Regular Article

Journal of Psychoeducational Assessment
2024, Vol. 0(0) 1–20

DOI: 10.1177/07342829241290277
journals.sagepub.com/home/jpa
Sage

Development of a Cognitive Assessment Checklist for First-Grade Mathematics: Utilizing Hierarchical Cognitive Diagnostic Modeling in Elementary Education

Marzieh Haghayeghi[1], Ali Moghadamzadeh[1], Hamdollah Ravand[2], Mohamad Javadipour[1], and Hossein Kareshki[3]

Abstract

This study aimed to address the need for a comprehensive assessment tool to evaluate the mathematical abilities of first-grade students through cognitive diagnostic assessment (CDA). The primary challenge involved in this endeavor was to delineate the specific cognitive skills and sub-skills pertinent to first-grade mathematics (FG-M) and to determine the most suitable model for assessment. Employing a mixed-methods approach, the research identified nine cognitive attributes essential for FG-M and developed a Q-matrix delineating the relationship between these attributes. A preliminary version of the FG-M checklist comprising 74 items was formulated. Subsequently, the assessment tool was administered to 1018 first-grade students, with their teachers utilizing the FG-M checklist for evaluation. Results indicated a commendable accuracy in classifying the nine cognitive attributes, ranging from 0.68 to 0.92, with an average accuracy of 0.76. Furthermore, the validity of the scores was corroborated by a substantial correlation (r = 0.78) between scores from the FG-M checklist and the arithmetic subtest of the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). Overall, the study concluded that the scores from the FG-M checklist are reliable for assessing the mathematical skills of first-grade students. The findings underscored the students' proficiency in fundamental mathematical concepts such as cardinality, addition, subtraction, weight, and statistics. However, areas of weakness were identified in concepts related to multiplication, time, symmetry, and geometry. This research contributes significantly to the advancement of CDA in educational contexts,

[1]University of Tehran, Tehran, Iran
[2]Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran
[3]Ferdowsi University of Mashhad, Mashhad, Iran

Corresponding Author:
Ali Moghadamzadeh, Department of Curriculum Development & Instruction Methods, Faculty of Psychology and Education, University of Tehran, Tehran 9178166342, Iran.
Email: amoghadamzadeh@ut.ac.ir

providing educators with a valuable tool for identifying students' strengths and weaknesses in mathematical cognition.

Keywords
cognitive diagnostic assessment, first-grade mathematics, mathematical knowledge and skills, elementary education

Introduction

Early assessment of mathematical skills plays a critical role for educators and caregivers in understanding young learners' grasp of fundamental mathematical concepts. Identifying any potential gaps in learning at this early stage is crucial for creating targeted interventions that can prevent long-term academic challenges. Early assessments provide valuable insights into a child's cognitive development, spatial reasoning, and problem-solving abilities, allowing educators to design personalized educational plans that cater to individual learning needs (Sjoe et al., 2019; Verbruggen et al., 2021). Recognizing the importance of early mathematics education in shaping a child's cognitive development and academic journey, the first grade of primary school marks a pivotal period for laying the foundation of essential mathematical skills (Rittle-Johnson et al., 2019). Assessing young learners' mathematical abilities during this crucial phase helps identify strengths and weaknesses, empowering educators to adapt instructional strategies and interventions accordingly.

Certainly, several approaches are available for evaluating a child's abilities and achievements. Some of these techniques involve the utilization of observational tools, comprehensive narrative assessments, and direct evaluations (Li et al., 2020). To assess children's mathematical learning and development within the classroom setting, observational methodologies, such as rating scales and checklists, along with narrative assessments, including storytelling and portfolios, have been implemented (Rajkumar & Hema, 2019; Öztürk et al., 2020). For example, Charlesworth and Leali (2012) adopted a combination of observation, informal dialogues, and interviews to document the problem-solving approaches of preschool, kindergarten, and primary school children during various educational activities (Charlesworth & Leali, 2012). Similarly, Reikerås et al., 2012 conducted observations on a significant sample of toddlers engaged in problem-solving play and routine activities that required the application of mathematical language and logical reasoning (Reikerås et al., 2012). Presently, available assessment tools are grounded in either classical test theory (CTT) or item response theory (IRT). These evaluations typically yield an aggregate test score that reflects a child's ability or achievement and is often utilized to compare a child with others or against specific benchmarks (de la Torre et al., 2018). Consequently, children are classified into distinct groups based on whether they pass or fail, with the passing threshold (e.g., a score of 70) determining those who meet the criteria for success (Ravand & Robitzsch, 2015). However, it is worth noting that children with the same test scores may possess diverse knowledge structures. Therefore, relying solely on overall test scores might not be the most effective approach for mathematics teaching and learning (Ma et al., 2020). It is imperative that children's internal knowledge systems serve as the cornerstone for the instruction and acquisition of mathematical concepts.

Recent research highlights the significance of CDA in understanding students' distinct cognitive strengths and weaknesses by examining their individual knowledge structures and cognitive processing abilities (Chin & Chew, 2023; Ravand & Baghaei, 2020). CDA deconstructs test tasks into the strategies, processes, and knowledge necessary for successful task performance, thus aiding teachers in rectifying students' ineffective strategies (Embretson, 1984). Contrary to conventional educational psychometric models like IRT, which rely on the investigator's presumptions about the cognitive processes employed by test-takers in problem-solving during test situations, cognitive diagnostic models (CDMs) are grounded in empirical evidence of the actual processes and strategies utilized by test-takers (Ravand, 2015). CDMs evaluate test-takers competency across a spectrum of multiple discrete/dichotomous skills, predicting the likelihood of an observable categorical response from unobservable (i.e., latent) categorical variables. These discrete latent variables are variably referred to as skill, subskill, attribute, knowledge, ability, processes, and strategies. The initial step in the CDA process involves the construction of a cognitive model, which delineates the specific knowledge and skills essential for students to proficiently solve test problems. Subsequently, a cognitive diagnostic test is developed based on this cognitive model (Ravand & Baghaei, 2020). A cognitive diagnostic test yields patterns of attribute mastery and probabilities of attribute mastery. Patterns of mastery reveal which specific attributes students have acquired, while mastery probabilities indicate the extent of their proficiency (Leighton & Gierl, 2007). Insights garnered from the patterns and probabilities of mastery provide additional context for understanding students' academic achievements. This approach has enabled the integration of learning theories, cognition, and pedagogy with measurement theories, not only for assessing student learning but also for facilitating it (Leighton & Gierl, 2007; Ravand & Baghaei, 2020). In essence, the application of CDA in the assessment of FG-M in primary schools helps educators gain a deeper understanding of each student's cognitive strengths and weaknesses. This understanding, in turn, facilitates the implementation of targeted interventions and personalized learning plans, ultimately leading to improved academic outcomes and a more engaging and effective learning experience for the students (Ravand & Robitzsch, 2018). The current study, centered on evaluating the foundational mathematical skills of first-grade elementary students, employs a cognitive diagnostic approach to identify both their strengths and weaknesses in mathematical learning.

