intra n=30 ; version 12 items	Oui test-retest avec IC ; inter/intra avec faible pondération
Garn, A.C. (2021)	TGMD-3	862	cohérence interne, validité de construit	α/ω : α FMS= 922 ; Loc= 848 ; BS= 888 ; ω total FMS= 916 ; ω -h FMS= 797 ; sous-éche... ; Structure : CFI= 986 (ESEM) ; .967 (CFA), RMSEA= .043 (ESEM) ; .066 (CFA)	Suffisant (+) pour le facteur général ; sous-échelles à interpréter avec prudence	Population mixte 11% handicap ; doublon Webster ; pas d'ICC	Non
Temple, V.A. (2016)	TGMD-3	277	validité de construit	Changement grade 3-4 ; POMP ; différences sexe	Non cotable pour les propriétés retenues	Population mixte ; aucune propriété COSMIN cotable	Non
Rizkyanto, W.I. (2024)	TGMD-3	290	fidélité test-retest, cohérence interne	Test-retest : Pearson r= 819 (pas ICC) ; α/ω : α = 99 (élevé = interprété avec prudence)	Indéterminé (?)	Hors tranche cible partiellement : méthodes	Non

Étude	Version	N	Propriétés analysées	Résultats principaux	Cotation du résultat	Risque de biais COSMIN	Utilisation dans la synthèse
			validité de construit			incohérentes	
Valentini, N.C. (2017)	TGMD-3	597	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : Alpha niveau skill :.85-.99 (pas ICC subtest/total) ; Test-retest : Pearson : Loc= .93 ; BS= .81 ; Total= .90 ; Intra : Alpha :.60-.90 ; intervalle 2 mois ; α/ω : α skills-to-test total=.74 ; Loc=.63 ; BS= .76 ; α criteria-to-test total=.93 ... ; Structure : CFI=.94, RMSEA=.04 (.03-.05) Inter : « excellent > .85 », valeur exacte NC ; Intra : « excellent > .85 », valeur exacte NC, intervalle 1 mois ; α/ω : α =.91 ; ω =.86 ; ω h=.856 ; ω subscales BS=.11 / LS=.18 ; Structure : CFI=.986, RMSEA=.041 (.027-.053) Inter : Intra-évaluateur (vidéo, 2 évaluateurs) : ICC=.93-.98 ; Test-retest : N/A (TGMD-2 non testé en test-retest dans cet article) ; Intra : Voir col. 12 (ICC vidéo qualifié d'intra-évaluateur par les auteurs) Inter : « Sous-test global=.93 ; dribble=.94 ; strike=.85 ; overhand throw=.84 ; underhand throw=.84 ; Test-retest : N/A (pas de design test-retest) ; Intra : N/A (pas de design intra-rater séparé) ; α/ω : N/A (non rapporté) Inter : N/A (pas de design inter-rater — seul un évaluateur dans toute l'étude) ; Test-retest : Loc=.65 ; OC=.85 ; GMQ=.81 ; Intra : Loc=.95 ; OC=.99 ; GMQ=.97 ; α/ω : Alpha par groupe d'âge (8 tranches, range .65-.88) moyenné : Loc=.74 ; OC=.78... ; Structure : CFI=.98, RMSEA=.05 (.05-.06) Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Mixte : fidélité globale favorable, cohérence interne variable		Non pour ICC : oui synthèse qualitative
