

Exploring Changes in Early Number Sense Strategies from Preschool to First Grade: A Longitudinal Multi-Case Study

Nuran Tutkun¹, Sare Şengül²

Article Info

Keywords

Number sense, longitudinal study, early childhood, mathematics education

Article Type

Research Article



Received: 21 Jun 2026

Accepted: 18 Sep 2026

Abstract

This exploratory longitudinal multi-case study examines individual changes in number sense performance and strategy use across the transition from preschool to first-grade. The participants were seven students who had been assessed using the Number Sense Skills Scale (NSSS) during preschool and were reassessed with the same instrument approximately one year later, at the end of first grade. The study examined both response accuracy and the strategies underlying students' responses, with particular attention to number sense-based reasoning, rule-procedure-based approaches, and non-number sense responses. Quantitative analyses included descriptive statistics and the Wilcoxon signed-rank test, while students' responses and explanations were analyzed qualitatively according to response accuracy and strategy type. The median NSSS score increased from 12 (IQR = 3) at the preschool assessment to 14 (IQR = 6) at the first-grade assessment; however, the median individual difference score was -1, and the Wilcoxon signed-rank test indicated no statistically significant difference between scores at the two assessment occasions ($Z = -0.170$, $p = .813$; $r = .064$). Item-level analysis revealed heterogeneous patterns of strategy continuity and change across students. Some students maintained number-sense-based strategies, whereas others shifted between number-sense-based and rule-procedure-based approaches. Importantly, changes in response accuracy did not always correspond to changes in strategy type, highlighting the value of examining how children reason rather than relying on correctness alone. These findings should be interpreted cautiously because the assessment modalities differed across occasions: preschool items were read aloud by the researcher, whereas first-grade students independently read and responded to the items. Therefore, the observed changes cannot be attributed to literacy acquisition, formal mathematics instruction, or Philosophy for Children (P4C) participation alone. The findings provide exploratory, hypothesis-generating evidence that number sense development across the preschool to first-grade transition may involve heterogeneous individual trajectories and that assessment conditions should be considered when interpreting longitudinal changes in young children's mathematical reasoning.

1. Introduction

Mathematical competence extends beyond the accurate execution of calculations. It involves understanding quantities and numerical relationships, reasoning about mathematical situations, selecting appropriate approaches, and evaluating whether a result is meaningful (Howden, 1989; McIntosh et al., 1992). These competencies begin to develop well before children enter formal schooling, as young children encounter numbers and quantities through counting, comparing, estimating, sharing, and everyday problem-solving activities (Griffin, 2004). Number sense constitutes an important foundation of this early mathematical development because it encompasses interconnected understandings of number, number relations, and number operations (Jordan et al.,

¹ Türkiye Employment Agency, İstanbul, Türkiye, nurantutkun@hotmail.com, <https://orcid.org/0000-0001-7588-7839>

² Marmara University, İstanbul, Türkiye, zsengul@marmara.edu.tr, <https://orcid.org/0000-0002-1069-9084>

Corresponding Author: Nuran Tutkun, Türkiye Employment Agency, İstanbul, Türkiye

Tutkun, N., & Şengül, S. (2026). Exploring changes in early number sense strategies from preschool to first grade: A longitudinal multi-case study. *Curricular Studies & Perspectives*, 2(2), 93-110. <https://doi.org/10.29329/csp.2026.XXXX.X>

2022). Research on early numeracy further suggests that these components develop in relation to one another while also showing substantial individual variation across children (Jordan et al., 2022; Devlin et al., 2022).

Number sense, however, is not adequately characterized by whether a child arrives at a correct numerical answer. An important aspect of number sense concerns how children use their understanding of quantities and numerical relationships to select, adapt, and execute strategies. Flexible strategy use allows learners to approach a numerical situation in ways that are sensitive to the numbers and relationships involved rather than relying exclusively on a predetermined procedure. Research on mathematical strategy use indicates that children may possess multiple strategies, may vary in which strategies they use spontaneously, and may differ in their ability to adapt a strategy to the demands of a particular task (Hickendorff, 2022). Thus, examining children's strategy use can reveal aspects of mathematical understanding that are not fully captured by response accuracy alone.

This perspective places mathematical reasoning at the center of early number-sense development. Young children begin to reason mathematically when they compare quantities, identify numerical relationships, make predictions, generate solutions, and provide explanations or justifications for their decisions. Early mathematical reasoning is therefore closely connected to the strategies children construct and use when engaging with mathematical tasks. Recent research in early childhood mathematics similarly emphasizes problem solving, reasoning, and strategy development as important dimensions of young children's mathematical thinking rather than treating early mathematics as the acquisition of isolated numerical facts or procedures (Torres-Peña et al., 2025).