This study introduces a novel assessment tool, the FG-M checklist. The FG-M checklist is distinct from typical math tests, which primarily measure students' ability to solve specific problems under timed conditions. Instead, the FG-M checklist evaluates a broader range of mathematical skills and competencies. Unlike conventional tests that often focus on problem-solving speed and accuracy in a controlled setting, the FG-M checklist is designed to be used as an observational tool within the natural classroom environment. This allows for a more comprehensive assessment of a student's abilities. By focusing on these diverse aspects of mathematical understanding, the FG-M checklist provides a detailed profile of each student's strengths and areas for improvement. This holistic approach ensures that educators can identify not only whether a student can solve a particular problem but also understand how they approach mathematical concepts and processes. This nuanced insight is crucial for developing personalized educational strategies that address each student's unique needs.

Furthermore, the FG-M checklist is grounded in cognitive diagnostic modeling (CDM), a cognitive-psychometric modeling approach that provides valid classifications of human performance in specific domains, such as ability, trait, or competency (Rupp & Templin, 2008). By employing the CDM, the FG-M generates mastery classifications or sub-scores that are highly informative for diagnostic purposes. These classifications enable teachers to pinpoint specific areas where students excel or struggle, facilitating targeted instructional interventions. The FG-M checklist allows educators to recognize such patterns and tailor their teaching methods accordingly. By providing teachers with an assessment tool based on diagnostic classifications, this research has the potential to significantly enhance instructional strategies, consequently enriching students' mathematical learning experiences. The FG-M checklist not only aids in the early detection of learning gaps but also supports the development of individualized learning plans that promote long-term academic success.

CDMs

CDMs are used to assess learner abilities by pinpointing core operations and creating questions that require these skills. This allows researchers to identify which operations students can successfully employ (DiBello et al., 2006; Lee & Sawaki, 2009). CDMs provide detailed diagnostic insights, helping educators tailor instructional approaches and enhance learning environments (Huff & Goodman, 2007). These models combine cognitive psychology with statistical techniques to analyze the relationship between cognitive processes, test performance, and responses (Ravand & Robitzsch, 2015). CDMs classify test-takers into latent classes based on mastery and non-mastery patterns (Hagenaars & McCutcheon, 2002). They categorize individuals into 2^k latent classes, where k represents the number of attributes necessary for test performance, each attribute being binary (mastered or not). For example, with nine attributes, examinees are grouped into $2^9 = 512$ latent classes. CDMs estimate the likelihood of each person belonging to any one of the latent classes.

CDMs are categorized according to the relationships they assume among the attributes of test. CDMs can be either specific or general. Specific CDMs, in turn, are either conjunctive or disjunctive. In conjunctive models, such as the deterministic-input noisy-and-gate (DINA) model (Junker & Sijtsma, 2001), all required attributes must be mastered for a correct response. In disjunctive models, such as the deterministic-input noisy-or-gate (DINO) model (Templin & Henson, 2006), proficiency in any one attribute can lead to a correct response. General models such as the generalized DINA (G-DINA) model allow both conjunctive and disjunctive relationships within the same test. CDMs provide detailed insights into students' strengths and weaknesses in specific cognitive skills. Unlike IRT models, which assign a single continuous score, CDMs categorize individuals into multidimensional skill profiles, identifying them as masters or non-masters of each skill. This results in a complex loading structure known as within-item multidimensionality (Ma et al., 2020; Ravand & Baghaei, 2020; Ravand & Robitzsch, 2015, 2018).

A key element in CDMs is a Q-matrix, which clarifies the relationship between test items and cognitive attributes in a cognitive diagnostic test (Tatsuoka, 1983). Each row in the Q-matrix represents a test item, and each column represents a cognitive attribute. Entries are binary, with "1" indicating that an attribute is necessary for correctly answering the item and "0" indicating that it is not (Li & Suen, 2013). This binary framework ensures a comprehensive evaluation of the cognitive attributes involved. It is recommended that each attribute is assessed by at least three items to ensure reliability (de la Torre & Chiu, 2016). See Table 1 for the Q-matrix.

Method

Determination of the FG-M Attribute

Identifying cognitive attributes is essential in CDMs. Rupp and Templin (2008) argue that misspecifications of Q-matrix would result in misclassification of examinees. Accurate test-taker classification, the core goal of CDMs, requires meticulous attention to all factors impacting classification accuracy. Identifying FG-M attributes followed a two-phase approach: theoretical framework analysis and expert discussions, followed by structured interviews. This process identified nine key cognitive attributes for FG-M, validated by a panel of five mathematics

Table 1. The Q-Matrix of the FG-M Checklist.

	Attributes									
Items	V1	V2	V3	V4	V5	V6	V7	V8	V9	Number
1	1	0	0	0	0	0	0	0	0	1
2	1	0	0	0	0	0	1	1	0	3
3	1	1	0	0	0	0	1	1	1	5
4	1	0	0	0	1	0	0	0	0	2
5	1	0	0	0	0	0	0	1	1	3
6	1	1	0	0	0	0	1	0	0	3
7	1	0	0	0	0	0	0	0	0	1
8	1	0	0	0	0	0	0	0	0	1
9	1	1	0	0	0	0	0	0	0	2
10	1	0	0	0	0	0	1	0	0	2
11	1	0	0	0	0	1	0	0	0	2
12	1	1	0	0	0	0	1	1	0	4
13	1	0	0	0	1	0	0	1	0	3
14	1	0	0	0	1	0	1	0	0	3
15	1	0	0	0	0	0	0	0	0	1
16	1	1	0	0	0	0	0	1	0	3
17	1	1	0	0	0	0	1	0	0	3
18	1	1	1	0	0	0	1	0	0	4
19	1	1	0	0	0	0	1	1	1	5
20	1	1	0	0	0	0	1	0	0	3
21	1	1	0	0	0	0	1	1	0	4
22	1	1	1	0	0	0	1	1	0	5
23	1	1	0	1	0	0	1	0	0	4
24	1	0	1	0	0	0	0	0	0	2
25	1	1	1	0	0	0	0	1	0	4
26	1	0	1	0	0	0	1	1	1	5
27	1	1	1	0	0	0	0	0	0	3
28	1	0	1	0	0	0	1	0	0	3
29	1	0	1	0	0	0	0	1	1	4
30	1	1	1	1	0	0	1	0	0	5
31	1	0	1	0	0	0	1	1	0	4
32	1	1	1	0	1	0	1	0	0	5
33	1	1	1	0	0	0	1	0	0	4
34	1	0	1	0	0	0	1	0	0	3
35	1	0	1	0	0	0	1	1	1	5
36	1	1	1	1	0	0	0	1	1	6
37	1	0	1	0	0	1	1	0	0	4
38	1	1	1	1	0	0	0	0	0	4
39	1	1	0	1	0	0	1	1	0	5
40	1	1	0	1	0	0	0	0	0	3
41	1	1	0	1	0	0	0	1	1	5
42	1	1	0	1	0	0	0	1	1	5
43	1	0	0	0	1	0	0	1	1	4
44	1	1	0	1	1	0	0	0	0	4

