

Hang Le Truong Thu

SYSTEMATIC REVIEW OF DYNAMIC ASSESSMENT IN ELEMENTARY INCLUSIVE EDUCATION

Faculty of Education and Culture
Master Thesis
September 2021

ABSTRACT

Hang Le: Systematic Review of Dynamic Assessment in Elementary Inclusive Education
Master Thesis
Tampere University
Teacher Education
September 2021

Endeavouring to integrate assessment and teaching, Dynamic Assessment (DA) has been recognized its pivotal role in inclusive education. Previous empirical and non-empirical reviews in this area have not addressed the implications of DA in elementary inclusive education. Consequently, this study aims to provide an overall and comprehensive picture of the applications of DA in elementary inclusive education, evidencing the methodologies and DA tools applied in previous studies, as well as the challenges and recommended solutions to develop DA practice in this area.

To achieve the research's aims, a systematic literature review was conducted with the results summarized narratively. Twenty-five studies carried out between 2010 and 2021 composed the sample for this review. Different methodologies were employed to conduct studies, in which the majority investigated DA as a sole approach to assess students or as a supplement to other static tests. DA was commonly used to predict students' special educational needs, future performance, and academic development, and to predict students' responsiveness in a Response to Intervention (RTI) decision-making process. However, it was rarely applied to guide intervention, inform short and long-term instruction, or develop individual educational planning. This paper also indicated some novel DA applications that was not mentioned in previous reviews. Regarding the assessors, trained research assistants and researchers were reported to carry out this type of assessment in most included studies, using different types of feedback, mediation and assistance techniques during the assessment procedures. This review indicated the scarcity of research on teachers and school psychologists as the examiners of DA in classroom settings. Key challenges when employing DA in elementary inclusive education and recommended solutions for developing its practice were also identified. The review concluded by discussing the contribution of this review to the theory and practice of assessment in inclusive education and offered recommendations for not only future reviews on DA but also future research investigating DA in this field.

Keywords: Dynamic assessment, elementary inclusive education, systematic review, dynamic testing.

The originality of this thesis has been checked using the Turnitin Originality Check service.

CONTENTS

1	INTRODUCTION	5
2	WHAT IS DYNAMIC ASSESSMENT?	7
2.1	The epistemology of dynamic assessment.....	7
2.2	Applicability of Dynamic Assessment	10
3	INCLUSIVE EDUCATION IN BASIC EDUCATION	12
3.1	Inclusive education	12
3.2	Assessments in inclusive basic education	15
4	METHOD	17
4.1	Research Questions	17
4.2	Research Methods and Design.....	17
4.3	Data gathering.....	18
4.3.1	<i>Search Terms and Procedures</i>	18
4.3.2	<i>Inclusion and Exclusion Criteria</i>	19
4.3.3	<i>Study selection process</i>	20
4.4	Data analysis and synthesis	22
4.4.1	<i>Data extraction</i>	22
4.4.2	<i>Data appraisals</i>	22
4.4.3	<i>Synthesis</i>	24
5	RESULTS	25
5.1	Descriptive findings	25
5.2	Research question 1: What does empirical literature reveal about the applicability of DA in elementary inclusive education?	34
5.2.1	<i>How is DA used?</i>	34
5.2.2	<i>When is DA used?</i>	35
5.2.3	<i>Who participates in the process?</i>	35
5.3	Research question 2: What are the tools, techniques, and methodologies applied?	36
5.4	What are the challenges for DA in elementary inclusive education and the solutions for developing its practice?	38
6	DISCUSSION	41
6.1	Main findings	41
6.2	Limitations, validity, and reliability of the current systematic review	47
7	CONCLUSION	49
	REFERENCES	51
	APPENDICES	59
	Appendix A: The original version of MIXED METHODS APPRAISAL TOOL (MMAT), VERSION 2018 (Hong et al., 2018, p.2)	59
	Appendix B: Instructions for categorizing studies to rate in the MMAT* (Hong et al., 2018, p.8)	61
	TABLES	
TABLE 1.	Inclusion and exclusion criteria.....	20

TABLE 2.	Adapted Mixed Methods Appraisal Tool (MMAT).....	23
TABLE 3.	Summary of included studies.....	27
TABLE 4.	Challenges and solutions.....	40
FIGURES		
FIGURE 1.	The rights framework for inclusion (UNESCO, 2005, p.14).....	13
FIGURE 2.	Flowchart for study inclusion	21

1 INTRODUCTION

There is a substantial body of literature that recognizes the significance of Dynamic Assessment (DA) in education, especially in inclusive education (Kirschenbaum, 1998; Regalla & Peker, 2017; Al-Hroub & Whitebread, 2019; Gellert & Elbro, 2018). A previous meta-analysis focusing on the main effects of DA in inclusive education concluded that DA procedures have a substantially positive impact on students' testing performance over static testing conditions (Swanson & Lussier 2001). DA offers a means for measuring the untapped potential and cognitive skills of children with disabilities, developmental disorders or learning difficulties in a large number of empirical studies (Kirschenbaum, 1998; Regalla & Peker, 2017; Al-Hroub & Whitebread, 2019; Gellert & Elbro, 2018).

Most of the research delving into the applicability of DA in education has heightened the need for systematic reviews to unveil a panorama of the implications of DA in inclusive education. A comprehensive understanding of the implications of DA in inclusive practices can provide teachers and educators with a set of guidelines that would improve teaching and learning practices. Additionally, previous studies indicate scarcity of reviews on DA which includes students with disabilities, developmental disorders or high abilities (Ghanbarpour, 2017; Rezaee & Ghanbarpour, 2016), or that synthesizes relevant findings from studies with a wide range of research designs (Swanson & Lussier, 2001). In addition, Swanson and Lussier (2001)'s meta-analysis targeted primarily on the effect sizes of research investigating DA rather than on the implications of DA in inclusive education. Beside the aforementioned empirical studies, there were non-empirical research reviewing the implications of DA in education in general (Haywood & Tzuriel, 2002; Elliott, 2003; Elliott et al., 2018). Thus, the implications of DA in elementary inclusive education for the learning and development of children with disability or other types of special educational needs due to

developmental disorders, psychological conditions or high abilities have not yet been the subject of systematic reviews.

Therefore, this study makes a major contribution to research on dynamic assessment by conducting a systematic review on empirical studies exploring the implications of DA in elementary inclusive education. This study provided an overall picture of how DA is used in this field, evidencing what are the tools and methodologies employed in previous research, and the challenges and recommended solutions to improve DA practice in elementary inclusive education.

The present study gathered empirical studies investigating DA in elementary inclusive education with the timeframe from 2010 to 2021. The theoretical framework includes two parts: (1) the epistemology of Dynamic Assessment, and (2) applicability of Dynamic Assessment. The data extracted across the studies were synthesized to answer the research questions.

2 WHAT IS DYNAMIC ASSESSMENT?

Dynamic Assessment (DA) is an interactive approach to performing assessment that integrates components of instruction for the purpose of measuring how well a learner can acquire with assistance (Caffrey et al., 2008; Grigorenko, 2009; Wagner & Compton, 2011). The term dynamic expresses the idea that assessment happens through a process between individuals, thus not localized in the individual, and happens in a continuum (Al-Hroub & Whitebread, 2019). The significant discrepancy between DA and traditional, or *static* assessments, is the object of assessment.; the latter assess students' independent performance focusing on the product of learning (Cho et al., 2020), while DA focuses on the learner's learning potential and cognitive skills to respond to intervention (Al-Hroub & Whitebread, 2019). Dynamic approaches address assessment by teaching the skills assessed in the test directly and evaluating learners' responsiveness to that instruction (Cho et al., 2020). Therefore, DA has found to be useful to provide an assessment which not only assess students' learning but also effect change to improve students' effectiveness at learning and task mastery especially in inclusive settings (Lawrence & Cahill, 2014).

2.1 The epistemology of dynamic assessment

DA integrates interaction into the structure of a test-intervene-retest approach to psychoeducational and psychological assessment (Haywood & Lidz, 2006), in which teaching or learning phase plays a crucial role in the assessment procedure. The field of DA has developed in many divergent directions and varied applications. Different DA models have been employed in educational, speech/language, neuropsychological, and psychological contexts (Haywood & Lidz, 2006).

Different types of DA are theoretically based on Vygotsky's concept of the Zone of Proximal Development (ZPD) (Vygotsky, 1978) and Feuerstein's concept

of structural cognitive modifiability (Feuerstein et al., 1979, as cited in Kozulin, 2015). Vygotsky's concept of ZPD conceptualizes the discrepancy between children's previously formed mental functions and the potential ones (Kozulin, 2011), focusing on the fundamental role of joint activity of learners and teachers and the appropriation of mediation in the development of higher mental functions. Vygotsky's discussion of ZPD happened in three major interrelated contexts including developmental, assessment related and educational contexts (Kozulin, 2011 & Kozulin, 2015). The first one concentrated on the need to investigate the emergent mental functions of the child under development, which is impossible to be detected when the student conducts the activity independently (Kozulin, 2015). The assessment related context indicated that under conditions of performance with assistance or in collaboration with others, children with the same intelligence quotient (IQ) scores may illustrate different ZPDs (Kozulin, 2015; Kozulin & Garb, 2016). The educational context addressed the interpersonal interaction between educators and students and students with their peers since Vygotsky had a belief that the basis nature of a learning activity is a meaningful interpersonal communication between educators and learners as well as learners with their peers (Kozulin, 2015; Kozulin & Garb, 2016). Within this interaction through learning situations, new learning abilities and cognitive functions are commenced and later internalized and transformed, becoming a cycle of students' mental development (Kozulin & Garb, 2016). Albeit Vygotsky offered various valuable and broad recommendations relating to evaluating students' ZPD and stated a diverse array of potential interactive interventions to be applied amid ZPD assessment, e.g. modelling, asking questions to lead the learners, asking students to continue, to name but a few, he did not generate standardized procedure or models for the ZPD assessment (Kozulin & Garb, 2016). Nevertheless, ZPD undoubtedly offers "a comprehensive dynamic paradigm" (Alony & Kozulin, 2007, p.324) including these following features for the DA method: (1) a direct interaction between the teacher and students amid the assessment; (2) attention to students' potential for change rather than their current performance level; as well as (3) the gain score measured by comparison of performance with and without assistance (Kozulin, 2011 & Kozulin & Garb, 2016).

For several reasons in relation to society, politics and science little attention was paid to the concept of dynamic assessment in a time frame from the 1930s to 1960s (Kozulin & Garb, 2016). Not until the late 1960s that the surge in criticising standardized psychometric tests occurred was the renewed interest of DA stimulated (Kozulin & Garb, 2016). The first completely versions of dynamic cognitive assessment were evolved by Budoff and Friedman (1964) and Feuerstein and Shalom (1968) (as cited in Kozulin & Garb, 2016). The tasks utilized in dynamic assessment were similar to the tasks of standardized psychometric tests; however, the assessment procedure was fundamentally different by consisting of a teaching session (Budoff & Friedman, 1964). Moreover, the ideology of assessment has also been changed, focusing on students' learning potential rather than their previous experience (Budoff & Friedman, 1964).

On the other hand, Jensen and Feuerstein (1987, p.380)'s version of DA was formulated on the "theory of structural cognitive modifiability" which assumes that "the individual is an *open system* susceptible to influences that can produce *structural* changes in cognitive functioning". Feuerstein (Feuerstein, 1980; Feuerstein, Jensen, Hoffman, & Rand, 1985 as cited in Jensen & Feuerstein, 1987, p.380) argued that a common procedure of learning is normally portrayed by the direct exposure of the individual to "sources of stimuli", but that this process itself cannot explain the observed discrepancies in intellectual functioning or the phenomenon of culture. The theory assumes that the essential element determining the progression in humans of higher levels of cognitive functioning is dependent on the growing child's opportunity to benefit from mediated learning experiences (MLE) (Jensen & Feuerstein, 1987). In this progress of learning, stimuli are chosen by a mediator, framed by goals and attributes, and developed through being connected by purpose and imbued with meaning (Jensen & Feuerstein, 1987). The function of the mediator is to assist the student in structuring his or her learning experience (Jensen & Feuerstein, 1987). These assumptions are theoretical base for Feuerstein's version of DA. The tasks used in DA should address merely those issues whose solutions are dependent on implementing cognitive strategies rather than tasks requiring prior knowledge, memory, or highly automatic skills (Kozulin & Garb, 2016). According to Feuerstein, the capacity of the teacher or evaluator is to discern the learners'

issues within the pre-test and to supply the mediation or assistance needed during the learning session.