Number-sense development begins during the preschool years and does not follow a uniform trajectory across children. Longitudinal evidence indicates that key number-sense skills can show nonlinear patterns of growth during preschool, with meaningful differences in both initial performance and rates of change across children (Hojnoski et al., 2023; Palabıyık, 2022). More recent evidence likewise indicates that individual differences in early number-sense competencies remain important across the transition into formal schooling (Beliakoff et al., 2025). These findings suggest that children's early mathematical development is better understood as a set of individual developmental pathways than as a single, uniform progression from simple to more advanced numerical knowledge.

The transition into formal schooling may also be accompanied by changes in the strategies children use to solve mathematical tasks. Strategy development is not necessarily a simple replacement of informal strategies by formal procedures. Rather, children's mathematical learning may involve constructing new strategies, increasing reliance on previously developed strategies, improving strategy selection, and becoming more efficient in executing them (Clements et al., 2024; Siegler, 2006). Longitudinal research has also shown that the diversity and development of mathematical strategies during the early school years are related to subsequent mathematical growth, suggesting that the strategies children use provide an informative window into mathematical development (Gözüm et al., 2024; Jordan et al., 2010; National Council of Teachers of Mathematics [NCTM], 2000).

The transition from preschool to first grade is also important because mathematical activity occurs within changing instructional and assessment contexts. Preschool children who are not yet independent readers are commonly assessed through orally presented tasks, whereas first-grade students increasingly encounter written mathematical prompts that they are expected to read and interpret independently. Consequently, the same mathematical task may not constitute an equivalent assessment experience across the two developmental periods. Differences in performance may reflect changes in children's mathematical knowledge and strategies, but they may also be influenced by differences in task presentation, language demands, response requirements, and the broader context in which the task is completed.

Literacy is therefore relevant to the present study primarily as a contextual feature of assessment rather than as a causal explanation for changes in number-sense performance. Reading a mathematical

prompt requires children to coordinate numerical information with linguistic and task-processing demands, whereas an orally presented task may reduce the need for independent decoding of written information. However, the present study does not directly measure reading proficiency, reading comprehension, classroom instruction, or changes in instructional practices. Accordingly, differences observed between preschool and first grade cannot be interpreted as evidence that literacy acquisition caused changes in children's number-sense strategies. Instead, literacy-related task demands are considered alongside developmental change and differences in assessment modality when interpreting the observed trajectories.

Although early number sense has been widely studied, fewer longitudinal studies have examined how individual children's strategy use changes across the transition from preschool to first grade while simultaneously distinguishing strategy use from response accuracy. Existing research has provided important evidence concerning the development and predictive value of early numerical competencies, but less attention has been given to the continuity and transformation of individual strategy patterns across changing developmental and assessment contexts. This issue is particularly relevant when the same or comparable numerical tasks are administered under different conditions, because apparent changes in performance may reflect a combination of mathematical development and changes in how children access, process, and respond to the tasks. A longitudinal multi-case approach can therefore provide a more detailed account of individual trajectories than comparisons based solely on group-level accuracy scores.

Against this background, the present study adopts an exploratory longitudinal multi-case design to examine how children's number-sense response accuracy and strategy use are maintained, modified, or transformed across the transition from preschool to first grade. Rather than testing a causal relationship between literacy acquisition, formal schooling, and number-sense development, the study focuses on describing individual patterns of continuity and change while considering the changing conditions under which the numerical tasks were administered. By examining both response accuracy and the strategies underlying children's responses, the study seeks to provide a more nuanced account of early number-sense development across an important educational transition. Specifically, the study addresses the following research questions:

1. How do individual children's response accuracy and number-sense strategy use differ between oral preschool administration and self-read first-grade administration of the NSSS?
2. What patterns of continuity and change in children's number-sense strategy use emerge across the two assessment waves?
3. How are changes in strategy use related to changes in response accuracy at the individual level?

2. Methodology

2.1. Research Design

This study employed an exploratory longitudinal multi-case design incorporating quantitative and qualitative analyses to examine individual patterns of continuity and change in children's number sense performance and strategy use from preschool to the end of first grade. The same cohort of children was followed across two assessment occasions approximately one year apart. The longitudinal multi-case design was selected because it allows developmental changes to be examined within the same individuals while also enabling the identification of similarities and differences in individual trajectories (Cohen et al., 2009; Yin, 2018). In the present study, quantitative analyses were used to examine response accuracy and changes in performance across the two assessment occasions, whereas qualitative analyses focused on the strategies underlying children's responses. The integration of these two forms of evidence allowed the study to examine not only whether children's performance changed, but also how their strategy use was maintained, modified, or transformed over time.