(continued)

Table 1. (continued)

	Attributes									
Items	V1	V2	V3	V4	V5	V6	V7	V8	V9	Number
45	1	1	0	1	1	0	1	0	0	5
46	1	1	1	1	1	0	0	0	0	5
47	1	0	1	0	1	0	0	0	0	3
48	1	0	1	0	1	0	1	0	0	4
49	1	1	0	1	1	0	0	1	0	5
50	1	0	0	0	0	1	0	0	0	2
51	1	0	0	0	0	1	1	0	0	3
52	1	0	1	0	0	1	0	0	0	3
53	1	1	0	0	0	1	0	0	0	3
54	1	1	1	0	0	1	0	0	0	4
55	1	0	0	0	0	0	1	1	1	4
56	1	0	0	0	0	0	1	0	0	2
57	1	1	0	0	0	0	1	0	0	3
58	1	0	0	0	1	1	1	0	0	4
59	1	0	0	0	0	0	1	1	0	3
60	1	0	0	0	0	0	0	1	0	2
61	1	1	1	0	0	0	0	1	1	5
62	1	0	0	0	0	0	0	1	0	2
63	1	0	1	0	0	0	0	1	0	3
64	1	0	0	0	0	0	0	1	1	3
65	1	1	0	0	0	0	0	1	1	4
66	1	0	0	0	0	0	0	1	1	3
67	1	0	0	0	0	0	1	1	1	4
68	1	0	0	0	0	0	1	1	1	4
69	1	1	0	0	0	0	0	1	1	4
70	1	1	0	0	0	0	0	1	1	4
71	1	1	0	0	0	0	1	1	1	5
72	1	1	1	0	0	0	0	1	1	5
73	1	0	0	0	0	0	0	1	1	3
74	1	1	1	0	0	0	0	1	1	5
Frequency	74	37	26	12	12	8	34	37	23	

Notes. Number = the number of attributes measured by an item. Frequency = frequency of cognitive attribute being measured.

education experts using a 5-point scale. Each attribute received high ratings, confirming their significance. For an overview of these attributes, refer to Table 2.

Our theoretical framework for identifying first-grade mathematical concepts is grounded in cognitive development theory, educational standards, prior research, and expert consultations. We based our framework on Piaget's preoperational stage, where first-graders understand symbols and simple logical operations. We referenced the Common Core State Standards (CCSS) (Kamii, 2015) and National Council of Teachers of Mathematics (NCTM) (Clarke et al., 2014) recommendations, which outline key mathematical skills such as basic arithmetic and geometry. Additionally, we reviewed literature from Clements and Sarama (2009) and Ginsburg et al. (2008), highlighting essential areas of early mathematics education like number sense and counting

Table 2. The Cognitive Attributes of the FG-M.

Cognitive attributes	Codes	Description
Cardinality concept	**V1**	Cardinality refers to the understanding of the quantity or number of objects in a set. It's the ability to know how many items are present in a group without needing to count them one by one. Developing cardinality involves grasping that the last number counted in a sequence represents the total amount in the set. For instance, if a child can see a group of four apples and says "four" without counting each apple, they are demonstrating an understanding of cardinality.
Addition concept	**V2**	In the context of first-grade elementary school math, addition is the mathematical operation of combining two or more quantities to find a total or sum. It involves bringing together separate groups or numbers to calculate their combined value. In simple terms, it's understanding how to add numbers to find out "how many in total." For example, adding 3 and 2 means combining three items with two items to get a total of five items. First-graders typically start with basic addition problems using small numbers to build a foundational understanding of this concept
Subtraction concept	**V3**	In first-grade elementary school math, subtraction is a fundamental mathematical operation. It involves taking away or removing a certain quantity from a larger group or number. It's the process of finding the difference between two quantities. Subtraction helps students understand concepts like "how many are left" after some items have been removed. For instance, if there are 7 candies and 3 are taken away, subtraction helps determine that 4 candies are remaining. In the context of early math education, students begin with simple subtraction problems using small numbers to build their foundational understanding of this concept.
Multiplication concept	**V4**	By the conclusion of the first grade, students are expected to develop the ability to count in regular intervals or increments, such as counting by 10s (10, 20, 30) or by 5s (5, 10, 15), and even by smaller numbers such as 3s (3, 6, 9). This skill is important because it introduces the foundational idea of repeated addition, where students begin to see that numbers can be grouped and added in a structured way. Mastering this ability helps students transition smoothly into understanding multiplication as combining equal groups, which is a critical mathematical skill as they progress in their education.
Hour concept	**V5**	In the first grade of elementary school math, the concept of "hour" is introduced. An hour is a unit of time that helps us measure and understand the passing of time. Students typically learn how there are 24 hours in a day, and the clock's hour hand moves around the clock face twice a day. They begin to identify and read the numbers on the clock to tell the time in terms of hours. While they might not delve into more complex time concepts, this foundational understanding of hours sets the stage for further time-related learning in later grades.

(continued)

Table 2. (continued)

Weight concept	**V6**	In the first grade of elementary school math, the concept of “weight” is introduced as a measure of how heavy or light an object is. Students learn to compare the weight of objects using terms like “heavier” and “lighter.” They might use balance scales or their hands to make these comparisons. While the precise measurements might not be emphasized at this stage, students begin to develop a basic understanding of weight and how it’s used to differentiate between the heaviness of various objects. This initial understanding forms the basis for more detailed weight-related learning in later grades.
Statistics concept	**V7**	In the first grade of elementary school math, the concept of “statistics” is introduced in a basic manner. Students begin to gather and organize simple data, such as counting the number of students with a certain attribute (like hair color or favorite color) in their class. They learn to create simple pictographs or bar graphs to visually represent these data. While the idea of averages and more complex statistical concepts is not typically covered at this level, this introductory exposure to statistics helps students understand how information can be collected and presented visually.
Symmetry concept	**V8**	In the first grade of elementary school math, the concept of “symmetry” is introduced in a basic manner. Students learn about symmetry in shapes and objects. They explore how a shape can be divided into two equal halves that mirror each other. They might practice drawing lines of symmetry or identifying symmetrical objects in their environment, like butterflies or simple geometric shapes. While more complex symmetrical patterns might not be covered, this initial introduction helps students develop an awareness of balance and symmetry in everyday objects and shapes.
Geometry concept	**V9**	In the first grade of elementary school math, the concept of “geometry” is introduced with a focus on basic shapes and their properties. Students learn to identify and name common shapes like circles, squares, triangles, and rectangles. They explore characteristics of these shapes, such as the number of sides and corners. Additionally, they might learn about concepts like “above,” “below,” “beside,” and “inside” as they relate to the spatial relationships between objects. While more complex geometry topics aren’t typically covered, this foundational understanding of shapes and spatial concepts lays the groundwork for further geometry learning in later grades.

principles (Ginsburg et al., 2008; Sarama & Clements, 2009). Expert interviews, involving doctoral candidates in mathematics education and experienced first-grade teachers, validated the attributes. These experts assessed the significance and interconnectedness of the attributes, ensuring their relevance. This two-phase approach ensures our framework is theoretically sound and the scores are empirically validated, capturing critical mathematical concepts for first-graders.

