The Application of Dynamic Assessment for Performance Prediction in Education: A Scoping Review

Cecile Wiersema

Master Thesis – Talent Development & Creativity

S4392299

June, 2025

Department of Psychology

University of Groningen

Examiner/Daily supervisor:

Henderien Steenbeek

Declaration

I, Cecile Wiersema, hereby declare that I have written this Master's thesis titled "The Application of Dynamic Assessment for Performance Prediction in Education: A Scoping Review" independently. This thesis is my original work and has and will not be presented for a degree at another university. All sources used in this work have been acknowledged. To aid in the screening process an AI tool called Rayyan has been used. However, due to the relatively small quantity of studies that required screening, all articles have been personally screened by me, utilizing Rayyan as a data collection platform. Also, generative AI, like ChatGPT, has been used to improve readability. It has not been used to generate the content of this thesis, neither was it used to write text from scratch. This thesis is in agreement with the recommendations regarding AI use in theses, as clarified in the "Master Thesis, Internship, and Graduation Guide" (Psychology Department, 2024).

Abstract

Important educational decisions, like transitions such as the move from primary to secondary school, are guided by traditional, static assessments. However, because these assessments focus only on static skills and knowledge, there are concerns that they may not accurately capture a student's potential for future performance beyond their current abilities. In contrast, Dynamic Assessment (DA) offers a process-oriented approach that assesses a student's learning potential through interactive mediation. This scoping review aimed to examine how DA is applied in educational contexts for performance prediction across general and diverse student populations, including students with learning difficulties and multilingual learners. Twenty-six quantitative studies were analyzed using descriptive statistics and qualitative content analysis. The findings show that DA is a significant predictor of scholastic performance in both the general and diverse student populations. The qualitative and quantitative information acquired by DA can play a role in both short-term and long-term performance prediction, supporting tailored interventions and educational planning. However, practical barriers, like time consumption and the need for specialized training, limit its widespread application. While DA offers an informative and equitable assessment tool, due to its implementation challenges, it should be used as a supplement to static measures rather than replace them. Future research should focus on standardizing DA tools, and exploring various populations and contexts, while striving to improve its feasibility.

Keywords: dynamic assessment, learning potential, performance prediction, primary and secondary education, scoping review

Contents

The Application of Dynamic Assessment for Performance Prediction in Education

The role of assessments in educational transitions, like between primary and secondary education, is crucial because they often give educators an idea of students' academic trajectories. In the Netherlands, the "doorstroomtoets" (transition test) plays a key role in recommending secondary education levels (Ministerie van Onderwijs, Cultuur en Wetenschap, 2024). Currently, in the Netherlands, only traditional, static assessment methods are allowed for this transition test. Traditional assessments, also referred to as static assessments, measure a student's current level of knowledge or skills at a fixed point in time (Haywood & Lidz, 2006). These standardized tests do not allow for interaction, feedback, or mediation between the assessor and student during the assessment process. Research suggests that traditional, static assessments may not give an accurate prediction of a student's long-term academic performance (Caffrey et al., 2008; Dixon et al., 2023; Tzuriel, 2021a). Because these assessments measure static skills and knowledge, they are not always able to capture a student's potential for growth.

Dynamic Assessment

Dynamic Assessment (DA), also called Learning Potential Assessment (LPA), is an approach used in educational settings to measure a student's capacity to learn and adapt to their environment (Feuerstein et al., 1979; Hamers et al., 1993; Stad et al., 2018). This approach focuses mainly on the possibility for future learning, also called learning potential. In the educational context, learning potential refers to the ability of a student to gain new knowledge, skills, or abilities when provided with appropriate instruction (Stad et al., 2018). This concept is closely related to cognitive modifiability, which is an individual's ability to learn from experiences and change one's cognitive processes (Kozulin, 2011; Tzuriel, 2021a).

Whereas traditional assessment methods focus more on measuring static skills or knowledge, DA takes a more dynamic, process-oriented approach by measuring a student's capacity for growth (Feuerstein et al., 1979; Lidz, 1987). Through DA, it is possible to evaluate how students respond to instruction, resulting in a more nuanced understanding of a child's strengths and weaknesses. This information enables educators to tailor educational interventions or strategies to individual students, especially those from diverse or underrepresented populations. Furthermore, DA might be able to paint a clearer picture of the student's future performance compared to static measures (Dixon et al., 2023; Tzuriel, 2021a).

Theoretical Background

DA is grounded in Vygotsky's Sociocultural Theory of Cognitive Development (Tzuriel, 2021b; Vygotsky, 1978), which emphasizes the importance of mediation and social interaction on the child's cognitive development. An important concept in this theory is the Zone of Proximal Development: the difference between a child's actual developmental level (individual performance) and their potential development (supported performance). DA is also based on Feuerstein's theory of Structural Cognitive Modifiability (SCM) (Feuerstein et al., 1979), stating that similar to Vygotsky's theory, intelligence and cognitive structures are modifiable through intentional, guided interactions, also called Mediated Learning Experiences (MLEs). One of the first operationalizations of DA was Feuerstein's Learning Potential Assessment Device (LPAD), which allows the assessment of the child's cognitive modifiability, also called learning potential, by determining their responsiveness to structured mediation during cognitive tasks (Feuerstein et al., 1979). Lastly, DA is closely related to constructivist theories of learning, which emphasize that student actively build their understanding and knowledge by engaging with their environment through experiences, context, and interaction (Bada, 2015). Based on this perspective, educational assessments should not only evaluate outcomes, like in static measures but also focus on the learning processes (e.g., DA). By doing this, DA provides a more holistic and inclusive assessment method, especially for students from underrepresented groups.

Methods and Techniques of Dynamic Assessment

A variety of related methods and techniques fall under the dynamic assessment approach, for example, mediated learning, inquiry-based learning, and analogical reasoning. Mediated learning, a concept based on the theory of Mediated Learning Experience (Feuerstein et al., 1991), is a method that involves a mediator or instructor who guides the student through a task to increase the student's understanding and internalization of novel concepts and strategies. This method assesses a student's responsiveness to instruction. Inquiry-based learning is a teaching method that encourages students to explore and ask questions, promoting active engagement in the learning process (Pedaste et al., 2015). This method is closely related to DA because it involves the interaction between the student and the teacher to improve the student's understanding of a concept. Lastly, analogical reasoning is a technique that makes use of analogies to increase an individual's comprehension of a new concept or situation (Stevenson et al., 2013). By determining how many analogies the individual solves, it is possible to assess the student's learning potential or cognitive modifiability.

Dynamic Assessment versus Traditional Assessments

DA seems to play a significant role in performance prediction by assessing a child's ability to adapt and grow when given the appropriate support and guidance (Hamers et al., 1993). In this review, performance prediction refers to estimating a student's future scholastic outcomes, ranging from one day to multiple years. In the educational context, when performance prediction is determined by static measures, it is possible that students' learning capacities are overlooked. If a child shows great learning potential, it suggests that, with continued support, the child is likely to show improved performance in future tasks or other educational settings (Caffrey et al., 2008). Whereas traditional assessments focus on current knowledge and abilities, DA focuses more on future-oriented, giving it predictive capabilities (Hamers et al., 1993). This predictive capability might help educators identify students who might underperform on traditional assessment but are able to achieve high levels of success in the right learning environment (Caffrey et al., 2008; Elliott et al., 2018; Tiekstra et al., 2016).

Because DA focuses on learning potential and responsiveness to mediation, it is highly relevant to both the general student population and students from more diverse backgrounds. While DA might be beneficial for all students, it might be particularly valuable for diverse student populations, which are often misrepresented by traditional assessments (Cho et al., 2014; Hasson et al., 2012; Navarro et al., 2018). In this review, diverse student populations include students from culturally, linguistically, and socioeconomically diverse backgrounds, those with learning disabilities, and high-performing students. Traditional assessments might not fully capture these student's capabilities because they are affected by factors such as culture, language, or learning needs (Meijer, 2001; Petersen et al., 2018; Tzuriel, 2021a). Because DA takes a student's learning potential and responsiveness to instruction into consideration, it is less likely to be biased and offers a more equitable assessment tool (Peña et al., 2006; Salas et al., 2013). Therefore, this review examines DA's applicability in the general student population while also taking a look at its role in providing a fair assessment for diverse learners.

Educational Transitions

Educational transitions refer to critical periods where a student moves from one educational stage to another, for example, from primary to secondary school (Jindal-Snape, 2023). During these periods, significant changes in social environments, academic expectations, and personal development play an important role. Especially during these periods, the predictive validity of DA can be helpful (Caffrey et al., 2008; Tzuriel, 2021a). DA may provide insights into which students may need additional resources, allowing educators to tailor interventions that help bridge the gaps between educational stages but also facilitate smoother transitions while maximizing the student's potential for success in the new educational environment (Feuerstein et al., 1979; Hamers et al., 1993).

Challenges of Practical Application

Even though DA has been recognized for its potential benefits in theory and research, its practical application seems to be limited. (De Beer, 2006; Elliott et al., 2018; Haywood & Lidz, 2006; Mohammadi & Babaii, 2022). This gap between research and practice raises critical questions about DA's real-world applicability and the obstacles to its utilization by educators and policymakers. There are several factors that potentially contribute to the gap between research and practice, including DA's resource intensity, lack of knowledge, and limited research into its utilization, specifically during educational transitions (Elliott et al., 2018; Mohammadi & Babaii, 2022). Many educators may be unfamiliar with dynamic assessment methods or do not know how to integrate them into the current curricula (Haywood & Tzuriel, 2002; Karpov & Tzuriel, 2009). Legal requirements, such as the Americans with Disabilities Act that requires tests to accommodate students with disabilities, might make educators hesitant to apply new forms of assessment to their policies (U.S. Department of Justice Civil Rights Division, n.d.).

Research Questions

Following the gap between research and practice, the overarching research question of this scoping review is: "What is the current application of dynamic assessment regarding performance prediction in education?". This main question is divided into two sub-questions:

1. "How is dynamic assessment applied in the general student population in relation to performance prediction?";
2. "How is dynamic assessment applied in diverse student populations, such as students with disabilities, gifted students, or students with a migration background, in relation to performance prediction?".

Application, in this scoping review, refers to the practical implementation of DA as a predictive tool in educational settings. This includes the methods and tools utilized, the contexts and populations in which it was practiced, the types of outcomes predicted, and the overall purpose of DA, such as identification, future performance prediction, or instructional planning. Moreover, the advantages and challenges of implementation of DA into the educational context will be considered.

This scoping review is part of a broader, ongoing, practice-focused research project by Het Landelijke Expertisecentrum PO-VO (LEPOVO) and Hanze University of Applied Sciences. The project aims to support schools in easing the transition of students between primary and secondary education in the Netherlands. By providing a solid scientific foundation, this project aims to contribute to the previously mentioned issues, including growing inequality, overlooking students' abilities, and misrepresentation of diverse student groups. This scoping review extends beyond examining the role of DA in the context of educational transitions. It also explores the application of DA for scholastic performance prediction across a range of educational settings. By taking a broader perspective, this scoping review highlights the variety of DA's applications, for instance, as a diagnostic, informative, or predictive tool, across various populations.

Methodology

Study Design and Research Aim

This scoping review follows the methodological framework outlined by Arksey and O'Malley (2005) and adheres to the reporting guidelines provided by the PRISMA-ScR statement to enhance transparency (see Appendix A; Tricco et al., 2018). The following steps from the Arksey and O'Malley framework were taken: 1) Identifying the research question, 2) Identifying relevant studies, 3) Study selection, 4) Data charting, and 5) Collating, Summarizing, and Reporting Results.

This scoping review aims to answer the research question, providing clarity on the application of DA. Understanding how DA is currently implemented in real-world settings may reveal insights about its applicability with regard to educational transitions. Furthermore, this scoping review seeks to identify gaps in the literature, which could guide future research to gain a greater understanding of the application of DA. Through synthesizing existing research, this review aims to bridge the gap between research and practice, potentially offering valuable information to educators and policymakers. A scoping review is particularly relevant for this study since it allows an extensive exploration of the literature surrounding DA. Generally, systematic reviews aim to answer highly specific questions based on the PICO model, which helps to structure a research question by specifying the Population, Intervention(s), Comparison(s) (if any), and Outcomes of interest (Eriksen & Frandsen, 2018). In contrast, scoping reviews are intended to map key concepts, theories, methodologies, or applications of a topic across various studies (Munn et al., 2018). As a result, scoping reviews typically address broader research questions and provide a comprehensive overview of the existing literature.