As well as MLE-based dynamic assessment, a wide variety of dynamic cognitive assessment mechanisms is presently accessible (Kozulin & Garb, 2016). There are considerable differences among these DA techniques in many aspects, except for their dependence upon the test-teach-test paradigm (Kozulin & Garb, 2016). The first difference lies in their goals. Some of them are presented as supplemental methods to be utilized in addition to static test, whereas others occur as a substitute for standardized psychometric assessment (Kozulin & Garb, 2016). MLE-based assessments concentrate on analysing the students' instructional needs and disapprove the goal of prediction, whereas other learning tests generated in Germany (see Guthke and Stein, 1996) was stated to have a stronger predictive power than standardized tests (Kozulin & Garb, 2016). Another difference is in the assessment process. While some dynamic tests demand of a quite large amount of initiative on the role of the evaluator, others consist of standardized rules of intervention in the learning phase (Kozulin & Garb, 2016).

2.2 Applicability of Dynamic Assessment

Most prominent researchers studying DA (e.g., Jensen & Feuerstein, 1987; Haywood, 1997; Lidz, 1991 as cited in Haywood & Lidz, 2006) have recommended that DA is particularly beneficial when (1) scores on static tests are poor, and particularly when they do not correspond to information from other sources; (2) learning seems to be limited by lack of motivation, learning disabilities, intellectual disabilities, personality disorders, emotional and behavioral disorders; (3) students encounter problems in language, such as developmental language disorders, restricted vocabulary, differences between the learner's mother tongue and the language of instruction used in school. In addition, DA is recommended to be used when there are notable distinctions in culture between minority students such as immigrants and their peers who come from the dominant culture. In these circumstances, static testing is prone to produce poor results and associated negative predictions of learning development and academic achievement. However, the main role of DA is not to

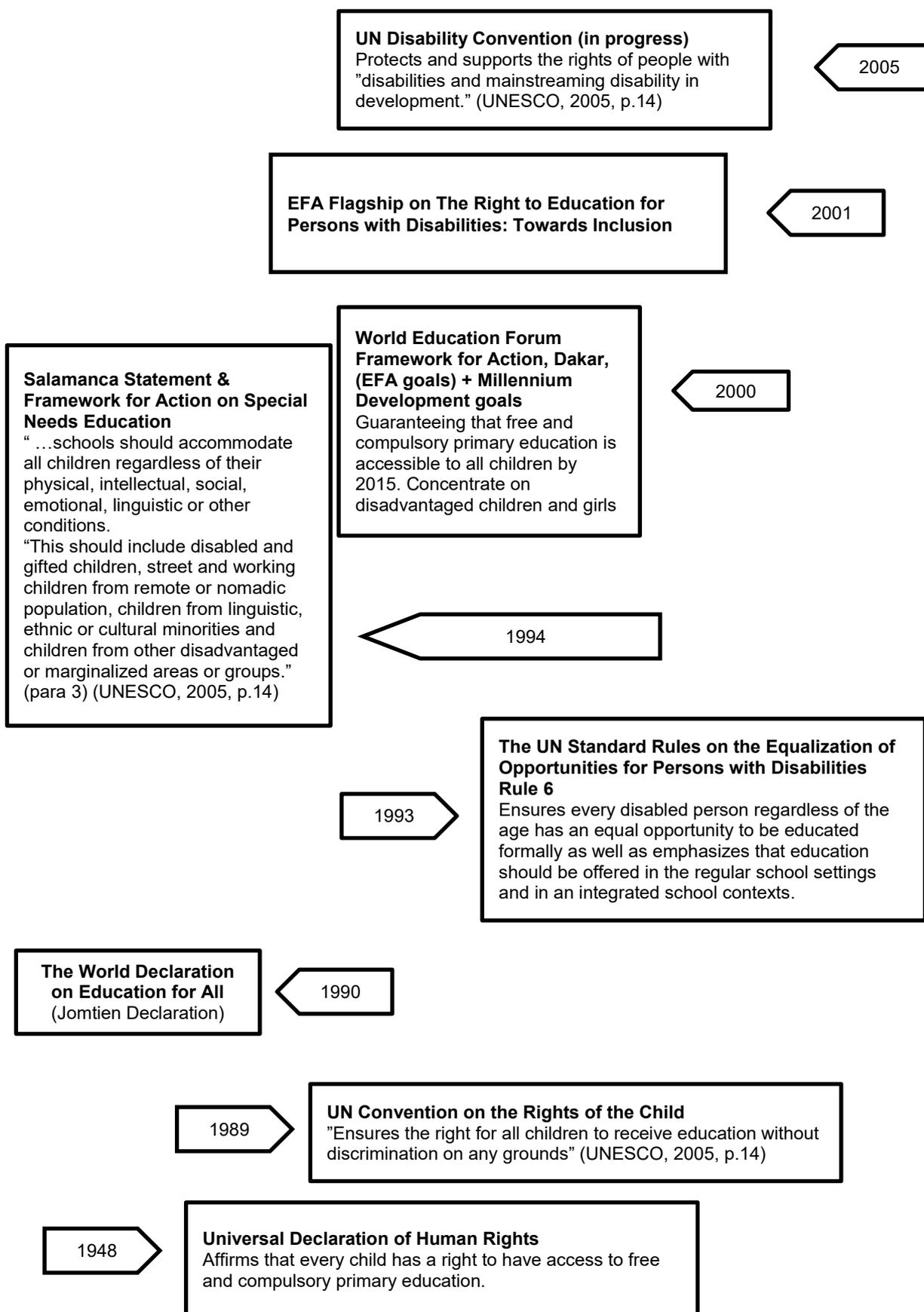
doubt those prophecies; instead, its role is to identify hindrances to more successful learning and performance, to discover strategies to surmount those hindrances, and to evaluate the effects of removing those hindrances on the effectiveness of future learning and performance (Haywood & Lidz, 2006). By extending that function, the main aim of DA is to offer suggestions on what can be done to vanquish the negative prophecies that are normally created based on the results of static tests, consisting of extrapolating the kinds and appropriate amount of instructional or educational psychology intervention that will be needed to generate important enhancement and the potential impacts of that intervention (Haywood & Lidz, 2006).

3 INCLUSIVE EDUCATION IN BASIC EDUCATION

3.1 *Inclusive education*

The term “inclusive education” was first stated officially in the Declaration of Salamanca (UNESCO, 1994). The Statement starts with a commitment to Education for All, emphasizing the necessity of educating all learners within the mainstream educational systems, no matter the disability (UNESCO, 1994). The Declaration is a reaffirmation that every person has a right to receive free primary education, as heightened in the Universal Declaration of Human Rights. It also regenerated the commitment made at the “World Conference on Education for All” held in 1990 to assure that right for all children, irrespective of their individual differences and recalled many “United Nations declarations” climaxing in the 1993 “United Nations Standard Rules on the Equalization of Opportunities for Persons with Disabilities” Rule 6 (UNESCO, 2005, p.14). In a nutshell, albeit the term “inclusive education” was first mentioned officially in 1994, this notion had been implicitly referred to since 1948 and has been appeared in many major UN Declarations and Conventions (as seen in Figure 1 The Rights Framework for Inclusion (UNESCO, 2005, p.14)). Since many international agreements and documents emphasized the significance of inclusive education, the educational systems in multiple nations have made the attempt to decrease discrimination in education practices and lend supplementary support to assist individual students fulfil their potential (Kiuppis, 2013).

FIGURE 1. “The rights framework for inclusion” (UNESCO, 2005, p.14)



Inclusive education is formed on the values of social justice and equity that every child has the right to learn with their peers in regular classrooms (Madureira Ferreira, 2018). The definition of inclusive education, however, varies among researchers. For example, on the one hand, some pay more attention to disability and define this concept as the endeavour to provide education to students with mental retardation by making an attempt to integrate them into regular schools (Michailakis & Reich, 2009). On the other hand, others define this concept broadly as including and appreciating all individual differences embracing culture, gender, race, disability, and socio-economic status (Mariga et al., 2014; Ainscow, Dyson, Hopwood, & Thompson, 2016). In this current study, the broader definition was chosen. This review targeted particularly on students who have or may have special educational needs in regular classrooms.

According to UNESCO (2005), there are four key elements conceptualizing the notion of inclusion. First, inclusion is an endless process of searching to find better methods in response to diversity (UNESCO, 2005). It is about learning how to harmonize with individual differences and learning how to appreciate the diversity. Second, inclusion is regarding to the process of identifying as well as removing the barriers (UNESCO, 2005). Accordingly, it includes gathering, collating, and appraising information from a wide variety of sources such as school psychologists, students' peers, and parents, to mention a few in order to make plans for improvements in practice and policy (UNESCO, 2005). It is concerned with utilizing evidence from different sources to enhancing creativity and problem-solving. Third, inclusion is about "the presence, participation, and achievement" of all learners (UNESCO, 2005, p.15). 'Presence' relates to where learners are provided education, and how reliable and punctual they are when attending. 'Participation' is about the quality of the learners' educational experiences and thus, must integrate with the learner's perspectives. 'Achievement' is concerned with learning outcomes across a curriculum, not targeting only on the results from tests or exams. Last but not least, inclusion emphasizes particularly on groups of learners who are prone to low performance, exclusion, or marginalization. This illustrates the ethical responsibilities to warrant that those aforementioned portions of population are paid more attention to, and when necessary, actions are taken to warrant their presence, participation, and achievement in the educational system.

3.2 Assessments in inclusive basic education

Leeber, Candeias, and Grácio (2011), when taking various cultural contexts into consideration, held a thorough discussion about the significance of analysing critically the utility of standardized tests to assess students with disabilities, and to a greater extent, merely employing such measurements to identify and justify the need for educational support or special education. Indeed, Leeber et al. (2013) claimed that one of the hindrances to inclusive education is generated by the way learners, who display some signs of developmental or learning difficulties, are examined before and through formal instruction. In their paper, Leeber et al. (2013) raised the questions regarding the utility and interpretation of static testing. They emphasized the possible challenges correlating with employing only this instrument to label students, which embraces a threat of arousing low educational expectations or self-fulfilling prophecies (Leeber et al., 2013).

Additionally, when the goal of assessing is to develop therapeutic or educational plans or to place a child to a mainstream or special school, the results from those static tests exert a negative effect (Leeber et al., 2013). According to Leeber et al. (2013), parents and teachers complained that the test outcomes solely provide them a list of individual student's inadequacies, presenting what the individual is incapable of doing in comparison to his/her peers. Test reports fail to provide any indication of how to understand the learner's issues, how to teach and cooperate with him/her, what the student's potential is, or what the student really needs to reach their potential. Thus, in some cases, it results in "wrong placements, low-profile programming, and parental depression" (Leeber et al., 2013, p.5).

Therefore, Leeber et al. (2013) indicated the necessity for novel forms of assessments that consider a child from a different perspective. According to Leeber et al. (2013), the goal of the novel assessment should be to comprehend a learner's functioning in an interactive and dynamic way, how to enhance and develop her/his functioning, participation and acquiring and what might be the obstacles for his/her development. Rather than focusing on the value of predicting and labelling dysfunction and learning difficulties, this new assessment should target on the value of identifying individual's educational needs in order to guide

interventions and support teachers in individual educational planning that enables the learner to be fully included.