The baseline data were collected during the 2023-2024 academic year as part of the researcher's doctoral dissertation entitled "*The effect of philosophy for children practices on number sense skills of preschool students*". During the doctoral study, children's number sense skills were assessed using the Number Sense Skills Scale (NSSS), a measurement tool developed by the researcher. Approximately one year later, at the end of the 2024-2025 academic year, the same instrument was administered again to the same participants. No experimental treatment or intervention was introduced during the follow-up period. Accordingly, the present study does not aim to estimate the effects of literacy acquisition, formal mathematics instruction, or Philosophy for Children (P4C) participation. Instead, it adopts an exploratory longitudinal perspective to describe naturally occurring patterns of change in number sense performance and strategy use across the transition from preschool to first grade.

The two assessment occasions differed in both developmental context and administration modality. During the preschool assessment, all NSSS items were read aloud by the researcher, who documented the children's oral responses. During the first-grade assessment, children independently read and responded to the items. Consequently, the two assessment occasions cannot be considered identical measurement conditions. Differences observed across the two waves were therefore interpreted cautiously, with attention to both individual developmental change and changes in assessment conditions. Because literacy skills, classroom instructional practices, and other potentially relevant experiences during the intervening period were not independently measured, the observed changes cannot be attributed to any single factor.

Within this design, the individual child constituted the primary case of analysis. The study first examined quantitative patterns of response accuracy across the two assessment occasions and then examined the qualitative characteristics of the strategies evident in children's responses. Cross-case comparisons were subsequently used to identify recurring patterns of continuity and change across the seven children. This design was intended to provide an integrated account of individual developmental trajectories rather than to produce population-level estimates or causal inferences.

2.2. Participants

The participants were children enrolled in a public preschool located in the Ataşehir district of Istanbul. The initial assessment was conducted during the 2023-2024 academic year with 11 kindergarten children ($M = 5.6$ years). These children had previously participated in the administration of the Number Sense Skills Scale (NSSS) as part of the researcher's doctoral research. A follow-up assessment was conducted approximately one year later, at the end of the 2024-2025 academic year, after the children had completed first grade ($M = 6.6$ years). Of the original 11 children, seven (4 girls, 3 boys) participated in both assessment occasions and constituted the final analytic sample.

Four children were lost to follow-up. Two children returned to kindergarten because they were in the 61-month age group, whereas the families of two children could not be contacted. Criterion sampling, a purposive sampling technique, was used to identify children who met the inclusion criterion of having participated in both the initial preschool assessment and the follow-up assessment at the end of first grade (Yıldırım & Şimşek, 2016). Accordingly, the final sample consisted of seven children who completed both measurement occasions.

Given the longitudinal multi-case design, the purpose of participant selection was not to obtain a statistically representative sample or to support population-level generalization. Rather, the inclusion criterion ensured that each child had data available at both assessment occasions, allowing individual trajectories of response accuracy and strategy use to be examined longitudinally.

This study was derived from Nuran Tutkun's doctoral dissertation, conducted under the supervision of Sare Şengül. Ethical approval was obtained from the Institute of Educational Sciences Publication Ethics Committee at Marmara University on November 1, 2023 (Approval No. 651808). At the time of the initial assessment, parents/guardians were informed that the study could include a subsequent follow-up assessment and that the NSSS might be re-administered at a later stage. Before the follow-

up assessment, this information was reiterated to the parents/guardians, and renewed informed consent was obtained for their children's continued participation in the study.

2.3. Data Collection Tool

The Number Sense Skills Scale (NSSS), developed by the researcher, was used as the primary data collection instrument (Tutkun, 2025). The NSSS was originally developed to assess number sense skills among preschool children. The ten-item scale consists of three sub-dimensions reflecting the theoretical structure of number sense: (1) Knowing the meaning and size of numbers, (2) Mental calculation and flexibility in calculation, and (3) Using reference points and engaging in quantitative reasoning and inference. The first, second, and third sub-dimensions contain three, four, and three items, respectively.

During the development of the NSSS, an initial 15-item draft was administered to 300 children attending preschool education in Istanbul, and exploratory factor analysis was conducted. Five items were subsequently removed, resulting in a three-factor structure explaining 46.71% of the total variance. Factor loadings ranged from .435 to .762. