Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identifying Early Mathematics Learning Difficulty Risk: A Cross-Sectional Evaluation of Rating- and Performance-Based Tiered Screening

Muhammet Yasin Yassıkaya

muhammetyasin.yassikaya@erdogan.edu.tr

Recep Tayyip Erdoğan University

İbrahim Halil Diken

Anadolu University

Sinan Olkun

Ankara University

Veysel Aksoy

Anadolu University

Mehmet Okur

Kafkas University

Research Article

Keywords: mathematics learning difficulties, screening, early childhood

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10041509/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Early identification of mathematics learning difficulties (MLD) during the preschool years requires screening approaches that are both practical for large-scale implementation and capable of capturing children's multidimensional learning profiles. This study examined the screening validity and discriminative utility of rating-based and performance-based assessment measures for identifying early risk for MLD and broader learning difficulties (LD). We further evaluated an evidence-informed tiered screening framework integrating behavioral, cognitive, and academic indicators.

Using a cross-sectional correlational design, we examined associations among the Learning Difficulty Early Signs Screening Scale (LDESSS), cognitive indicators derived from the Anadolu-Sak Intelligence Scale (ASİS), and early mathematics competence measured by the Test of Early Mathematics Ability–Third Edition (TEMA-3). The sample included 111 preschool children and their parents (mean age = 65.9 months).

LDESSS total scores were significantly associated with TEMA-3 performance ($r = -.380$) and explained additional variance in mathematics competence beyond preschool attendance duration ($\Delta R^2 = .142$, $p < .001$). Screening analyses indicated that the LDESSS showed high sensitivity for detecting both MLD and LD risk (.90 and .91, respectively), although its discriminative accuracy was modest (AUC = .658 and .685). In contrast, performance-based cognitive indicators demonstrated stronger discriminative accuracy for identifying MLD risk, particularly the Memory Capacity Index (MCI; AUC = .855) and the VSM–WoM composite (AUC = .838).

Overall, these findings highlight the complementary roles of rating-based and performance-based assessment approaches. Integrating behavioral observations with cognitive and academic assessments within a tiered screening framework may improve early identification processes and support more targeted preventive interventions for children at risk of learning difficulties.

1. Introduction

Researchers widely regard early mathematical competence as one of the strongest predictors of later academic development. During the preschool years, children acquire foundational skills such as number knowledge, quantity understanding, comparison, pattern recognition, and basic calculation, and these early competencies are closely associated with later mathematics achievement. Conversely, weaknesses in these areas often serve as early precursors of later learning difficulties (Aunio, 2019; Duncan et al., 2007; Hannula-Sormunen et al., 2015; Nelson & Powell, 2018). For this reason, identifying risk for mathematics learning difficulties (MLD) early matters not only for preventing later academic failure but also for recognizing children's developmental needs in time to plan appropriate support (Geary, 2013; Gersten et al., 2012; Watts et al., 2014).

At the same time, early identification of MLD risk is methodologically complex. Preschool children develop rapidly, and indicators of difficulty may emerge differently across behavioral, cognitive, and academic domains (Aunio, 2019; Nelson & Powell, 2018). A child may perform one way on a standardized test and quite differently in everyday learning situations (Snow & Van Hemel, 2008). Likewise, adults may observe learning-related difficulties before stable academic deficits become clearly established (Reid et al., 2014). These features make it difficult to rely on a single assessment approach. In particular, the distinction between rating-based and performance-based assessment becomes especially important in preschool screening (Vessonen et al., 2023). Performance-based measures provide direct and structured evidence of children's cognitive and academic competencies, whereas

rating-based measures allow parents or teachers to make learning-related signs visible across natural contexts (Reid et al., 2014; Snow & Van Hemel, 2008; Vessonen et al., 2023).

1.1. Conceptualizing MLD Risk in Early Childhood

Researchers generally define MLD as persistent and marked difficulty in acquiring age- and developmentally appropriate mathematical knowledge and skills that cannot be explained solely by inadequate instruction, general intellectual disability, or clear environmental deprivation (Geary, 2013; Kaufmann & von Aster, 2012). Children with MLD typically show weaknesses in number concepts, quantitative relations, symbolic understanding, basic operations, and the use of calculation strategies (Aunio, 2019; Geary, 2013; Geary et al., 2017). Importantly MLD is heterogeneous. Some children struggle mainly with number sense and quantity processing, whereas others show difficulty because of weaknesses in working memory, executive functioning, attentional regulation, or mathematical language (Geary, 2013; Kaufmann & von Aster, 2012; Murphy et al., 2007). Recent reviews and longitudinal studies further show that mathematical difficulties often co-occur with broader domain-general cognitive vulnerabilities and that early cognitive precursors contribute meaningfully to later MLD and persistent low mathematics achievement (Agostini et al., 2022; Geary et al., 2017; Passolunghi & Lanfranchi, 2012; Peng & Fuchs, 2016; Zhang et al., 2020).

Furthermore, the developmental nexus between early language competence and subsequent learning vulnerabilities is well-documented, as early communicative or linguistic delays often serve as robust precursors to broader domain-general cognitive risks, including memory impairments and academic difficulties (Aunio, 2019; Peng & Fuchs, 2016). Because mathematical logic and memory encoding in early childhood are heavily mediated by verbal and conceptual sub-systems, language-related behavioral indicators are increasingly recognized not merely as domain-specific signs, but as vital markers of overarching learning difficulty risks (Kaufmann & von Aster, 2012; Zhang et al., 2020).

In the preschool years, however, formal diagnosis is often inappropriate because developmental variability remains high and performance patterns may not yet have stabilized (Fuchs et al., 2012; Geary, 2013). According to clinical standards, a formal diagnosis of a specific learning disorder typically requires a history of persistent difficulties during formal school-age instruction, making explicit diagnostic labeling invalid in early childhood (American Psychiatric Association [APA], 2022). For this reason, the concept of risk is far more appropriate than that of diagnosis during these formative years (Kaufmann & von Aster, 2012). A risk-oriented perspective focuses on identifying early signs of difficulty and initiating preventive support before those difficulties become structurally more established (Fuchs et al., 2012; Geary, 2013).

1.2. The Importance of Early Risk Identification

Early childhood is a critical developmental period during which the foundations of mathematical thinking emerge. During this stage, children begin to recognize numbers, compare quantities, identify patterns, and engage in basic problem solving, and these competencies form the basis of later formal mathematics learning (Aunio, 2019; Hannula-Sormunen et al., 2015). Research also points to strong continuity between early mathematics competence and later academic performance, suggesting that early weaknesses do not necessarily resolve on their own (Duncan et al., 2007; Geary, 2013; Nelson & Powell, 2018). When children do not receive adequate support, early weaknesses in mathematics may develop into more persistent difficulties during the primary school years (Fuchs et al., 2012; Gersten et al., 2012).

Accordingly, early identification serves not only to document children's current functioning but also to inform prevention and intervention planning. In other words, screening in early childhood serves both an assessment function and an intervention-planning function.

1.3. Assessment Approaches to Early MLD Risk

Researchers generally group assessment approaches used to identify MLD risk into two broad categories: performance-based direct measures and rating-based indirect measures. Performance-based instruments assess children's academic or cognitive functioning through standardized tasks and tests and therefore provide direct evidence regarding mathematics competence and the processes that support learning. In early mathematics research, these measures remain central to both screening and diagnostic work (Geary et al., 2007; Outhwaite et al., 2024; Seethaler & Fuchs, 2010). However, direct testing also requires individual administration, trained assessors, controlled conditions, and considerable time, which may limit its practicality as a universal first-stage screener in preschool settings.

By contrast, rating-based measures rely on evaluations provided by parents or teachers who observe the child in natural contexts over time. These tools offer important practical advantages, especially in early childhood, because they allow researchers and practitioners to gather information efficiently without requiring extensive individual assessment sessions (Reid et al., 2014; Snow & Van Hemel, 2008; Vessonen et al., 2023). This advantage becomes especially valuable in preschool screening, where limited time, limited personnel, and children's attentional capacities often constrain large-scale direct assessment.

At the same time, the two approaches differ in both strengths and limitations. Although performance-based measures may appear more direct and objective, they may reflect children's momentary performance and may not fully capture behavioral patterns that emerge in daily life (Snow & Van Hemel, 2008). Rating-based measures, in turn, incorporate information from natural contexts but may also reflect informants' perceptions, expectations, and opportunities for observation (De Los Reyes & Kazdin, 2005). Studies of teacher judgments and rating scales have shown that such measures can provide meaningful information about early academic functioning, but their accuracy varies depending on child and classroom factors (Baker et al., 2015; Furnari et al., 2016; Kowalski et al., 2018). Related evidence from executive function research likewise suggests that performance-based tasks and rating measures often capture overlapping but non-identical constructs (Garon et al., 2016; Hamamcı et al., 2023). Similarly, parent ratings of preschool numeracy show meaningful but only moderate correspondence with direct mathematics assessments (Lin et al., 2021).