After pinpointing the cognitive attributes in FG-M, we delved into establishing a hierarchy among these attributes. To do so, we enlisted the insights of ten experts, including three Ph.D. candidates specializing in mathematics education and seven seasoned teachers who have excelled in teaching first-grade math. Their task was to map out the hierarchical relationships among these attributes. We explored all potential hierarchical models and eventually settled on a model that gained the approval of 80% of our expert panel (8 individuals). The resulting model, depicted in Figure 1, visually represents the hierarchical connections among the nine identified attributes.

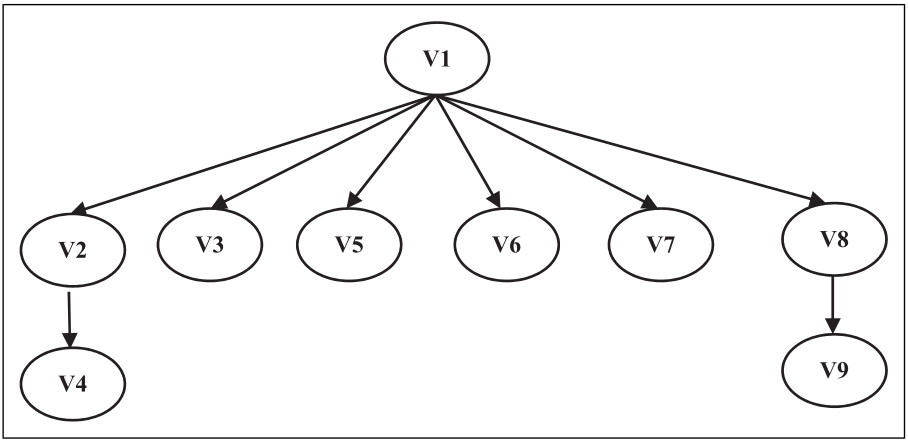

Figure 1. Hierarchical model of cognitive attributes of FG-M.

Development of a Cognitive Diagnostic Checklist for the FG-M

The Q-matrix was constructed using nine attributes specified in Table 2 (refer to Table 1). The Q-matrix of the FG-M checklist shows that Attribute V1 is the most frequently measured, appearing in 56 items, highlighting its foundational importance. Other attributes are measured less frequently, with V2 and V3 appearing 21 and 18 times, and V4, V5, and V6 only 5 to 7 times each. Attributes V7, V8, and V9 are measured in 15, 13, and 12 items, respectively. The items vary in the number of attributes they measure: 20 items measure a single attribute, 17 items assess two attributes, 13 items evaluate three attributes, and 2 items measure four attributes. This distribution ensures comprehensive coverage of all cognitive attributes while maintaining alignment with the hierarchical nature of the skills assessed. The detailed Q-matrix is provided in Table 1.

After developing the Q-matrix, test items were designed to match each pattern. This followed a top-down approach, starting with conceptualizing measurement patterns and developing corresponding items. The process involved collaboration among three doctoral students in mathematics education and seven experienced teachers. Together, they created items for the first-grade FG-M checklist, aligning them with predetermined cognitive attributes. Items were organized systematically based on the Q-matrix. The kappa coefficient was used to gauge consensus among experts on item attributes, with a threshold of 0.60 for selection. This rigorous process identified 74 checklist items for the FG-M assessment.

The Pilot Experiment of the FG-M Checklist

A group of twenty first-grade teachers were requested to assess a minimum of five students using the initial FG-M checklist and provide feedback on its items for the pilot phase. During this evaluation, it was observed that three of the items led to confusion among the teachers. Consequently, these items were revised to enhance clarity.

Testing the Quality of the Initial FG-M Checklist

Participants. The study encompassed 1018 first-grade students enrolled in elementary schools located in Mashhad, Iran. The educational system in Mashhad is divided into eight well-defined educational districts, ranging from District 1 to District 7, alongside an additional area known as

Tabadkan. To ensure comprehensive coverage, five classrooms were randomly selected from each district, resulting in the incorporation of a total of 41 classrooms in the study. A systematic sampling approach was used to ensure equitable representation across all districts. Once the specific classes were identified, teachers were requested to assess their respective students' mathematical competencies using the FG-M checklist. Detailed information about the descriptive statistics for these 1018 students is provided in Table 3.

Instruments

FG-M Checklist

The initial FG-M checklist was utilized to evaluate the mathematical proficiency of first-grade students. This checklist encompassed a total of 74 items, each with a scoring system of 0 to 1. As a result, by the conclusion of the 2022–2023 academic year, the teacher conducted an assessment of their students' mathematical skills using the FG-M checklist criteria. A score of 1 indicated that a student had exhibited full competence across all attributes measured by the item, whereas a score of 0 indicated that mastery was lacking in at least one attribute. These evaluations were conducted through individualized interviews, with an average duration of approximately 20 minutes per student. The collected data from the checklist encompassed the cumulative test score, the proficiency levels displayed across 9 cognitive attributes, and the corresponding probabilities of mastery for each of these attributes. The arrangement of mastery across attributes was depicted in a binary vector format, with a "1" denoting successful mastery by the first-grade students and a "0" indicating non-mastery. The average likelihood of mastery was computed through a model-data fitting analysis, utilizing the chosen CDM, which accurately represented the mastery status of the 9 cognitive attributes.

The Arithmetic Subtest of the WISC-IV

The Arithmetic Subtest of the WISC-IV is a key component of this cognitive assessment tool, designed for children aged 6 to 16. The WISC-IV consists of 15 scales: 10 primary scales and 5 supplementary scales. Each subscale is standardized with a mean score of 10 and a standard deviation of 3. The overall evaluation yields a general intelligence score, along with four specific indexes: verbal comprehension, conceptual reasoning, processing speed, and working memory. The full-scale IQ for the WISC-IV is 100, with a standard deviation of 15. The validity of the

Table 3. The Descriptive Statistics of the 1018 First-Graders.