Search Strategy

This scoping review will focus on research that examines the use of DA in educational settings for performance prediction. DA is closely linked to various methodologies, such as learning potential assessment, mediated learning, inquiry-based learning, and analogical reasoning. These assessment methods measure learning potential as its outcome. However, some studies do not measure learning potential but related constructs, for example, cognitive modifiability, zone of proximal development, or responsiveness to instruction. Since these concepts are similar, they are also included in the scope of this study. Furthermore, various terms are used to describe performance prediction, including academic achievement, predictive validity, selection, and school tracking, which are all included in the scope of this study.

The search for this scoping review was conducted through the following databases: *ERIC*, *PsycINFO*, and *Web of Science*. The search string consists of the following terms: ("learning potential assess*"OR “dynamic test*" OR "dynamic assess*" OR "mediated learn*"

OR "inquiry* based learning" OR "analog* problem* solving" OR "process assess*" OR "figural analog*" OR "analog* reason*") AND (“learning potential*” OR "learning abilit*" OR "talent*" OR "cogniti* potential*" OR "cogniti* flexib*" OR "modifiabilit*" OR "cogniti* develop*" OR "achievement*" OR "zone of proximal development" OR "potential for learning" OR "responsive* to intervention*" OR "responsive* to instruct*" OR "intervention effect*") AND (“child*” OR "student*" OR "adolescen*") AND ("prediction*" OR "performance*" OR " predict* achievement*" OR "predictive validity" OR "construct validity" OR "select*" OR "tracking"). These terms were searched in titles, abstracts, and keywords. As this review focuses on research that is currently available, unpublished work or preprints are not included.

Eligibility Criteria

Before screening the search results, eligibility criteria were set to ensure the inclusion of relevant literature in line with the research questions. Studies were screened and selected based on the inclusion and exclusion criteria presented in Table 1.

Study Selection Process

Figure 1 presents the flowchart of the scoping review process. Using the previously described search string, a database search was conducted. Filters were applied to include only peer-reviewed studies published in English between 2000 and 2025. During this database search, 581 records were identified and collected until February 11th, 2025. Following, 181 duplicates were identified and removed using the software Rayyan. As part of the screening phase, the title and abstract of 400 records were screened based on the previously mentioned inclusion and exclusion criteria, again utilizing the software Rayyan. In this phase, 38 out of 400 records were selected for full-text screening. Of the 38 reports, 12 reports were excluded due to various reasons, including 1) not regarding performance prediction, 2) wrong population, 3) sample size too small, 4) full text not available. After assessing the reports for eligibility, 26 studies were included in this scoping review.

Table 1

Eligibility Criteria

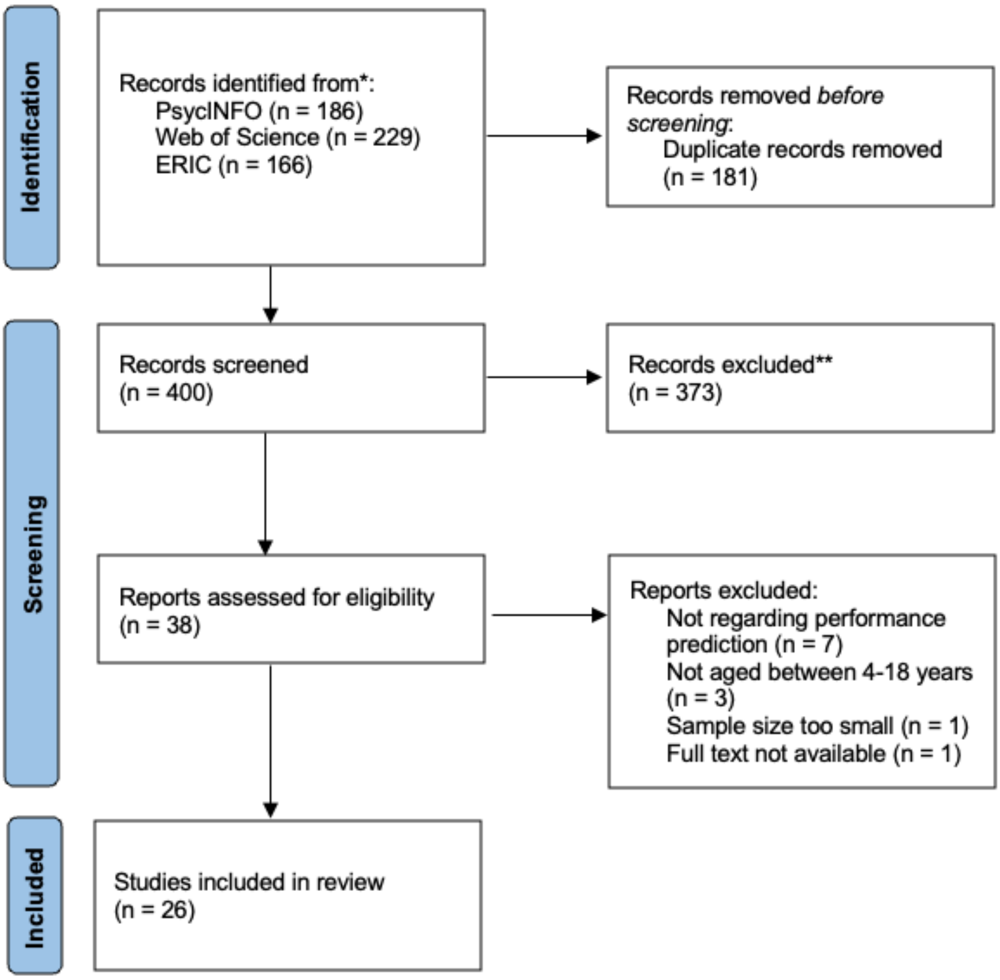
Eligibility Criteria	Inclusion Criteria	Exclusion Criteria
Concept	Studies discussing Dynamic Assessment (DA) in relation to performance prediction	Studies that do not focus on DA and/or its predictive value
Population	Kindergarten, primary and secondary education (4-18 years old); general student populations and diverse student groups (e.g., students with disabilities, gifted students, students with a migration background)	Studies focusing on non-educational populations (e.g., workplace setting, clinical patients) or with participants out of the age range 4-18 years old.
Publication Type	Peer-reviewed, empirical studies	Theoretical papers, books, clinical studies, opinion pieces, non-peer-reviewed sources
Language	English studies	Non-English studies (unless translated)
Publication year	2000-2025	Studies published before 2000
Sample size	>15	<15

Data Extraction, Synthesis, and Analysis

A total of 26 articles were selected to be further analyzed in this scoping review. The data were extracted in relation to three key areas: 1) study characteristics, including geographical location, study design, and sample details; 2) the scope of dynamic assessment applications, including terminology, assessment domains, tools used, prediction period, and type, and predicted outcomes; 3) qualitative findings focusing on the predictive validity, key results, and factors related to implementation. The extracted data were organized and presented in a structured table using a standardized charting form. Following the data extraction, descriptive statistics and qualitative content analysis were conducted. Descriptive statistics were applied to summarize important characteristics of the studies. The qualitative content analysis was conducted to identify themes, patterns, and gaps in the literature regarding the utilization of DA for performance prediction. The data synthesis was guided by this scoping review's objectives and research questions and focused on a comprehensive overview of existing literature.

Figure 1

Flowchart of The Scoping Review Process



Note. PRISMA, 2020.

Results

Descriptive Overview of Included Studies

This section provides a descriptive summary of the 26 studies included in this scoping review. An overview of the main characteristics of the studies is presented, including their geographical location, study design, sample demographics, and aspects of DA studies, such as DA terminology, domain focus, assessment tool, and prediction period. A summary of these descriptive results can be found in Appendix C.

Study Characteristics

The 26 studies included in this review were conducted in various areas of the world. Half of the studies were conducted in Europe (N=13, 50.0%), followed by North America (N=8, 30.7%). Further, three studies were conducted in Asia (11.5%), one study in Africa (3.8%), and one study in South America (3.8%).

All 26 studies were quantitative. A combination of methodological designs was used in these studies, including the following features: longitudinal (N=12, 46.2%), pretest-training-posttest (N=11, 42.3%), quasi-experimental (N=10, 38.5%), correlational (N=9, 34.6%), control group (N=7, 26.9%), experimental (N=5, 19.2%), and repeated measures (N=4, 15.4%). Most studies used a combination of these design features (N=23, 88.5%) with very few repeated designs, for example, longitudinal correlational design with stratified random sampling (N=2, 7.7%).

One of the inclusion criteria of this scoping review was that studies examine the application of DA for performance prediction. Of the included studies, 21 (80.8%) focused primarily on DA's predictive validity. Out of these 21, eight studies also included a secondary research aim, including categorization (N=3, 37.5%) and the influence of other factors, such as language (N=5 62.5%). Of the 26 studies, the remaining 5 studies had a variety of research aims, including solely categorization (N=3, 11.5%) and other objectives like tool development (N=2, 7.7%)

Sample Characteristics. The sample sizes varied greatly, ranging from 24 to 368 participants. Of the 26 studies, 10 (38.4%) had a total sample size of fewer than 100 participants and 16 (61.5%) had a total sample size of more than 100 participants. The participants can be categorized into three different age groups: 4-6 years old (N=3, 11.5%), 6-12 years old (N=22, 84.6%), and 13-18 years old (N=6, 23.1%). Some researchers included multiple experimental groups of different ages or conducted a longitudinal study over an extended period of time, causing participants to fall into multiple age categories.

Regarding the population, 12 out of the 26 studies (46.2%) examined the application of DA for performance prediction in a general student population. The other 14 studies (53.8%) conducted research on a diverse student population. These 14 studies analyzed one (N=12, 85.7%) or a combination (N=2, 14.3%) of diverse student populations, namely students from low-income families with limited English (L2) proficiency. Of the 14 studies, eight (57.1%) focused on students with or at risk for learning impairments, including language impairments (N=2, 25.0%), reading and/or math difficulties (N=4, 50.0%), and general learning difficulties (N=2, 25.0%). Three out of 14 studies (21.4%) examined students from low-income families, two (14.3%) looked into the predictive validity of DA for low-and high-performing students, and two studies (14.3%) focused on students with limited second language proficiency. One study (7.1%) had a female-only sample.

Dynamic Assessment Aspects

Terminology. A variety of terminology was used in the included studies to describe dynamic assessment (DA). A majority of the studies used ‘dynamic assessment’ as their main term (N=20, 76.9%). Another five studies (19.2%) used the term ‘dynamic testing’ and one study (3.8%) used the term “dynamic measures”. The definitions used by the included studies can be categorized into different dimensions. All studies (N=26, 100.0%) mention a form of interaction or mediation between the assessor and student. Out of 26 studies, 22 (84.6%) explicitly mentioned that DA is used to assess the learning potential of the student. Next, 17 articles (65.4%) described the instruction part of DA as being embedded in the assessment itself. About half of the studies (N=14, 53.8%) clearly emphasized the learning process during DA over the product of the assessment. Furthermore, nine studies (34.6%) pointed out the instructional use of DA results in their definition. For example, insights into a student's learning potential can help to create or refine psychoeducational interventions (Navarro et al., 2018). Four studies (15.4%) included the structure and/or format of DA in the definition. Lastly, three studies (11.5%) explicitly mentioned the Vygotskian foundations of DA.

Domain. With regard to the domain that the studies focused on, DA was most frequently applied in the domain of language (N=19, 73.1%), including reading (N=11, 57.9%) and spelling (N=2, 10.5%). Two out of 19 studies (10.5%) included both reading and spelling. The language domain is closely followed by mathematics (N=13, 50.0%). Six out of 26 studies (23.1%) look into the predictive validity of DA in relation to general school performance. Five studies (19.2%) examined other domains, including reasoning, attention, and 'Life Orientation'. Most studies focused on one domain (N=16, 61.5%), while other studies applied DA in several domains simultaneously (N=10, 38.5%). For example, Touw et al. (2019) examined the predictive validity of DA in the domains of mathematics, technical reading, and spelling.