Taken together, this literature suggests that researchers and practitioners should not treat rating-based and performance-based measures as interchangeable. Instead, these approaches appear to provide different kinds of information about children's functioning. Rating-based tools are well suited to detecting broad observable learning-related difficulties, whereas performance-based tools may provide more precise evidence regarding cognitive vulnerabilities and mathematics competence. Recent reviews of early mathematics screeners similarly suggest that no single assessment approach is sufficient on its own and that combining tools may provide the most psychometrically and practically useful basis for early identification (Clarke et al., 2023; Outhwaite et al., 2024).

Furthermore, the integration of these two approaches is driven by critical economic and logistical necessities in educational systems. Even in developed nations, universally administering comprehensive intelligence tests or extensive direct mathematics assessments to every child is highly impractical due to substantial costs, severe

time constraints, and the shortage of clinically trained personnel required for individual testing (Seethaler & Fuchs, 2010). Conversely, rating-based measures capitalize on parents' and teachers' unique opportunities to observe children's learning behaviors across diverse, natural, and longitudinal contexts (Reid et al., 2014; Snow & Van Hemel, 2008). Utilizing high-sensitivity parent ratings as an initial filter drastically reduces unnecessary referrals for resource-intensive direct evaluations. Consequently, a multi-tiered framework is not merely a psychometric preference, but a pragmatic and economically viable solution for sustainable, large-scale preschool screening.

1.4. The Present Study

Although the literature on identifying MLD risk in early childhood has expanded considerably, much of the existing work has focused either on performance-based measures or on rating-based assessments alone. As a result, important questions remain insufficiently addressed. We still know relatively little about how these different assessment approaches relate to one another, what kind of information each provides, and how they might best be organized within an early screening process. This gap is especially noticeable in preschool contexts, where developmental variability is high and direct comparisons of the relative and complementary contributions of rating-based and performance-based tools remain limited.

We designed the present study to address this gap. By examining the relationships among a rating-based parent-report tool (LDESSS) and two performance-based measures (ASİS and TEMA-3), we investigated not only whether these tools are associated, but also how they differ in screening validity and discriminative utility. In doing so, we moved beyond simple comparison of measures and instead considered how behavioral, cognitive, and academic indicators may contribute distinct forms of evidence within a sequential early screening system.

We also sought to contribute to the literature by treating MLD risk identification not as a single-step decision process, but as part of a tiered, multi-source screening framework. Within this framework, rating-based measures function as efficient first-stage tools for broad detection of potential risk, whereas performance-based cognitive and academic measures provide greater discriminative precision at later stages when the goal is to refine and confirm risk.

Against this background, we aimed to examine the screening validity and discriminative utility of rating-based and performance-based assessment measures for identifying risk for mathematics learning difficulties (MLD) and learning difficulties (LD) in early childhood, and to propose an application-oriented tiered screening model informed by empirical findings. Specifically, we examined how a parent-report screening tool, the Learning Difficulty Early Signs Screening Scale (LDESSS), relates to performance-based cognitive indicators derived from the Anadolu-Sak Intelligence Scale (ASİS) and to early mathematics competence as measured by the Test of Early Mathematics Ability–Third Edition (TEMA-3). In addition, we investigated how accurately these rating-based and performance-based indicators distinguished children at risk for LD and MLD within a preschool sample.

Accordingly, we addressed the following research questions:

(1) What are the nature and strength of the relationships between LDESSS scores and children’s performance on ASİS and TEMA-3, and to what extent do LDESSS scores explain variance in performance on these measures?

(2) To what extent does the LDESSS accurately distinguish children at risk for learning difficulties (LD) as defined by ASİS-based indicators and children at risk for mathematics learning difficulties (MLD) as defined by TEMA-3 performance, in terms of sensitivity, specificity, predictive values, and ROC-based discrimination?

(3) To what extent do ASİS cognitive indicators, specifically the Memory Capacity Index (MCI) and the VSM–WoM composite, distinguish children at risk for mathematics learning difficulties (MLD) as defined by TEMA-3 performance, in terms of sensitivity, specificity, predictive values, and ROC-based discrimination?

2. Method

2.1. Research Design

This observational study employed a cross-sectional correlational design to examine the relationships among behavioral, cognitive, and academic indicators of early risk for mathematics learning difficulties (MLD). A multi-method assessment approach was adopted, integrating rating-based behavioral reports with performance-based cognitive and academic measures. In addition to correlational and regression analyses, classification-based analyses were used to evaluate the screening accuracy of these indicators in identifying children at risk for learning difficulties.

2.2. Participants and Recruitment

Participants were preschool children enrolled in kindergartens in a city located in northeastern Türkiye and their parents. A total of 122 children and their parents participated in the study, including 11 children in the pilot-testing phase and 111 children in the main data-collection phase. The final analytic sample consisted of 111 preschool children, including 48 boys (43.2%) and 63 girls (56.8%), with a mean age of 65.9 months (*SD* = 7.0; range = 49–80 months). Participants were recruited from three preschool settings: an independent public kindergarten (*n* = 39), a kindergarten affiliated with a public primary school (*n* = 42), and a private kindergarten (*n* = 30).

Table 1. *Demographic Characteristics of the Participants*

Variable	N	%
Gender		
Girls	63	56.8
Boys	48	43.2
School Type		
Independent public kindergarten	39	35.1
Public kindergarten (affiliated)	42	37.8
Private kindergarten	30	27.1

Note. *N* = number of participants; % = percentage of the final analytic sample.

We used a convenience sampling approach. Children were eligible if they were enrolled in preschool and did not have a diagnosis of intellectual disability. Children were excluded if their scores fell below 70 across both the general intelligence index and all sub-indices simultaneously. Accordingly, even if a child's general IQ was below 70, they were retained in the study as long as they scored above 70 on any of the sub-indices. Official research permission was obtained from the provincial educational authorities, and parents were invited through school-based communication channels. After exclusions due to test refusal, incomplete assessment, absence, or the

intellectual disability criterion, the final analytic sample consisted of 111 children who completed both ASİS and TEMA-3. The …… University Social and Human Sciences Research Ethics Committee approved the study (Approval No: 455091)

Table 2. *Participant Recruitment and Inclusion Process*

Stage	Number of Participants
Children whose parents volunteered and provided consent	118
Excluded due to incomplete test administration or absence	4
Excluded due to test refusal	2
Excluded due to intellectual disability criterion (IQ < 70)	1
Final analytic sample	111

Note. Values indicate the number of participants at each stage of recruitment and screening.

2.3. Measures

To represent behavioral, cognitive, and academic indicators associated with learning difficulty risk in early childhood, we used three instruments: a rating-based behavioral screening tool (Learning Difficulty Early Signs Screening Scale; LDESSS), a performance-based cognitive assessment (Anadolu-Sak Intelligence Scale; ASİS), and a performance-based academic mathematics assessment (Test of Early Mathematics Ability–Third Edition; TEMA-3).

Anadolu-Sak Intelligence Scale (ASİS): ASİS is a standardized intelligence test developed by Sak et al. (2016) for children aged 4–12 years. The scale was standardized using a large-scale and nationally representative standardization sample of 4,641 children across Türkiye, ensuring high structural and psychometric validity within the Turkish cultural context. It includes three standard index scores—the Verbal Potential Index, the Visual Potential Index, and the Memory Capacity Index (MCI)—and seven subtests assessing visual and verbal reasoning and memory processes. ASİS is administered individually and typically requires 30–45 minutes. Sak et al. (2016) reported internal consistency coefficients ranging from .95 to .99 for the index scores and test–retest reliability coefficients ranging from .89 to .95.

In the present study, we focused on two ASİS-based indicators of LD risk that have been clinically validated in Türkiye: (a) the Memory Capacity Index (MCI) and (b) a VSM–WoM composite score. The MCI, a standard ASİS index, is based on Visual Simultaneous Memory (VM), Visual Sequential Memory (VSM), and Verbal Short-Term Memory (VSTM) performance. Additionally, a specific VSM–WoM composite was created by summing the standardized T-scores of the VSM and Words–Meanings (WoM) subtests. Post-normative profile analyses reported by Cirik et al. (2023) showed that lower MCI scores and deficits in VSM are characteristic of students with LD, and that the VSM–WoM combination is a robust predictor of LD risk in Türkiye.