Salami, S. (2021)	TGMD-3	496	fidélité inter-évaluateur, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : « excellent > .85 », valeur exacte NC ; Intra : « excellent > .85 », valeur exacte NC, intervalle 1 mois ; α/ω : α =.91 ; ω =.86 ; ω h=.856 ; ω subscales BS=.11 / LS=.18 ; Structure : CFI=.986, RMSEA=.041 (.027-.053) Inter : Intra-évaluateur (vidéo, 2 évaluateurs) : ICC=.93-.98 ; Test-retest : N/A (TGMD-2 non testé en test-retest dans cet article) ; Intra : Voir col. 12 (ICC vidéo qualifié d'intra-évaluateur par les auteurs) Inter : Sous-test global=.93 ; dribble=.94 ; strike=.85 ; overhand throw=.84 ; underhand throw=.84 ; Test-retest : N/A (pas de design test-retest) ; Intra : N/A (pas de design intra-rater séparé) ; α/ω : N/A (non rapporté) Inter : N/A (pas de design inter-rater — seul un évaluateur dans toute l'étude) ; Test-retest : Loc=.65 ; OC=.85 ; GMQ=.81 ; Intra : Loc=.95 ; OC=.99 ; GMQ=.97 ; α/ω : Alpha par groupe d'âge (8 tranches, range .65-.88) moyenné : Loc=.74 ; OC=.78... ; Structure : CFI=.98, RMSEA=.05 (.05-.06) Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Mixte : structure suffisante, fidélité ICC indéterminée	ICC non chiffrés ; pas de test-retest	Non ICC ; oui structure/alpha
Soares, I.A. (2020)	TGMD-2	251	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, validité de construit	Inter : Intra-évaluateur (vidéo, 2 évaluateurs) : ICC=.93-.98 ; Test-retest : N/A (TGMD-2 non testé en test-retest dans cet article) ; Intra : Voir col. 12 (ICC vidéo qualifié d'intra-évaluateur par les auteurs) Inter : Sous-test global=.93 ; dribble=.94 ; strike=.85 ; overhand throw=.84 ; underhand throw=.84 ; Test-retest : N/A (pas de design test-retest) ; Intra : N/A (pas de design intra-rater séparé) ; α/ω : N/A (non rapporté) Inter : N/A (pas de design inter-rater — seul un évaluateur dans toute l'étude) ; Test-retest : Loc=.65 ; OC=.85 ; GMQ=.81 ; Intra : Loc=.95 ; OC=.99 ; GMQ=.97 ; α/ω : Alpha par groupe d'âge (8 tranches, range .65-.88) moyenné : Loc=.74 ; OC=.78... ; Structure : CFI=.98, RMSEA=.05 (.05-.06) Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Suffisant (+) pour la fidélité vidéo ; validité critériée modérée	N=251 pour la cotation vidéo ; IC95% non fournis ; TGMD-2 = critère, pas l'instrument étudié	Non (en l'état) — ICC vidéo utilisable comme donnée de fidélité secondaire avec réserve, pas comme étude de fidélité primaire poolable