Educational districts	Total *N*	Girl *N*	Boy *N*	Age M	Age SD	ANSC
Districts 1	134	65	69	88.89	5.61	26.83
Districts 2	132	69	63	86.02	4.34	26.51
Districts 3	119	59	60	81.68	3.67	24.01
Districts 4	117	61	56	80.98	2.01	23.41
Districts 5	129	63	66	89.06	2.80	25.83
Districts 6	123	65	58	79.98	2.04	24.61
Districts 7	120	58	62	84.27	4.09	24.01
Tabadkan districts	144	69	75	88.91	5.83	28.89

N: number, M: mean, SD: standard deviation, ANSC: Average Number of Students in each Class. The age mean is based on the months.
Note. Age is reported in months. All students were between 6.5 and 7.5 years of age.

WISC-IV's scores is well-established, with factor analyses consistently supporting its four-factor model. Furthermore, the WISC-IV scores demonstrate strong correlations with the scores of other intelligence tests (e.g., Kaufman Assessment Battery for Children, Second Edition (KABC-II) and Differential Ability Scales), highlighting its robust convergent validity (Baron, 2005; Watkins & Canivez, 2022). In this study, the Arithmetic Subtest of the WISC-IV was specifically used to assess the convergent validity of the scores from the FG-M checklist. This subtest, which contributes to the Working Memory domain, plays a key role in evaluating children's cognitive abilities. In the present study, the Persian version of the WISC-IV was used. This version was adapted and validated by Hassanpur and Minaei (2018) at the request of the Iranian Ministry of Education for Persian-speaking children. The adaptation process involved translating the test items from English to Persian and making necessary cultural adjustments to ensure the items were contextually appropriate for Iranian students. The translated version underwent pilot testing with a sample of Persian-speaking children to assess its reliability and validity. The results confirmed that the adapted version preserved the psychometric integrity of the original WISC-IV.

Data Analysis

In our research, selecting the appropriate model in the CDA process is critical yet complex, particularly due to the debate over whether fundamental mathematical skills at the elementary level are conjunctive or disjunctive. This study assessed the fit of three models—GDINA, DINA, and DINO—to determine the best alignment with our data, focusing on identifying the most suitable CDMs for evaluating first-grade students' mathematical abilities. Parameters were estimated using the GDINA R package (Ma et al., 2016). We utilized both absolute and relative fit indices to assess model suitability. Absolute fit indices, including the mean absolute difference for item-pair correlations (MADcor), the mean residual covariance (MADRES), standardized root mean square residual (SRMSR), and Max χ^2, measure how well a model fits the data independently, with lower values indicating a better fit. If the null hypothesis is rejected ($p < .05$), the reduced model is rejected. If multiple reduced models are retained and DINA or DINO is among them, the model with the largest p-value is selected. If neither DINA nor DINO is retained, the reduced model with the largest p-value is chosen. Notably, when several reduced CDMs have p-values larger than .05, DINA or DINO are preferred due to their statistical simplicity (Rupp & Templin, 2008). Relative fit indices, including −2 log-likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC), compare the fit of different models. Model selection criteria focused on lower AIC, BIC, and −2 log-likelihood values, better absolute fit indices, and higher p-values for Max χ^2. Following model selection, root mean square error of approximation (RMSEA) indices were used to assess item alignment with the selected model. Items with an RMSEA value above 0.1 indicate a poor fit, those with values between 0.05 and 0.1 suggest a moderate fit, and items with a value below 0.05 are considered to have a good fit with the model (Kunina-Habenicht et al., 2009).

Classification consistency is an important measure that indicates whether a person would be assigned to the same category if they were to retake the same test or its parallel form. To evaluate classification accuracy and consistency, this study used a cut-off of 0.60, based on established guidelines for CDAs to ensure a balance between reliability and practical applicability (Rupp & Templin, 2008). The rationale for selecting the 0.60 cut-off is grounded in the need to ensure that the classifications are both reliable and interpretable, reflecting a reasonable threshold for acceptable levels of accuracy and consistency in educational assessments. This threshold helps to ensure that the classifications provide meaningful insights into students' cognitive strengths and weaknesses without being overly stringent, which could exclude potentially useful items. The algorithms did not explicitly account for chance in computing classification consistency and accuracy. For convergent validity evidence, the correlation between the checklist's scores and the scores from the Arithmetic Subtest of the WISC-IV was calculated. In pursuit of another research objective, the study utilized the expected a posteriori (EAP) method to gauge the likelihood of each participant's proficiency in various skills. Individuals exceeding a mastery probability of 0.6 were classified as "mastered," while those falling below this threshold were deemed "non-mastered" (Embretson & Reise, 2013). Eventually, to illustrate the utility of CDMs, we presented a few student profiles as examples of individual-level learning profiles.

Missing data were handled using the expectation-maximization (EM) algorithm, which estimated missing values based on observed data, reducing bias and preserving dataset integrity. Sensitivity analyses compared results with and without imputed values, ensuring robustness. An expert panel reviewed the imputed data and analyses, validating the methods. This comprehensive approach ensured the accuracy and reliability of the FG-M checklist findings.

Results

Model Selection

Detailed results, including absolute fit indices for assessing model-data fit and relative fit indices for model comparisons, can be found in Tables 4 and 5, respectively.

In this study, the GDINA model demonstrates a superior fit compared to both the DINA and DINO models. Notably, the DINA conjunctive model, as indicated by the significant p-value for the Max $\chi 2$ index, exhibits a poor fit. Furthermore, the -2 log-likelihood for the GDINA model is lower than that of the DINA and DINO models. However, the values for the DINO are acceptable and closer to those of the GDINA. Theoretically, general CDMs invariably exhibit superior data fit compared to reduced CDMs due to their more intricate parameterization. However, the preference for saturated models is not unequivocal. One reason is that general CDMs necessitate larger sample sizes for precise estimation. Another reason is that reduced CDMs are more straightforward and easier to interpret, provided that their fit is not significantly inferior. Consequently, the DINO model which had a comparable fit to the GDINA is the simplest model fitting the data in the present study.

Item Fit

Table 6 displays the RMSEA indexes, which are used to assess how well the items align with the DINO model.

Referring to Tables 6, it was observed that six items, 6, 21, 30, 50, 51, and 56, had RMSEA values greater than 0.05. As a result, the research team decided to exclude these items from the FG-M checklist. All other items demonstrated a strong alignment with the DINO model. The removal of the six items, due to their poor RMSEA values, has not adversely affected the desired representation between items and attributes in the FG-M checklist. Each attribute remains sufficiently

Table 4. The Absolute Fit Indices for the FG-M.

	MADcor	MADRES	SRMSR	Max χ^2	P
GDINA	0.0056	0.0051	0.0079	4.6035	**0.99**
DINA	0.0753	0.0526	0.0931	702.085	**0.000**
DINO	0.0143	0.0262	0.0286	38.428	**0.513**

Table 5. The Relative Fit Indices.

	−2 log-likelihood	BIC	AIC
GDINA	−22439.47	43732.29	41702.95
DINA	−26068.67	55701.30	51455.34
DINO	−22980.67	51725.31	48479.34

Table 6. Item Fit Indices for FG-M Items.