Test of Early Mathematics Ability–Third Edition (TEMA-3): TEMA-3 (Ginsburg & Baroody, 2003) is a standardized measure designed to assess early mathematics competence in children aged 3–8 years. The test includes 72 items assessing early numeracy skills such as counting, number recognition, quantity comparison, number relations, and basic arithmetic reasoning. Bliss (2006) reported internal consistency coefficients ranging from .92

to .96 and test–retest reliability coefficients of .82 for Form A and .93 for Form B. Erdoğan and Baran (2006) conducted the Turkish adaptation and validation of the instrument. In the present study, Form A was administered individually, and the obtained standard scores were used as the primary criterion measure of early mathematics competence.

Learning Difficulty Early Signs Screening Scale (LDESSS): The LDESSS is a rating-based behavioral screening instrument developed by Okur and Aksoy (2024) to identify early indicators of learning difficulty risk in children aged 4–6 years. The scale relies on parent reports of children's observable learning-related behaviors in everyday contexts. It consists of 52 items rated on a five-point Likert scale and includes four domains: language and communication, cognitive skills, psychomotor skills, and social–emotional functioning. Total scores range from 52 to 260, with higher scores indicating greater risk of learning difficulties. Okur and Aksoy (2024) reported strong psychometric properties, including internal consistency (α = .95), content validity (.92), criterion validity (.82), and test–retest reliability (.86).

2.3.1. Roles of the Measures in the Analytic Framework

Within the analytic framework of the study, the LDESSS functioned as a first-stage rating-based behavioral screening indicator, whereas ASİS MCI and the VSM–WoM composite functioned as performance-based cognitive indicators. TEMA-3 served as the primary academic criterion measure. Although the instruments are described here according to their conceptual roles in the proposed screening framework, the order of data collection did not follow this sequence.

2.4. Procedure

A trained assessor (the first author) administered all assessments individually following the standardized procedures described in the test manuals. The assessor held a master's degree in special education, was a doctoral candidate at the time of the study, and had prior professional experience working with children in special education settings. The assessor had completed formal training and certification for the administration of ASİS, TEMA-3, and LDESSS. The use of a single trained assessor helped ensure consistency in administration and minimized assessor-related variability.

Before the main data collection, we conducted pilot testing with 11 children to evaluate the feasibility of the assessment procedures. Based on the observations obtained during this phase, the main data collection took place in the children's school environments rather than in the assessor's office, thereby providing more familiar and developmentally appropriate testing conditions.

Performance-based assessments were carried out in quiet rooms within the participating schools. To minimize fatigue, no more than one assessment was administered to a child on a given day. ASİS was administered first, followed by TEMA-3 on a separate day. In addition to providing the cognitive indicators examined in the study, ASİS also served to screen for possible intellectual disability. Parent-reported LDESSS data were collected via Google Forms. Parents were contacted by phone before participation and received brief feedback after completion. All data were anonymized and stored using coded identifiers.

Although rating-based instruments are typically positioned as first-stage screening tools in applied contexts, performance-based assessments preceded the parent-report measure in the present study in order to prevent

prior knowledge of parent-reported risk indicators from influencing the administration or scoring of the performance-based measures.

2.5. Data Analysis

All data were compiled into a master dataset and analyzed using IBM SPSS Statistics (Version 29). Before the analysis, the dataset was screened for missing values and data-entry errors. No substantial missing data were detected. An independent observer trained in the scoring procedures of ASİS and TEMA-3 also checked data accuracy.

The analyses proceeded in several stages. First, descriptive statistics were calculated. Second, Pearson product–moment correlation coefficients (two-tailed) were computed to examine associations among rating-based indicators (LDESSS), cognitive performance measures (ASİS), and early mathematics competence (TEMA-3). Third, hierarchical regression and multiple linear regression analyses were conducted to examine predictive associations among variables. Before these analyses, assumptions of normality, linearity, multicollinearity, and independence of errors were evaluated. Finally, screening validity was examined using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and receiver operating characteristic (ROC) curve analyses.

In the ROC curve analyses, continuous scores were utilized to evaluate risk classification accuracy against established risk criteria, whereas subsequent predictive utility analyses (sensitivity, specificity, PPV, and NPV) relied on dichotomous classifications. In the context of these ROC analyses, the risk identification performance was evaluated across three distinct frameworks:

1. Within the TEMA-3-based MLD risk framework, ROC analyses examined how effectively the continuous LDESSS scores (total raw score and individual subtest scores) discriminated between TEMA-3-based risk-present and risk-absent children.
2. Within the ASİS-based risk framework, ROC analyses evaluated the discriminative power of the continuous LDESSS scores against the risk status established by the ASİS indicators.
3. Lastly, to test the screening utility of the cognitive indices themselves against the mathematical criterion, ROC analyses evaluated how well the continuous standard scores of the ASİS MCI and VSM–WoM composite predicted the TEMA-3-based MLD risk status.

In all models, ROC analyses directly utilized these continuous scores, whereas the subsequent predictive utility metrics were calculated based on the respective dichotomous risk classifications—whether derived from the test manual or literature-based cutoffs.

2.6. Risk Classification Criteria

TEMA-3–based MLD risk was classified into high (≤10th percentile), moderate (11th–25th percentile), and mild (26th–40th percentile) risk groups based on percentile ranks. This inclusive classification was informed by prior research and early screening frameworks recommending broader cutoffs at initial screening stages to maximize sensitivity and minimize the risk of false negatives (Fuchs & Fuchs, 2006; Geary, 2013; Jenkins et al., 2007; Murphy et al., 2007; Stock et al., 2010). In the absence of Turkish normative data for TEMA-3, percentile cutoffs were derived from the U.S. normative sample.

ASİS-based Learning Disability (LD) risk status was established using an inclusive clinical classification approach based on two complementary indicators: the Memory Capacity Index (MCI) and a customized VSM–WoM composite. The MCI was obtained directly from the standardized ASİS scoring system. To develop the VSM–WoM composite, standardized T-scores (M = 50, SD = 10) from the VSM and WoM subtests were summed to yield an aggregated metric (M = 100). Informed by the clinical profiles identified in Cırık et al. (2023), thresholds for the VSM–WoM composite were classified as high risk (≤78), moderate risk (79–84), or low risk (85–90), while MCI scores were categorized as high risk (≤83), moderate risk (84–89), or low risk (90–95). A child's overall ASİS risk status was initially determined using a higher-risk criterion across these two indicators.

The LDESSS manual specifies score ranges corresponding to very low, low, moderate, and high risk on total and subscale scores. These manual-based cutoffs were applied directly. For screening-validity analyses, all risk classifications were dichotomized into "at risk" and "not at risk." Children identified at any level of risk (mild, moderate, or high) were coded as "at risk," whereas those classified as having no risk were coded as "not at risk."

For the purposes of clinical screening efficiency and diagnostic utility analyses, the multi-tier risk categories across all measures (ASİS, TEMA-3, and LDESSS) were operationally dichotomized into a binary outcome framework: Risk Present versus Risk Absent. To maximize screening sensitivity and ensure that early indicators of vulnerability were not overlooked, the "Risk Present" group was defined inclusively across all instruments. Specifically, for both the ASİS indicators and the TEMA-3, children classified within the low, moderate, and high-risk tiers were combined into the Risk Present category. Similarly, for the LDESSS, which yielded a substantial prevalence of low-risk classifications, the mild, moderate, and high-risk tiers—along with the low-risk threshold—were designated as Risk Present. Consequently, the Risk Absent category was strictly reserved for children, demonstrating performance completely within the typical/no-risk boundaries. This binary transformation was utilized to execute robust receiver operating characteristic (ROC) curve analyses and to calculate clinical predictive metrics, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

3. Results

3.1. Descriptive Statistics

Table 3.1 presents descriptive statistics for the main study variables. Overall, participants' mean scores were close to the normative means of the respective instruments. The distributions of ASİS IQ, MCI, the VSM–WoM composite, and TEMA-3 scores were approximately normal, whereas the LDESSS total score showed a non-normal distribution.

Table 3

Descriptive Statistics for the Main Study Variables

Variable	N	Mean	SD	Range
ASİS IQ	111	102.9	14.1	64–142
MCI	111	99.6	14.5	65–131
VSM–WoM Composite	111	98.6	16.3	39–128
TEMA-3	111	96.9	16.0	59–146
LDESSS Total Score	111	98.0	26.2	53–226

Note. VSM–WoM Composite = composite indicator calculated as the sum of the standardized T-scores of the Visual Sequential Memory and Words–Meanings subtests

3.2 Relationships Among LDESSS, ASİS, and TEMA-3 Scores and LDESSS-Based Regression Findings

To address the first research question, associations among children's LDESSS total and subscale scores, ASİS full-scale IQ and cognitive indicators, and TEMA-3 scores were examined. Table 3.2 presents the correlation coefficients.