Barnett, L.M. (2013)	TGMD-2	37	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne	Inter : Sous-test global=.93 ; dribble=.94 ; strike=.85 ; overhand throw=.84 ; underhand throw=.84 ; Test-retest : N/A (pas de design test-retest) ; Intra : N/A (pas de design intra-rater séparé) ; α/ω : N/A (non rapporté) Inter : N/A (pas de design inter-rater — seul un évaluateur dans toute l'étude) ; Test-retest : Loc=.65 ; OC=.85 ; GMQ=.81 ; Intra : Loc=.95 ; OC=.99 ; GMQ=.97 ; α/ω : Alpha par groupe d'âge (8 tranches, range .65-.88) moyenné : Loc=.74 ; OC=.78... ; Structure : CFI=.98, RMSEA=.05 (.05-.06) Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Suffisant (+)	Population locomotrice non testée ; N modeste ; kappa par composante affecté par le paradoxe prévalence (3/24 composantes concernées, signalé par les auteurs) Un seul évaluateur dans toute l'étude (pas d'inter-rater testé) ; $\chi^2/df=5.70$ élevé (mais attribué à la taille d'échantillon, cf. discussion des auteurs) ; corrélation inter-facteurs très élevée	Oui — ICC inter-évaluateur avec IC95% complets pour le sous-test global et les 6 habiletés (sous-test contrôle d'objet)
Farrokhi, A. (2014)	TGMD-2	1438 (normalif) ; 63 (test-retest) ; 32 (intra-rater)	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : N/A (pas de design inter-rater — seul un évaluateur dans toute l'étude) ; Test-retest : Loc=.65 ; OC=.85 ; GMQ=.81 ; Intra : Loc=.95 ; OC=.99 ; GMQ=.97 ; α/ω : Alpha par groupe d'âge (8 tranches, range .65-.88) moyenné : Loc=.74 ; OC=.78... ; Structure : CFI=.98, RMSEA=.05 (.05-.06) Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Mixte : test-retest locomoteur insuffisant (-), autres scores suffisants (+)	Oui — ICC test-retest ET intra-rater poolables avec IC95% complets pour les 3 scores (Loc, OC, GMQ) ; alpha également poolable (moyenné par transformation z sur 8 groupes d'âge)	
Maeng, H. (2017)	TGMD-3	10	fidélité inter-évaluateur, fidélité intra-évaluateur	Inter : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run=.66 ; gallop=.66 ; hop=.92 ; skip=.90 ; horizontal jump=.81 ... ; Intra : Tous évaluateurs combinés (Tableau 2, colonne 'All raters') : Locomoteur : run...	Mixte : scores globaux suffisants (+), certains items insuffisants (-)	Oui — inter-rater Total/Loco/BS/tous items avec IC95% complets (candidat pour ajout Fig.1, k=3-4) ; intra-rater également poolable, IC95% complets tous items et sous-échelles (colonne 1)	
Magistro, D. (2020)	TGMD-3	5210	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, validité de construit	Inter : Total=.0.973 (couples d'évaluateurs, p<0.001) ; Test-retest : Total=.0.996 ; Loco=.0.996 ; BS=.0.997 (n=450) ; par tranche d'âge 3-11 ans : To... ; Intra : Non chiffré en ICC — Cohen's kappa >0.80 pour tous les critères (Table S3 non... ; Structure : CFI=.0.955, RMSEA=.0.050 [90%CI 0.048-0.053])	Suffisant (+)	Oui — ICC inter-rater Total ET ICC test-retest Total/Loco/BS, tous avec IC95% complets ; candidat majeur pour renforcer le pooling Fig.1 (inter-rater, p actuel=.99,9% non interprété...)	

Annexe 6 : Évaluation détaillée des propriétés psychométriques et du risque de biais COSMIN des études portant sur le MABC-2

Étude	Version	N	Propriétés analysées	Résultats principaux	Cotation du résultat	Risque de biais COSMIN	Utilisation dans la synthèse
Brown, T. (2009)	MABC-2	N/A (revue)	Aucune propriété psychométrique cotable	Aucune valeur quantitative exploitable	Non applicable	N/A	Non
Smits-Engelsman, B.C.M. (2011)	MABC-2 AB1	50	fidélité inter-évaluateur, fidélité test-retest, intra-évaluateur, cohérence interne, erreur de mesure	Inter : Inter-testeur (≠ inter-rater classique) : .76 (n=19, éch. propre) / .79 (n=22... ; Test-retest : Intra-testeur (même évaluateur) Total : .83 (n=28, éch. complet) / .94 (n=23... ; Intra : N/A — étude ne mesure pas l'intra-rater classique (enregistrement par le même... ; α/ω : .70-.87	Suffisant (+) pour les ICC et la cohérence interne rapportés	Modéré (N modeste)	Non — ICC95% confirmés absents pour tous les ICC exclu du pooling quantitatif...
Kokstijn, J. (2018)	MABC-2 AB1	510	fidélité test-retest, validité de construit	Inter : N/A (non étudié) ; Test-retest : N/A (non étudié) ; Intra : N/A (non étudié) ; α/ω : Non rapporté (non étudié)	Mixte : structure suffisante dans certains sous-groupes, insuffisante dans d'autres	Faible (N=510, design solide)	Non — aucune donnée ICC/alpha à pooler ; structure exploitable en synthèse qu...
Jaikaew, R. & Saliensukpong, N. (2019)	MABC-2 AB2	29	fidélité inter-évaluateur, validité de construit	Inter : Placing Pegs pref. = .99(.98-.99) ; non-préf. = .99(.99-.99) ; Threading Lace = .99... ; Test-retest : N/A (non étudié — pas de design test-retest) ; Intra : N/A (non étudié) ; α/ω : Non rapporté (non étudié)	Globalement suffisant (+), avec un item non calculable en raison d'un effet plafond	Modéré (N=29, mais design méthodologiquement pur ; scoring simultané en 1 séa...	Oui — ICC par item avec ICC95% complets (seule étude AB2 du corpus avec cette ...
Jaikaew, R. & Saliensukpong, N. (2021)	MABC-2 AB2	360	fidélité inter-évaluateur	Inter : N/A pour CET article (voir note)	Non cotable	N/A	Non pour cet article
Banátová, K. et al. (2022)	MABC-2 Checklist	Non rapporté	validité de construit	Validité de contenu jugée satisfaisante (2e ronde de panel d'experts) ; pas d...	Indéterminé (?) avec les données disponibles	N/A	Non — forme Checklist (questionnaire), non poolable avec le test de performan...
Verbecque, E. et al. (2025)	MABC-2 AB2	3025	validité de construit	Structure à 3 domaines testées (résultats détaillés non retrouvés dans les ex...	Indéterminé (?) avec les données disponibles	N/A	Non en l'état (aucune valeur numérique confirmée)
Wu, O.K. et al. (2025)	MABC-3 AB1	24	fidélité inter-évaluateur, fidélité test-retest, validité de construit	Inter : N/A (pas d'inter-rater rapporté) ; Test-retest : α = .563-.897 (Performance Test, Spearman — PAS un ICC)	Indéterminé pour la fidélité ; validité convergente variable	Élevé (N=24)	=> c'est le MABC-3, version différente (règle de h...