Assessment Tools. The assessment tools used to test DA varied greatly across the studies. The majority of the studies (N=15, 57.7%) utilized a domain-specific DA tool, such as decoding DA (measure focusing on decoding skills, N=3, 11.5%), the EDPL battery (measure focusing on reading skills; N=2, 7.7%), or BEDA (measures focusing on novel math content; N=2, 7.7%). Out of 15 studies, 11 studies (73.3%) used a language-specific DA tool, including seven (63.6%) studies that used a reading-specific DA tool. The remaining four domain-specific

DA tools (26.7%) were focused on mathematics. Other researchers applied tools that measure learning potential by taking a domain-general approach (N=11, 42.3%). Also, out of 26 studies, seven studies (26.9%) made use of a computerized format, including both domain-general and domain-specific tools.

Prediction Moment, Horizon, and Type. Table 2 summarizes the 26 studies by prediction moment, horizon, and type. To examine the predictive validity of DA for school performance, 22 studies (84.6%) employed predictive measures, meaning that the criterion measures were collected after DA administration. Three studies (11.5%) used concurrent measures, which are measured prior to DA, and one study (3.8%) combined both types. Since concurrent measures refer to previously assessed outcomes, DA's prediction horizon is irrelevant for these studies.

Table 2

Predictive Validity in DA-studies (N=26)

Category	Subcategory	Number of Studies	Percentage
Prediction moment	Predictive	22	84.6%
	Concurrent	3	11.5%
	Mixed	1	3.8%
Prediction horizon (of 22 studies)	Intra-session	3	13.6%
	Short-term	2	9.1%
	Medium-term	6	27.3%
	Long-term	8	36.4%
	Very long-term	2	9.1%
	Unknown	2	9.1%
Prediction type	Continuous	19	73.1%
	Classification	7	26.9%

Note. The studies included in the prediction horizon are the 22 studies with a predictive prediction moment.

Among the 22 studies with predictive measures, the prediction horizons varied. Three studies (13.6%) collected outcomes during the same session as DA, two (9.1%) within one day to two months (short-term), six (27.3%) within two to six months (medium-term), and eight (36.4%) within seven months to two years (long-term). Two out of 22 studies (9.1%) used very long-term horizons of over two years. Two other studies (9.1%) did not specify the length of their prediction period but noted that DA and outcome measures were collected in the same academic year. There seems to be no relation between the prediction horizon and the distribution of effect sizes.

Furthermore, studies were categorized by prediction types: 1) continuous outcome prediction, where DA scores predicted numerical outcomes, such as test scores, grades, or teacher ratings, and 2) classification prediction, where DA differentiated between groups (e.g., at-risk vs. not at risk). Of the 26 studies, 19 studies (73.1%) focus on continuous outcomes, while seven (26.9%) analyzed classification accuracy.

Statistical Analysis. The included studies utilized various statistical methods to examine the predictive validity of DA compared to conventional measures. Half used a single analysis type (N=13, 50.0%), while the other half used multiple (N=13, 50.0%). Most studies (N=20, 76.9%) applied a form of regression analysis, including hierarchical, stepwise, multiple, logistic, and ordinal regression. In eight studies (30.8%), researchers used correlation analyses. Seven studies (26.9%) employed advanced modeling techniques, such as path analysis, growth modeling, multilevel modeling, structural equation modeling, and growth mixture modeling. Four studies (15.4%) utilized classification or discriminant analysis, and one study (3.8%) applied a two-way ANCOVA.

Applications of Dynamic Assessment for Performance Prediction

This section presents the key findings of the included articles that concern the application of DA to predict scholastic performance. This way it directly addresses the research question "What is the current application of dynamic assessment regarding performance prediction?". Besides the overall findings, this section also takes a look at key findings in the general student population versus diverse student populations, such as multilingual learners, children from low socioeconomic backgrounds, and children with learning difficulties or disabilities.

Predictive Validity and Key Findings

Almost all studies (N=25, 96.2%) found DA to be a significant predictor of (a domain) of school performance. For instance, Fabio (2005) reported significant effects in both language and mathematics across primary and secondary education. Only Lauchlan & Elliott (2001) did not find DA to significantly predict reading, math, and non-verbal reasoning abilities overall, except for a subgroup with high learning potential receiving cognitive intervention.

Effect Sizes. Across the 25 studies with significant results, effect sizes varied. Ten studies (40.0%) reported multiple effect sizes. Two studies (8.0%) found significant but small effects regarding DA's predictive validity, eight (32.0%) reported small to medium effects, and another eight studies (32.0%) observed medium effects. Medium to large effects were demonstrated in 11 studies (44.0%), while another 11 studies (44.0%) noted large effects. Overall, DA seemed to show slightly larger effect sizes among diverse student populations, including multilingual learners, children from low socioeconomic backgrounds, and those with learning difficulties or disabilities, compared to general student populations. Exploratory analysis revealed a median effect size of medium for general samples and medium to large for diverse samples. Furthermore, studies using domain-specific tools seemingly showed a slight difference in overall effect sizes compared to studies using domain-general tools, namely medium-large compared to medium median effect sizes.

Confounding Factors. Nine studies (34.6%) explored factors that potentially influence the DA's effectiveness, including language (N=3, 11.5%), sociocultural background (N=2, 7.7%), socio-economic status (N=1, 3.8%), gender (N=1, 3.8%) previous learning experience (N=1, 3.8%), prior reading instruction (N=1, 3.8%) test anxiety (N=1, 3.8%), and types of scaffolding (N=1, 3.8%). Two studies (7.7%) examined multiple factors. Language and types of scaffolding were found to significantly affect DA outcomes, while sociocultural background only moderated static measures. Additionally, one study found that cognitive flexibility is a significant mediator between DA and educational outcomes.

Comparison with Static Measures. All (N=26, 100.0%) of the included studies compared DA to static measures. Of these, 22 studies (84.6%) demonstrated that DA had additional value over static measures. In 12 of those 22 studies (55.5%), DA uniquely explained additional variance in educational performance. Six of 22 studies (27.3%) reported greater predictive validity or classification accuracy for DA, and four studies (18.2%) identified DA as the strongest predictor in their models. Only one out of the 26 studies (3.8%) showed that static measures had greater predictive validity than the dynamic measure used, with the dynamic measure showing no significant predictive value.

Implementation Advantages, Challenges, and Authors' Recommendations

All 26 articles mentioned at least two advantages of DA over static measures. These advantages can be grouped into eight categories, namely DA 1) provides additional information (N=17, 65,4%), particularly about instructional needs (N=13, 50,0%), 2) shows improved group differentiation (N=14, 53,8%), 3) offers equal or better predictive power (N=12, 46,2%), 4) captures learning potential rather than momentary performance (N=12, 46,2%), 5) is less biased (N=10, 38,5%), 6) is time-/cost-effective (N=5, 19,2%), 7) reduces false positives (N=4, 15,4%), and 8) other advantages (N=7, 26,9%). The advantages of DA that fall into the "Others" category include DA: provides contextualized test content related to the domain assessed, is tailored to the child's level, compensates for lack of experience, facilitates classroom transfer, is easy to interpret when it is domain-specific, allows for repeated testing, and its processes and results can be generalized.

Next, 16 out of 26 studies (61.5%) discussed the limitations of implementing DA in education, with 75% mentioning only one. The most common concern (N=9, 34.6%) is about the time-consuming nature of DA. Other noted issues included questions about its validity, reliability, and generalizability (23.1%), the need for extensive training (11.5%), and unclear definitions (7.7%). Additional single-mention limitations included challenges for younger children, neglect of affective factors, the need for clear-cut results for policymakers, the complexity of DA, potential masking of learning potential by extraneous circumstances, and the risk of labeling students leading to lowered expectations and resourcing.

The majority of studies (N=24, 92.3%) gave advice about implementing DA in education, with some addressing multiple areas (N=6, 23.1%). Eight studies (30.8%) recommended using DA as a supplement to static measures to enhance efficiency and reduce false positives. Another eight (N=8, 30.8%) highlighted the application of DA to gain insight into the instructional needs of the students. Five studies (19.2%) emphasized making DA more practical and time-efficient, with three (60.0%) suggesting computerized formats. Seven studies already used computerized DA in their research. Out of the 26 studies, three studies (11.5%) emphasized contextualizing the DA measures to ensure applicability, while two (7.7%) recommended its use to improve the prediction and identification accuracy. Another two (7.7%) implied that more research was needed before widespread implementation. Three studies (11.5%) offered unique advice, including targeting cognitive flexibility in DA measures to increase its applicability and predictive value, prioritizing learning potential over static scores to encourage students to optimize their potential, and combining qualitative and quantitative measures for a deeper understanding of students' learning potential.

Discussion

This chapter presents a critical analysis of the key findings in relation to the overarching research question and its associated sub-questions. The primary aim of this review was to determine how DA is currently used in the educational context in regard to performance prediction of students, comparing general and diverse student populations.

Key Trends Related to Performance Prediction

This scoping review has found that DA has been widely used across educational domains, ranging from reading and mathematics to logical reasoning and general scholastic performance. Across the included studies, there was a close-to-even distribution of general versus diverse student populations taken as samples. Most studies have shown that DA has significant predictive validity over several school performance measures. The majority of these studies showed that DA has additional value beyond conventional, static measures in relation to performance prediction.

Overall Findings

The consistent finding that DA is a significant predictor for later school performance illustrates its utility in educational assessment. The added value of assessing learning potential beyond static measures seems to be demonstrated by the various studies that found that DA uniquely explained some variance that was not explained by static measures. This suggests that it captures elements of a child's learning capacity, for instance, cognitive flexibility or modifiability, that are overlooked by static measures. Furthermore, effect sizes differed greatly across studies. This may indicate that the effectiveness of DA depends on the specific DA tool used or the population assessed.

Confounding Factors

Language dominance has been identified as a potential confounding factor in at least two of the included studies (Cho et al., 2020; Seethaler et al., 2016). When DA was conducted in a child's non-dominant language, the predictive value seemed to be reduced. It is possible that language processing leads to a different interpretation of provided instruction or that increased cognitive demands have an influence on the student's performance level (Meyer, 2000; Pozzan & Trueswell, 2016). Language dominance is an important factor that should be taken into account when administering DA to linguistically diverse populations. Language-appropriate adaptations might be necessary, such as reducing the language complexity of the assessment (Li & Suen, 2012).

Additionally, the type of scaffolding affected the predictive validity in one study (Luković et al., 2022). This study found that DA focusing on motivational-affective scaffolding is a better predictor of language, mathematics, and overall school performance compared to DA with cognitive scaffolding. Because only one study looked into the effect of scaffolding type, further research should be conducted to determine the effects on DA's effectiveness. Other potentially confounding factors examined in the included studies did not show a significant relation with DA.

Short-Term and Long-Term Dynamic Assessment

Among the studies with predictive measures, great temporal diversity was found, ranging from intra-session to very long-term periods. This diversity demonstrates the wide utility of DA in education. On the one hand, studies with a short-term prediction horizon illustrate the potential of DA as a practical tool for identifying the immediate needs of children. By focusing on how a child responds to support in the moment, short-term DA can help educators tailor interventions that address current learning challenges and help students reach their full potential (Bosma & Resing, 2012; Gustafson et al., 2014; Navarro & Mora, 2011).

On the other hand, studies with a long-term prediction horizon highlight the utilization of DA as a prediction tool that allows schools to determine a student's future potential and track their learning potential over time. Predicting students' developmental trajectories can aid in educational planning, selection, and efficient resource allocation (Tzuriel, 2021a). Additionally, the long-term predictive validity of DA can be important for guiding high-stakes transitions, such as in the Dutch education system, where students are placed into different secondary education tracks at the end of primary school. The potential ability of DA to forecast academic performance two or more years ahead opens up the possibility for DA to be used as a tool to support more equitable and individualized placements. This is especially valuable for students whose performance is underestimated by traditional, static tests due to cultural, linguistic, or socioeconomic factors (Meijer, 2001; Petersen et al., 2018; Tzuriel, 2021a). By looking at learning potential rather than current achievement alone, DA can provide a more informed decision, allowing the placement of students in an educational pathway that matches their capabilities and growth trajectories (Tzuriel, 2021a).