Inclusive assessment is defined as an approach to assessment in mainstream education where practice and policy are established to enhance the learning of all learners regardless of their disabilities (Watkins, 2007). The primary purpose of inclusive assessment is that all policies relating to assessment and assessing procedures should assist as well as promote the effective inclusion and participation of every learner at risk of exclusion and marginalisation (Watkins, 2007). To attain the aim of inclusive assessment, the principles underpinning inclusive education should be taken into consideration. Assessment processes should be employed to inform and enhance learning for all learners (Watkins, 2007). All students should be involved in every assessment procedure and the needs of learners with special educational needs should be taken into consideration (Watkins, 2007). All assessment procedures should be supplementary, inform each other and target to 'celebrate' differences by discerning and appreciating all learners' own learning progress and achievements (Watkins, 2007). This approach to assessment explicitly aims to diminish isolation by avoiding labelling students and instead concentrating on teaching and learning practice that furthers inclusion in mainstream contexts.

4 METHOD

4.1 Research Questions

Striving to identify, analyse, and summarize literature on the implications of DA in elementary inclusive education with the timeframe from 2010 to 2021, and provide a set of strategies that could be used as guidelines for teachers on developing assessment in inclusive classrooms, the current study addresses three main questions.

- 1) What does empirical literature reveal about the applicability of DA in elementary inclusive education?
 - a. How is DA used?
 - b. When is DA used?
 - c. Who participates in the process?
- 2) What are the tools, techniques, and methodologies applied?
- 3) What are the challenges for DA in elementary inclusive education and the solutions for developing its practice?

4.2 Research Methods and Design

The methodological approach for the current research is a systematic literature review with the results summarized narratively. A systematic review is defined as a literature review associated with a specific research question that employs explicit, systematic methods to locate, appraise, and synthesize the findings of all primary research relating to the question at hand (Boland et al., 2017; Zawacki-Richter et al., 2020). The use of pre-specified research questions and application of rigorous and systematic methods aim to minimize the bias so as to reach conclusions based on all the included studies in a comprehensive and objective way (Zawacki-Richter et al., 2020). Therefore, systematic review approach was chosen for this current study to provide an impartial and overall picture of the applicability of dynamic assessment in inclusive education, the tools, techniques,

and methodologies within this scope of work as well as the challenges and solutions for developing its practices.

The research questions in this study directed towards investigating scholarly peer reviewed studies investigating the implications of dynamic assessment in inclusive settings. This data was used to discern the implications of DA in elementary inclusive education. A critical theoretical approach yielded a foundation for the research. Comprehending the practice of DA in inclusive education as well as its challenges and solutions for advancing its practices assisted in the understanding of the overall benefits of DA in assessing students, the practical strategies, and guidelines on how to implement DA effectively in inclusive settings.

4.3 *Data gathering*

4.3.1 Search Terms and Procedures

The literature is purposefully searched for by retrieving related articles from nine databases: Education Resources Information Center (ERIC), ScienceDirect, SAGE Journals Online, Taylor & Francis Online, SpringerLink, PubMed, Web of science, PsycINFO, and Wiley Online Library. Additionally, *Journal of Cognitive Education and Psychology*, which devotes particular attention to this approach to assessment, was searched.

The search string was developed through the extraction of search terms employed in previous reviews carried out in the area of dynamic assessment (Swanson & Lussier, 2001; Rezaee & Ghanbarpour, 2016; Ghanbarpour, 2017). The search terms ["Special education" OR "Inclusive education" OR "special class"] AND ["dynamic assessment" OR "sociocultural theory" OR "activity theory" OR "zone of proximal development" OR "ZPD-based assessment" OR "mediated learning experience" OR "scaffolding" OR "classroom assessment" OR "classroom teacher assessment" OR "Vygotsky"] were used to locate and retrieve potential research articles for this systematic review.

Initially, the searches were limited to articles published since the declaration of Salamanca statement in 1994 (UNESCO, 1994) when "inclusive education" was first mentioned as an official term in this statement. However, in order to

provide an up-to-date summary of the state of research knowledge on dynamic assessment in inclusive education, the researcher decided to deal with the most recent studies from the past ten years, i.e., from 2010 to 2021. The search with the time frame since 1994 yielded a total of 50497 titles, of which 46094 remained after deduplication. After setting a new time frame, 24951 titles remained for the study selection process.

4.3.2 Inclusion and Exclusion Criteria

A significant matter in selecting data for a systematic review is that the criteria for including and excluding primary research studies should be clearly stated (Boland et al., 2017). To assure that the included studies had an adequate methodological rigor, only research papers that were published in peer-reviewed journals were taken into consideration (Wilder, 2014). Therefore, conference presentations, dissertations, and book reviews were all omitted.

Having in mind the focal points of the research questions and contemplating the need to integrate comparable results and data types, the researcher comprised only the articles that (1) reported primary research studies (2) conducted in inclusive settings, which means some of the participants have to be students with special needs or students with disabilities, and (3) examined the empirical findings on practical implications of DA in inclusive settings, as well as applications of DA to inclusive classroom assessment and pedagogy. Also, those research that investigated Computerized Dynamic Assessment (C-DA) were eliminated (see Table 1 for inclusion criteria).

TABLE 1. Inclusion and exclusion criteria

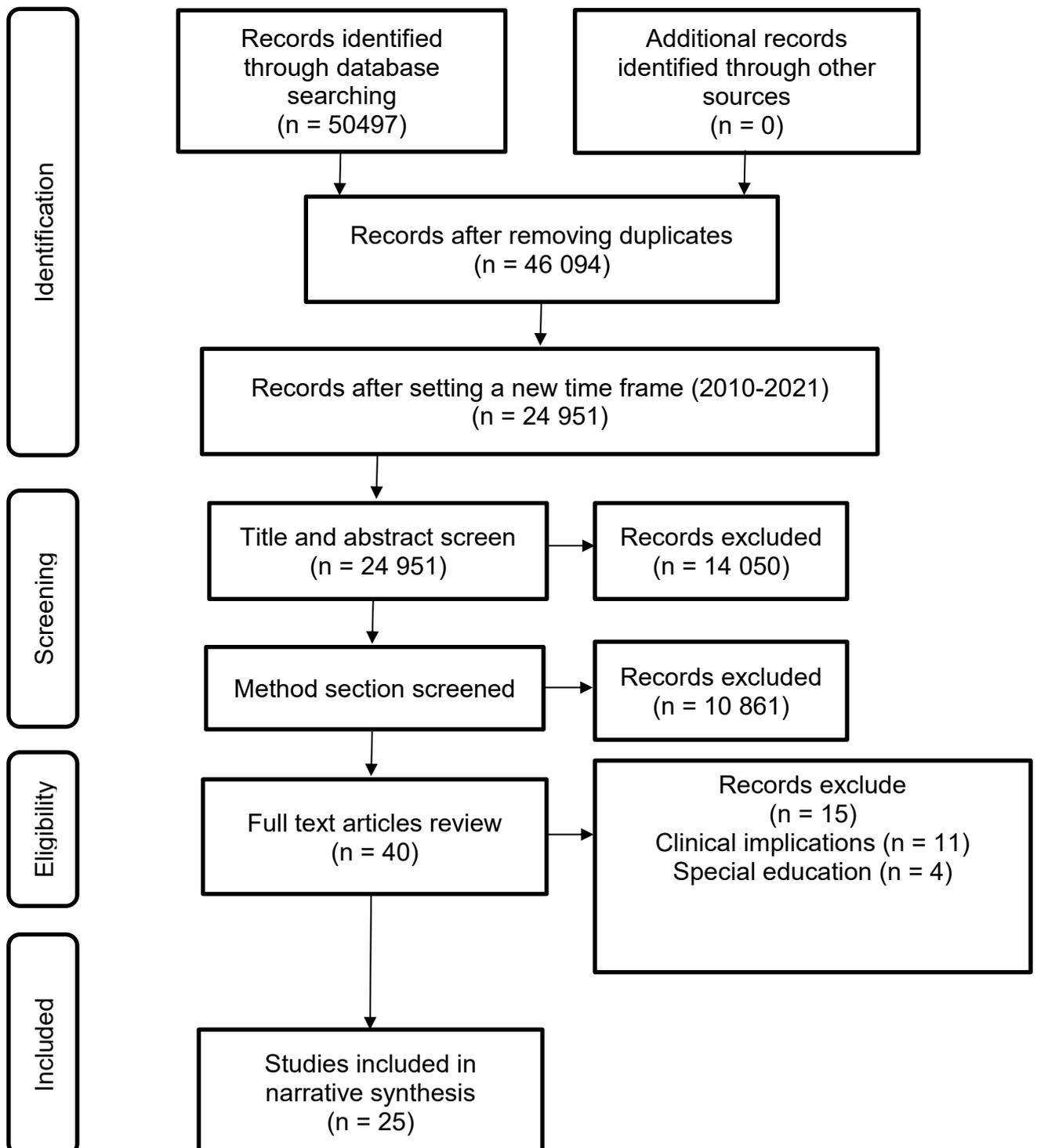
Criterion	Included	Excluded
Type of Dynamic Assessment	Dynamic Assessment as a classroom assessment in inclusive settings. Dynamic Assessment within the field of speech or language. Dynamic Assessment within the domain of psychology.	Computerized Dynamic Assessment
Research Site	Classroom settings Teacher-directed or small-group instruction settings	Consultation rooms
Participants	Student with permanent or temporary special educational needs such as students with disabilities, developmental disorders, gifted abilities, or students encountering educational challenges due to adaptation to cultural differences or low-performing students Teachers of students with special educational needs in regular classrooms Parents of students with special educational needs.	The articles addressing any other target sample
Levels of Education of students as participants	Primary Education	Early childhood Education Lower Secondary Education Upper Secondary Education Higher education
Findings	Implications for teaching and learning practice in inclusive education	Implications for clinical practice and research.

4.3.3 Study selection process

During title and abstract screening, the researcher used Rayyan software (Ouzzani et al., 2016) to support the procedure of deciding whether the studies identified met any of the exclusion criteria for the review or appeared to fit the inclusion criteria. In this stage, 10901 studies seemed to meet the inclusion

criteria. In the next step, the researcher reviewed the methods section of each of these studies to determine if the articles investigated dynamic assessment in inclusive education. Following this, 10861 articles were eliminated, leaving 40 studies for reviewing the whole text. Of the 40 articles, 25 studies met all criteria for study inclusion. Fig. 1 shows a flowchart illustrating the study identification process according to PRISMA guidelines (Moher et al., 2009).

FIGURE 2. Flowchart for study inclusion



4.4 *Data analysis and synthesis*

4.4.1 Data extraction

The data extraction in form of tables was employed to summarize included studies. Data relating to research questions on all selected studies were coded across the following categories: (a) authors and year published, (b) location of the study, (c) research setting, (d) characteristics of participants, (e) DA procedure, (f) the utility of DA, (g) methodologies employed, (h) DA tools and techniques, (i) challenges, and (j) solutions. Data extraction tables were first experimented on three of included studies to check whether all the necessary data for answering the research questions were being captured (Boland et al., 2017). After actually extracting all relevant data, the researcher redid the data extraction after three days and cross-checked that both sets of extracted data were the same (Boland et al., 2017). The summaries of included studies were displayed in Table 2 and Table 3 alphabetically by the first author's last name and chronologically by publishing years.

4.4.2 Data appraisals

In this current research, the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018) was employed to appraise the reports included in this systematic review. The MMAT is a quality assessment tool developed to provide critical appraisal of the methodological quality of empirical studies (Hong et al., 2018). Since the data in this current review consisted of a wide range of various study designs, MMAT was considered as an appropriate tool to assess different categories of study designs including (a) mixed methods studies, (b) quantitative descriptive, (c) randomized controlled trial, (d) nonrandomized and (e) qualitative. The latest version of the MMAT (version 2018) encompasses 25 criteria (5 criteria for each category), and 2 screening questions. Two screening questions are used to indicate whether the article is an empirical study in order to be appraised via the MMAT. Since the selection criteria of this current review were limited to empirical research, it was unnecessary to respond to these questions (Hong et al., 2018).