Arbi, Z.E. et al. (2025)	MABC-2 AB1	62	fidélité inter-évaluateur, fidélité test-retest, cohérence interne, erreur de mesure	Inter : N/A — design test-retest à un seul examinateur ; aucune mesure inter-évaluat... ; Test-retest : Items (AB1, n=62) : MD1p= .99(.99-1.00) ; MD1np= .99(.99-1.00) ; MD2= .99(.99-1.00) ; Intra : N/A (non distingué du test-retest dans cet article — un seul examinateur, pas... ; α/ω : Alpha GLOBAL par groupe (pas par item) : 3-4 ans= .79 ; 5-6 ans= .65 ; AB11...	Mixte : fidélité souvent suffisante, alpha insuffisant chez les 5-6 ans	Modéré (N=62 modeste mais correct ; attrition 22.5% au retest ; aucune compar...	Oui — ICC test-retest poolable avec ICC95% pour 7/10 items (AB1, n=62) ; les 3...
Valentini, N.C., Ramalho, M.H. & Oliveira, M.A. (2014)	MABC-2 (original)	844 (440G/404F) ; test-retest : n=188 ; longitudinal : n=41 (3-6 ans) ; concouran...	fidélité inter-évaluateur, fidélité test-retest, validité de construit	Inter : Dextérité manuelle (MD) AB-AC-BC-ABC = .99/.99/.99 ; Visée-réception (BS) = ... ; Test-retest : Score standard ICC = .85 ; Pearson r score standard = .74 (p<.0001) ; sous-échel... ; Intra : Intra-évaluateur (A-A) : MD = .81 ; BS = .71 ; Équilibre = .72 ; Score total = .88 ; α/ω : 3 sous-échelles ensemble α = .78 ; MD α = .77 ; BS α = .52 (sous le seuil usuel de ...	Mixte : fidélité globale favorable, cohérence interne et validité concurrente faibles pour certains domaines	Modéré à élevé (grand N global, mais sous-échantillons de fidélité non détaill...	Non en l'état (absence d'ICC95% empêche la transformation z de Fisher)
Serbetar, I., Loffenes, J.M. & Mamen, A. (2019)	MABC-2 AB1	683 (366G/317F) ; n=650 items complets ; test-retest analysé : n=173 ; inter-év...	fidélité inter-évaluateur, fidélité test-retest, intra-évaluateur, validité de construit, erreur de mesure	Inter : Items (4 évaluateurs, n=36) : .86 (catching beanbag) à .99 (threading beads / ... ; Test-retest : Score standard total = .79 (n=173) ; Items = .66 (jumping) à .83 (drawing trail) ... ; Intra : Non distingué de l'inter-évaluateur dans cet article (un seul design « inter... ; Structure : CFI=Non rapporté (l'article utilise NFI/AGFI/SMRM, pas le CFI), RMS...	Mixte : scores globaux suffisants, certains items insuffisants	Faible à modéré (grand N global, mais sous-échantillons de fidélité plus rest...	Oui — ICC test-retest et inter-évaluateur avec ICC95% complets (AB1)
Scarfó, E., Ranucci, M., Berardi, A., Smeon, R. & Galeoto, G. (2024)	MABC-2 (original)	108 (69G/39F)	fidélité inter-évaluateur, fidélité test-retest, intra-évaluateur, cohérence interne, erreur de mesure	Inter : Appelé « Inter-operator test-retest » dans l'article. Bande 1 : .964-1.00 ; B... ; Test-retest : Intra-évaluateur (48n). Bande 1 : .823-.975 (8 items) ; Bande 2 : .839-.940 s... ; Intra : .975 (8 items) ; CFI : colonne ICC test-retest (terminologie de l'article : « intra-rater test-r... ; α/ω : Non rapporté, limite reconnue par les auteurs eux-	Mixte : suffisant en bande 1, résultats instables en bandes 2-3	Faible en bande 1 ; élevé en bandes 2-3 (échantillons réduits, ICC proches de...	Oui pour la bande 1 (réserve) ; Non pour les bandes 2-3 (N insuffisant, ICC95%...