Dynamic Assessment in General and Diverse Student Populations

No major differences in DA's effectiveness have been observed between the general and diverse student populations, including multilingual learners, children from low socioeconomic backgrounds, and children with learning difficulties or disabilities. In both populations, DA has been shown to be a significant predictor of academic performance. Overall, DA seemed to show slightly larger effect sizes among diverse student populations compared to general student populations, suggesting the need for fair assessment tools in these populations. Whereas static measures often underrepresent the abilities of children from diverse backgrounds, DA's wide applicability allows for a more accurate and equitable assessment (Orellana et al., 2019; Petersen et al., 2018; Salas et al., 2013).

Implementation of Dynamic Assessment

Besides evaluating the predictive validity of DA, it is also important to consider how DA can be implemented in educational settings. This section addresses a key dimension of the main research question by examining the drivers and barriers of the application of DA for predicting student performance in education.

Advantages of Dynamic Assessment

Various practical and theoretical benefits of DA have been highlighted by the included studies. Several studies indicated that DA is less biased than static measures. This is in line with past research which has shown that DA is less influenced by cultural and socioeconomic factors compared to static measures (Meijer, 2001; Petersen et al., 2018; Tzuriel, 2021a). Assessing learning potential through guided support rather than measuring current knowledge or skill sets, bypasses confounding factors that influence static measures (Le et al., 2023). This helps to reduce the risk of under- or over-identification of students from diverse backgrounds, reducing false negatives and false positives (Cho et al., 2020; D. Fuchs et al., 2011; L. S. Fuchs et al., 2011; Hasson et al., 2012). These results demonstrate the usefulness of DA as a less biased information source that helps to provide a more nuanced understanding of children's developmental trajectory, whether they are typically or non-typically developing students.

Furthermore, half of the studies emphasized that DA provides additional information in regard to a child's instructional needs. Its process-oriented approach reveals how a student learns, including the intensity and type of instruction needed, which provides a more nuanced understanding of the child's cognitive processes (Tzuriel, 2021a). This allows for more individualized and effective teaching strategies, which might be especially beneficial for diverse student groups, as shown by previous research (Gustafson et al., 2014; Navarro & Mora, 2011; Resing, 2013). However, DA administration requires specialized training and can be more time-consuming than conventional methods, limiting its feasibility in educational settings. Nonetheless, the potential for DA to provide unique information about student's academic trajectories represents a significant contribution to educational practice.

Although DA has been criticized for being resource-intensive (Afshari et al., 2020; Haywood & Tzuriel, 2002; Petersen et al., 2017), several studies in this review have found that recent adaptations, like computerized delivery, have improved its efficiency (de Beer, 2011;

Elliott et al., 2018; Resing et al., 2012; K. W. J. Touw et al., 2019). These advancements offer a solution to this implementation barrier. This is especially valuable in schools with limited resources. Future research should explore further adaptations to enhance the feasibility of DA in educational settings.

Barriers Regarding the Implementation of Dynamic Assessment

Besides DA’s resource intensity, additional limitations include the validity, reliability, and generalizability of DA, the need for specialized training, and ethical considerations. First, this review illustrates the wide variation in DA approaches, which limits comparability and generalizability (Holzmeister et al., 2024). Studies differ in their formats, assessment tools, scaffolding, and scoring criteria, which makes it difficult to compare results across studies. This variability also leads to uncertainty about whether findings from one context or population will be replicated in others. This complexity makes it challenging to develop clear and universal conclusions and guidelines that can be confidently applied across different students, ages, and educational systems.

Second, implementing DA requires specialized training, which can be a barrier, especially in under-resourced contexts. While many educators are trained in traditional assessments, DA involves interactive and adaptive techniques that require a deeper understanding of the underlying principles. Concepts such as graduated prompts, types of scaffolding, and responsive feedback, must be applied with skill to ensure reliable outcomes (Grigorenko, 2009).

Finally, the interpretation of DA results raises critical ethical concerns. If a student is labeled as having ‘low learning potential’, it is possible that negative assumptions about their capabilities are promoted (Bosma & Resing, 2008; Elliott, 2003). As a result, teachers or other adults might lower their expectations of the students and, consequently, provide fewer learning opportunities. This is opposite to the desired effect of DA, where the student receives support where they need it. Therefore, it is essential that DA results are interpreted carefully and constructively to promote positive educational outcomes and avoid the reinforcement of limiting stereotypes.

Recommendations About Implementation

The majority of the studies (92.3%) included in this review provided practical recommendations on the implementation of DA in educational settings. Firstly, DA was recommended to be used as a supplement rather than a replacement for static assessments. As a supplement, it might improve early identification of students at risk for developing learning difficulties by unmasking their latent learning potential overlooked by static measures (Cho et al., 2020; L. S. Fuchs et al., 2011; Gan et al., 2023; Stad et al., 2018; Swanson & Howard, 2005). Additionally, combining static measures with DA may enhance predictive accuracy while reducing the practical constraints of DA, such as time consumption and other associated costs (de Beer, 2011; D. Fuchs et al., 2011; Lu & Hu, 2019; Seethaler et al., 2016). Other included studies also mentioned that DA offers valuable additional information regarding the child's learning needs and strategies that will aid in instructional planning on group and individual levels (de Beer, 2011; Gan et al., 2023; Petersen & Gillam, 2015; Touw et al., 2017). This illustrates not only its potential as an assessment tool but also as an information source for teaching approaches. Another way to reduce the practical constraints that come with the implementation of DA, various studies emphasized the need to make DA more time- and cost-efficient. As mentioned before, this can be achieved by adopting computerized versions of DA. If well-designed, these formats can match the efficiency of static measures, while maintaining the informational depth (Hidri & Roud, 2020; Siengyen & Wasanasomsithi, 2024).

General Versus Diverse Student Populations

As shown throughout this review, DA offers predictive value across both general and diverse student populations, but its application and impact vary. To conclude previously mentioned applications and challenges, in the general student population, DA might aid in instructional planning by highlighting individual learning potential (Bosma & Resing, 2012; Gustafson et al., 2014; Navarro & Mora, 2011). However, barriers such as resource intensity and the need for specialized training can hinder widespread implementation. Adaptations like computerized DA or using DA as a supplement may help overcome these obstacles.

For diverse student populations, particularly those from culturally and socioeconomically varying backgrounds, DA's emphasis on learning potential and process-based evaluation helps lower the risk of misclassification and inappropriate placement compared to static measures (Cho et al., 2020; D. Fuchs et al., 2011; L. S. Fuchs et al., 2011; Hasson et al., 2012). These advantages illustrate the value of DA as a prediction tool for these groups. However, the influence of language dominance on DA outcomes should be considered, since students with limited proficiency in the language of instruction might be at a disadvantage. Furthermore, proper interpretation of DA outcomes is crucial to avoid reinforcing stereotypes, particularly for students who show low responsiveness to instruction (Bosma & Resing, 2008; Elliott, 2003). Overall, the different applications of DA across populations showcase the importance of contextualized implementation and highlight the need for standardized, but adaptive tools that match the needs of diverse learners while maintaining DA's predictive value (de Beer, 2011; Navarro et al., 2018; Navarro & Mora, 2011; Petersen & Gillam, 2015; Touw et al., 2019).

Limitations of This Scoping Review

This scoping review examined the current application of DA in relation to educational performance prediction. However, several limitations should be acknowledged. Due to the nature of the scoping review, a formal quality appraisal or risk of bias assessment has not been conducted. Consequently, the methodological rigor of the included studies remains uncertain, which has limited the ability to draw solid conclusions about the overall strength of the evidence.

Next, much heterogeneity can be found in the studies' assessment tools, target populations, and conceptualizations of DA. While this allows for a broad understanding of the topic, it makes it hard to generalize the results. This heterogeneity in the studies might be associated with the use of varied search terms in this scoping review. This broad coverage of search terms was necessary to capture the diversity in terminology but likely introduced conceptual variability into the results. Future research could focus on the effectiveness of specific aspects or methods of DA to avoid this heterogeneity.

Lastly, the review only included peer-reviewed studies that are published in English. As a result, relevant research in other languages or found in gray literature might be overlooked. Furthermore, due to the time constraints of this Master's thesis, the database selection was limited to ERIC, PsycINFO, and Web of Science, resulting in the exclusion of potentially relevant studies indexed elsewhere. Additionally, the number of studies that were screened on title and abstract was restricted due to practical time constraints, which may have further narrowed the scope of this review. To avoid publication bias, future research should include gray literature besides published work.

Limitations of The Included Studies

While the included studies provide valuable insights into the current application of dynamic assessment, several limitations need to be taken into account when interpreting their findings. The majority of the studies focused on students in primary education. While providing useful information, early detection of learning difficulties and performance prediction at later educational stages seem to be underexplored. Early identification is important since it allows for timely intervention, resulting in improved, long-term educational outcomes (Cho et al., 2020; L. S. Fuchs et al., 2011; Gan et al., 2023; Stad et al., 2018; Swanson & Howard, 2005). Similarly, the limited research on the predictive validity of DA in later educational stages prevents a deeper understanding of students' developmental trajectories. In turn, this restricts

DA's informative value regarding educational placements and critical transition periods. Research across the full educational span is important to get the full picture of the possibilities of DA as an assessment tool.

Moreover, the diverse student populations examined by the included studies mostly included low-performing children, like children with learning difficulties. Only two studies looked at high-performing children (Calero et al., 2011; Touw et al., 2017). This imbalance creates uncertainty about DA's performance across the full spectrum of students. Additionally, the included research has predominantly focused on academic domains like reading and mathematics, while other domains, such as geography and history, remain underexplored. Future research should include a broader range of student profiles and domains to better evaluate DA's applicability and benefits across various student populations.

Several studies reported having a small sample size as a limitation, which might have affected the reliability and generalizability of the results. Furthermore, methodological variability, as mentioned before, is common among the included studies. Differences in assessment formats and reporting can be found, which complicates the comparison of study results and makes it hard to draw a definite conclusion.

Implications for Future Research

These limitations highlight areas for future research that can deepen our understanding of DA in education. Studies should include larger and more diverse samples to improve reliability and generalizability. Research on young children could clarify DA's value as an early detection tool, while studies on its role in educational placements and transitions may reveal insight into DA's impact on long-term academic success. Expanding research beyond reading and mathematics will help to determine the extent of DA's applicability. Additionally, exploring potential confounding factors, like language dominance, further will help to refine the limits of DA as an assessment tool. Efforts to standardize DA tools can improve consistency across studies, aiding in the generalizability of results (Navarro et al., 2018; Navarro & Mora, 2011; Touw et al., 2019). Finally, research into adaptations of DA might reduce implementation barriers and provide a fair assessment.

Implications for Educational Practice

Dynamic Assessment, as a more interactive, equitable, and informative tool, holds significant potential to reshape educational assessment practices. With its predictive validity and reduced risk of bias, it complements traditional, static assessments. DA supports early identification, instructional planning, and placement decisions, especially for diverse or underrepresented students whose potential may be underestimated by conventional tests (Bosma & Resing, 2012; Cho et al., 2020; L. S. Fuchs et al., 2011; Gustafson et al., 2014; Navarro & Mora, 2011; Stad et al., 2018; Tzuriel, 2021a).

Beyond prediction, DA provides rich qualitative insight into how a child learns, offering a deeper understanding of a child's cognitive processes and instructional needs (Luković et al., 2022). However, the application of DA in high-stakes situations, for instance, school placement decisions, should be approached with caution. Labeling students as having "low learning potential" without appropriate support and careful interpretation might reinforce inequalities (Bosma & Resing, 2008; Elliott, 2003). Differences in school systems and instructional methods mean that low or high responsiveness may not directly translate into a specific educational track.

Therefore, DA is best applied as a supplemental tool, as suggested by various studies included in this review. While DA might not be ideal to use as the sole determinant of future performance, it can help tailor interventions for struggling students and support nuanced, data-informed educational decisions.

Conclusion

Dynamic Assessment appears promising as a more equitable and informative tool for predicting student performance and supporting educational decision-making, compared to static measures. However, some limitations have also been presented, including methodological and conceptual heterogeneity, limited focus on older students, and underrepresentation of certain diverse populations, such as gifted students. To provide further evidence for the application of DA in the prediction of educational performance, future studies should focus on larger and more diverse samples, look at various educational domains, and prioritize the standardization of DA administration. From this review, it can be concluded that DA should be used to complement rather than replace conventional, static assessments. Its ability to capture a student's learning potential, produce insight into individual needs, and reduce bias accentuates its value in promoting personalized and fair educational evaluations. To realize DA's potential, it must be applied thoughtfully, keeping in mind practical constraints and ethical considerations.