ITEM	RMSEA	ITEM	RMSEA	ITEM	RMSEA	ITEM	RMSEA	ITEM	RMSEA
1	0.0423	**16**	0.0236	**31**	0.0441	**46**	0.0446	**61**	0.0254
2	0.0264	**17**	0.0422	**32**	0.0445	**47**	0.0432	**62**	0.0234
3	0.0436	**18**	0.0421	**33**	0.0420	**48**	0.0442	**63**	0.0306
4	0.0326	**19**	0.0324	**34**	0.0432	**49**	0.0382	**64**	0.0333
5	0.0342	**20**	0.0343	**35**	0.0421	**50**	0.0953	**65**	0.0126
6	0.1495	**21**	0.0644	**36**	0.0421	**51**	0.0945	**66**	0.0345
7	0.0421	**22**	0.0406	**37**	0.0424	**52**	0.0431	**67**	0.0353
8	0.0255	**23**	0.0334	**38**	0.0443	**53**	0.0413	**68**	0.0412
9	0.0255	**24**	0.0433	**39**	0.0436	**54**	0.0241	**69**	0.0424
10	0.0414	**25**	0.0431	**40**	0.0320	**55**	0.0351	**70**	0.0426
11	0.0345	**26**	0.0403	**41**	0.0443	**56**	0.0944	**71**	0.0435
12	0.0343	**27**	0.0302	**42**	0.0330	**57**	0.0441	**72**	0.0424
13	0.0343	**28**	0.0426	**43**	0.0364	**58**	0.0403	**73**	0.0431
14	0.0404	**29**	0.0341	**44**	0.0445	**59**	0.0434	**74**	0.0405
15	0.0245	**30**	0.0744	**45**	0.0413	**60**	0.0265		

Table 7. Classification Consistency and Accuracy Indices.

Cognitive attributes	Consistency	Accuracy
Cardinality concept	0.933	0.921
Addition concept	0.872	0.798
Subtraction concept	0.743	0.731
Multiplication concept	0.821	0.797
Hour concept	0.691	0.684
Weight concept	0.716	0.701
Statistics concept	0.814	0.805
Symmetry concept	0.748	0.737
Geometry concept	0.731	0.713

measured by at least three items, allowing the checklist to continue offering valuable diagnostic insights for targeted instructional interventions.

Classification Consistency and Accuracy

In this study, all skills demonstrated consistency and accuracy indices exceeding 0.60 (Table 7). This suggests that the classifications are quite reliable and dependable. Notably, the "Cardinality

Concept" feature exhibited the highest accuracy, while the "Concept of Weight" feature had the lowest accuracy among the skills assessed.

Regarding convergent validity evidence, the scores from the FG-M checklist displayed a strong correlation of .78 with the Arithmetic Subtest of the WISC-IV.

The Strengths and Weaknesses of the Participants

The classification divided participants into two distinct groups: "mastered" and "non-mastered." The study then determined the frequency and percentage of individuals in each group, along with their corresponding skill sets. These results are compiled and presented in Table 8.

According to the data presented in Table 8, the mastery levels of participants were evaluated using the FG-M checklist. The percentages of participants who demonstrated mastery across nine attributes, namely, Concepts of Cardinality, Addition, Subtraction, Multiplication, Hour, Weight, Statistics, Symmetry, and Geometry, were 72.79, 68.17, 57.66, 46.45, 18.70, 61.00, 62.57, 40.1, and 35.26, respectively. Based on the provided data, it can be inferred that the participants' notable strength lies in their grasp of the cardinality concept. Conversely, their relative weakness is evident in their understanding of the hour concept.

Individual-Level Learning Profiles

One of the key advantages of the CDM is its ability to generate detailed individual-level learning profiles. These profiles provide insights into each student's specific cognitive strengths and weaknesses, enabling targeted feedback and personalized interventions. Below are graphical representations of the learning profiles of three students, showcasing their mastery of various cognitive attributes. Each profile is depicted in a radar chart, where each axis represents a different cognitive attribute, and the value indicates the level of mastery (ranging from 0 to 1).

Referring to Figure 2, Student A demonstrates high mastery in most cognitive attributes, particularly in the Cardinality and Addition Concepts. However, there is room for improvement in the Hour Concept and Weight Concept. This profile suggests that while Student A excels in numerical operations, targeted interventions could be beneficial in enhancing their understanding of time and measurement concepts. Student B shows moderate mastery across various attributes, with notable strengths in the Hour Concept and Multiplication Concept. The lower scores in Addition and Subtraction Concepts indicate a need for focused practice and instruction in basic

Table 8. The Frequency and Percentage of the "Master" and "Non-Master" Persons in Each Skill.

	Master		Non-master	
Skills	Frequency	Percentage	Frequency	Percentage
Cardinality concept	741	72.79	277	27.21
Addition concept	694	68.17	324	31.83
Subtraction concept	587	57.66	431	42.34
Multiplication concept	473	46.45	545	53.54
Hour concept	394	18.70	624	61.29
Weight concept	621	61.00	397	39.00
Statistics concept	637	62.57	381	37.43
Symmetry concept	409	40.17	609	59.82
Geometry concept	359	35.26	659	64.73

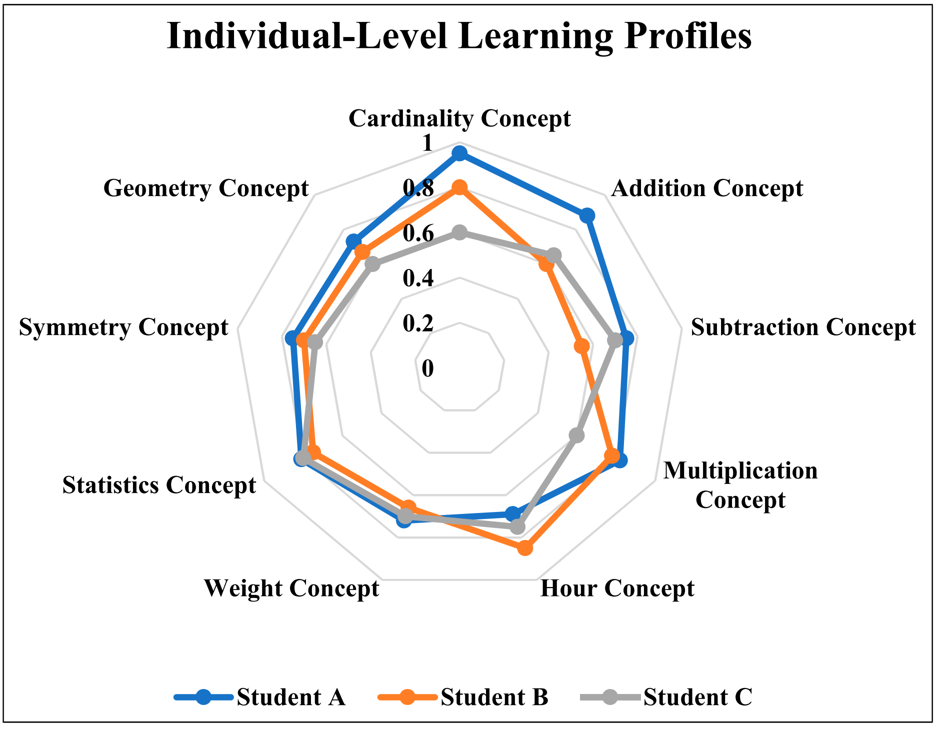

Figure 2. Learning profiles of three example students.