Étude	Version	N	Propriétés analysées	Résultats principaux	Cotation du résultat	Risque de biais COSMIN	Utilisation dans la synthèse
Ellinoudis, T. et al. (2011)	MABC-2 AB1	183	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : N/A (étude ne mesure pas l'inter-rater classique = fidélité intra-évaluateur ... ; Test-retest : Items (n=60, IC95%) : Posting coins préf.=.93(.89-.96) ; non-préf.=.94(.91-.9...) ; Intra : Préliminaire seulement : ICC moyen=.96 « across items », aucun détail par it... ; α/ω : .51 (dextérité manuelle) ; .70 (visée-réception) ; .66 (équilibre ; Struct... Inter : Items (Table 3, n=30) : .57(placing pegs préf.) ; .68(non-préf.) ; .53(readi... ; Test-retest : Items (Table 2, n=29) : .50(placing pegs préf.) ; .77(non-préf.) ; .35(readi... ; Intra : N/A — non distingué du design test-retest dans cet article (un seul examine... ; α/ω : Non rapporté (non étudié) Inter : Rapporté PAR ÉVALUATEUR individuel (pas un ICC omnibus multi-évaluateurs) — l... ; Test-retest : N/A (pas de design test-retest, seulement intra-évaluateur) ; Intra : Intra-évaluateur, 2 mesures (Tableaux 3-4) : MD1p=.98/0.98 ; MD2=.99/0.98 ... ; α/ω : 0.65 (évaluateur 1) ; 0.63 (évaluateur 2) ; 0.66 (évaluateur 3) — score total... α/ω : 0.602 (8 items), qualifié d'« acceptable » par les auteurs (sans IC95%) ; Structure : CFI=Non extrait précisément des passages disponibles (critère >0.95 atteint selon... ; RMSEA=Non extrait précisément des passages disponibles (critère <0.05 atteint selon... α/ω : Non rapporté dans les extraits disponibles ; Structure : CFI=Non extrait précisément (ajustement jugé bon selon critères a priori : AGFI=.0... ; RMSEA=.048 (modèle 5-6 ans ; valeur jeunes non extraite des passages disponibles)	Mixte : fidélité suffisante, cohérence interne insuffisante pour certains domaines	Faible (N=183, design solide, bootstrap pour robustesse AFC)	Oui — ICC test-retest par item et par domaine, avec IC95% complets ; alpha ég...
Holm, I. (2013)	MABC-2 AB2	45	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, erreur de mesure	Inter : Items (Table 3, n=30) : .57(placing pegs préf.) ; .68(non-préf.) ; .53(readi... ; Test-retest : Items (Table 2, n=29) : .50(placing pegs préf.) ; .77(non-préf.) ; .35(readi... ; Intra : N/A — non distingué du design test-retest dans cet article (un seul examine... ; α/ω : Non rapporté (non étudié) Inter : Rapporté PAR ÉVALUATEUR individuel (pas un ICC omnibus multi-évaluateurs) — l... ; Test-retest : N/A (pas de design test-retest, seulement intra-évaluateur) ; Intra : Intra-évaluateur, 2 mesures (Tableaux 3-4) : MD1p=.98/0.98 ; MD2=.99/0.98 ... ; α/ω : 0.65 (évaluateur 1) ; 0.63 (évaluateur 2) ; 0.66 (évaluateur 3) — score total... α/ω : 0.602 (8 items), qualifié d'« acceptable » par les auteurs (sans IC95%) ; Structure : CFI=Non extrait précisément des passages disponibles (critère >0.95 atteint selon... ; RMSEA=Non extrait précisément des passages disponibles (critère <0.05 atteint selon... α/ω : Non rapporté dans les extraits disponibles ; Structure : CFI=Non extrait précisément (ajustement jugé bon selon critères a priori : AGFI=.0... ; RMSEA=.048 (modèle 5-6 ans ; valeur jeunes non extraite des passages disponibles)	Insuffisant (-) pour plusieurs domaines et pour le score total	15 enfants chevauchants entre échantillons intra (n=29) et inter (n=30) sur N...	Oui — ICC inter-testeur et intra-testeur (test-retest) avec IC95% complets pa...