References

Afshari, H., Amirian, Z., & Tavakoli, M. (2020). Applying group dynamic assessment procedures to support EFL writing development: Learner achievement, learners' and teachers' perceptions. *Journal of Writing Research*, *11*(3), 445–467. https://doi.org/10.17239/jowr-2020.11.03.02

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bada, D. S. O. (2015). Constructivism Learning Theory: A Paradigm for Teaching and Learning. *Journal of Research & Method in Education*, *5*(6), 66–70. https://doi.org/10.9790/7388-05616670

Bosma, T., & Resing, W. C. M. (2008). Bridging the Gap Between Diagnostic Assessment and Classroom Practice. *Journal of Cognitive Education and Psychology*, *7*(2), 174–198. https://doi.org/10.1891/194589508787381854

Bosma, T., & Resing, W. C. M. (2012). Need for instruction: Dynamic testing in special education. *European Journal of Special Needs Education*, *27*(1), 1–19. https://doi.org/10.1080/08856257.2011.613599

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The Predictive Validity of Dynamic Assessment: A Review. *The Journal of Special Education*, *41*(4), 254–270. https://doi.org/10.1177/0022466907310366

Calero, M. D., Belen, G.-M. M., & Robles, M. A. (2011). Learning Potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. *Learning and Individual Differences*, *21*(2), 176–181. https://doi.org/10.1016/j.lindif.2010.11.025

Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to Tier 2 intervention. *Journal of Learning Disabilities*, *47*(5), 409–423. psyh. https://doi.org/10.1177/0022219412466703

Cho, E., Fuchs, L. S., Seethaler, P. M., Fuchs, D., & Compton, D. L. (2020). Dynamic assessment for identifying Spanish-speaking English learners' risk for mathematics disabilities: Does language of administration matter? *Journal of Learning Disabilities*, *53*(5), 380–398. psyh. https://doi.org/10.1177/0022219419898887

De Beer, M. (2006). Dynamic testing: Practical solutions to some concerns. *SA Journal of Industrial Psychology*, *32*(4), Article 4. https://doi.org/10.4102/sajip.v32i4.245

de Beer, M. (2011). The role of the Learning Potential Computerised Adaptive Test (LPCAT) in the vocational guidance assessment of adolescents. *Educational and Child Psychology*, *28*(2), 114–129. psyh.

Dixon, C., Oxley, E., Gellert, A. S., & Nash, H. (2023). Dynamic assessment as a predictor of reading development: A systematic review. *Reading and Writing: An Interdisciplinary Journal*, *36*(3), 673–698. psyh. https://doi.org/10.1007/s11145-022-10312-3

Elliott, J. G. (2003). Dynamic Assessment in Educational Settings: Realising potential. *Educational Review*, *55*(1), 15–32. https://doi.org/10.1080/00131910303253

Elliott, J. G., Resing, W. C. M., & Beckmann, J. F. (2018). Dynamic assessment: A case of unfulfilled potential? *Educational Review*, *70*(1), 7–17. https://doi.org/10.1080/00131911.2018.1396806

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: A systematic review. *Journal of the Medical Library Association : JMLA*, *106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *The Journal of General Psychology*, *132*(1), 41–64. https://doi.org/10.3200/GENP.132.1.41-66

Feuerstein, R., Klein, P. S., & Tannenbaum, A. J. (1991). *Mediated Learning Experience (MLE): Theoretical, Psychosocial and Learning Implications*. Freund Publishing House Ltd.

Feuerstein, R., Rand, Y., Hoffman, M. B., & Vig, E. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore: University Park Press.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The Construct and Predictive Validity of a Dynamic Assessment of Young Children Learning to Read: Implications for RTI Frameworks. *Journal of Learning Disabilities*, *44*(4), 339–347. https://doi.org/10.1177/0022219411407864

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011). Two-Stage Screening for Math Problem-Solving Difficulty Using Dynamic Assessment of Algebraic Learning. *Journal of Learning Disabilities*, *44*(4), 372–380. https://doi.org/10.1177/0022219411407867

Gan, Y., Zhang, J., Kharabi-Yamato, L., Su, Y., Zhang, J., Jiang, Y., Hui, Y., & Li, H. (2023). The Unique Predictive Value of Dynamic Assessment of Character Decoding in Reading Development of Chinese Children from Grades 1 to 2. *Scientific Studies of Reading*, *27*(3), 215–231. https://doi.org/10.1080/10888438.2022.2143271

Gellert, A. S., & Elbro, C. (2017). Try a Little Bit of Teaching: A Dynamic Assessment of Word Decoding as a Kindergarten Predictor of Word Reading Difficulties at the End of Grade 1. *Scientific Studies of Reading*, *21*(4), 277–291. https://doi.org/10.1080/10888438.2017.1287187

Grigorenko, E. L. (2009). Dynamic Assessment and Response to Intervention. *Journal of*

Learning Disabilities, *42*(2), 111–132. https://doi.org/10.1177/0022219408326207

Gustafson, S., Svensson ,Idor, & and Fälth, L. (2014). Response to Intervention and Dynamic Assessment: Implementing Systematic, Dynamic and Individualised Interventions in Primary School. *International Journal of Disability, Development and Education*, *61*(1), 27–43. https://doi.org/10.1080/1034912X.2014.878538

Hamavandi, M., Rezai ,Mohammad Javad, & and Mazdayasna, G. (2017). Dynamic assessment of morphological awareness in the EFL context. *Cogent Education*, *4*(1), 1324254. https://doi.org/10.1080/2331186X.2017.1324254

Hamers, J. H. M., Sijtsma, K., & Ruijssenaars, A. J. J. M. (1993). *Learning potential assessment: Theoretical, methodological and practical issues*. Amsterdam: Swets & Zeitlinger.

Hasson, N., Dodd, B., & Botting, N. (2012). Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for the assessment of syntax in children with language impairments. *International Journal of Language & Communication Disorders*, *47*(3), 285–299. https://doi.org/10.1111/j.1460-6984.2011.00108.x

Haywood, H. C., & Lidz, C. S. (2006). *Dynamic Assessment in Practice: Clinical and Educational Applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511607516

Haywood, H. C., & Tzuriel, D. (2002). Applications and Challenges in Dynamic Assessment. *Peabody Journal of Education*, *77*(2), 40–63. https://doi.org/10.1207/S15327930PJE7702_5

Hidri, S., & Roud, L. F. P. (2020). Developing and using hints in computerized dynamic assessment of a TOEFL *iBT* reading exam. *Heliyon*, *6*(9), e04985. https://doi.org/10.1016/j.heliyon.2020.e04985

Holzmeister, F., Johannesson, M., Böhm, R., Dreber, A., Huber, J., & Kirchler, M. (2024). Heterogeneity in effect size estimates. *Proceedings of the National Academy of Sciences*, *121*(32), e2403490121. https://doi.org/10.1073/pnas.2403490121

Jindal-Snape, D. (2023). Multiple and multi-dimensional educational and life transitions: Conceptualization, theorization and XII pillars of transitions. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education (Fourth Edition)* (pp. 530–543). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.14060-6

Karpov, Y. V., & Tzuriel, D. (2009). Dynamic Assessment: Progress, Problems, and Prospects. *Journal of Cognitive Education and Psychology*, *8*(3), 228–237. https://doi.org/10.1891/1945-8959.8.3.228

Kozulin, A. (2011). Learning potential and cognitive modifiability. In *Addressing Issues of Access and Fairness in Education through Dynamic Assessment* (2nd ed., Vol. 18, pp. 169–181). Routledge. https://doi-org.proxy-ub.rug.nl/10.1080/0969594X.2010.526586

Lauchlan, F., & Elliott, J. (2001). The psychological assessment of learning potential. *British Journal of Educational Psychology*, *71*(4), 647–665. psyh. https://doi.org/10.1348/000709901158712

Le, H., Ferreira, J. M., & Kuusisto, E. (2023). Dynamic assessment in inclusive elementary education: A systematic literature review of the usability, methods, and challenges in the past decade. *European Journal of Special Education Research*, *9*(3), 94–125. https://doi.org/10.46827/ejse.v9i3.5009

Li, H., & Suen, H. K. (2012). The Effects of Test Accommodations for English Language Learners: A Meta-Analysis. *Applied Measurement in Education*, *25*(4), 327–346. https://doi.org/10.1080/08957347.2012.714690

Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to evaluating learning potential*. New York: Guilford Press.

Lu, Y.-Y., & Hu, C.-F. (2019). Dynamic assessment of phonological awareness in young foreign language learners: Predictability and modifiability. *Reading and Writing: An Interdisciplinary Journal*, *32*(4), 891–908. psyh. https://doi.org/10.1007/s11145-018-9894-x

Luković, S., Marinković, B., & Kostić, M. Z. (2022). The zone of actual and the zone of proximal development measured through preschool dynamic assessment as predictors of later school performance—A longitudinal study. *Psihologija*, *55*(1), 89–105. psyh.

Meijer, J. (2001). Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. *Anxiety, Stress & Coping*, *14*(3), 337–362. https://doi.org/10.1080/10615800108248361

Meijer, J., & Elshout, J. J. (2001). The predictive and discriminant validity of the zone of proximal development. *British Journal of Educational Psychology*, *71*(1), 93–113. psyh. https://doi.org/10.1348/000709901158415

Meyer, L. M. (2000). Barriers to Meaningful Instruction for English Learners. *Theory Into Practice*, *39*(4), 228–236. https://doi.org/10.1207/s15430421tip3904_6

Ministerie van Onderwijs, Cultuur en Wetenschap. (2024, April 29). *Verplichte doorstroomtoets basisonderwijs—Schooladvies en doorstroomtoets in groep 8—Rijksoverheid.nl* [Onderwerp]. Ministerie van Algemene Zaken. https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-doorstroomtoets-basisschool/verplichte-doorstroomtoets-basisonderwijs

Mohammadi, S., & Babaii, E. (2022). The Dynamic Assessment of Reading Comprehension: An Exploration of EFL Teacher's Perception and Practice. *Journal of Modern Research in English Language Studies*, *Online First*. https://doi.org/10.30479/jmrels.2020.12998.1607

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018).

Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Navarro, J.-J., & Mora, J. (2011). Analysis of the implementation of a dynamic assessment device of processes involved in reading with learning-disabled children. *Learning and Individual Differences*, *21*(2), 168–175. https://doi.org/10.1016/j.lindif.2010.11.008

Navarro, J.-J., Mourgues-Codern, C., Guzmán, E., Rodríguez-Ortiz, I. R., Conejo, R., Sánchez-Gutiérrez, C., de la Fuente, J., Martella, D., & Saracostti, M. (2018). Integrating Curriculum-Based Dynamic Assessment in Computerized Adaptive Testing: Development and Predictive Validity of the EDPL-BAI Battery on Reading Competence. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01492

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The Use of Dynamic Assessment for the Diagnosis of Language Disorders in Bilingual Children: A Meta-Analysis. *American Journal of Speech-Language Pathology*, *28*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic Assessment of School-Age Children’s Narrative Ability: An Experimental Investigation of Classification Accuracy. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1037–1057. https://doi.org/10.1044/1092-4388(2006/074)

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017).

Dynamic Assessment of Narratives: Efficient, Accurate Identification of Language Impairment in Bilingual Students. *Journal of Speech, Language, and Hearing Research: JSLHR*, *60*(4), 983–998. https://doi.org/10.1044/2016_JSLHR-L-15-0426

Petersen, D. B., & Gillam, R. B. (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, *48*(1), 3–21. psyh. https://doi.org/10.1177/0022219413486930

Petersen, D. B., Gragg, S. L., & Spencer, T. D. (2018). Predicting Reading Problems 6 Years Into the Future: Dynamic Assessment Reduces Bias and Increases Classification Accuracy. *Language, Speech, and Hearing Services in Schools*, *49*(4), 875–888. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism (Cambridge, England)*, *19*(3), 636–643. https://doi.org/10.1017/S1366728915000838

Resing, W. C. M. (2013). Dynamic Testing and Individualized Instruction: Helpful in Cognitive Education? *Journal of Cognitive Education and Psychology*, *12*(1), 81–95. https://doi.org/10.1891/1945-8959.12.1.81

Resing, W. C. M., Stevenson, C. E., & Bosma, T. (2012). Dynamic testing: Measuring inductive reasoning in children with developmental disabilities and mild cognitive impairments. *Journal of Cognitive Education and Psychology*, *11*(2), 159–178. psyh. https://doi.org/10.1891/1945-8959.11.2.159

Salas, N., González, F., & Assael, C. (2013). The contribution of dynamic assessment to promote inclusive education and cognitive development of soclo-economically deprived children with learning disabilities. *Erdélyi Pszichológiai Szemle*, *Spec Issue*, 207–222. psyh.

Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2016). Does the Value of Dynamic

Assessment in Predicting End-of-First-Grade Mathematics Performance Differ as a Function of English Language Proficiency? *The Elementary School Journal*, *117*(2), 171–191. https://doi.org/10.1086/688870

Siengyen, C., & Wasanasomsithi, P. (2024). Development of a Computerized Dynamic Reading Assessment Program to Measure English Reading Comprehension of Thai EFL Undergraduate Students. *rEFLections*, *31*(3), 1216–1248. https://doi.org/10.61508/refl.v31i3.277414

Stad, F. E., Van Heijningen, C. J. M., Wiedl, K. H., & Resing, W. C. M. (2018). Predicting school achievement: Differential effects of dynamic testing measures and cognitive flexibility for math performance. *Learning and Individual Differences*, *67*, 117–125. https://doi.org/10.1016/j.lindif.2018.07.006

Stevenson, C. E., Hickendorff, M., Resing, W. C. M., Heiser, W. J., & de Boeck, P. A. L. (2013). Explanatory item response modeling of children’s change on a dynamic test of analogical reasoning. *Intelligence*, *41*(3), 157–168. https://doi.org/10.1016/j.intell.2013.01.003

Swanson, H. L., & Howard, C. B. (2005). Children with Reading Disabilities: Does Dynamic Assessment Help in the Classification? *Learning Disability Quarterly*, *28*(1), 17–34. psyh. https://doi.org/10.2307/4126971

Tiekstra, M., Minnaert, A., & Hessels, M. G. P. (2016). A review scrutinising the consequential validity of dynamic assessment. *Educational Psychology*, *36*(1), 112–137. https://doi.org/10.1080/01443410.2014.915930

Touw, K., Vogelaar, B., Verdel, R., Bakker, M., & Resing, W. (2017). Children’s progress in solving figural analogies: Are outcomes of dynamic testing helpful for teachers? *Educational and Child Psychology*, *34*(1), 21–38. https://doi.org/10.53841/bpsecp.2017.34.1.21

Touw, K. W. J., Vogelaar, B., Thissen, F., & Resing, W. C. M. (2019). Dynamic testing of children with a developmental language disorder: Progression in series completion performance and need for instruction. *Educational and Child Psychology*, *36*(1), 27–39. https://doi.org/10.53841/bpsecp.2019.36.1.27

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., Garritty, C., … Straus, S. E. (2018). *PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation*. https://doi.org/10.7326/M18-0850

Tzuriel, D. (2021a). Dynamic Assessment (DA) of Learning Potential. In D. Tzuriel (Ed.), *Mediated Learning and Cognitive Modifiability* (pp. 69–88). Springer International Publishing. https://doi.org/10.1007/978-3-030-75692-5_4

Tzuriel, D. (2021b). The Socio-Cultural Theory of Vygotsky. In D. Tzuriel (Ed.), *Mediated Learning and Cognitive Modifiability* (pp. 53–66). Springer International Publishing. https://doi.org/10.1007/978-3-030-75692-5_3

U.S. Department of Justice Civil Rights Division. (n.d.). *Americans with Disabilities Act of 1990, As Amended*. ADA.Gov. Retrieved June 4, 2025, from https://www.ada.gov/law-and-regs/ada/

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Appendix A

PRISMA-ScR Checklist

Table A

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	1
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
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.	7-12
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.	11-12
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.	12-13
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.	15
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.	15-16
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	13-14
Selection of sources of evidence	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	14-15
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	15-16

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
		done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	15-16
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).	N.A.
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	15-16
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.	14-15
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	Appendix C + D
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	N.A.
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.	Appendix E
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	17-24
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.	25-31
Limitations	20	Discuss the limitations of the scoping review process.	31-32
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.	25-35
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.

Appendix B

Search String Per Database

Web of Science:

TS=(("learning potential assess*"OR “dynamic test*" OR "dynamic assess*" OR "mediated learn*" OR "inquiry* based learning" OR "analog* problem* solving" OR "process assess*" OR "figural analog*" OR "analog* reason*") AND (“learning potential*” OR "learning abilit*" OR "talent*" OR "cogniti* potential*" OR "cogniti* flexib*" OR "modifiabilit*" OR "cogniti* develop*" OR "achievement*" OR "zone of proximal development" OR "potential for learning" OR "responsive* to intervention*" OR "responsive* to instruct*" OR "intervention effect*") AND (“child*” OR "student*" OR "adolescen*") AND ("prediction*" OR "performance*" OR " predict* achievement*" OR "predictive validity" OR "construct validity" OR "select*" OR "tracking"))

ERIC en PsychINFO:

(TI("learning potential assess*"OR “dynamic test*" OR "dynamic assess*" OR "mediated learn*" OR "inquiry* based learning" OR "analog* problem* solving" OR "process assess*" OR "figural analog*" OR "analog* reason*") OR AB("learning potential assess*"OR “dynamic test*" OR "dynamic assess*" OR "mediated learn*" OR "inquiry* based learning" OR "analog* problem* solving" OR "process assess*" OR "figural analog*" OR "analog* reason*") OR SU("learning potential assess*"OR “dynamic test*" OR "dynamic assess*" OR "mediated learn*" OR "inquiry* based learning" OR "analog* problem* solving" OR "process assess*" OR "figural analog*" OR "analog* reason*"))

AND

(TI(“learning potential*” OR "learning abilit*" OR "talent*" OR "cogniti* potential*" OR "cogniti* flexib*" OR "modifiabilit*" OR "cogniti* develop*" OR "achievement*" OR "zone of proximal development" OR "potential for learning" OR "responsive* to intervention*" OR "responsive* to instruct*" OR "intervention effect*") OR AB(“learning potential*” OR "learning abilit*" OR "talent*" OR "cogniti* potential*" OR "cogniti* flexib*" OR "modifiabilit*" OR "cogniti* develop*" OR "achievement*" OR "zone of proximal development" OR "potential for learning" OR "responsive* to intervention*" OR "responsive* to instruct*" OR "intervention effect*") OR SU(“learning potential*” OR "learning abilit*" OR "talent*" OR "cogniti* potential*" OR "cogniti* flexib*" OR "modifiabilit*" OR "cogniti* develop*" OR "achievement*" OR "zone of proximal development" OR "potential for learning" OR "responsive* to intervention*" OR "responsive* to instruct*" OR "intervention effect*"))

AND

(TI(“child*” OR "student*" OR "adolescen*") OR AB(“child*” OR "student*" OR "adolescen*") OR SU(“child*” OR "student*" OR "adolescen*"))

AND

(TI("prediction*" OR "performance*" OR " predict* achievement*" OR "predictive validity" OR "construct validity" OR "select*" OR "tracking") OR AB("prediction*" OR "performance*" OR " predict* achievement*" OR "predictive validity" OR "construct validity" OR "select*" OR "tracking") OR SU("prediction*" OR "performance*" OR " predict* achievement*" OR "predictive validity" OR "construct validity" OR "select*" OR "tracking"))

Appendix C

Table of Dynamic Assessment Definitions

Table C

Dynamic Assessment Definitions

Reference	Interaction	Assess learning potential	Embedded instruction	Process	Instructional use	Format of DA	Vygotskian foundations
(Calero et al., 2011)	✓	✓	-	-	-	✓	-
(Cho et al., 2014)	✓	✓	✓	-	-	-	-
(Cho et al., 2020)	✓	✓	-	✓	-	-	✓
(de Beer, 2011)	✓	✓	✓	✓	-	-	-
(Fabio, 2005)	✓	✓	✓	-	-	-	-
(D. Fuchs et al., 2011)	✓	-	-	✓	-	-	-
(L. S. Fuchs et al., 2011)	✓	✓	✓	-	-	-	-
(Gan et al., 2023)	✓	✓	✓	✓	-	-	-
(Gellert & Elbro, 2017)	✓	✓	✓	-	-	-	-
(Hamavandi et al., 2017)	✓	-	✓	-	-	-	-
(Hasson et al., 2012)	✓	✓	✓	✓	-	✓	-
(Lauchlan & Elliott, 2001)	✓	✓	✓	-	-	-	✓
(Lu & Hu, 2019)	✓	✓	✓	-	-	-	✓
(Luković et al., 2022)	✓	✓	✓	-	✓	-	-

Reference	Interaction	Assess learning potential	Embedded instruction	Process	Instructional use	Format of DA	Vygotskian foundations
(Meijer & Elshout, 2001)	✓	-	✓	-	-	✓	-
(Navarro & Mora, 2011)	✓	-	-	✓	✓	-	-
(Navarro et al., 2018)	✓	✓	✓	✓	✓	-	-
(Peña et al., 2006)	✓	-	✓	✓	-	-	-
(Petersen & Gillam, 2015)	✓	✓	✓	✓	-	✓	-
(Resing et al., 2012)	✓	✓	-	-	✓	-	-
(Seethaler et al., 2016)	✓	✓	-	-	✓	-	-
(Stad et al., 2018)	✓	✓	✓	✓	✓	-	-
(Stevenson et al., 2013)	✓	✓	✓	✓	✓	-	-
(Swanson & Howard, 2005)	✓	✓	-	✓	✓	-	-
(Touw et al., 2017)	✓	✓	-	✓	-	-	-
(Touw et al., 2019)	✓	✓	✓	✓	✓	-	-

Appendix D

Descriptive Results Tables

Table D1

Study Characteristics of the Included Studies

Reference	Location	Design	Sample Size	Age range	Population	Research Aim
(Calero et al., 2011)	Spain	Correlational two-group comparative design with pretest-training-posttest structure	127 (75F, 52M)	7-11 years	General	Predictive validity
(Cho et al., 2014)	USA	RCT with longitudinal predictive	134 (65F, 67M, 2 unidentified)	6-7 years*	Diverse: at-risk for reading disability	Predictive validity
(Cho et al., 2020)	USA	RCT with predictive correlational	368 (179F, 189M)	6-7 years*	Diverse: low-income and limited L2 proficiency	Predictive validity + Other factors
(de Beer, 2011)	South Africa	Non-experimental correlational predictive	262 (161F, 100M, 1 unidentified)	14-15 years*	General	Predictive validity
(Fabio, 2005)	Italy	Quasi-experimental pretest-intervention-posttest	150 (68F, 82M)(Exp 1), 287 (157F, 130M)(Exp 2), 198 (?F, ?M)(Exp 3)	4-18 years	General	Predictive validity + Other factors
(D. Fuchs et al., 2011)	USA	Longitudinal predictive correlational	318 (50.6%F, 49.4%M)	6-7 years*	Diverse: low-performing	Predictive validity + Categorization

Reference	Location	Design	Sample Size	Age range	Population	Research Aim
(L. S. Fuchs et al., 2011)	USA	Longitudinal predictive experimental design with a two-stage screening network	122 (55F, 67M)	8-13 years*	General	Categorization
(Gan et al., 2023)	China	Longitudinal predictive correlational	135 (62F, 73M)	6-8 years*	General	Predictive validity + Categorization
(Gellert & Elbro, 2017)	Denmark	Longitudinal predictive correlational	171 (?F, ?M)	6-8 years*	Diverse: at-risk for reading disability	Predictive validity
(Hamavandi et al., 2017)	Iran	Quasi-experimental pretest-intervention-posttest design with control group	25 (25F, 0M)	14-18 years	Diverse: female-only	Predictive validity
(Hasson et al., 2012)	UK	Quasi-experimental repeated-measures	24 (3F, 21M)	8-10 years	Diverse: language impairment	Tool development
(Lauchlan & Elliott, 2001)	UK	Quasi-experimental pretest-intervention-posttest design with control group	30 (17F, 13M)	M = 9.0 years**	Diverse: learning difficulties	Other factors
(Lu & Hu, 2019)	Taiwan	Correlational	50 (?F, ?M)	9-10 years*	General	Predictive validity
(Luković et al., 2022)	Serbia	Longitudinal correlational	114 (60F, 54M)	4-14 years*	General	Predictive validity + Other factors

Reference	Location	Design	Sample Size	Age range	Population	Research Aim
(Meijer & Elshout, 2001)	Netherlands	Longitudinal experimental within-subject design	Between 158 and 305, depending on data completeness (51%F, 49%M)	14-17 years	General	Predictive validity
(Navarro & Mora, 2011)	Spain	Quasi-experimental pretest-intervention-posttest design with control group	60 (23F, 37M)	9-16 years	Diverse: reading disability	Predictive validity
(Navarro et al., 2018)	Chile	Correlational longitudinal	324 (46%F, 54%M)	8-12 years	General	Predictive validity
(Peña et al., 2006)	USA	Pretest-intervention-posttest control group experimental design	58 (37F, 21M)(Exp 1), 71 (?F, ?M)(Exp 2)	6-8 years*	Diverse: low-income	Categorization
(Petersen & Gillam, 2015)	USA	Longitudinal predictive correlational	63 (34F, 29M)	5-6 years*	Diverse: at-risk for language impairment	Predictive validity + Categorization
(Resing et al., 2012)	Netherlands	Repeated measures randomized blocking	30 (1F, 29M)	8-10 years*	Diverse: learning impairment	Predictive validity
(Seethaler et al., 2016)	USA	Longitudinal predictive correlational	292 (153F, 139M)	6-7 years*	Diverse: low-income and limited L2 proficiency	Predictive validity
(Stad et al., 2018)	Netherlands	Experimental pretest-training-posttest control group design	152 (76F, 76M)	6-8 years*	General	Predictive validity + Other factors

Reference	Location	Design	Sample Size	Age range	Population	Research Aim
(Stevenson et al., 2013)	Netherlands	Longitudinal predictive correlational design with repeated measures	188 (100F, 88M)	5-8 years*	General	Predictive validity
(Swanson & Howard, 2005)	USA	Cross-sectional quasi-experimental design	70 (39F, 31M)	10-12 years*	General	Predictive validity + Other factors
(Touw et al., 2017)	Netherlands	Quasi-experimental pretest-training-posttest design	80 (?F, ?M)	6-7 years*	Diverse: low- and high-performing	Categorization
(Touw et al., 2019)	Netherlands	Experimental pretest-training-posttest control group design	164 (89F, 75M)	7-8 years*	General	Predictive validity

Table D2

Dynamic Assessment Aspects of the Included Studies

Reference	Terminology	Domain	Assessment Tool	Prediction moment	Prediction horizon	Prediction type	Statistical Analysis
(Calero et al., 2011)	Dynamic Assessment	Logic and reasoning	Learning Potential Assessment Device (LPAD) (domain-general)	Concurrent	N.A.	Classification	Discriminant Analysis
(Cho et al., 2014)	Dynamic Assessment	Reading	Decoding Dynamic Assessment (domain-specific: reading)	Predictive	Medium-term	Continuous	Individual Growth Modeling
(Cho et al., 2020)	Dynamic Assessment	Mathematics	Balancing Equations Dynamic Assessment* (domain-specific: mathematics)	Predictive	Medium-term	Continuous	Multigroup Path Analysis
(de Beer, 2011)	Dynamic Assessment	English, Life Orientation, Mathematics, and general academic ability	Learning Potential Computerized Adaptive Test (LPCAT) (domain-general)	Predictive	Did not specify	Continuous	Multiple Regression Analysis + Correlation Analysis
(Fabio, 2005)	Dynamic Testing	Attention and General school performance	Self-created dynamic test (domain-general)	Mixed	Intra-session	Continuous	Correlational Analysis
(D. Fuchs et al., 2011)	Dynamic Assessment	Reading	Decoding Dynamic Assessment (domain-specific: reading)	Predictive	Medium-term	Continuous	Correlation Analysis + Multilevel Modeling

Reference	Terminology	Domain	Assessment Tool	Prediction moment	Prediction horizon	Prediction type	Statistical Analysis
(L. S. Fuchs et al., 2011)	Dynamic Assessment	Mathematics	Dynamic Assessment for Algebra Skills (domain-specific: mathematics)	Predictive	Very long-term	Classification	Logistic Regression Analysis
(Gan et al., 2023)	Dynamic Assessment	Reading	Dynamic Assessment of character decoding (domain-specific: reading)	Predictive	Long-term	Continuous	Latent Growth Modeling + Growth Mixture Modeling
(Gellert & Elbro, 2017)	Dynamic Assessment	Reading	Decoding Dynamic Assessment (domain-specific: reading)	Predictive	Long-term	Continuous	Hierarchical Stepwise Logistic Regression Analysis
(Hamavandi et al., 2017)	Dynamic Assessment	Reading	Dynamic Assessment Task of Morphological Analysis (DATMA) (domain-specific: language)	Predictive	Medium-term	Continuous	Exploratory Multiple Regression Analysis
(Hasson et al., 2012)	Dynamic Assessment	Language	Dynamic Assessment of Sentence Structure (DASS) (domain-specific: language)	Predictive	Long-term	Continuous	Correlational Analysis
(Lauchlan & Elliott, 2001)	Dynamic Assessment	Academic ability, reading, arithmetic, and non-verbal reasoning ability	Children's Analogical Thinking Modifiability test* (CATM) (domain-general)	Predictive	Long-term	Classification	Two-way ANCOVA

Reference	Terminology	Domain	Assessment Tool	Prediction moment	Prediction horizon	Prediction type	Statistical Analysis
(Lu & Hu, 2019)	Dynamic Assessment	Spelling	Dynamic Phonological Awareness* (domain-specific: language)	Predictive	Intra-session	Continuous	Hierarchical Regression Analysis
(Luković et al., 2022)	Dynamic Assessment	Language, mathematics, general school performance	TIP-1 (domain-general)	Predictive	Very long-term	Continuous	Multiple Stepwise Logistic Regression Analysis + Correlation Analysis
(Meijer & Elshout, 2001)	Dynamic Assessment	Mathematics	Mathematics learning test (domain-specific: mathematics)	Predictive	Long-term	Continuous	(Stepwise) Multiple Regression Analysis
(Navarro & Mora, 2011)	Dynamic Assessment	Academic performance and language	EDPL (domain-specific: reading)	Predictive	Short-term	Continuous	Hierarchical Multiple Regression Analysis + Stepwise Regression Analysis
(Navarro et al., 2018)	Dynamic Assessment	Reading	EDPL-BAI battery (domain-specific: reading)	Predictive	Medium-term	Continuous	Hierarchical Linear Regression Analysis
(Peña et al., 2006)	Dynamic Assessment	Narration (language)	Dynamic Assessment of Narratives (domain-specific: language)	Predictive	Short-term	Classification	Classification Analysis + Discriminant Function Analysis
(Petersen & Gillam, 2015)	Dynamic Assessment	Reading	Dynamic Assessment of Nonsense Word Decoding (domain-specific: reading)	Predictive	Long-term	Classification	Linear Multiple Regression Analysis + Classification Analysis

Reference	Terminology	Domain	Assessment Tool	Prediction moment	Prediction horizon	Prediction type	Statistical Analysis
(Resing et al., 2012)	Dynamic Testing	General school performance, mathematics, language	AnimaLogica test* (domain-general)	Concurrent	N.A.	Continuous	Correlational Analysis
(Seethaler et al., 2016)	Dynamic Assessment	Mathematics	Balancing Equations Dynamic Assessment (BEDA) (domain-specific: mathematics)	Predictive	Long-term	Continuous	Regression Analysis for each group
(Stad et al., 2018)	Dynamic Testing	Mathematics	Dynamic series completion test (domain-general)	Predictive	Did not specify	Continuous	Hierarchical Multinomial Regression Analysis
(Stevenson et al., 2013)	Dynamic Testing	Mathematics and reading	AnimaLogica test (domain-general)	Predictive	Medium-term	Continuous	Multilevel Modeling
(Swanson & Howard, 2005)	Dynamic Assessment	Mathematics and reading	S-CPT* (domain-general)	Predictive	Intra-session	Classification	Hierarchical Regression Analysis
(Touw et al., 2017)	Dynamic Testing	Mathematics, language, reasoning	AnimaLogica test* (domain-general)	Concurrent	N.A.	Classification	Correlational Analysis
(Touw et al., 2019)	Dynamic Testing	Mathematics, technical reading, spelling	Dynamic series completion test* (domain-general)	Predictive	Long-term	Continuous	Ordinal Regression Analysis + Correlational Analysis

Appendix E

Qualitative Content Analysis Tables

Table E1

Summary of Findings of the Included Studies Regarding the Predictive Validity of Dynamic Assessment