arithmetic operations. By identifying these specific areas, educators can tailor their teaching strategies to address Student B's weaknesses more effectively. Student C's profile reveals a balanced but lower overall mastery across the cognitive attributes compared to Students A and B. The highest mastery is observed in the Statistics Concept, whereas other areas such as Multiplication and Geometry Concepts are relatively weak. This balanced yet lower mastery profile suggests a need for comprehensive support across multiple areas, potentially through a more individualized learning plan that addresses multiple concepts simultaneously.

Discussion

The primary objective of this study was to develop an instrument for evaluating the mathematical abilities of first-graders using the CDA approach. By leveraging CDA, the study aimed to offer more precise and comprehensive insights into the mathematical capabilities of children, providing valuable information for future studies in early childhood education.

In developing the FG-M checklist, we identified nine cognitive attributes related to the mathematical abilities of first-graders through educational content, research findings, and expert insights in early mathematics. These attributes are documented in Table 2. Establishing hierarchical relationships among these attributes to form the Q-matrix was a crucial step.

In this study, the GDINA model fits the data better than both the DINA and DINO models. However, the DINO model's fit is acceptable and closely aligns with the GDINA. General CDMs typically offer better data fit due to their complex parameterization, but they require larger sample sizes for accurate estimation and are more difficult to interpret. Reduced CDMs, like the DINO, are easier to interpret if their fit is close to that of the more complex models. Therefore, in this study, the DINO model, which fits comparably to the GDINA, is the simplest model that adequately fits the data. The fitting of first-graders' mathematics abilities with the DINO model highlights the disjunctive nature of mathematical skill acquisition at this early stage of education. The DINO model assumes that mastering a single requisite skill can compensate for the lack of others in successfully solving a problem. This aligns well with the developmental characteristics of first-graders, who often demonstrate uneven mastery across different mathematical domains. Previous studies have shown that young children frequently leverage various intuitive strategies and partial knowledge to arrive at correct answers (Carpenter et al., 1999; Hoffman & Grialou, 2005), which supports the appropriateness of the DINO model for this age group. For instance, research has documented similar findings, showing that elementary students often succeed in problem-solving using partial skill sets rather than complete mastery of all required skills (Hiebert & Grouws, 2007; Siegler & Opfer, 2003). The congruence of our findings with these studies suggests that the DINO model effectively captures the cognitive processes underpinning FG-M, thereby providing a robust framework for diagnosing and supporting early mathematical development.

The RMSEA indexes were used to evaluate item conformity with the model. Six items were removed due to poor RMSEA values exceeding 0.05 to improve the overall fit and ensure the reliability and validity of the scores produced by the assessment tool. Despite this reduction, the integrity of the FG-M checklist was maintained, with each attribute measured by at least three items. This approach enhances the precision and effectiveness of the FG-M checklist, ensuring that only items with strong psychometric properties are included. Classification consistency and accuracy were thoroughly examined, confirming the reliability of all identified skills. The findings highlighted participants excelled in cardinality, addition, subtraction, weight, and statistics concepts, while showing weaknesses in multiplication, hour, symmetry, and geometry concepts. Consequently, it is imperative for the FG-M curriculum to capitalize on students' strengths while addressing and reinforcing weaker areas.

The individual learning profiles generated by the CDM offer several advantages. They enable educators to provide targeted feedback based on each student's unique strengths and weaknesses, rather than relying on generalized test scores. This allows for the design of personalized interventions tailored to specific needs, leading to more effective learning experiences. The CDM also enhances instructional planning by identifying common areas of difficulty, guiding resource allocation and teaching strategies. Additionally, these profiles facilitate monitoring of student progress, providing clear insights into learning trajectories and intervention effectiveness.

This study's primary contribution lies in its detailed breakdown of specific knowledge, skills, and abilities inherent in the mathematical proficiency of first-graders. Unlike existing assessment tools, such as the Early Numeracy Test-Revised (ENT-R) (Wright et al., 2006), Woodcock-Johnson IV (Woodcock et al., 2014), and WISC-IV (Wechsler, 2003), which often lack clarity in pinpointing specific dimensions of mathematics being assessed, the FG-M checklist provides a comprehensive evaluation of each student's mathematical abilities. This two-fold reporting strategy offers a fundamental building block for children's future advancement in mathematics.

While the scores produced by the FG-M checklist demonstrated strong reliability and validity, there are limitations to this study. The use of a cognitive checklist rather than a comprehensive cognitive test may limit the depth of assessment. Although the checklist covers the entire FG-M curriculum, using a cognitive test to cover all educational content is impractical for young children due to potential fatigue. Therefore, it is advisable to assess the reliability of the checklist through the integration of alternative measurement methods.

In summary, the FG-M checklist is a robust tool for assessing first-graders' mathematical skills, offering detailed insights that facilitate targeted feedback and personalized interventions. Future research should focus on further validating the scores generated by this instrument and exploring additional methods to enhance their reliability and applicability in various educational settings.

Authors' Contributions

First author (M. H.) contributed to the conception and design of the study, as well as the acquisition, analysis and interpretation of data, drafting the article, and writing the entire manuscript. Second author (A. M.) contributed to developing the study design and to final approval of the version to be submitted. Third author (H. R.) contributed to developing the study design and to supervising the research project. Forth author (M. J.) contributed to the conception and design of the study. Fifth author (H. K.) contributed to advising the research project.

Declaration of Conflicting Interests

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements) or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This manuscript's version has been approved by all authors.

Ethical Statement

Ethics Approval

Because the study's instruments were questionnaires and the participants were drawn from community-dwelling adults (teachers), no ethical approval was required.

Consent to Participate

The participants were informed that they could withdraw from the study at any time and without explanation.

Consent for Publication

We hereby declare that we participate in the study and the development of the manuscript titled "The Mediating role of Teacher Efficacy in the Association between Teacher Self-Concept and Burnout: A Moderated Mediation Approach." We have read the final version and give our consent for the article to be published in the journal of public health. Thank you for your support.

ORCID iD

Ali Moghadamzadeh https://orcid.org/0000-0002-3960-5037

Data Availability Statement

The authors will make this study data available to any qualified researcher, without undue reservation.