Niño-Cruz, G.I. (2019)	MABC-2 AB1	29	fidélité inter-évaluateur, fidélité test-retest, fidélité intra-évaluateur, cohérence interne, validité de construit	Inter : Rapporté PAR ÉVALUATEUR individuel (pas un ICC omnibus multi-évaluateurs) — l... ; Test-retest : N/A (pas de design test-retest, seulement intra-évaluateur) ; Intra : Intra-évaluateur, 2 mesures (Tableaux 3-4) : MD1p=.98/0.98 ; MD2=.99/0.98 ... ; α/ω : 0.65 (évaluateur 1) ; 0.63 (évaluateur 2) ; 0.66 (évaluateur 3) — score total... α/ω : 0.602 (8 items), qualifié d'« acceptable » par les auteurs (sans IC95%) ; Structure : CFI=Non extrait précisément des passages disponibles (critère >0.95 atteint selon... ; RMSEA=Non extrait précisément des passages disponibles (critère <0.05 atteint selon... α/ω : Non rapporté dans les extraits disponibles ; Structure : CFI=Non extrait précisément (ajustement jugé bon selon critères a priori : AGFI=.0... ; RMSEA=.048 (modèle 5-6 ans ; valeur jeunes non extraite des passages disponibles)	Mixte : fidélité élevée, cohérence interne insuffisante	N=29 sous le seuil de 50 ; échantillonnage de convenance ; ICC rapportés par ...	Oui — items MD1p, MD2, MD3, CA1 poolables avec les 3 études déjà en Fig.7a/7b...
Kita, Y. (2016)	MABC-2 AB2	132	cohérence interne, validité de construit	Inter : Rapporté PAR ÉVALUATEUR individuel (pas un ICC omnibus multi-évaluateurs) — l... ; Test-retest : N/A (pas de design test-retest, seulement intra-évaluateur) ; Intra : Intra-évaluateur, 2 mesures (Tableaux 3-4) : MD1p=.98/0.98 ; MD2=.99/0.98 ... ; α/ω : 0.65 (évaluateur 1) ; 0.63 (évaluateur 2) ; 0.66 (évaluateur 3) — score total... α/ω : 0.602 (8 items), qualifié d'« acceptable » par les auteurs (sans IC95%) ; Structure : CFI=Non extrait précisément des passages disponibles (critère >0.95 atteint selon... ; RMSEA=Non extrait précisément des passages disponibles (critère <0.05 atteint selon... α/ω : Non rapporté dans les extraits disponibles ; Structure : CFI=Non extrait précisément (ajustement jugé bon selon critères a priori : AGFI=.0... ; RMSEA=.048 (modèle 5-6 ans ; valeur jeunes non extraite des passages disponibles)	Mixte : validité structurelle favorable, cohérence interne insuffisante	Aucune donnée de fidélité — limite explicitement reconnue par les auteurs eux...	Non — aucune donnée ICC/alpha avec IC95%. Utile en synthèse qualitative uniqui...
Psotta, R. (2016)	MABC-2 AB1	399	cohérence interne, validité de construit	Inter : Rapporté PAR ÉVALUATEUR individuel (pas un ICC omnibus multi-évaluateurs) — l... ; Test-retest : N/A (pas de design test-retest, seulement intra-évaluateur) ; Intra : Intra-évaluateur, 2 mesures (Tableaux 3-4) : MD1p=.98/0.98 ; MD2=.99/0.98 ... ; α/ω : 0.65 (évaluateur 1) ; 0.63 (évaluateur 2) ; 0.66 (évaluateur 3) — score total... α/ω : 0.602 (8 items), qualifié d'« acceptable » par les auteurs (sans IC95%) ; Structure : CFI=Non extrait précisément des passages disponibles (critère >0.95 atteint selon... ; RMSEA=Non extrait précisément des passages disponibles (critère <0.05 atteint selon... α/ω : Non rapporté dans les extraits disponibles ; Structure : CFI=Non extrait précisément (ajustement jugé bon selon critères a priori : AGFI=.0... ; RMSEA=.048 (modèle 5-6 ans ; valeur jeunes non extraite des passages disponibles)	Suffisant (+) pour la validité structurelle ; fidélité non étudiée	Aucune donnée de fidélité ; double saturation mineure de l'item Threading bea...	Non — aucune donnée ICC/alpha. Utile en synthèse qualitative (plus grand N du...