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Calero et al., 2011)	DA effectively discriminated between gifted and non-gifted students, with gifted student showing greater learning potential. The most reliable predictor was all three DA subtests taken together.	Small	-	Yes
(Cho et al., 2014)	DA of decoding significantly predicted both final performance in word identification fluency over a 14-week Tier 2 intervention. DA accounted for 3-13% unique variance in responsiveness, outperforming static decoding measures and Tier 1 growth indicators. Also, it remained predictive after controlling for prereading skills.	Medium-large	-	Yes
(Cho et al., 2020)	DA significantly predicts mathematics outcomes, explaining 5-8% unique variance beyond that accounted for by static measures. Language dominance is a significant moderator.	Small-medium	Language	Yes
(de Beer, 2011)	DA post-test scores correlated significantly with aggregate academic performance. Combining the DA and static scores predicted 35.3% of the variance in academic performance. DA post-test alone accounted for 12.9%, while the static measure accounted for 29.2%. DA did not outperform the static measure.	Medium	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Fabio, 2005)	Kindergarten (Exp1): The DA scores correlated more strongly with attention than static measures and was not influenced by sociocultural background. Primary school (Exp2): DA scores had a stronger correlation with attention and school performance than static measures. Sociocultural background only affected the static measures. Teenagers (Exp3): DA scores significantly correlated with mathematics scores but not language scores.	Medium (3x), Medium-large (2x)	Sociocultural background	Yes
(D. Fuchs et al., 2011)	High correlations between DA and reading outcomes were observed, but correlations between the static measure and reading outcomes were even higher. Both the static and dynamic measures are significant predictors of the reading outcomes. Together, they explained between 48.5-63.6%. DA is a significant predictor for word identification and reading comprehension, explaining 3.3-5.6% unique variance.	Small-medium, Large	-	Yes
(L. S. Fuchs et al., 2011)	Using DA in a two-stage model resulted in better classification accuracy (specificity = 70.4%, hit rate = 73.8%). The two-stage model outperformed the single-stage models that used static screeners alone (specificity = 48.0%, hit rate = 55.7%) in predicting which students would perform poorly at the end of the year. The two-stage model significantly reduced false positives by 57%, improving the efficiency of screening.	Large	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Gan et al., 2023)	DA of character decoding in Grade 1 uniquely predicted the final level and growth rate of Chinese character reading from Grade 1 to Grade 2, even after controlling for static predictors. DA showed stronger predictive power than static measures, with rapid naming being the only static predictor significantly associated with character reading outcomes. Learning potential subgroups could be mapped onto distinct reading trajectories, with slow gainers at risk for persistent difficulties.	Small-medium, Medium-large	-	Yes
(Gellert & Elbro, 2017)	DA of decoding significantly predicted Grade 1 reading difficulties, even after controlling for traditional predictors, adding 4% unique variance. When entered after only the most common predictors, DA contributed an additional 11% variance. DA of decoding was highly correlated with reading-specific skills but weakly with vocabulary and nonverbal IQ. DA was the single strongest predictor of Grade 1 reading outcomes, predicting 48% variance in reading status. When combined with static measures, this improved to 53%.	Medium, Large	-	Yes
(Hamavandi et al., 2017)	DA is the strongest predictor of reading comprehension compared to static predictors, explaining 36% variance. When entered in the regression together, the static measure was excluded as a non-significant predictor, confirming the greater predictive power of DA.	Large	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Hasson et al., 2012)	DA scores at the first period (T1) were significantly correlated with the change score (CELF-3 T1-T4), unlike the static measures, which showed non-significant correlations with the change score. DA is only correlated to the T4 score, not T2 or T3 scores, indicating longer-term outcome prediction.	Large	-	Yes
(Lauchlan & Elliott, 2001)	No significant differences were found in DA pre-test or post-test scores when comparing intervention vs. non-intervention or high vs. low potential groups. DA was only able to differentiate the "high potential group with cognitive intervention" group from the other seven groups. DA is not a significant predictor of school performance.	Non-significant	-	Yes
(Lu & Hu, 2019)	Static assessment scores explain 21% of variance in real word spelling and 17% in pseudoword spelling. DA predicts an additional 8% in real word spelling and 12% in pseudoword spelling beyond the static measure. When DA was added first to the regression, it explained 28% of variance in real word spelling and 29% in pseudoword spelling. Adding the static measure did not provide any additional explained variance. After controlling for experience, DA contributed an additional 14% and 21% respectively, while the static measure did not have additional value. DA is a better predictor of spelling performance.	Medium-large	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Luković et al., 2022)	The static measure was a better predictor of later school performance than DA. The static scores explained between 9-18% of variance in school performance measures. DA explained 6% of variance in language trial test performance and 8% in mathematics final trial test performance. Affective-motivational scaffolding was a better predictor than cognitive scaffolding, explaining between 7-15% of variance in school performance measures.	Small-medium	Types of scaffolding	Yes
(Meijer & Elshout, 2001)	Conventional mathematics pretest score is the best predictor for conventional mathematics posttest performance, explaining 44% of variance. DA test performance in Grade 3 explains an additional 4% of variance. Conventional pretest was not a significant predictor for learning potential posttest performance, with DA is the main predictor (explained 41% of variance).	Small-medium, Large	Test anxiety	Yes
(Navarro & Mora, 2011)	DA significantly predicts academic performance and progress, beyond various static measures, explaining between 19-58% of variance.	Medium-large	-	Yes
(Navarro et al., 2018)	DA scores significantly predict reading competence, explaining 68% of variance together with the static measures. DA shows an incremental significant contribution to the three reading measures (46%).	Medium-large, Large	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Peña et al., 2006)	Static assessment score provided low classification scores (specificity = 70.5%, sensitivity = 78.6%), with race as a moderator. DA posttest scores had greater classification accuracy compared to the static measure (specificity = 83%, sensitivity = 64%). Modifiability scores (difference between DA pre-test and post-test scores) provided the best classification accuracy (specificity = 96%, sensitivity = 93%), and is thus the best predictor of narrative skills. Combining DA posttest scores with modifiability scores provided 100% classification accuracy for the "Story Components".	Large	Sociocultural background	Yes
(Petersen & Gillam, 2015)	DA modifiability score significantly predicts first-grade reading outcomes, explaining between 19-24% of variance across reading outcomes. A high classification accuracy of DA for predicting reading difficulties in bilingual Latino children was found ($0.80 \leq$ specificity $\leq$ 0.88; $0.86 \leq$ sensitivity ≤ 1.00). DA was a stronger predictor of end-of-year English reading outcomes than static measures. In regression analyses, when both static and DA variables were included, the DA gain score emerged as the only significant predictor, suggesting it provided unique and valuable information that static tests did not capture	Medium-large	Socioeconomic status, previous learning experience, and language	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Resing et al., 2012)	Teacher ratings of school performance and learning potential are highly correlated with static IQ scores and showed lower but still moderate to high correlations with DA scores. The authors mentioned that the teachers likely based their judgment on the initial diagnostic information they received when the children entered the special education school. The mathematics and language scores had low-to-zero correlations with IQ scores, but moderate to high correlations with DA. This likely indicates that DA scores give a better view of the objectively measured scholastic achievements of the children.	Medium-large, Large	-	Yes
(Seethaler et al., 2016)	1.Calculations outcomes: For Limited English Proficiency (LEP) students, the static mathematics test and DA significantly predicted Calculations outcomes, together explaining 47.8% of variance. The static test had significantly higher predictive validity . For non-LEP students, only the static measure was a significant predictor of Calculations outcomes. 2. Word Problem outcomes: for LEP students, DA and the static test were significant predictors of Word Problem outcomes, together explaining 39.2% of variance. DA was a significantly better predictor compared to the static test. For non-LEP students, the static test and DA significantly predicted Word Problem outcomes, together explaining 47.2% variance. There was no significant difference in predictive power between DA and the static test.	Medium	Language	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Stad et al., 2018)	DA scores, specifically post-test scores and the number of prompts required during training, significantly predicted children's mathematics achievement beyond static pre-test performance, explaining an additional 7-12% of variance. Cognitive flexibility was identified as a significant factor influencing children's performance on DA, explaining 35% of variance in DA outcomes.	Small-medium, Medium	Cognitive flexibility	Yes
(Stevenson et al., 2013)	Static test and DA both significantly predicted children's performance in mathematics and reading. When DA and static scores were both included in the regression model, they each contributed unique variance in predicting achievement in mathematics and reading. DA was the strongest additional predictor of reading achievement even after controlling for the static scores and gender. For mathematics achievement, both DA and static assessment were significant predictors.	Small-medium	Gender	Yes
(Swanson & Howard, 2005)	Static assessment significantly predicts reading and mathematics outcomes, explaining 10% and 30% respectively. DA is a significant additional predictor of both reading and mathematics performance, explaining an additional 6% and 25% respectively.	Small-medium, Large	-	Yes
(Touw et al., 2017)	There is a strong correlation between dynamic measures and scholastic achievement, similar to static measures. Also, significant correlations between DA and teacher ratings of school performance were found.	Medium-large (3x), Large (1x)	-	Yes

Reference	Key Findings	Effect Size(s)	Confounding Factor(s)	Comparison to Static
(Touw et al., 2019)	DA scores significantly predicted school achievement one year later, specifically the number of prompts required. DA of word reading and mathematics contributed significantly to the explained variance in the respectively school achievements. DA added incremental validity over static test scores and teacher judgments in predicting school performance in reading and mathematics.	Small	-	Yes

Table E2

Summary of Advantages Mentioned

Reference	Additional information	Improved differentiation	Predictive power	Captures learning potential	Less biased	Time/cost-effective	Reduce false positives	Other advantages
(Calero et al., 2011)	-	✓	-	✓	✓	-	-	-
(Cho et al., 2014)	-	✓	✓	✓	-	-	-	-
(Cho et al., 2020)	-	✓	-	-	-	-	✓	-
(de Beer, 2011)	✓	-	-	-	-	✓	-	-
(Fabio, 2005)	-	✓	-	✓	✓	-	-	-
(D. Fuchs et al., 2011)	✓	-	-	✓	-	-	✓	-
(L. S. Fuchs et al., 2011)	-	✓	-	✓	-	-	-	-
(Gan et al., 2023)	✓	✓	✓	-	-	-	-	-
(Gellert & Elbro, 2017)	✓	-	✓	-	-	✓	-	-
(Hamavandi et al., 2017)	-	-	✓	✓	-	-	-	-
(Hasson et al., 2012)	-	-	✓	-	✓	-	✓	-
(Lauchlan & Elliott, 2001)	✓	✓	✓	-	-	-	-	-
(Lu & Hu, 2019)	✓	-	-	-	✓	-	-	✓
(Luković et al., 2022)	✓	✓	✓	-	-	-	-	-

Reference	Additional information	Improved differentiation	Predictive power	Captures learning potential	Less biased	Time/cost-effective	Reduce false positives	Other advantages
(Meijer & Elshout, 2001)	-	-	✓	-	✓	-	-	-
(Navarro & Mora, 2011)	✓	-	-	-	-	-	-	✓
(Navarro et al., 2018)	✓	✓	-	-	-	-	-	✓
(Peña et al., 2006)	✓	✓	-	-	✓	-	-	-
(Petersen & Gillam, 2015)	✓	-	✓	-	✓	✓	-	✓
(Resing et al., 2012)	✓	✓	-	✓	-	✓	-	-
(Seethaler et al., 2016)	-	-	-	✓	-	-	✓	✓
(Stad et al., 2018)	✓	✓	-	✓	-	-	-	✓
(Stevenson et al., 2013)	✓	-	✓	✓	-	-	-	-
(Swanson & Howard, 2005)	-	✓	-	✓	-	-	-	-
(Touw et al., 2017)	✓	-	✓	-	-	-	-	✓
(Touw et al., 2019)	✓	✓	✓	-	-	✓	-	-

Table E3

Summary of Limitations Mentioned

Reference	Time-consuming	Validity, reliability, generalizability	Need for extensive training	Unclear definitions	Other limitations
(Calero et al., 2011)	-	✓	-	-	-
(Cho et al., 2014)	-	-	✓	-	-
(Cho et al., 2020)	-	-	-	-	-
(de Beer, 2011)	✓	-	-	-	-
(Fabio, 2005)	✓	✓	✓	✓	-
(D. Fuchs et al., 2011)	✓	-	-	-	-
(L. S. Fuchs et al., 2011)	✓	-	-	-	-
(Gan et al., 2023)	-	-	-	-	-
(Gellert & Elbro, 2017)	-	-	-	-	-
(Hamavandi et al., 2017)	-	-	-	-	-
(Hasson et al., 2012)	-	-	-	-	✓
(Lauchlan & Elliott, 2001)	-	✓	-	-	✓
(Lu & Hu, 2019)	✓	-	-	-	-
(Luković et al., 2022)	-	-	-	✓	✓
(Meijer & Elshout, 2001)	-	✓	-	-	-

Reference	Time-consuming	Validity, reliability, generalizability	Need for extensive training	Unclear definitions	Other limitations
(Navarro & Mora, 2011)	-	-	-	-	-
(Navarro et al., 2018)	-	-	-	-	-
(Peña et al., 2006)	✓	-	-	-	-
(Petersen & Gillam, 2015)	-	-	-	-	-
(Resing et al., 2012)	-	✓	-	-	-
(Seethaler et al., 2016)	✓	-	-	-	-
(Stad et al., 2018)	-	-	-	-	-
(Stevenson et al., 2013)	-	-	-	-	-
(Swanson & Howard, 2005)	-	✓	-	-	-
(Touw et al., 2017)	✓	-	-	-	-
(Touw et al., 2019)	✓	-	-	-	-

Table E4

Summary of Advice Given

Reference	Supplement	Instructional needs	Make more practical/efficient	Contextualize the assessment	Use to improve accuracy	More research needed	Other advice
(Calero et al., 2011)	-	-	-	-	-	-	-
(Cho et al., 2014)	✓	-	✓	-	✓	-	-
(Cho et al., 2020)	-	-	-	-	-	-	-
(de Beer, 2011)	✓	✓	✓	-	-	✓	✓
(Fabio, 2005)	-	-	-	-	-	-	-
(D. Fuchs et al., 2011)	✓	-	-	-	-	-	-
(L. S. Fuchs et al., 2011)	✓	-	-	-	-	-	-
(Gan et al., 2023)	-	✓	-	-	-	✓	-
(Gellert & Elbro, 2017)	✓	-	-	-	-	-	-
(Hamavandi et al., 2017)	-	-	-	-	-	-	-

Reference	Supplement	Instructional needs	Make more practical/efficient	Contextualize the assessment	Use to improve accuracy	More research needed	Other advice
(Hasson et al., 2012)	-	-	-	-	-	-	✓
(Lauchlan & Elliott, 2001)	-	-	-	-	-	-	✓
(Lu & Hu, 2019)	✓	-	-	-	-	-	-
(Luković et al., 2022)	-	-	-	-	-	-	✓
(Meijer & Elshout, 2001)	✓	-	-	-	-	-	-
(Navarro & Mora, 2011)	-	-	-	✓	-	-	-
(Navarro et al., 2018)	-	-	-	✓	-	-	-
(Peña et al., 2006)	-	-	✓	-	-	-	-
(Petersen & Gillam, 2015)	-	✓	-	-	-	-	✓
(Resing et al., 2012)	-	✓	✓	-	-	-	-

Reference	Supplement	Instructional needs	Make more practical/efficient	Contextualize the assessment	Use to improve accuracy	More research needed	Other advice
(Seethaler et al., 2016)	✓	-	-	-	-	-	-
(Stad et al., 2018)	-	-	-	-	-	-	✓
(Stevenson et al., 2013)	-	-	-	-	-	-	-
(Swanson & Howard, 2005)	-	-	-	-	✓	-	-
(Touw et al., 2017)	-	✓	-	-	-	-	-
(Touw et al., 2019)	-	✓	✓	-	-	-	-