Across all studies, there are 12 reports mentioning the research questions. Seven studies stated the hypotheses clearly without research questions. The

remaining six studies only presented the purpose of the studies. Therefore, the researcher decided to modify the MMAT checklists to suit the data in this current study (see Table 2 for the adapted MMAT and the original version can be seen in Appendix A). For each included study, the researcher first classified the study design using a set of instructions for categorizing studies to rate in the MMAT (Hong et al., 2018) (the algorithm can be seen in Appendix B) in order to place it in the suitable category to give appraisal. After being categorized, the study was rated following the criteria of the selected category. For instance, if the article was a mixed methods study, the researcher only rated the five criteria in the mixed methods category. Each criterion was rated on a scale of yes, no and can't tell. The 'Can't tell' response category signified that the article did not provide suitable information to answer 'Yes' or 'no', or that provide insufficient or vague information regarding the criterion. Overall quality of the selected studies was found to be adequate, with clearly stated research questions/ hypotheses/ aims and appropriate used methods.

TABLE 2. Adapted Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018)

Category of study designs	Methodological quality criteria	Responses		
		Yes	No	Can't tell
1. Qualitative	1.1. Is the qualitative approach appropriate to answer the research question/ test the hypothesis/ achieve the aim?			
	1.2. Are the qualitative methods to collect data adequate to answer the research question/ hypothesis/ aim?			
	1.3. "Are the findings adequately derived from the data?" (Hong et al., 2018, p.2)			
	1.4. "Is the interpretation of results sufficiently substantiated by data?" (Hong et al., 2018, p.2)			
	1.5. Is the coherence between qualitative data sources, collection, analysis, and interpretation? (Hong et al., 2018, p.2)			
2. "Quantitative randomized controlled trials" (Hong et al., 2018, p.2)	2.1. Is randomization appropriately conducted?			
	2.2. "Are the groups comparable at baseline?" (Hong et al., 2018, p.2)			
	2.3. Are the outcome data completed?			
	2.4. "Are outcome assessors blinded to the intervention provided?" (Hong et al., 2018, p.2)			
	2.5. Did the participants adhere to the assigned intervention?			
3. "Quantitative non-randomized" (Hong et al., 2018, p.2)	3.1. Do the participants represent the target population?			
	3.2. Are measurements suitable in terms of both the outcome and intervention/ DA process?			
	3.3. Are the outcome data completed?			
	3.4. "Are the confounders accounted for in the design and analysis?" (Hong et al., 2018, p.2)			
	3.5. During the research period, is the intervention/the DA process conducted according to the plan?			

4. Quantitative descriptive	4.1. Does the sampling strategy relate to answer the research question/ hypothesis/ aim?			
	4.2. Does the sample represent the target population?			
	4.3. Are the measurements suitable?			
	4.4. Is the risk of nonresponse bias poor?			
	4.5. Is the statistical analysis suitable to answer the research question/ test the hypothesis/ achieve the aim?			
5. Mixed methods	5.1. Is there an adequate explanation for utilizing a mixed methods design to address the research question/ hypothesis/ aim?			
	5.2. Are the different components of the study effectively integrated to answer the research question/ test the hypothesis/ achieve the aim?			
	5.3. "Are the outputs of the integration of qualitative and quantitative components adequately interpreted?" (Hong et al., 2018, p.2)			
	5.4. "Are divergences and inconsistencies between quantitative and qualitative results adequately addressed" (Hong et al., 2018, p.2)			
	5.5. "Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?" (Hong et al., 2018, p.2)			

4.4.3 Synthesis

In this current systematic review, the data extracted across included studies was synthesized narratively to present a portrait of the participants in the included studies, the settings, DA procedures (for research question 1), the certain DA tools, techniques, and methodologies applied (for research question 2), and the challenges along with solutions for developing DA's practice in inclusive education (for research question 3). Narrative synthesis refers to presenting the results of included studies utilizing words merely in relation to data tables (Boland et al., 2017). Due to the range of diverse study designs of the studies selected in this systematic review, a narrative synthesis is considered as an appropriate instrument to synthesize the findings of studies (Schwarz et al., 2019).

5 RESULTS

This systematic literature review aims to explore the implications of DA in elementary inclusive education with the timeframe from 2010 to 2021. The study was conducted by employing explicit, systematic methods to identify, give critical appraisals, and synthesize narratively the results of all primary research associating with the research questions. In this chapter, the researcher presents the results by answering the research questions evidencing the applicability of DA in elementary inclusive education in the past ten years. The results presented are based on narrative synthesis of the findings of all included studies.

5.1 *Descriptive findings*

Of the entire pool of 25 included studies, 11 (44%) had been conducted in the U.S, 3 (12%) in the Netherlands, and 2 in Denmark. Three of the studies (12%) did not mention the location. The remaining five studies had been carried out in other countries such as the UK, Jordan, Singapore, South Africa, and Spain. In terms of characteristics of the participants, a minority of the studies (N = 2) included teachers in their sample (Bosma et al., 2012; Jeltova et al., 2011). The former study focused specifically on teacher population, while the latter had both teachers and students as participants. In the remain studies, there is a wide variety of special educational needs and ages (from kindergarten to the last year of elementary education) among students. Four studies used DA to assess students with high abilities (Al-Hroub, 2011; Al-Hroub & Whitebread, 2019; Calero et al., 2011; Vogelaar et al., 2016). Other six studies include students struggling with reading or at risk for reading disabilities (Gellert & Elbro, 2017; Gellert & Elbro, 2018) and for math difficulties (Kong & Orosco, 2016; Orosco et al., 2011; Orosco, 2014; Seethaler et al., 2012) in their inclusion criteria. Poor performance in reading (Compton et al., 2010; Elleman et al., 2011; Fuchs et al., 2011; Cho et al., 2014) and in math (Bosma et al., 2017; Moscardini & Moscardini, 2020) were

also the objects of DA in this field. DA was likewise beneficial to students with limited English proficiency (Cho et al., 2020; Seethaler et al., 2016). A small number of studies (Aljunied & Frederickson, 2011; Hasson & Dodd, 2014; Jeltova et al., 2011; Kong & Orosco, 2016) reported to examine students with autism (N = 1), with language impairment (N = 1), and ethnic minorities (N = 2) respectively. One study targeted on students with a range of barriers to learning but did not state the participants' special educational needs in detail (Amod et al., 2017). The remaining study (Lee Swanson, 2011) included subgroups of children with reading disabilities.

To meet the criteria for inclusive education, all included studies included both students with typical development and students with special educational needs, or teachers and students from both mainstream and special schools. Table 2 displays description of the study location, setting, characteristics of participants, DA procedure, DA utility, DA tools, techniques, and methodologies applied.

TABLE 3. Summary of included studies

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Al-Hroub, 2011	UK	3 Primary schools	5 “mathematically gifted children with learning difficulties” (Al-Hroub, 2011, p. 7) (nominated by teachers) Grade: 4-6	Pretest-teaching-posttest Individually Implemented by: Teacher	To assess “the untapped mathematical potential development of the participants” (p. 36)	DA was added in “a multidimensional assessment” to “identify mathematical giftedness and LDs in the same students.” (Al-Hroub, 2011, p.15)	Immediate feedback and additional mediation
Al-Hroub & Whitebread, 2019	Jordan	not mentioned	30 students with high potential in mathematics and learning difficulties Grade: 5-6	Test–intervention–test Implemented by: Mathematics teacher	To predict the students’ mathematical giftedness in twice-exceptional learners	Dynamic Mathematics Assessment Test	Mediated assistance
Aljunied & Frederickson, 2011	Singapore	Participants ' respective schools	52 children with autism from both mainstream and special schools in Singapore Age: 8-12	Pre-test- teaching (with mediation) - post-test Individually Implemented by: Researcher (a trained educational psychologist)	To assess gains in learning	The scores from gains in learning (assessed by means of DA) were compared with the scores from central coherence to investigate the relationship between “central coherence” and “gains in learning”. (Aljunied & Frederickson, 2011, p.178)	Explicit teaching Improvement of “reflective and analytic processes” (Aljunied & Frederickson, 2011, p.176) Giving assistance with cues, prompts, and reminders Feedback on students’ learning process
Amod et al., 2017	South Africa	a private remedial school	51 students presenting a variety of obstacles to learning from a			DA was combined with a model of cognitive	

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
			private remedial school Grade: 4-5			processing to develop a MLE intervention programme (Amod et al., 2017, p.428)	
Bosma et al., 2012	The Netherlands	not mentioned	188 elementary teachers The vast majority of the teachers taught at regular schools and the rest taught at schools for students with special needs.	Dynamic assessment was not conducted	To develop individual educational planning.		
Bosma et al., 2017	The Netherlands	not mentioned	120 low-performing students Grade: 2	Pretest-dynamic training-posttest Individually	To identify differences in instructional needs of 2nd grade students struggling with arithmetic	DA including multiple graduated prompts protocols	Multiprotocol-graduated prompts
Calero et al., 2011	Spain	not mentioned	127 children 64 high-IQ and 63 average-IQ Age range: 7-11	Pretest-training-posttest: 2 tests Dynamic training within test: 1 test 3 individual sessions lasting 50 minutes each Group	To identify gifted children, regardless of the IQ they show	Different dynamic tests	Assistance with increasingly precise strategic clues Directing children toward ways of solving problems

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Cho et al., 2014	not mentioned	not mentioned	134 low performing students in reading Grade: 1	Mid-November After Tier 1 in RTI model Dynamic training within the test (one-session DA) Individually Implemented by: trained examiners	Identify tier 2 nonresponders	DA of decoding in an RTI decision-making process	Instructional scaffolding
Cho et al., 2020	U.S.A.	not mentioned	368 students with LEP in which 15 students received special education services Grade: 1	November Implemented by: Research team members completing the training and practice in conducting the measures accurately	To predict later-emerging mathematics disabilities as a supplement screener	The Balancing Equations DA in L1 and in English (L2)	Scripted instructional scaffolding with visual representation
Compton et al., 2010	U.S.A.	not mentioned	355 children The majority were low performing students in reading. Grade: 1	The fall of first grade Individually Implemented by: Trained examiners	To predict second grade RD status	DA was added in the 2 nd step of a "two-step gated screening procedure" in an RTI framework (Compton et al., 2010, p.1).	Graduated prompts
Elleman et al., 2011	U.S.A.	not mentioned	100 children including high, average, and low performers in reading 12% received special education services. Grade: 2	Early May 25 min-1 hour Individually Procedure: Pretest-Training - Posttest Implemented by: Trained graduate students	To predict students' reading disabilities	DA was used in combination with comprehension measures	Graduated prompts
Fuchs et al., 2011	U.S.A.	school buildings	318 high, average, and low performing students in reading Grade: 1	The fall of first grade A one-session DA (20-30 min) Individually Implemented by: 9 trained master's students	To predict responsiveness to reading instruction	DA of decoding learning	Instructional scaffolding

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Gellert & Elbro, 2017	Denmark	Participants ' schools	160 kindergarten students including 76 children at risk for RD and 84 not-at-risk children	In the fall of kindergarten Individually Implemented by: Trained assistants and the researcher	To predict children's early reading development (6 to 7 months after "the administration of the prereading measures" (Gellert & Elbro, 2017, p. 234) (not as a supplement to standardized testing)	DA of phonological awareness before the onset of formal reading instruction	Scaffolding Prompting
Gellert & Elbro, 2018	Denmark	Participants ' schools	158 kindergarten students. The majority were students at risk for reading difficulties.	1st time: Prior to the beginning of formal reading instruction 2nd time: After a few months of formal reading instruction Implemented by: Trained assistants and the researcher	To predict difficulties with reading accuracy (as a supplement to traditional measures)	DA of decoding before and after the onset of reading instruction	Corrective feedback
Hasson & Dodd, 2014	not mentioned	not mentioned	"A 9-year-old boy with a specific language impairment, who was receiving therapy in the language base attached to his school." (Hasson & Dodd, 2014, p.353)	DA: 2 times 1st Time: at the beginning of the fall academic term 2nd time: after one term Individually	To evaluate learning potential to guide intervention for a child with language impairment	2 parallel versions of DA of sentence structure	A hierarchy of cues