Annexe 7 : Tableau des indicateurs pour les synthèses quantitatives et Script R

Indicateur	Transformation avant pooling	Estimation de la précision
ICC	Transformation z de Fisher	IC95 % rapporté ou erreur standard reconstruite à partir de l'IC
Corrélation inter-facteurs	Transformation z de Fisher	Formule exacte de Fisher à partir de (r) et de (N)
Alpha de Cronbach	Méthode précisée dans le script R	IC rapporté ou estimation selon Feldt et al. (1987)
RMSEA	Transformation logarithmique	Bornes de l'IC90 % ou de l'IC rapporté
CFI	Synthèse exploratoire	Précision approximée à partir de (N) en l'absence d'IC rapporté

Note : Les valeurs regroupées ont été retransformées dans leur échelle originale pour la présentation. Les synthèses du CFI et du RMSEA ont été considérées comme exploratoires.

#

Script R consolidé : formules et méthode de pooling

Package : meta (v8.5-0), méthode REML, transformation z de Fisher

=====

library(meta)

--- 1. ICC et alpha : transformation de Fisher (atanh) ---

z_transform <- function(x) atanh(x)

se_z_transform <- function(x, se) se / (1 - x^2)

SE depuis IC95 % rapporté (méthode prioritaire)

se_from_ci95 <- function(lo, up) (up - lo) / (2 * 1.96)

SE de l'alpha estimée à partir de l'alpha, de l'effectif et du nombre d'items (Feldt et al., 1987)

k = nombre d'items, n = effectif

se_alpha_N <- function(alpha, n, k) {

 sqrt(2 * (1 - alpha)^2 * (1 + alpha * (k - 1))^2 / (k * (k - 1) * n))

}

--- 2. Corrélation (r inter-facteurs) : formule EXACTE de Fisher ---

Pas d'approximation : SE = 1/sqrt(N-3)

se_r_fisher <- function(n) 1 / sqrt(n - 3)

--- 3. CFI : approximation depuis N (type binomial) quand IC absent ---

se_cfi_from_N <- function(cfi, n) sqrt(cfi * (1 - cfi) / n)

--- 4. RMSEA : échelle [0,∞), transformation log, IC90% (pas 95%) ---

se_from_ci90 <- function(lo, up) (up - lo) / (2 * 1.645)

se_rmsea_from_N <- function(rmse, n) sqrt((1 - rmse^2)^2 / (2 * n))

z_rmsea <- function(rmse) log(rmse)

se_z_rmsea <- function(rmse, se) se / rmse

--- 5. Traitement des bornes à 0 ou 1 ---

atanh(1) = Inf : un ICC/alpha/r rapporté exactement à 1,000 est exclu

du pooling. Règle appliquée : conserver la précision native de la source

(ex. ICC=0,998 et non 1,00) ; si la source rapporte réellement 1,000,

l'étude est écartée du pooling quantitatif et signalée en synthèse

qualitative plutôt que forcée dans le modèle.

--- 6. Pooling générique (effets aléatoires, REML) ---

run_pool <- function(z, se_z, etude, titre) {

 m <- metagen(TE = z, seTE = se_z, studlab = etude,

```

        method.tau = "REML", fixed = FALSE, random = TRUE,
        title = titre)
summary(m)
}
# --- Exemple d'application : ICC inter-évaluateur TGMD-3 (Figure 2, k=5) ---
donnees <- data.frame(
  etude = c("Rintala 2017", "Maiano 2022", "Marinšek 2023",
    "Maeng 2017", "Magistro 2020"),
  icc = c(0.620, 0.998, 0.900, 0.960, 0.973),
  ic_inf = c(0.589, 0.997, 0.878, 0.910, 0.969),
  ic_sup = c(0.649, 0.998, 0.918, 0.990, 0.977)
)
donnees$se_icc <- se_from_ci95(donnees$ic_inf, donnees$ic_sup)
donnees$z <- z_transform(donnees$icc)
donnees$se_z <- se_z_transform(donnees$icc, donnees$se_icc)
run_pool(donnees$z, donnees$se_z, donnees$etude, "ICC inter-evaluateur TGMD-3")
# → Résultat : ICC poolé = 0,960 [0,786-0,993], I² = 99,8%

```

Résumé

Objectif : Cartographier les instruments standardisés utilisés pour évaluer les manifestations fonctionnelles du contrôle et du développement moteur chez les enfants de 1 à 12 ans au développement typique, et examiner leurs propriétés psychométriques.

Méthode : Une scoping review a été menée dans PubMed, Embase et Scopus. Les études incluses ont fait l'objet d'une synthèse descriptive. Une évaluation méthodologique selon COSMIN et des synthèses quantitatives ciblées ont été réalisées pour le TGMD et le MABC-2.

Résultats : 74 publications ont été incluses. Les estimations de fidélité du TGMD-3 étaient généralement élevées, mais présentaient une forte hétérogénéité entre les études. Les résultats de validité structurelle soutenaient l'importance d'une dimension générale, invitant à interpréter les sous-scores avec prudence. Pour le MABC-2, la reproductibilité des items était globalement élevée mais variable, notamment pour *Posting Coins* et *Drawing Trail*.

Conclusion : Aucun instrument ne peut être considéré comme universellement supérieur. Leur choix doit dépendre de l'âge, du domaine moteur évalué et du contexte d'utilisation. Des validations locales et un rapport plus standardisé des propriétés psychométriques restent nécessaires.