References

Baron, I. S. (2005). Test review: Wechsler intelligence scale for children-fourth edition (WISC-IV). *Child Neuropsychology*, *11*(5), 471–475. https://doi.org/10.1080/09297040590951587

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Heinemann Portsmouth, NH.

Charlesworth, R., & Leali, S. A. (2012). Using problem solving to assess young children's mathematics knowledge. *Early Childhood Education Journal*, *39*(6), 373–382. https://doi.org/10.1007/s10643-011-0480-y

Chin, H., & Chew, C. M. (2023). Cognitive diagnostic assessment with ordered multiple-choice items for word problems involving 'Time. *Current Psychology*, *42*(20), 17042–17061. https://doi.org/10.1007/s12144-022-02965-8

Clarke, B., Haymond, K., & Gersten, R. (2014). Mathematics screening measures for the primary grades. In *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 199–221). American Psychological Association. https://doi.org/10.1037/14316-008

de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, *81*(2), 253–273. https://doi.org/10.1007/s11336-015-9467-8

de la Torre, J., van der Ark, L. A., & Rossi, G. (2018). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296. https://doi.org/10.1080/07481756.2017.1327286

DiBello, L. V., Roussos, L. A., & Stout, W. (2006). 31A review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (pp. 979–1030). Elsevier. https://doi.org/10.1016/S0169-7161(06)26031-0

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Embretson Whitely, S. (1984). A general latent trait model for response processes. *Psychometrika*, *49*(2), 175–186. https://doi.org/10.1007/BF02294171

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report: Society for Research in Child Development*, *22*(1), 1–24. https://doi.org/10.1002/j.2379-3988.2008.tb00054.x

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge University Press.

Hassanpur, M., & Minaei, A. (2018). *Examination of the factor structure of the fourth edition of the Wechsler Intelligence Scale for Children based on the Cattell-Horn-Carroll (CHC) theory. Paper presented at the National Conference on New Research in Iran and the World in Psychology*. Iran: Educational Sciences, and Social Studies.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning*, *1*(1), 371–404.

Hoffman, H., & Grialou, T. (2005). Test of early mathematics ability (3rd ed.) by Ginsburg, H. P., & baroody, A. J. In H. P. Ginsburg & A. J. Baroody (Eds.), *Assessment for effective intervention*, (3rd ed.), pp. 57–60). Austin, TX: PRO-ED. https://doi.org/10.1177/073724770503000409

Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. https://doi.org/10.1017/CBO9780511611186.002.1960

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258–272. https://doi.org/10.1177/01466210122032064

Kamii, C. (2015). Selected standards from the common core state standards for mathematics, grades K-3: My reasons for not supporting them. *Defending the Early Years*, *17*(1), 21–27.

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies In Educational Evaluation*, *35*(2-3), 64–70. https://doi.org/10.1016/j.stueduc.2009.10.003

Lee, Y.-W., & Sawaki, Y. (2009). Cognitive diagnosis and Q-matrices in language assessment. *Language Assessment Quarterly*, *6*(3), 169–171. https://doi.org/10.1080/15434300903059598

Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment*, *18*(1), 1–25. https://doi.org/10.1080/10627197.2013.761522

Li, L., Zhou, X., Huang, J., Tu, D., Gao, X., Yang, Z., & Li, M. (2020). Assessing kindergarteners' mathematics problem solving: The development of a cognitive diagnostic test. *Studies In Educational Evaluation*, *66*(1), 100879. https://doi.org/10.1016/j.stueduc.2020.100879

Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, *40*(3), 200–217. https://doi.org/10.1177/0146621615621717

Ma, W., Minchen, N., & de la Torre, J. (2020). Choosing between CDM and unidimensional IRT: The proportional reasoning test case. *Measurement: Interdisciplinary Research and Perspectives*, *18*(2), 87–96. https://doi.org/10.1080/15366367.2019.1697122

Öztürk, M., Akkan, Y., & Kaplan, A. (2020). Reading comprehension, Mathematics self-efficacy perception, and Mathematics attitude as correlates of students' non-routine Mathematics problem-solving skills in Turkey. *International Journal of Mathematical Education in Science & Technology*, *51*(7), 1042–1058. https://doi.org/10.1080/0020739X.2019.1648893

Rajkumar, R., & Hema, G. (2019). Factors affecting mathematical problem solving competence of undergraduate students in facing competitive examinations. *Impact: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL)*, *7*(2), 319–328.

Ravand, H. (2015). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *34*(8), 782–799. https://doi.org/10.1177/0734282915623053

Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing*, *20*(1), 24–56. https://doi.org/10.1080/15305058.2019.1588278

Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research and Evaluation*, *20*(1), 11. https://doi.org/10.7275/5g6f-ak15

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology*, *38*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1489524

Reikerås, E., Løge, I. K., & Knivsberg, A. M. (2012). The mathematical competencies of toddlers expressed in their play and daily life activities in Norwegian kindergartens. *International Journal of Early Childhood*, *44*(1), 91–114. https://doi.org/10.1007/s13158-011-0050-x

Rittle-Johnson, B., Zippert, E. L., & Boice, K. L. (2019). The roles of patterning and spatial skills in early mathematics development. *Early Childhood Research Quarterly*, *46*(1), 166–178. https://doi.org/10.1016/j.ecresq.2018.03.006

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, *6*(4), 219–262. https://doi.org/10.1080/15366360802490866

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243. https://doi.org/10.1111/1467-9280.02438

Sjoe, N. M., Bleses, D., Dybdal, L., Tideman, E., Kirkeby, H., Sehested, K. K., Nielsen, H., Kreiner, S., & Jensen, P. (2019). Short Danish version of the tools for early assessment in math (TEAM) for 3–6-year-olds. *Early Education & Development*, *30*(2), 238–258. https://doi.org/10.1080/10409289.2018.1544810

Tatsuoka, K. K. (1983). Rule space: an approach for dealing with misconception based on itemresponse theory. *Journal of Education Measurement*, *20*(1), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Verbruggen, S., Depaepe, F., & Torbeyns, J. (2021). Effectiveness of educational technology in early mathematics education: A systematic literature review. *International Journal of Child-Computer Interaction*, *27*(2), 100220. https://doi.org/10.1016/j.ijcci.2020.100220

Watkins, M. W., & Canivez, G. L. (2022). Assessing the psychometric utility of IQ scores: A tutorial using the wechsler intelligence scale for children–fifth edition. *School Psychology Review*, *51*(5), 619–633. https://doi.org/10.1080/2372966X.2020.1816804

Wechsler, D. (2003). *Wechsler intelligence scale for children – Fourth edition (WISC-IV)*. Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Riverside Insights.

Wright, R. J., Martland, J., & Stafford, A. K. (2006). *Early Numeracy. Assessment for teaching*. & interventionSAGE Publications.