Table 4

Correlations Among LDESSS, ASİS, and TEMA-3 Scores

Variables	1	2	3	4	5	6	7	8	9	10
1. ASİS IQ	1									
2. TEMA-3	.714**	1								
3. LDESSS T	− .433**	− .380**	1							
4. Preschool D	.168	.250**	− .016	1						
5. VSM–WoM	.797**	.627**	− .322**	.191*	1					
6. MCI	.877**	.666**	− .374**	.145	.724**	1				
7. LDESSS C	− .273**	− .291**	.851**	.055	− .200*	− .229*	1			
8. LDESSS LC	− .411**	− .353**	.846**	− .039	− .287**	− .377**	.591**	1		
9. LDESSS P	− .416**	− .336**	.740**	− .149	− .373**	− .301**	.501**	.545**	1	
10. LDESSS S	− .224*	− .144	.519**	.071	− .120	− .253**	.249*	.389**	.170	1

Note. *N* = 111 for all variables; * = *p* < .05; ** = *p* < .01 LDESSS T = LDESSS Total score, LDESSS C = LDESSS Cognitive Subtest score, LDESSS LC = LDESSS Language and Communication Subtest score, LDESSS P = LDESSS Psychomotor Subtest score, LDESSS S = LDESSS Social and Emotional Subtest score

3.2.1. Correlations Between LDESSS, TEMA-3, and ASİS

Correlation analyses revealed a significant negative association between TEMA-3 scores and the LDESSS total score, *r* = − .380, *p* < .01. Similar negative relationships emerged between TEMA-3 scores and several LDESSS subdomains, including Cognitive (*r* = − .291), Language and Communication (*r* = − .353), and Psychomotor (*r* = − .336). These findings indicate that higher levels of parent-reported learning difficulty indicators were associated with lower mathematics performance. By contrast, the Social–Emotional domain was not significantly related to TEMA-3 scores. A modest but significant positive correlation also emerged between preschool attendance duration and TEMA-3 scores, *r* = .250, *p* < .01, suggesting that longer preschool experience was associated with higher mathematics performance.

Strong positive associations also emerged between performance-based cognitive indicators derived from ASİS and mathematics competence. TEMA-3 scores correlated significantly with ASİS full-scale IQ (*r* = .714), the VSM–WoM composite (*r* = .627), and MCI (*r* = .666), all *ps* < .01. These findings indicate that stronger cognitive functioning was associated with better mathematics performance. Preschool attendance duration showed a small but significant positive association with the VSM–WoM composite (*r* = .191, *p* < .05), whereas no significant relationships were observed between preschool duration and ASİS full-scale IQ or MCI.

Consistent with these findings, ASİS cognitive indicators were negatively associated with LDESSS scores. Higher ASİS full-scale IQ scores were significantly related to lower parent-reported learning difficulty indicators across the LDESSS total score ($r = -.433$) and its subdomains, including Cognitive ($r = -.273$), Language and Communication ($r = -.411$), Psychomotor ($r = -.416$), and Social–Emotional ($r = -.224$). Similarly, both the VSM–WoM composite and MCI were negatively correlated with LDESSS total and subscale scores, indicating that higher cognitive performance was associated with fewer observed indicators of learning difficulty.

3.2.2. LDESSS-Based Regression Findings

3.2.2.1. Prediction of TEMA-3 Scores From LDESSS Total Scores

A hierarchical regression analysis examined whether LDESSS total scores predicted TEMA-3 performance after controlling for preschool duration. In Step 1, preschool duration significantly predicted TEMA-3 scores, $R^2 = .063$, $F(1, 109) = 7.28$, $p = .008$. In Step 2, adding LDESSS total scores significantly improved the model, $\Delta R^2 = .142$, $\Delta F(1, 108) = 19.21$, $p < .001$, increasing the explained variance to 20.4% ($R^2 = .204$). These findings indicate that LDESSS total scores explained additional variance in TEMA-3 performance beyond preschool duration.

Table 5

Hierarchical Regression Analysis Predicting TEMA-3 Scores From LDESSS Total Scores

Model / Predictor	R	R²	SE	ΔR²	ΔF	p	DW
1. Preschool Duration	.250	.063	15.6	.063	7.28	.008	
2. LDESSS Total Score	.452	.204	14.4	.142	19.21	< .001	1.340

Note. Step 1 included preschool attendance duration. Step 2 added LDESSS total score. SE = standard error of estimate; DW = Durbin–Watson statistic

3.2.2.2. Prediction of the VSM–WoM Composite From LDESSS Total Scores

A hierarchical regression analysis also examined whether LDESSS total scores predicted the ASİS VSM–WoM composite after controlling for preschool duration. In Step 1, preschool duration significantly predicted the VSM–WoM composite, $R^2 = .036$, $F(1, 109) = 4.11$, $p = .045$. In Step 2, adding LDESSS total scores significantly improved the model, $\Delta R^2 = .102$, $\Delta F(1, 108) = 12.79$, $p < .001$, increasing the total explained variance to 13.8% ($R^2 = .138$). These findings indicate that LDESSS total scores significantly predicted the VSM–WoM composite even after controlling for preschool duration.

Table 6

Hierarchical Regression Analysis Predicting the VSM–WoM Composite From LDESSS Total Scores

Model / Predictor	R	R²	SE	ΔR²	ΔF	p	DW
1. Preschool Duration	.191	.036	16.074	.036	4.111	.045	
2. LDESSS Total Score	.372	.138	15.269	.102	12.792	< .001	1.201

Note. Step 1 included preschool attendance duration. Step 2 added LDESSS total score. SE = standard error of estimate; DW = Durbin–Watson statistic

3.2.2.3. Prediction of MCI Scores From LDESSS Subscale Scores

A multiple linear regression analysis examined whether the LDESSS subscale scores predicted MCI. The model included the Cognitive, Language and Communication, Psychomotor, and Social–Emotional subscales as predictors. The overall model was statistically significant and explained 17.1% of the variance in MCI scores (R = .413, R^2 = .171, $F(4, 106) = 5.45$, $p < .001$). Among the predictors, only the Language and Communication subscale made a unique and statistically significant contribution to the prediction of MCI scores (β = −.263, p = .035). The Cognitive, Psychomotor, and Social–Emotional subscales did not remain significant when all predictors were entered simultaneously.

These findings suggest that parent-reported language- and communication-related learning behaviors were particularly associated with children's memory capacity performance.

Table 7

Multiple Regression Analysis Predicting ASİS MCI Scores From LDESSS Subscales

Predictor	B	SE	β	t	p	VIF
Constant	118.247	5.145		22.981	< .001	
LDESSS Cognitive	.044	.139	.036	.315	.753	1.656
LDESSS Language and Communication	− .442	.207	− .263	-2.130	.035	1.950
LDESSS Psychomotor	− .285	.204	− .153	-1.396	.166	1.535
LDESSS Social–Emotional	− .323	.233	− .133	-1.383	.170	1.184

Note. Dependent variable = ASİS Memory Capacity Index (MCI). LDESSS = Learning Difficulty Early Signs Screening Scale; B = unstandardized regression coefficient; SE = standard error; β = standardized regression coefficient; VIF = variance inflation factor. R = .413, R^2 = .171, $F(4, 106) = 5.453$, $p < .001$.

3.3. Screening Validity of Rating-Based and Performance-Based Measures in Identifying LD and MLD Risk

The second and third research questions focused on the screening validity of the measures. Table 3.6 presents the distribution of children classified as at risk and not at risk for learning difficulties (LD) and mathematics learning difficulties (MLD) across the assessment measures.

Table 8

Number of Children Classified as At Risk and Not at Risk Across Measures

Measure / Score Type	High Risk	Moderate Risk	Mild Risk	Total At Risk	No Risk
TEMA-3	16	27	8	51	60
VSM–WoM Composite	15	2	8	25	86
MCI	16	11	15	42	69
ASİS Overall*	23	10	11	44	67
LDESSS Total	1	6	49	56	55
LDESSS Cognitive	3	17	60	84	27
LDESSS Language and Communication	1	6	25	32	79
LDESSS Psychomotor	1	4	24	29	82
LDESSS Social–Emotional	4	16	27	47	64
LDESSS Overall*	4	18	63	85	26

Note. VSM–WoM Composite = composite indicator calculated as the sum of the standardized T-scores of the Visual Sequential Memory and Words–Meanings subtests. ASİS Overall and LDESSS Overall indicate classification as at risk according to any of the relevant indicators or subscales. *N* = 111 for all values.