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Jeltova et al., 2011	U.S.A.	classrooms	1,332 students from 4 ethnic groups. Some of the participants have learning disabilities, or low SES. 63 classroom teachers	At the end of each mathematics unit. Pretest (15-20 min)-process training (20-30 min)-posttest (15-20 min) Group Implemented by: Teachers	To evaluate groups of learners in a regular education classroom context and prone to result in higher student achievement when compared to other classroom assessments without intervention	Group-administered DA while teaching mathematics is used as a posttest assessment in a pretest-intervention-posttest methodology.	Discussion as well as demonstration of the solutions of the items in the pretest
Kong & Orosco, 2016	U.S.A.	general classroom	8 minority students at risk for MD Grade: 3	Small group (4 students)		"A word-problem-solving strategy" called "dynamic strategic math" was developed, based on a DA framework (Kong & Orosco, 2016, p. 173) DA of working memory	Instructional scaffolding
Lee Swanson, 2011	U.S.A.	not mentioned	78 students including students with RD merely; students struggling with both reading and calculation; low verbal IQ readers, and skillful readers Mean age: 11.6 (SD = 2.28)	Initial (without assistance)- gain (with assistance)- maintenance (without assistance) testing conditions Individually Implemented by: Graduate students	To predict later reading comprehension performance		Graduated cues/probes

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Moscardini & Moscardini, 2020	not mentioned	classroom	1 low-performing student in math Age: 8	Three successive 30-minute sessions. Whole class or group Implemented by: A support teacher	To inform not only instant but also long-term instructional decisions.	Interaction between teacher and students as the core to the DA procedure	Reciprocal interaction between teacher and student Observation
Orosco et al., 2011	U.S.A.	not mentioned	6 Latino ELLs at risk for MD Grade: 2	Individually		The structure of DA was utilized to develop “a word-problem-solving strategy” called “dynamic strategic math” (DSM) formulated on the principles of DA (Orosco et al., 2011,	Performance-contingent feedback Generality probes Scaffolding support
Orosco, 2014	U.S.A.	not mentioned	6 Latino ELLs at risk for MD Grade: 3	Individually		A “word-problem-solving strategy called dynamic strategic math” was developed, based on a DA framework (Orosco, 2014, p. 46)	Probing
Petersen & Gillam, 2015	U.S.A.	not mentioned	63 bilingual Latino kindergarten students “at risk for language impairment” in general education classrooms (Petersen & Gillam, 2015, p.4)	The beginning of the school year Pretest-teaching-posttest Implemented by: Research team	To predict reading ability in bilingual students	DA of nonsense word decoding	Noncontingent feedback lasting less than 3 minutes

Study	Location	Setting	Participants	DA procedure	DA utility	Methodologies	Tools and techniques
Seethaler et al., 2012	not mentioned	not mentioned	184 high, average, and at-risk students in math, in which 9 students received special educational services a learning, speech, or language disability. Grade: 1	The beginning of 1st grade Training-test Individually Implemented by: Trained graduate students	To predict 1st-grade mathematics development, especially word problems as an alternative to traditional math assessment	Balancing Equations DA	Instructional scaffolding
Seethaler et al., 2016	U.S.A.	not mentioned	129 LEP qualified to receive Second Language services, and 163 non-LEP children Grade: 1	At the start of the first grade Instructional scaffolding-mastery testing (without any instructional scaffolding). Individually Implemented by: Research assistants and the project coordinator	To predict individual differences in year-end performance on WP and calculation In non- ELs: to predict calculation but not WP outcome In ELs: to predict both outcomes.	The Balancing Equations DA	Instructional scaffolding
Vogelaar et al., 2016	The Netherlands	not mentioned	113 gifted and average-ability children Age: 7 and 8 years	Pretest-training-posttest 20-30 min/each session	To measure the cognitive abilities of intellectually gifted children	DA of analogical reasoning	Graduated prompts

Note: *WP* word problem, *LEP* limited English proficiency, *ELs* English learners, *ELLs* English language learners, *RD* reading disabilities, *MD* mathematics difficulties, *low SES* low-socioeconomic status, *RTI* responsiveness to intervention, *L1* the first language, which in this context is Spanish, *DA* dynamic assessment, *MLE* Mediated Learning Experience, *LD* learning difficulties

5.2 *Research question 1: What does empirical literature reveal about the applicability of DA in elementary inclusive education?*

5.2.1 How is DA used?

A total of 21 studies reported the applicability of DA in the final data set. See Table 2 for a summary of the reported DA utility across studies.

DA in predicting special educational needs and future performance

One common reported contribution of DA was the ability to predict students' special educational needs (Al-Hroub & Whitebread, 2019; Calero et al., 2011; Elleman et al., 2011; Petersen & Gillam, 2015), and to assist classification when combined with other static assessments (Al-Hroub, 2011; Gellert & Elbro, 2018; Cho et al., 2020). Other reported value of DA in 4 studies was predicting future performance (Lee Swanson, 2011; Seethaler et al., 2016) and future development (Gellert & Elbro, 2017; Seethaler et al., 2012).

DA in guiding intervention, planning, and giving instruction

DA was likewise employed to support the teachers in teaching process by evaluating student's learning potential to guide intervention (Hasson & Dodd, 2014), providing information for individual educational planning (Bosma et al., 2012); informing immediate and long-term instructional decisions (Moscardini & Moscardini, 2020), and identifying the differences in the need for instruction (Bosma et al., 2017).

DA in an RTI decision-making process

When added to a Responsiveness-To-Intervention (RTI) decision-making process, DA was used as a predictor of students' responsiveness to identify who would not be in response to Tier 2 and therefore require Tier 3 (Cho et al., 2014; Fuchs et al., 2011; Compton et al., 2010). RTI is a multi-tier approach for early identifying and supporting students with learning disabilities (Fletcher et al., 2004; Fuchs & Vaughn, 2012). In these studies, participants had been previously identified as unresponsive to general instruction (Tier 1) and moved to more intensive and explicit instruction (Tier 2). After being assessed dynamically, students who were identified as non-responders to Tier 2 were moved to more intensive, individualized interventions (Tier 3) (Fuchs & Vaughn, 2012).

Other utility

The remaining studies utilized this approach to assess gains in learning (Aljunied & Frederickson, 2011), to measure intellectual gifted students' cognitive abilities (Vogelaar et al., 2016) and to assess groups of students in regular classrooms as a formative assessment (Jeltova et al., 2011).

5.2.2 When is DA used?

A total of 12 studies mentioned the time point DA took place in their reports. The majority of students were evaluated at the start of the school year (Petersen & Gillam, 2015; Seethaler et al., 2012; Seethaler et al., 2016) or in the first three months of the academic year (Compton et al., 2010; Fuchs et al., 2011; Gellert & Elbro, 2017; Cho et al., 2014; Cho et al., 2020).

Two studies reported to conduct DA two times: the first time was at the start of the autumn school term and the second time was after one term (Hasson & Dodd, 2014), or after a few months of formal instruction (Gellert & Elbro, 2018). Opposite to those studies mentioned above, Elleman et al. (2011) carried out DA near the end of the academic year, i.e., in early May.

To the last study, since the methodology of the study was a pretest-intervention-posttest and DA functioned as a posttest assessment, Jeltova et al. (2011) undertook DA session immediately after each intervention, i.e., at the end of each mathematics unit.

5.2.3 Who participates in the process?

Across all 21 studies, the study conducted by Bosma et al. (2012) was excluded. Bosma et al. (2012) carried out a survey to investigate teachers' preferences for information relating to individual educational planning, particularly information obtained from dynamic testing procedures. Thus, no actual DA was conducted in that study. Therefore, data from a total of 20 studies were collected for this sub-question.

One-to-one DA

One-to-one DA constituted the majority (N = 13) (see Table 2). In most cases (N = 9), members of research team implemented the procedure. They were

trained to be capable of giving instruction, feedback, and assistance to the students. Teacher was the assessor in one study (Al-Hroub, 2011; Moscardini & Moscardini, 2020). In the latter study, albeit the testee took the sessions in the regular classroom with her classmates, she was the only student attending the dynamic process with the support teacher. The remaining two studies did not mention who were responsible for carrying out the assessment (Bosma et al., 2017; Hasson & Dodd, 2014).

Group-administered DA

Only one study (Jeltova et al., 2011) investigated group-administered DA that was performed by teachers. Teachers were divided into three groups for different conditions. All teachers participated in a training workshop lasting for two days. The content of the workshop was associated with the condition of the research.

Not mentioned

Six studies did not state whether students were assessed individually or in groups in their reports. Mostly, trained research assistants and researchers conducted the process (Petersen & Gillam, 2015; Cho et al., 2020; Gellert & Elbro, 2018). In Al-Hroub & Whitebread (2019) 's study, subject teacher was in charge of carrying out the approach. Calero et al. (2011) as well as Vogelaar et al. (2016) did not mention the examiners' information.

5.3 Research question 2: What are the tools, techniques, and methodologies applied?

DA tools and techniques

Studies varied in the nature of teaching, the degree and type of assistance, and type of feedback offering to students in the dynamic phase. Two studies (Amod et al., 2017; Bosma et al., 2012) were excluded as neither of them conducted DA procedure nor mentioned the DA tools and techniques employed. Thus, only data extracted from 23 studies was utilized to answer this part of the second research question.

Among studies providing assistance to support students, instructional scaffolding was the most commonly reported technique (N = 7). Of these studies, Cho et al. (2020) utilized visual representation to facilitate the scripted

instructional scaffolding process. Another common method implemented across studies was graduated prompts (N = 5). While four of these studies utilized graduated prompts with a single structured protocol, the other employed the same type of technique with multiple structured protocols (Bosma et al., 2017). Another similar way to assist students was cueing/probing by means of a hierarchy of cues as in Hasson and Dodd (2014), increasingly precise strategic clues (Calero et al., 2011), graduated cues/probes (Lee Swanson, 2011), generality probes (Orosco et al., 2011), or only probes (Orosco, 2014) versus a combination of cues, prompts, and reminders as in Aljunied & Frederickson (2011). Albeit Al-Hroub and Whitebread (2019) likewise reported to provide students with mediated assistance in the intervention phase, they did not describe how teachers employed this technique in detail.

In addition to offering assistance, another standard method was giving feedback including immediate feedback (Al-Hroub, 2011), corrective feedback (Gellert & Elbro, 2018), feedback on whether students succeeded or failed in their learning (Aljunied & Frederickson, 2011), and performance-contingent feedback (Orosco et al., 2011) versus noncontingent feedback (Petersen & Gillam, 2015).

Other DA techniques reported in the remaining studies encompassed exchange of views, presentation, and illustration of the solutions for the tasks in the pretest (Jeltova et al., 2011), reciprocal interaction between teacher and student (Moscardini & Moscardini, 2020), explicit teaching and enhancement of reflective and analytic processes (Aljunied & Frederickson, 2011); and directing children toward ways of solving problems (Calero et al., 2011).

Methodologies applied

The vast majority of included studies investigated dynamic assessment as an approach to assess students' skills or measure their learning potential (N = 18). The remaining studies applied DA in a variety of methodologies. Al-Hroub (2011) included DA in a multidimensional assessment to identify mathematical students with LDs. Aljunied and Frederickson (2011) used DA to assess gains in learning. The scores from gains in learning were analyzed with the scores on central coherence to examine the correlation between these variables. Amod et al. (2017) combined DA with a specific model of cognitive processing to develop a Mediated Learning Experience (MLE) intervention programme. Compton et al. (2010) included DA in the second stage of a two-step gated screening process in

a Responsiveness to Intervention (RTI) framework to predict students' status. The last three studies (Kong & Orosco, 2016; Orosco, 2011; Orosco, 2014) utilized DA as a framework to develop a word-problem solving strategy.

5.4 What are the challenges for DA in elementary inclusive education and the solutions for developing its practice?

Across all included studies, a total of nine studies mentioned the challenges for applying this approach in elementary inclusive education (see Table 4 for information regarding challenges and solutions in included studies).

The whole approach

In terms of the entire process, the question of the construct validity is a pressing one (Cho et al., 2014). To explore the construct validity of decoding DA, Cho et al. (2014) recommended future research to apply a structural equation modelling approach by utilizing various methods to indicate students' responsiveness. An additional solution is to investigate whether decoding and responsiveness show a distinct or unitary construct by employing standardized and decoding tests to measure the same construct, decoding (Cho et al., 2014). Another unanswered question mentioned in this study is the influence of different types of instruction on DA's predictive validity. One possible solution is using multilevel modelling techniques to model this variance, i.e., classroom instruction provided for students in Tier 1 or Tier 2 of RTI framework and to examine whether the type or features of instruction used in the classroom affects DA's predictive value (Cho et al., 2014). Another reported problem is the floor effect of DA with few low-performing participants (Fuchs et al., 2011). The measure failed to provide evidence about how much these poor readers learned the CVC(C)ing skill (Fuchs et al., 2011). Two solutions were presented. The first one is to decrease the difficulty of the CVC(C)ing task (Fuchs et al., 2011). The second solution is to deal with the difficulty of the task by providing more useful guidance or more direct and straightforward probes earlier in a hierarchy of prompts (Fuchs et al., 2011).

The test

In terms of the test, one main problem is that the employment of static test as a criterion measure made a content-specific DA as a traditional static form of

assessment as in Jeltova et al. (2011). This issue has yet to be solved due to the dominance of this form of test in the state-based testing in the country where the study was conducted and other places (Jeltova et al., 2011). One minor problem that might notwithstanding be worth observing is that in Petersen and Gillam's (2015) study, during posttest, participants were occasionally uncommunicative and did not want to carry out the test unassisted. In that case, the tester offered some moderately neutral prompts such as "You can guess—what do you think this word says?" or "Remember what I told you?" to guide the students.

The assistance/mediation provided for students

Main challenges relating to the mediation and assistance offered for students during DA process were the absence of evaluations of the effectiveness of the prompts or their appropriate arrangement and the efficacy of the inference instruction and feedback (Elleman et al., 2011). Whether providing prompts supported children in making the inferences or whether time demanded for the inference instruction was necessary for the assessment to predict students' reading comprehension deficits has been still unanswered (Elleman et al., 2011). One more challenge regarding prompts was posed when graduated prompts with various structured protocols was used in Bosma et al. (2017)'s study. This technique was reported to make the examiners have difficulty in learning, performing, and scoring. One solution for this problem is offering teachers with a training in computer skills that can assist them in administrating and scoring the necessary prompts.

Another question was raised when conducting DA with autistic children in Aljunied and Frederickson's study (2011). Offering cues, probes, or reminders utilized in teaching and mediating students might set out greater requirements on central coherence as students encounter more difficulty. Students with weak central coherence may get low scores in gain in learning. Therefore, the employment of DA may underestimate the learning potential of children with autism. One vague solution recommended to this issue in the paper was developing particular mediation techniques that can limit the impacts of weak central coherence.

TABLE 4. Challenges and solutions

Study	Challenges	Solutions
<i>Problems relating to the whole DA process</i>		
Cho et al., 2014	Challenge 1: the construct validity of DA Challenge 2: whether the emphasis of instruction moderates DA's predictive validity.	Solution for challenge 1: 1. "Using a structural equation modeling approach to derive a latent construct of responsiveness by using multiple indicators of responsiveness" (Cho et al., 2014, p.421) 2. "Testing whether decoding and responsiveness represent a unitary or distinct construct using static and decoding measures of the same construct, decoding" (Cho et al., 2014, p.421).
Fuchs et al., 2011	the floor effect of DA with few low-performing participants	2 solutions: 1. To replace a less difficult task for the CVC(C)ing task. 2. To "keep the CVC(C)ing task but to give children more helpful guidance—more explicit information—earlier in the sequence of graduated prompts" (p.346).
<i>Problems relating to the tests</i>		
Jeltova et al., 2011	Using a static test, as a criterion measure, makes DA as a traditional static form of assessment.	not mentioned
Petersen & Gillam, 2015	During posttest, in some cases, children were reluctant and did not want to conduct the test independently.	Assessor offered some neutral prompts for instance "You can guess—what do you think this word says?" or "Remember what I told you?"
<i>Problems relating to the assistance/ mediation provided for students</i>		
Aljunied & Frederickson, 2011	"Providing prompts, cues, or reminders in mediation may impose greater demands on central coherence as increasing difficulty is experienced. Weak central coherence may make the gains less effective. The use of DA may underrate the learning potential of children with autism" (p.182)	Developing specific mediation strategies that are able to minimize the impacts of weak central coherence.
Bosma et al., 2017	The multiple protocols are difficult for testers to learn, apply and score.	A computerized training, such as a tangible table
Elleman et al., 2011	1. "An evaluation of the effectiveness of the inference instruction and feedback was not conducted. The inference training required time that may not be necessary for the test to predict comprehension deficits" (p.356). 2. "The effectiveness of the prompts or their proper ordering was not conducted. The researcher was not sure if providing clues helped the children make the inferences" (p. 356).	not mentioned

6 DISCUSSION

This study used systematic review methodology to investigate the implications of dynamic assessment in elementary inclusive education with the time frame from 2010 to 2021. This work aimed to answer research questions investigating (1) the applicability of DA in this scope of work, (2) the DA tools and techniques applied, (3) the methodologies applied, and (4) the challenges for applying DA in inclusive education and the recommended solutions for developing its practice.

6.1 Main findings

Methodologies applied in research concerning DA in inclusive education

Previous papers reviewing DA targeted primarily on the utilities of DA, DA challenges, and the effect sizes of research. These non-empirical and empirical research demonstrate the scarcity of reviews on methodologies applied in studies investigating DA in inclusive education. Therefore, the picture was less clear in terms of what methodologies have been adopted in this scope of work. The current review filled this research gap by extracting data concerning this aspect.

The most common methodologies implemented across studies are investigating DA as a sole approach to assess students' skills or measure their learning potential, and a supplement to static tests. Albeit the aforementioned method is typical in this field, there are also other different novel methodologies adopted. These methodologies encompass (1) adding DA to a multidimensional assessment, (2) employing DA as a measure to assess one of the variables, i.e., gains in learning, (3) merging DA with a cognitive processing model to develop an intervention program, (4) designing a word-problem solving strategy based on DA framework, and (5) adding DA to the second stage of a two-step gated screening process in an RTI framework. An overview of methodologies applied would facilitate the development of DA in both theory and practice, supplying future research with different methods of conducting studies on DA.

DA applications in elementary inclusive education

DA was reported to have practical value in predicting special educational needs, future performance and development, in guiding intervention, planning, and giving instruction, and in predicting students' responsiveness in an RTI decision-making process. Consistent with Elliott (2003), this current review supports the statement that this approach generates better predictions than static cognitive tests. However, there are some conflicts regarding the predictive power of DA. First, DA was reported to produce greater accuracy of prediction when employed as a sole predictor in some studies and as a supplement to standardized testing in other studies. Second, the duration of the prediction value of DA was also a controversial issue. While Gellert and Elbro (2017) stated that this unique value only lasted 6 to 7 months, Seethaler et al., 2016 evidenced the value of DA in predicting students' year-end performance. According to Gellert and Elbro (2017), there are many factors that affect the preciseness of long-lasting prediction consisting of the selection of participants, the measure's sensitivity, and educational practices, to mention a few.

Another reported way of utilizing DA's predictive power was to locate non-responders to Tier 2 in RTI model by predicting students' responsiveness to instruction. RTI is a 3-tier approach for early discerning and supporting learners who are prone to learning disabilities (Fletcher et al., 2004; Fuchs & Vaughn, 2012). A primary challenge for RTI models lies in the decision-making process of the amount and what kind of support each student required. In their non-empirical review, Elliot et al. (2018) reported the argument provoked among researchers about the validity of DA as a relatively quick method of forming a rational judgement about which tier is suitable for a given student. This current review gathered considerable pieces of evidence to support employing DA as a predictor in an RTI decision-making procedure. When using DA in an RTI model, the researcher should take the influence of different types of instruction on DA's predictive validity into consideration. It was suggested that the researchers and practitioners of DA use multilevel modelling techniques to model this variance, which is the classroom instruction offered for students in Tier 2 of RTI framework, and test whether the type or characteristics of instruction used in the classroom affects DA's ability to predict.

Despite of being considered as a primary contribution of DA in Elliott et al. (2018)'s review, the utility of DA in guiding intervention was only reported in one study. Other similar applications were supporting the instruction process both in short term and long term by identifying students' needs and providing information for individual educational planning. In studies investigating the first two utilities, the demonstrations of what type of information was derived from this approach and how that information was used in practice were provided. Nevertheless, the generalizability of the findings of these studies should be considered with a careful attitude since all of them are case studies. With the last application, however, there is a need for future research to explore how and what kinds of information educators really utilize and require for individual educational planning. Having the capability of providing information that can help teachers determine students' educational needs rather than merely predicting learning problems, DA deals with the problems that current standardized assessments in inclusive education have. Leeber et al. (2013) point out that reports from those static tests hardly provide any hints about "how to understand the child's problems, how to work with a child, what the child may be able to do, or what the real needs are" (Leeber et al., 2013, p.5), thus they fail to support teachers and school psychologists in designing educational or intervention plans. Nevertheless, compared to other utilities, there is a scarcity of the studies investigating these applications of DA. Therefore, more future research should be undertaken to investigate how the teachers, educators, and school psychologists use the information obtained from DA process in classroom contexts and the effectiveness of its application.

Furthermore, the present review discovered some novel utilities of DA that have not been reported in previous reviews. Beside measuring students' potential for the purpose of giving predictions and providing information for supporting the teaching process, DA was also employed to assess students' gains in learning, measure intellectual gifted students' cognitive abilities and assess groups of students in classroom contexts as a formative assessment. However, each application was reported in only one study. Thus, there is a need for future research to investigate extensively and thoroughly the validity of these DA utilities in inclusive settings.

Implementing DA in elementary inclusive education

For the vast majority of included studies, trained research assistants and researchers carried out the DA procedure. Teachers were reported to conduct DA in only three studies. All the examiners except researchers had to be trained in intensive workshops to have the capability to provide instruction, feedback, assistance, and mediation to the students, which is consistent with Haywood and Tzuriel (2002). As teachers, and school psychologist are the practitioners in school and work directly with students every school day, there is a need for future research to investigate more how teachers and school psychologists perform DA in school and classroom settings and the interprofessional collaboration in school.

The testers employed different types of feedback and assistance to support students in the dynamic phase. Across included studies, instructional scaffolding was the most widely used technique in DA. Instructional scaffolding is an interactive process through which teachers gradually provide supports for students so as to improve learning and assist with the mastery of tasks (Vygotsky, 1978; Walqui, 2006). The testers could use visual representation to facilitate the scaffolding procedure (Cho et al., 2020). There are approximately a maximum of four or five degrees of instructional scaffolding starting from the least explicit to the most explicit for each problem or task (Seethaler et al., 2012). In case the individual is unsuccessful to solve the task after the last level of scaffolding, the DA is ended (Seethaler et al., 2012).

Graduated prompts was also one of the most common instructional techniques of DA (Compton et al., 2010). The nature of this technique bears a relative resemblance to instructional scaffolding. If a student provides an incorrect answer, the tester supports the student by offering a hierarchy of prompts until the learner can master the test item (Campione & Brown, 1987). The studies employed graduated prompts with single structured protocol are greater in number than ones with multiple protocols of graduated prompts. In fact, there is only one study using the multiple protocols. The multiple protocols allowed the examiners to classify groups of students with different needs for instruction. Nevertheless, this specific technique was found to be difficult for the examiners to master, conduct, and score because to respond to each answer of an individual, the examiner had to select the appropriate protocol, and prompts in

some cases were not useful for all students (Bosma et al., 2017). One solution for this problem is providing teachers with a training in computer skills that can assist them in administering and scoring the necessary prompts.

Cueing and probing were likewise employed in some studies by means of a hierarchy of cues, increasingly explicit clues, graduated cues, generality probes, or a composite of cues, prompts, and reminders. Similar to above-mentioned techniques, the examiner provides assistance if the student finds it difficult to solve the task independently. These assistance strategies are mediated to the student as demanded by the individual student and in response to the student's endeavours. Albeit these strategies are reported to support the students to reach their potential in most cases, to some particular students such as autistic students, it may raise considerable challenges for them to understand the meaning of those cues, probes, and reminders in mediation and make meaningful connections to solve the tasks. Thus, there is a possibility that the employment of DA fails to measure the learning potential of autistic children. The solution recommended in the study was relatively vague. Thus, further attention needs to be given to develop specific mediation strategies that are appropriate for students with autism.

Different types of feedback were used in the teaching session in DA procedure, consisting of immediate feedback on students' problem-solving strategies, corrective feedback, feedback on success or failure in the learning process, performance-contingent feedback, and noncontingent feedback. In terms of performance-contingent feedback, the study applied this tool did not mention how it was employed in detail. Regarding noncontingent feedback, there is a considerable difference between this tool and other types of feedback. While other types of feedback targeted on individual's performance, noncontingent feedback was reported to involve teachers responding to a student in a standardized manner, paying no attention to individual student mistakes or errors (Petersen & Gillam, 2015).

In addition to instructional scaffolding, feedback, prompts, cues, and probes, there are other teaching techniques that are employed in DA procedure. They consist of discussion and demonstration of the solutions, reciprocal interaction between teacher and student, explicit teaching and enhancement of reflective and analytic processes, and directing children toward ways of solving problems.

Typically, students were assessed individually in the DA process so that the teachers could identify the need of that particular student or measure his/her potential effectively and accurately via providing individualized mediation and assistance. Despite its merits, one-on-one DA needs a great deal of time to conduct the procedure, which is, in some cases, impractical especially in regular classroom settings. Utilizing DA on a group level could be a partial solution for this. However, only one study conducted group-administered DA in a particular academic subject in a regular real-life classroom context, which yielded promising findings. Thus, future research should delve into group-administered DA in greater depth in order to increase the practicality but still ensure the efficacy of DA.

When implementing DA in elementary inclusive education, there are some problems that future researchers and practitioners of DA should take into consideration. In this part, the researcher only presented the challenges that was not above-mentioned in this chapter in order to avoid duplication. Regarding to the whole process, consistent with Haywood and Tzuriel (2002), the construct validity of DA is considered as a persistent issue. Construct validity is one way to test how well DA measures up to its claims (Westen & Rosenthal, 2003). If a test lacks construct validity, it will be difficult to make an interpretation based on the results gained from that test or procedure (Westen & Rosenthal, 2003). In order to establish construct validity of DA, Cho et al. (2014) suggested using traditional static testing of the same construct to investigate whether the results gained from DA show a distinct or unitary construct when compared to static testing. Another solution is utilizing multiple indicators of responsiveness.

Another problem regarding to the entire procedure that was not mentioned in previous reviews is the floor effect of DA with few low-performing participants. Floor effect is a situation in which many students perform very poorly when administered the test (Fuchs et al., 2011). Thus, the measure fails to tell the differences in students' performance (Fuchs et al., 2011). In order to minimize this effect, the examiner can reduce the difficulty of the task or provide more useful guidance or more direct and straightforward probes earlier in a hierarchy of prompts (Fuchs et al., 2011).

In terms of the instrument, one key issue is the use of static test as a criterion measure both in pre-test and post-test made DA as a traditional standardized

assessment. The question is whether it is necessary to specifically design a new test for DA especially when that static test is widely used, and significant decisions are made based on that test.

Regarding to the assistance/mediation provided for students, previous studies did not mention the problem relating to the effectiveness of DA tools and techniques. This current review indicated the lack of evaluations of the prompts or their appropriate arrangement and the efficacy of the inference instruction and feedback. These tools and techniques have a substantial influence on students' performance, thus affect the assessment results. This problem, however, has yet to be addressed seriously. Therefore, the future research should take this problem into consideration when conducting DA.

6.2 Limitations, validity, and reliability of the current systematic review

Some limitations were identified in this current systematic review. First, due to including only articles in peer-reviewed journals, the present review is prone to publication bias and may exaggerate the effect of DA (Shadish et al. 2016). Second, systematic reviews are characteristically conducted collaboratively by a group of researchers. Therefore, the fact that the current review was conducted by a sole author weakens its overall strength and credibility. Lack of the constituent part of the cooperation, discussion, and debate between researchers, the selection process of included studies, the interpretation and coding procedure are depicted in a more subjective way. This raises the possibility of omissions and errors. Last, albeit the researcher of this present study has attempted to make a synthesis of relevant findings from as extensive an array of eligible literature as possible, potential studies may be missed due to the search terms employed. There is possibility that the search strings ["Special education" OR "Inclusive education" OR "special class"] are relatively insufficient to capture all relevant papers investigating DA in inclusive education.

Albeit there are some limitations existing in several aspects, the reliability and validity of this current study are still guaranteed. Firstly, its theoretical framework is established cogently by going through from the epistemology of DA to the applicability of DA in general. Secondly, this systematic review was

designed and verified by explicit, well-defined, and rigorous methods for identifying, critically appraising, and synthesizing all relevant findings. In addition to the systematic methods adopted, the utility of a wide range of databases and the inclusion of a variety of study designs ensure the validity and reliability of the present review. Lastly, in general, the methodological quality of the included studies, which were assessed by MMAT, was acceptable. The findings across different studies are remarkably consistent as the vast majority of included paper reported the effectiveness of applying DA to assess learners with special needs or to support the teaching and learning practice in inclusive education.

Furthermore, this review makes a meaningful contribution to both theory and practice. This current study contributed to the field of alternative assessments in inclusive education by providing an overview and synthesis of the literature on the implications of DA. This study not only filled the research gap by targeting particularly on elementary inclusive education but also contributed to teaching practice by providing a set of strategies that could be used as guidelines for teachers on developing assessment in inclusive classrooms.

7 CONCLUSION

This present study aimed to provide a narrative systematic review of the literature on dynamic assessment in elementary inclusive education by focusing on the utilities of DA, the methodologies applied, DA tools and techniques, the challenges as well as recommended solutions from 2010 to 2021. The purpose was to guide both future research on and the practical implementation of this assessment in inclusive education.

The research questions were answered by collecting all studies on DA in elementary inclusive education, extracting the relevant data and narratively synthesizing the findings. In this conclusion chapter, the major findings of the present study are presented together with a set of guidelines for teachers on developing this approach to assessment in primary inclusive classrooms.

First, across included studies, there is a variety of methodologies applied in this scope of work, in which examining DA as a sole approach to assessment or as a supplement to other static tests is the most common one.

Second, regarding the applications of DA, DA was commonly used to predict students' special educational needs, future performance, and academic development, and to predict students' responsiveness in an RTI decision-making process. Only two studies reported to apply DA to guide intervention or to inform both short-term and long-term instruction. The value of providing information for individual educational planning was reported in one study. In addition, this review also indicated some novel DA applications that was not mentioned in previous reviews.

Last, regarding implementing DA in practice, trained research assistants and researchers were reported to conduct this approach in the majority of included studies. To support students in the teaching session and assess students dynamically, the examiners employed different types of feedback, mediation and assistance strategies. Among them, instructional scaffolding and graduated prompts are the most widely used techniques. The present review also

identified some key challenges when applying DA and recommended solutions for future research and DA practitioners.

Based on the findings of all included studies, a set of guidelines is provided on the implementation of DA in inclusive classroom context. Before the assessment, it is advisable for the teachers to get an overview of student's profile by interviewing parents, observing the student, and cooperating with other professionals in school such as school psychologists. By doing that, teachers can choose appropriate DA tools and techniques for the given student among many different types of feedback, mediation and assistance strategies. In addition to DA tools and techniques, teachers also need to find a suitable DA instrument in relation to the assessment goals and decide whether employing DA as a sole approach or combining DA with other static tests.

During the assessment procedure, teachers should create a positive collaborative environment to stimulate students as well as observe and take notes to modify the tools and techniques currently used. By means of observing and interacting with students, the teachers can evaluate to some extent the effectiveness of the feedback, prompts, and instruction implemented in the assessment procedure.

After the assessment process, the educators can utilize the information obtained from this approach to assessment to identify the differences in students' needs in order to support them in individual educational planning and to inform both short-term and long-term instruction.

The study makes a valuable contribution to theory and practice of DA in inclusive education. The study has provided a fresh insight into the methodologies applied across studies on DA in this scope of work and DA applications. In addition, a set of guidelines is also provided to support teachers in developing DA in classroom settings.

The research still needs to be extended by cooperating with other researchers from different disciplines such as educational psychology to get a more in-depth understanding of the methods to measure the outcomes and efficacy of DA, DA instruments, the effectiveness of DA from the perspective of teachers, school psychologists and students. Future research on DA should focus on how the teachers implement DA in regular classroom contexts and how they use the DA report to support their teaching process.

REFERENCES

- Ainscow, M., Dyson, A., Hopwood, L., & Thompson, S. (2016). *Primary schools responding to diversity: Barriers and possibilities*. York, UK: Cambridge Primary Review Trust.
- Al-Hroub, A. (2011). Developing assessment profiles for mathematically gifted children with learning difficulties at three schools in Cambridgeshire, England. *Journal for the Education of the Gifted*, 34(1), 7–44. <https://doi.org/10.1177/016235321003400102>
- Al-Hroub, A., & Whitebread, D. (2019). Dynamic assessment for identification of twice-exceptional learners exhibiting mathematical giftedness and specific learning disabilities. *Roeper Review*, 41(2), 129–142. <https://doi.org/10.1080/02783193.2019.1585396>
- Aljunied, M., & Frederickson, N. (2011). Does central coherence relate to the cognitive performance of children with autism in dynamic assessments? *Autism*, 17(2), 172–183. <https://doi.org/10.1177/1362361311409960>
- Alony, S., & Kozulin, A. (2007). Dynamic assessment of receptive language in children with Down syndrome. *International Journal of Speech-Language Pathology*, 9(4), 323–331. <https://doi.org/10.1080/14417040701291415>
- Amod, Z., Heafield, D., & Seabi, J. (2017). Assessing a REMEDIAL intervention programme in developing the planning skills of grade 4 and 5 Learners. *International Journal of Disability, Development and Education*, 65(4), 428–441. <https://doi.org/10.1080/1034912x.2017.1406067>
- Boland, A., Cherry, M. G., & Dickson, R. (Eds.). (2017). *Doing a systematic review: A student's guide* (2nd ed.). SAGE PUBLICATIONS.
- Bosma, T., Hessels, M. G. P., & Resing, W. C. M. (2012). Teachers' preferences for educational planning: Dynamic testing, teaching' experience and teachers' sense of efficacy. *Teaching and Teacher Education*, 28(4), 560–567. <https://doi.org/10.1016/j.tate.2012.01.007>

- Bosma, T., Stevenson, C. E., & Resing, W. C. (2017). Differences in need for instruction: Dynamic testing in children with arithmetic difficulties. *Journal of Education and Training Studies*, 5(6), 132.
<https://doi.org/10.11114/jets.v5i6.2326>
- Budoff, M., & Friedman, M. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology*, 28(5), 434–439. <https://doi.org/10.1037/h0040631>
- Caffrey, E., Fuchs, D., & Fuchs, L. (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>
- Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). Guilford Press.
- 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. <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.
<https://doi.org/10.1177/0022219419898887>
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk FIRST-GRADE readers for Early intervention: Eliminating false positives and exploring the promise of a two-stage GATED screening process. *Journal of Educational Psychology*, 102(2), 327–340.
<https://doi.org/10.1037/a0018448>

- Dyson, A. (1999). An international perspective on inclusion. In *Managing inclusive education*.
- Elleman, A. M., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities, 44*(4), 348–357. <https://doi.org/10.1177/0022219411407865>
- Elliott, J. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*(1), 15–32. <https://doi.org/10.1080/00131910303253>
- Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*(2), 304–331. <https://doi.org/10.1007/s11881-004-0015-y>
- 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., & Vaughn, S. (2012). Responsiveness-to-intervention. *Journal of Learning Disabilities, 45*(3), 195–203. <https://doi.org/10.1177/0022219412442150>
- Gellert, A. S., & Elbro, C. (2017). Does a dynamic test of phonological awareness predict early reading difficulties? *Journal of Learning Disabilities, 50*(3), 227–237. <https://doi.org/10.1177/0022219415609185>
- Gellert, A. S., & Elbro, C. (2018). Predicting reading disabilities using Dynamic assessment of decoding before and after the onset of reading instruction: A longitudinal study from kindergarten through Grade 2. *Annals of Dyslexia, 68*(2), 126–144. <https://doi.org/10.1007/s11881-018-0159-9>
- Ghanbarpour, M. (2017). A qualitative meta-synthesis of research on Dynamic assessment of second/foreign language Learning: Implications for language teachers. *Journal of Language Teaching and Research, 8*(4), 731. <https://doi.org/10.17507/jltr.0804.12>

- Glisan, E., & Donato, R. (2004). It's Not "Just a Matter of Time:" A Response to Rifkin. *Foreign Language Annals*, 37(3), 470–476.
<https://doi.org/10.1111/j.1944-9720.2004.tb02704.x>
- Grigorenko, E. (2009). Dynamic Assessment and Response to Intervention: Two Sides of One Coin. *Journal of Learning Disabilities*, 42(2), 111–132.
<https://doi.org/10.1177/0022219408326207>
- Guthke, J., & Stein, H. (1996). Are Learning Tests the Better Version of Intelligence Tests? *European Journal of Psychological Assessment* : Official Organ of the European Association of Psychological Assessment, 12(1), 1–13. <https://doi.org/10.1027/1015-5759.12.1.1>
- Harsh Suri, & David Clarke. (2009). Advancements in Research Synthesis Methods: From a Methodologically Inclusive Perspective. *Review of Educational Research*, 79(1), 395–430.
<https://doi.org/10.3102/0034654308326349>
- Hasson, N., & Dodd, B. (2014). Planning intervention using dynamic assessments: A case study. *Child Language Teaching and Therapy*, 30(3), 353–366. <https://doi.org/10.1177/0265659014521642>
- 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
- Haywood, H., & Lidz, C. (2006). *Dynamic Assessment in Practice: Clinical and Educational Applications*. In *Dynamic Assessment in Practice*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511607516>
- Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities*, 44(4), 381–395.
<https://doi.org/10.1177/0022219411407868>
- Kirschenbaum, R. (1998). Dynamic Assessment and Its Use With Underserved Gifted and Talented Populations. *Gifted Child Quarterly*, 42(3), 140–147.
<https://doi.org/10.1177/001698629804200302>
- Kiuppis, F. (2013). Why (not) associate the principle of inclusion with disability? Tracing connections from the start of the 'salamanca process.'

- International Journal of Inclusive Education*, 18(7), 746–761.
<https://doi.org/10.1080/13603116.2013.826289>
- Kong, J. E., & Orosco, M. J. (2016). Word-problem-solving strategy for minority students at risk for math difficulties. *Learning Disability Quarterly*, 39(3), 171–181. <https://doi.org/10.1177/0731948715607347>
- Kozulin, A. (2011). Learning potential and cognitive modifiability. *Assessment in Education : Principles, Policy & Practice*, 18(2), 169–181.
<https://doi.org/10.1080/0969594X.2010.526586>
- Kozulin, A. (2015). Dynamic Assessment of Adult Learners' Logical Problem Solving: A Pilot Study With the Flags Test. *Journal of Cognitive Education and Psychology*, 14(2), 219–230. <https://doi.org/10.1891/1945-8959.14.2.219>
- Kozulin, A., & Garb, E. (2016). Dynamic Assessment of EFL Text Comprehension. *School Psychology International*, 23(1), 112–127.
<https://doi.org/10.1177/0143034302023001733>
- Lawrence, N., & Cahill, S. (2014). The impact of dynamic assessment: an exploration of the views of children, parents and teachers. *British Journal of Special Education*, 41(2), 191–211. <https://doi.org/10.1111/1467-8578.12060>
- Lee Swanson, H. (2011). Dynamic testing, working memory, and reading Comprehension growth in children with Reading Disabilities. *Journal of Learning Disabilities*, 44(4), 358–371.
<https://doi.org/10.1177/0022219411407866>
- Leeber, J., Candeias, A., & Grácio, L. (2011). *With a different glance: Dynamic assessment of functioning of children oriented at development & inclusive learning*. Antwerp: Garant Publishing.
- Leeber, J., Partanen, P., Candeias, A., Grácio, M. L., Bohacs, K., Sonnesyn, G., Van de Veire, H., Van Trimpont, I., Orban, R., János, R., Demeter, K., Schraepen, B., & Dawson, L. (2013). The need for a more dynamic and ecological assessment of children experiencing barriers to learning to move towards inclusive education: A summary of results of the Daffodil Project. *Transylvanian Journal of Psychology, Special Issue*, 175-205.

- Madureira Ferreira, J. (2018). Inclusive early childhood education and the role of peer interaction: Brazil and Finland in dialogue. Tampere University Press.
- Mariga, L., MacConkey, R., & Myezwa, H. (2014). *Inclusive education in low-income countries*. Disability Innovations Africa.
- Michailakis, D., & Reich, W. (2009). Dilemmas of inclusive education. *Alter*, 3(1), 24–44. <https://doi.org/10.1016/j.alter.2008.10.001>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and META-ANALYSES: The PRISMA statement. *BMJ*, 339(jul21 1). <https://doi.org/10.1136/bmj.b2535>
- Moscardini, L., & Moscardini, C. (2020). Dynamic assessment and teachers' knowledge of children's mathematical thinking: A case study in children's mathematics. *Support for Learning*, 35(4), 522–541. <https://doi.org/10.1111/1467-9604.12331>
- Naicker, S. M. (1999). *Curriculum 2005: A space for all: An introduction to inclusive education*. Renaissance.
- Noblit, G., & Hare, R. (1988). *Meta-ethnography : synthesizing qualitative studies* . Sage.
- Orosco, M. J. (2014). Word problem strategy for latino english language learners at risk for math disabilities. *Learning Disability Quarterly*, 37(1), 45–53. <https://doi.org/10.1177/0731948713504206>
- Orosco, M. J., Swanson, H. L., O'Connor, R., & Lussier, C. (2011). The effects of dynamic strategic math on english language learners' word problem solving. *The Journal of Special Education*, 47(2), 96–107. <https://doi.org/10.1177/0022466911416248>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1). <https://doi.org/10.1186/s13643-016-0384-4>
- 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. <https://doi.org/10.1177/0022219413486930>
- Regalla, M., & Peker, H. (2017). Prompting All Students to Learn: Examining Dynamic Assessment of Special Needs and Typical Students in a Prekindergarten Inclusive French Program: Prompting All Students to

- Learn: Examining Dynamic. *Foreign Language Annals*, 50(2), 323–338. <https://doi.org/10.1111/flan.12261>
- Regalla, M., & Peker, H. (2017). Prompting All Students to Learn: Examining Dynamic Assessment of Special Needs and Typical Students in a Prekindergarten Inclusive French Program. *Foreign Language Annals*, 50(2), 323–338. <https://doi.org/10.1111/flan.12261>
- Rezaee, A. A., & Ghanbarpour, M. (2016). The Measurement Paradigm and Role of Mediators in Dynamic Assessment: A Qualitative Meta-Synthesis. *International Journal of Language Studies*, 10(4), 77–108.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. Springer.
- Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Predicting first graders' development of Calculation VERSUS word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology*, 104(1), 224–234. <https://doi.org/10.1037/a0024968>
- 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>
- Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis*, 49(3), 656–673. <https://doi.org/10.1002/jaba.308>
- Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, 71(2), 321–363. <https://doi.org/10.3102/00346543071002321>
- UNESCO (1994). *Salamanca Statement and Framework for Action on Special Needs Education*. Spain: Salamanca.
- UNESCO (2005). *Guidelines for Inclusion. Ensuring Access to Education for All (Orientaciones para la inclusión. Asegurar el acceso a la Educación para Todos)*
- Vogelaar, B., Bakker, M., Elliott, J. G., & Resing, W. C. (2018). Dynamic testing and test anxiety amongst gifted and average-ability children. *British*

Journal of Educational Psychology, 87(1), 75–89.

<https://doi.org/10.1111/bjep.12136>

Vygotsky, L., (1978). *Mind in Society*. Harvard University Press.

Wagner, R., & Compton, D. (2011). Dynamic Assessment and Its Implications for RTI Models. *Journal of Learning Disabilities*, 44(4), 311–312.

<https://doi.org/10.1177/0022219411407859>

Walqui, A. (2006). Scaffolding instruction for english language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism*, 9(2), 159–180. <https://doi.org/10.1080/13670050608668639>

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*, 50(2), 204–211.

<https://doi.org/10.1111/j.1365-2648.2005.03380.x>

Watkins, A. (2007). *Assessment in inclusive settings: Key issues for policy and practice*. Odense: European Agency for Development in Special Needs Education.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84(3), 608–618. <https://doi.org/10.1037/0022-3514.84.3.608>

Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*, 66(3), 377–397.

<https://doi.org/10.1080/00131911.2013.780009>

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (2020). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer VS.

APPENDICES

Appendix A: The original version of MIXED METHODS APPRAISAL TOOL (MMAT), VERSION 2018 (Hong et al., 2018, p.2)

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
"Screening questions (for all types)" (Hong et al., 2018, p.2)	S1. "Are there clear research questions?" (Hong et al., 2018, p.2)				
	S2. "Do the collected data allow to address the research questions?" (Hong et al., 2018, p.2)				
	<i>"Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions."</i> (Hong et al., 2018, p.2)				
1. Qualitative	1.1. "Is the qualitative approach appropriate to answer the research question?" (Hong et al., 2018, p.2)				
	1.2. "Are the qualitative data collection methods adequate to address the research question?" (Hong et al., 2018, p.2)				
	1.3. "Are the findings adequately derived from the data?" (Hong et al., 2018, p.2)				
	1.4. "Is the interpretation of results sufficiently substantiated by data?" (Hong et al., 2018, p.2)				
	1.5. "Is the coherence between qualitative data sources, collection, analysis, and interpretation?" (Hong et al., 2018, p.2)				
2. "Quantitative randomized controlled trials" (Hong et al., 2018, p.2)	2.1. "Is randomization appropriately performed?" (Hong et al., 2018, p.2)				
	2.2. "Are the groups comparable at baseline?" (Hong et al., 2018, p.2)				
	2.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)				
	2.4. "Are outcome assessors blinded to the intervention provided?" (Hong et al., 2018, p.2)				
	2.5. "Did the participants adhere to the assigned intervention?" (Hong et al., 2018, p.2)				
3. "Quantitative non-randomized" (Hong et al., 2018, p.2)	3.1. "Are the participants representative of the target population?" (Hong et al., 2018, p.2)				
	3.2. "Are measurements appropriate regarding both the outcome and intervention (or exposure)?" (Hong et al., 2018, p.2)				

	3.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)				
	3.4. "Are the confounders accounted for in the design and analysis?" (Hong et al., 2018, p.2)				
	3.5. "During the study period, is the intervention administered (or exposure occurred) as intended?" (Hong et al., 2018, p.2)				
4. Quantitative descriptive	4.1. "Is the sampling strategy relevant to address the research question?" (Hong et al., 2018, p.2)				
	4.2. "Is the sample representative of the target population?" (Hong et al., 2018, p.2)				
	4.3. "Are the measurements appropriate?" (Hong et al., 2018, p.2)				
	4.4. "Is the risk of nonresponse bias low?" (Hong et al., 2018, p.2)				
	4.5. "Is the statistical analysis appropriate to answer the research question?" (Hong et al., 2018, p.2)				
5. Mixed methods	5.1. "Is there an adequate rationale for using a mixed methods design to address the research question?" (Hong et al., 2018, p.2)				
	5.2. "Are the different components of the study effectively integrated to answer the research question?" (Hong et al., 2018, p.2)				
	5.3. "Are the outputs of the integration of qualitative and quantitative components adequately interpreted?" (Hong et al., 2018, p.2)				
	5.4. "Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?" (Hong et al., 2018, p.2)				
	5.5. "Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?" (Hong et al., 2018, p.2)				

Appendix B: Instructions for categorizing studies to rate in the MMAT* (Hong et al., 2018, p.8)