As Table 3.6 shows, the proportion of children identified as at risk varied across measures. According to TEMA-3, 45.9% of the sample was classified as being at risk for MLD. Depending on the indicator used, ASİS-based indicators identified between 22.5% and 39.6% of the sample as being at risk for LD. In contrast, the rating-based LDESSS identified a larger proportion of children as at risk when any subscale criterion was considered. For the screening-validity analyses, these multi-level classifications were dichotomized into “at risk” and “not at risk” categories.

3.3.1. Screening Validity of the LDESSS

3.3.1.1. Screening Accuracy of the LDESSS for MLD Risk (TEMA-3 Criterion)

The screening accuracy of the LDESSS for identifying MLD risk based on TEMA-3 criteria was examined using sensitivity, specificity, predictive values, and ROC analysis. The LDESSS demonstrated high sensitivity (.90) but low specificity (.35) in identifying children at risk for MLD. The positive predictive value was .54, and the negative predictive value was .81.

Table 9

Screening Accuracy of the LDESSS for TEMA-3–Based Mathematics Learning Difficulty Risk

Index	Value	Observed frequencies
Sensitivity	.90	46 / (46 + 5)
Specificity	.35	21 / (21 + 39)
Positive Predictive Value	.54	46 / (46 + 39)
Negative Predictive Value	.81	21 / (21 + 5)
AUC	.658	
Note. PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve.		

ROC curve analysis further indicated that the LDESSS total score showed modest but acceptable discriminative accuracy in distinguishing children with and without MLD risk, with an area under the curve (AUC) of .658. None of the LDESSS subscales produced a higher AUC than the total score for identifying TEMA-3–based MLD risk. Therefore, only the ROC results for the total score are reported here.

3.3.1.2. Screening Accuracy of the LDESSS for LD Risk (ASİS Criterion)

The screening validity of the LDESSS for identifying LD risk based on ASİS indicators (MCI and the VSM–WoM composite) was also examined. The LDESSS demonstrated high sensitivity (.91) but low specificity (.33) for identifying LD risk. The positive predictive value was .47, and the negative predictive value was .85.

Table 10

Screening Accuracy of the LDESSS for ASİS-Based Learning Difficulty Risk

Index	Value	Observed frequencies
Sensitivity	.91	40 / (40 + 4)
Specificity	.33	22 / (22 + 45)
Positive Predictive Value	.47	40 / (40 + 45)
Negative Predictive Value	.85	22 / (22 + 4)
AUC	.685	
Note. AUC = area under the curve.		

ROC analysis indicated that the LDESSS total score demonstrated modest discriminative accuracy for identifying LD risk (AUC = .685). However, the Language and Communication subscale yielded a higher discriminative accuracy (AUC = .716) than the total score. Therefore, the ROC results for this subscale are also reported.

3.3.2. Screening Validity of ASİS Cognitive Indicators for Identifying MLD Risk

The screening accuracy of cognitive indicators derived from ASİS—the MCI and the VSM–WoM composite—was examined in relation to TEMA-3–based MLD risk. Sensitivity, specificity, predictive values, and ROC analyses were used to evaluate their screening performance.

The MCI demonstrated moderate sensitivity (.67) and high specificity (.87) in identifying children at risk for MLD. Its positive predictive value was .81, and its negative predictive value was .75. ROC analysis further indicated strong discriminative accuracy for MCI (AUC = .855).

Similarly, the VSM–WoM composite demonstrated lower sensitivity (.48) but very high specificity (.98) and a high positive predictive value (.96) and lower negative predictive value (.69). ROC analysis indicated that this composite also showed strong discriminative accuracy in distinguishing children with and without MLD risk (AUC = .838).

Table 11
Screening Accuracy of ASİS Cognitive Indicators for TEMA-3–Based Mathematics Learning Difficulty Risk

Indicator	Sensitivity	Specificity	PPV	NPV	AUC
MCI	.67	.87	.81	.75	.855
VSM–WoM Composite	.48	.98	.96	.69	.838
Note. PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve.					

Observed frequencies for MCI: Sensitivity = 34 / (34 + 17), Specificity = 52 / (52 + 8) PPV = 34 / (34 + 8), NPV = 52 / (52 + 17)

Observed frequencies for VSM–WoM Composite: Sensitivity = 25 / (25 + 27), Specificity = 59 / (59 + 1), PPV = 25 / (25 + 1), NPV = 59 / (59 + 27)

These findings indicate that ASİS-based cognitive indicators, particularly MCI and the VSM–WoM composite, demonstrated substantially greater discriminative accuracy than the rating-based screening measure in identifying TEMA-3–defined MLD risk.

4. Discussion

In this study, we examined the screening validity and discriminative utility of rating-based and performance-based assessment measures for identifying early risk for mathematics learning difficulties (MLD) and learning difficulties (LD) in preschool children. Rather than treating informant-rating and direct performance measures as competing approaches, our findings suggest that they provide different but complementary information and work best when integrated within a sequential screening process. Across correlational, predictive, and classification-based analyses, rating-based parent reports captured broad observable learning-related difficulties and showed moderate but meaningful associations with both mathematics competence and cognitive performance, whereas performance-based cognitive indicators showed substantially stronger associations with mathematics competence and greater discriminative accuracy in identifying risk. In this respect, our findings add to a growing body of literature suggesting that early screening systems should not rely on a single type of measure but instead integrate multiple sources of information that capture different levels of children's functioning (Clarke et al., 2023; Outhwaite et al., 2024).

Our rating-based findings indicate that parent-reported screening tools can make a meaningful contribution to early identification. LDESSS total scores were moderately associated with TEMA-3 performance and significantly predicted mathematics competence even after controlling for preschool attendance duration, accounting for an additional 14.2% of variance. LDESSS total scores also predicted the VSM–WoM composite, and the Language and Communication subscale uniquely predicted MCI scores when all subdomains were considered simultaneously. These findings suggest that parent observations capture behaviorally expressed learning-related difficulties that are relevant not only to mathematics outcomes but also to broader cognitive vulnerabilities. More specifically, they suggest that language-related parent observations may be particularly informative for detecting broader learning-related cognitive vulnerability, not only mathematics-related difficulty. This interpretation aligns with previous work showing that informant-rating measures of early academic competence are psychometrically meaningful and educationally relevant, yet only moderately aligned with direct performance measures (Reid et al., 2014; Vessonen et al., 2023). It also aligns with evidence that parent ratings of preschool numeracy relate meaningfully—but not perfectly—to direct mathematics assessments (Lin et al., 2021).

At the same time, the moderate magnitude of these associations is theoretically expected. Informant-rating measures reflect children's learning-related behaviors in everyday contexts, whereas direct performance-based assessments evaluate specific skills under standardized conditions (Snow & Van Hemel, 2008; De Los Reyes & Kazdin, 2005). Studies of teacher judgments and rating scales in preschool and early school years have repeatedly shown that such measures can be useful but are also shaped by contextual and perceptual factors, including the nature of the behavior being observed and the setting in which judgments are formed (Kowalski et al., 2018; Furnari et al., 2016; Baker et al., 2015). Related evidence from executive function research likewise shows that ratings and performance-based tasks often tap overlapping but non-identical constructs (Garon et al., 2016; Hamamcı et al., 2023). Thus, rating-based tools are valuable less because they classify academic risk precisely and more because they efficiently signal children who may require further evaluation. In practical terms, this makes them especially useful for first-stage screening, where broad detection is essential.

The screening-validity findings make this distinction even clearer. The LDESSS demonstrated high sensitivity for both MLD risk defined by TEMA-3 (.90) and LD risk defined by ASİS (.91), indicating that it effectively detected children who may be at risk. However, its specificity and discriminative accuracy remained modest. For MLD risk, the LDESSS total score yielded an AUC of .658; for LD risk, the total-score AUC was .685, although the Language and Communication subscale produced a somewhat higher AUC (.716). From the perspective of ROC interpretation, these values indicate modest but meaningful discrimination rather than strong classification accuracy (Nahm, 2022). Accordingly, the LDESSS appears to function well as a high-sensitivity first-stage screener, but not as a stand-alone classification tool. This pattern is consistent with early screening logic, in which broad initial detection takes priority over precision in order to avoid overlooking children with emerging difficulties (Fuchs & Fuchs, 2006; Jenkins et al., 2007; Fuchs et al., 2012).

The fact that the LDESSS identified a larger proportion of children as at risk than the direct performance-based measures should therefore be interpreted in light of its first-stage screening function, particularly because many of these classifications fell within the mild-risk range. This pattern fits first-stage screening logic, in which mild-risk classifications increase sensitivity and reduce false negatives, even at the cost of greater over-identification. Specifically, the low specificity of .35 translates into a .65 false positive rate, meaning a majority of typically developing children were initially flagged as at risk. While such a high rate would overwhelm a single-step assessment system, this inflated false positive rate is methodologically acceptable and logistically manageable within a tiered screening framework (e.g., Fuchs & Fuchs, 2006; Jenkins et al., 2007). Because the LDESSS

functions merely as a zero-cost, parent-administered initial filter, these false positives do not trigger resource-intensive interventions. Instead, they are systematically and efficiently weeded out in the second tier by high-specificity performance-based cognitive indicators, such as the ASİS MCI (.87 specificity). Consequently, minimizing false negatives (.10) at the expense of elevated false positives represents a justified trade-off for a Tier 1 universal screener whose primary mandate is comprehensive risk capture rather than definitive classification

By contrast, the performance-based cognitive indicators were much more strongly aligned with children's mathematics competence. TEMA-3 scores were strongly associated with ASİS full-scale IQ (r = .714), MCI (r = .666), and the VSM–WoM composite (r = .627), and these effect sizes were substantially larger than the association between TEMA-3 and LDESSS total scores (r = − .380). This pattern is fully consistent with research showing that early numeracy depends strongly on domain-general cognitive processes, particularly reasoning, working memory, and language-related processing (Geary, 2013; Bull & Lee, 2014; Peng et al., 2016). It also accords with evidence that both MLD and broader mathematical difficulties are associated with weaknesses in memory, executive processes, and other domain-general cognitive skills (Passolunghi & Lanfranchi, 2012; Peng & Fuchs, 2016; Agostini et al., 2022). Longitudinal findings likewise suggest that cognitive precursors contribute substantially to later mathematics learning difficulties and persistent low achievement (Zhang et al., 2020; Geary et al., 2017). From a developmental perspective, cognitive resources such as maintaining information, recognizing patterns, integrating symbolic input, and coordinating multiple representations support children's ability to acquire and apply mathematical knowledge (De Smedt et al., 2009; Fuchs et al., 2010; LeFevre et al., 2010). The strong relations between ASİS indicators and TEMA-3 therefore suggest that performance-based cognitive measures provide more direct access to the processing capacities that underlie early mathematics learning than do rating-based reports.

At the same time, preschool attendance duration played only a modest role in our findings, suggesting that educational exposure matters but does not substitute for domain-general cognitive resources. Preschool duration was positively associated with TEMA-3 and, to a lesser extent, with VSM–WoM performance, yet it was not significantly related to broader indicators such as full-scale IQ or MCI. This pattern suggests that time spent in preschool may support certain aspects of early mathematics learning and task-related processing, while broader cognitive capacities remain shaped by a wider constellation of developmental factors. In this sense, our findings support a balanced interpretation in which early educational exposure contributes to development but does not account for the central role of domain-general cognitive processes in mathematics learning risk.

The screening profile of the ASİS indicators reinforces this interpretation. MCI showed strong discriminative accuracy for MLD risk (AUC = .855), together with high specificity (.87) and a positive predictive value of .81. Similarly, the VSM–WoM composite showed strong discriminative accuracy (AUC = .838), extremely high specificity (.98), and a very high positive predictive value (.96), although its sensitivity was lower (.48). These results indicate that the cognitive indicators were much more precise than the rating-based screening tool in distinguishing children with and without MLD risk. In practical terms, MCI appears to offer a more balanced refinement profile, whereas the VSM–WoM composite appears especially useful when the goal is to confirm risk with high certainty. This distinction matters because prior work has shown that the definition and cutoff criteria used to classify mathematics difficulty can substantially affect which children are identified and which cognitive characteristics appear most prominent (Murphy et al., 2007; Chan & Wong, 2019). Against that background, our findings show that performance-based cognitive indicators can provide a more discriminative second-stage screen even when first-stage behavioral indicators are broadly inclusive.

Taken together, our results support a tiered screening framework that integrates rating-based and performance-based assessments rather than treating them as alternatives. In the proposed model, the first stage involves universal rating-based behavioral screening, which is essential in preschool settings because it allows practitioners to collect information about large numbers of children quickly and without extensive individual testing. This function becomes especially important in early childhood, where time, personnel, and children's attentional limits often constrain large-scale direct assessment (Reid et al., 2014; Snow & Van Hemel, 2008; Clarke et al., 2023). The second stage involves brief performance-based cognitive screening, using indicators such as MCI and the VSM–WoM composite, to determine whether the behavioral risk signals identified in the first stage are accompanied by underlying cognitive vulnerabilities known to be strongly associated with mathematics learning. The third stage involves domain-specific academic assessment, such as TEMA-3, to confirm whether those behavioral and cognitive risk indicators are reflected in observable weaknesses in early mathematics competence. This sequencing is also compatible with recent reviews emphasizing that no single early mathematics screener is sufficient on its own and that multi-method approaches are likely to provide the most psychometrically and practically useful basis for early identification (Outhwaite et al., 2024).

This model also has an important practical implication for the ordering of cognitive and academic assessment. In later school years, practitioners often place academic screening earlier because classroom-based academic data are easier to obtain. Preschool contexts differ. For young children, brief cognitive tasks—especially those presented in puzzle-like or game-like formats—may be more feasible and engaging than comprehensive mathematics assessments. In our study, the ASİS-derived indicators used for screening were relatively brief, whereas a full TEMA-3 administration required considerably more time and sustained attention. From this perspective, positioning cognitive screening before comprehensive academic assessment may be both developmentally appropriate and practically efficient in preschool settings. In this sense, the proposed model shares the logic of multi-tiered systems such as RTI and MTSS but adapts that logic for early childhood by beginning with broad behavioral screening, followed by cognitive refinement, and then academic confirmation.

Several limitations should be acknowledged. First, we drew the sample from a limited regional context and from only three preschool settings, which constrains generalizability. Second, the cross-sectional design allows conclusions about concurrent associations, predictive relations within the same assessment period, and classification accuracy, but not about developmental ordering over time. Longitudinal studies are needed to test whether first-stage behavioral risk and second-stage cognitive risk predict later persistent mathematics difficulties or formal identification. Third, behavioral screening relied exclusively on parent reports. Although parent ratings are highly valuable in early childhood, future studies should also incorporate teacher reports to provide a broader picture of children's learning-related functioning across contexts. This point matters particularly because teacher ratings of early academic competence have shown meaningful validity in preschool populations, but their accuracy may vary depending on child and classroom factors (Reid et al., 2014; Kowalski et al., 2018; Furnari et al., 2016; Baker et al., 2015). Future research should also test the proposed tiered framework directly in applied preschool screening settings to determine whether it improves efficiency, reduces unnecessary referrals, and leads to more targeted follow-up assessment and intervention planning.

5. Conclusion

Our findings suggest that the key issue in early identification is not whether rating-based or performance-based measures should be preferred, but how practitioners and researchers should combine them. Rating-based parent reports provided meaningful predictive information and functioned as high-sensitivity first-stage screening tools,

but their relatively modest specificity and AUC values indicate that they are better suited to broad initial detection than to precise classification. In contrast, performance-based cognitive indicators—particularly the Memory Capacity Index (MCI) and the VSM–WoM composite—showed stronger associations with mathematics competence and substantially greater discriminative precision in identifying children at risk.

Taken together, these findings support a tiered screening framework in which rating-based behavioral indicators support efficient first-stage detection, performance-based cognitive indicators refine risk classification, and direct academic assessment confirms domain-specific mathematics difficulties. Such an approach may provide a more developmentally appropriate, practical, and evidence-informed basis for early identification of MLD risk in preschool settings.

Declarations

Ethics approval and consent to participate

This study was approved by the Anadolu University Social and Human Sciences Research Ethics Committee under approval number 455091. All procedures were conducted in accordance with the ethical standards of the institutional research committee and with the principles of the Declaration of Helsinki. Official permission was also obtained from the relevant provincial educational authorities before data collection.

Because the participants were preschool children, written informed consent was obtained from their parents or legal guardians prior to participation. Participation was voluntary, and parents were informed that their children could withdraw from the study at any stage without any negative consequences. All child assessment procedures were carried out individually in school settings by a trained assessor, and all data were anonymized using coded identifiers.

Consent for publication

Not applicable. The manuscript does not contain any identifiable individual information, images, or personal clinical details requiring consent for publication.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to ethical and privacy restrictions concerning data collected from preschool children. De-identified data may be made available from the corresponding author upon reasonable request and subject to approval by the relevant ethics committee and institutional regulations.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Authors' contributions

M.Y.Y. contributed to the conceptualization of the study, data collection, data analysis, interpretation of findings, and drafting of the manuscript. İ.H.D., S.O., V.A., and M.O. contributed to the conceptual development of the study, methodological supervision, interpretation of findings, and critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the participating children, their parents, school administrators, and teachers for their cooperation and support during the data collection process. The authors also thank the provincial educational authorities for granting permission to conduct the study in preschool settings.

References

1. Agostini, F., Zoccolotti, P., & Casagrande, M. (2022). Domain-general cognitive skills in children with mathematical difficulties and dyscalculia: A systematic review of the literature. *Brain Sciences, 12*(2), 239.
2. American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425763
3. Aunio, P. (2019). Early numeracy skills learning and learning difficulties: Evidence-based assessment and interventions. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Cognitive foundations for improving mathematical learning* (pp. 195–214). Academic Press. https://doi.org/10.1016/B978-0-12-815952-1.00008-6
4. Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis)perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of Educational Psychology, 107*(3), 805–820. https://doi.org/10.1037/edu0000008
5. Bliss, S. (2006). Test review: *Test of Early Mathematics Ability–Third Edition*. *Journal of Psychoeducational Assessment, 24*(1), 85–98. https://doi.org/10.1177/0734282905282839
6. Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36–41. https://doi.org/10.1111/cdep.12059
7. Chan, W. W. L., & Wong, T. T. Y. (2019). Subtypes of mathematical difficulties and their stability. *Journal of Educational Psychology, 112*(3), 649–666. https://doi.org/10.1037/edu0000383
8. Cirık, M., Sak, U., Arslan, D., Karaduman, E., & Opengin, E. (2023). Perfiles cognitivos de ASIS de niños con problemas de aprendizaje. *Anales de Psicología / Annals of Psychology, 39*(1), 72–80. https://doi.org/10.6018/analesps.466371
9. Clarke, B., Sutherland, M., Doabler, C. T., Lesner, T., Fainstein, D., Nolan, K., Landis, B., & Kosty, D. (2023). Developing and investigating the promise of early measurement screeners. *School Psychology Review, 52*(6), 696–708.
10. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
11. De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483
12. De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

13. Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428
14. Erdoğan, S. (2006). *Altı yaş grubu çocuklarında drama yöntemi ile verilen matematik eğitiminin matematik yeteneğine etkisinin incelenmesi* [Investigation of the effect of mathematics education given by drama method on mathematics ability in six-year-old children] [Doctoral dissertation, Ankara University].
15. Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage.
16. Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.
17. Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 92–99. https://doi.org/10.1598/RRQ.41.1.4
18. Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). The contributions of numerosity and domain-general abilities to school readiness. *Child Development, 81*(5), 1520–1533. https://doi.org/10.1111/j.1467-8624.2010.01489.x
19. Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3), 257–269.
20. Furnari, E. C., Whittaker, J., Kinzie, M., & DeCoster, J. (2016). Factors associated with accuracy in prekindergarten teacher ratings of students' mathematics skills. *Journal of Psychoeducational Assessment, 35*(4), 410–423.
21. Garon, N. M., Piccinin, C., & Smith, I. M. (2016). Does the BRIEF-P predict specific executive function components in preschoolers? *Applied Neuropsychology*: *Child, 5*(2), 110–118.
22. Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23–27.
23. Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359. https://doi.org/10.1111/j.1467-8624.2007.01069.x
24. Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680–693. https://doi.org/10.1037/edu0000159
25. Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*(4), 423–445.
26. Ginsburg, H., & Baroody, A. J. (2003). *TEMA-3: Test of Early Mathematics Ability* (3rd ed.). Pro-Ed.
27. Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children's spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning, 17*(2–3), 155–177.
28. Hamamcı, B., Acar, I., & Uyanik, G. (2023). Association between performance-based and ratings of Turkish children's executive function. *Current Psychology, 42*, 10053–10062. https://doi.org/10.1007/s12144-021-02307-0
29. Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582–600.

30. Kaufmann, L., & von Aster, M. (2012). The diagnosis and management of dyscalculia. *Deutsches Ärzteblatt International, 109*(45), 767–778. https://doi.org/10.3238/arztebl.2012.0767
31. Kowalski, K., Brown, R. D., Pretti-Frontczak, K., Uchida, C., & Sacks, D. F. (2018). The accuracy of teachers' judgments for assessing young children's emerging literacy and math skills. *Psychology in the Schools, 55*(9), 997–1012.
32. LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x
33. Lin, J., Napoli, A. R., Schmitt, S. A., & Purpura, D. J. (2021). The relation between parent ratings and direct assessments of preschoolers' numeracy skills. *Learning and Instruction, 71*, 101375. https://doi.org/10.1016/j.learninstruc.2020.101375
34. Murphy, M. M., Mazzocco, M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458–478.
35. Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology, 75*(1), 25–36.
36. Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities, 51*(6), 523–539.
37. Okur, M., & Aksoy, V. (2024). Öğrenme Güçlüğü Erken Belirtileri Tarama Ölçeğinin psikometrik niteliklerinin belirlenmesi. *Journal of Uludag University Faculty of Education, 37*(2), 534–554. https://doi.org/10.19171/uefad.1401110
38. Outhwaite, L. A., Aunio, P., Leung, J. K. Y., et al. (2024). Measuring mathematical skills in early childhood: A systematic review of the psychometric properties of early maths assessments and screeners. *Educational Psychology Review, 36*, 110. https://doi.org/10.1007/s10648-024-09950-6
39. Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*(1), 42–63.
40. Peng, P., & Fuchs, D. (2016). A meta-analysis of working memory deficits in children with learning difficulties: Is there a difference between verbal domain and numerical domain? *Journal of Learning Disabilities, 49*(1), 3–20. https://doi.org/10.1177/0022219414521667
41. Peng, P., Namkung, J. M., Barnes, M. A., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473. https://doi.org/10.1037/edu0000079
42. Reid, E. E., DiPerna, J. C., Missall, K., & Volpe, R. J. (2014). Reliability and structural validity of the teacher rating scales of early academic competence. *Psychology in the Schools, 51*(5), 535–553. https://doi.org/10.1002/pits.21769
43. Sak, U., Bal Sezerel, B., Ayas, B., Tokmak, F., Özdemir, N., Demirel Gürbüz, Ş., & Öpengin, E. (2016). *Anadolu Sak Zekâ Ölçeği (ASİS) uygulayıcı kitabı*. Anadolu Üniversitesi ÜYEP Merkezi.
44. Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*(1), 37–59.

45. Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. National Academies Press.
46. Stock, P., Desoete, A., & Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: Evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *Journal of Learning Disabilities, 43*(3), 250–268.
47. Vessonen, T., Widlund, A., Hakkarainen, A., & Aunio, P. (2023). Validating the early numeracy teacher rating scale for preschoolers (TRS–EN). *European Early Childhood Education Research Journal, 31*(2), 205–224. https://doi.org/10.1080/1350293X.2022.2081350
48. Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. https://doi.org/10.3102/0013189X14553660
49. Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2020). Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A 5-year longitudinal study. *Child Development, 91*, 7–27. https://doi.org/10.1111/cdev.13123

Figures

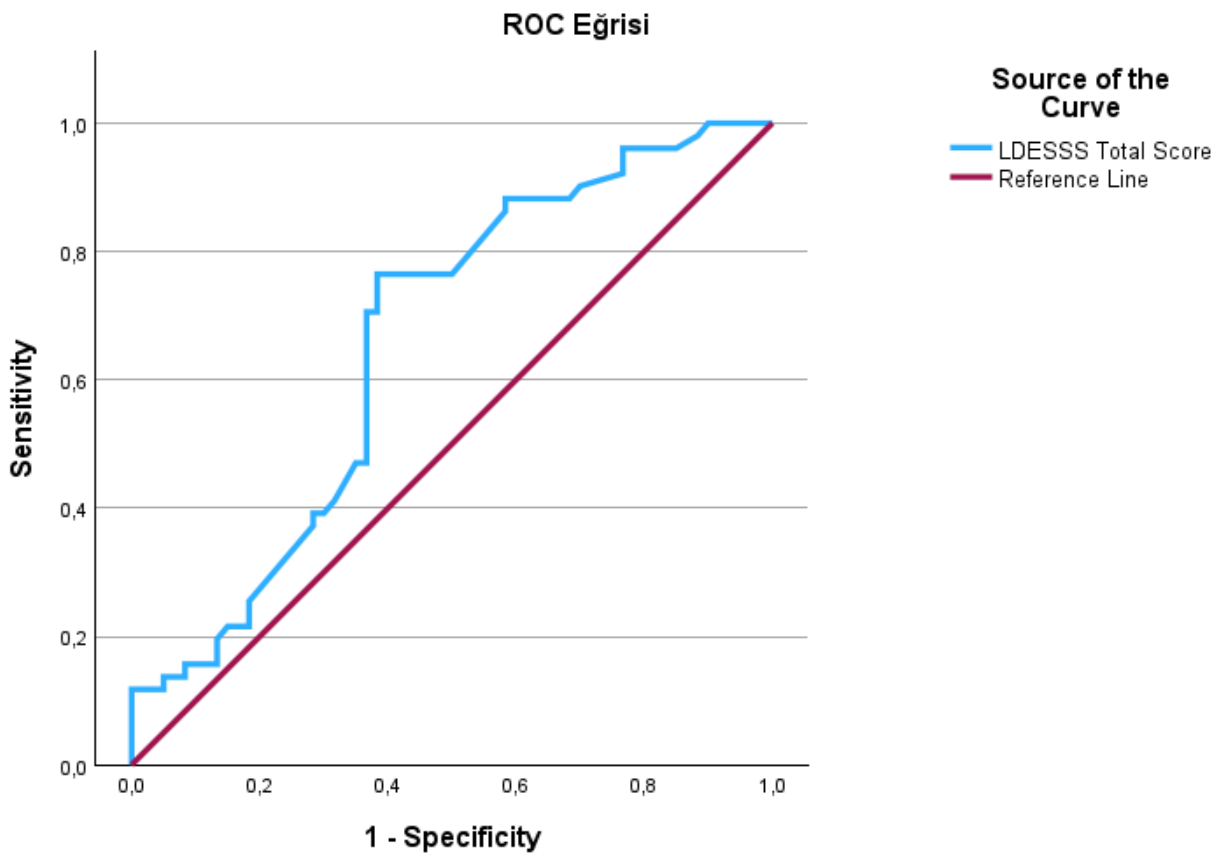


Figure 1

ROC Curve of LDESSS Total Scores in Determining the Risk of MLD

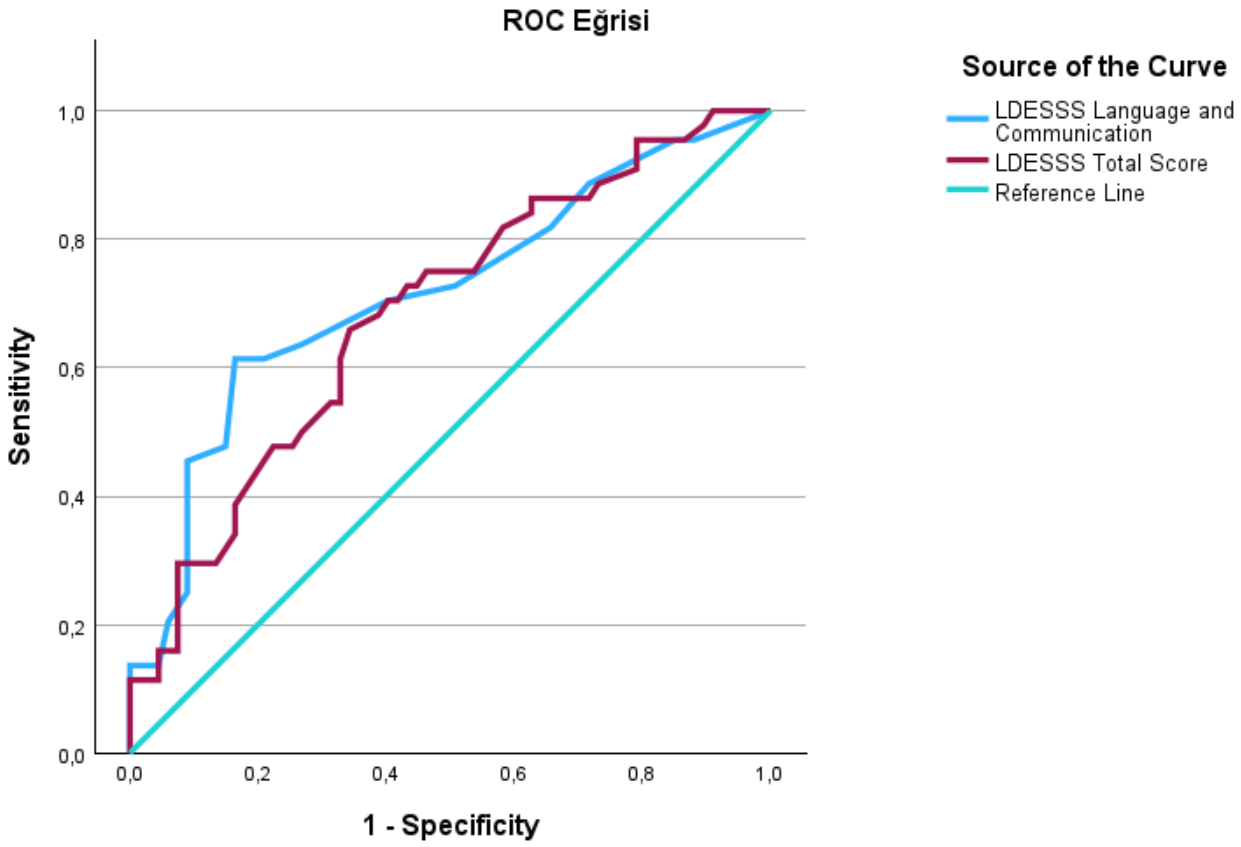


Figure 2

ROC Curve of LDESSS Total Scores and LDESSS Language and Communication Subtest Scores in Determining the Risk of LD

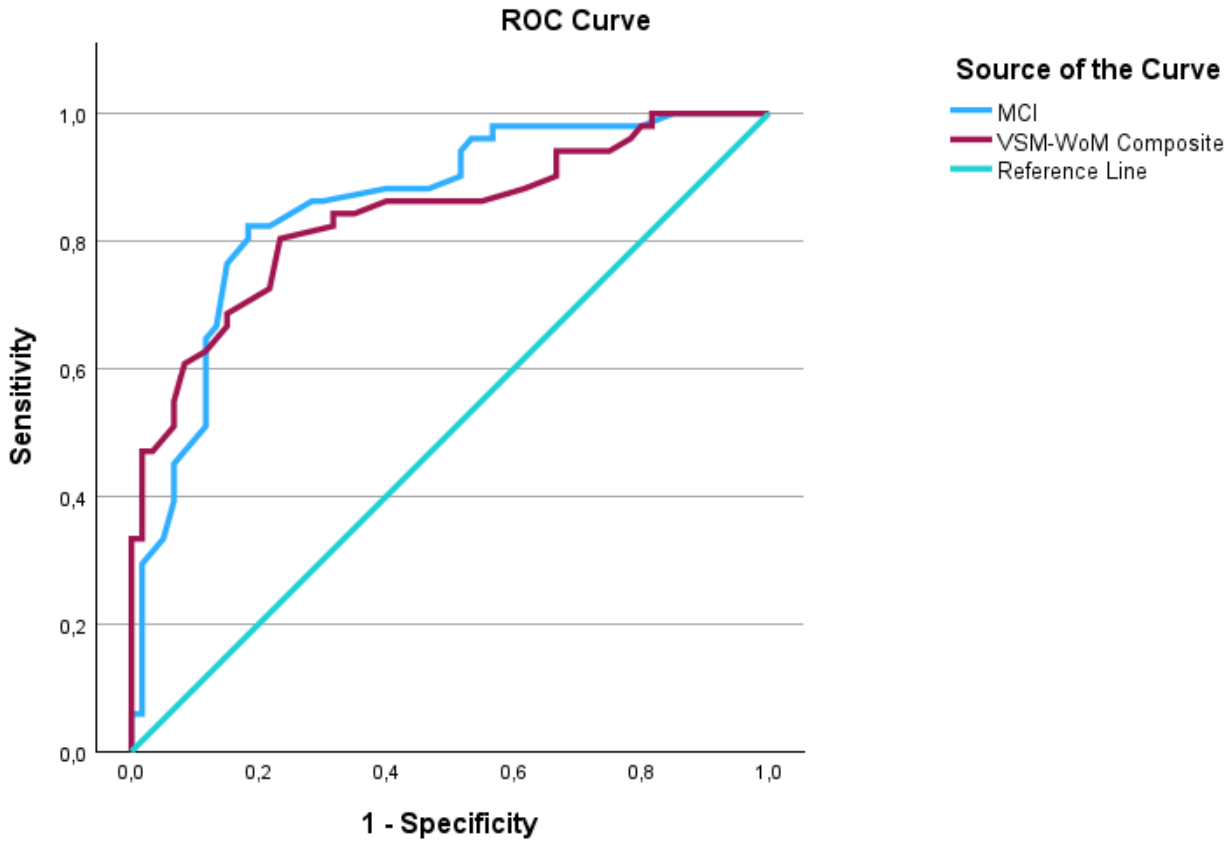


Figure 3

ROC Curve of MCI and VSM-WoM Composite Scores in Determining the Risk of MLD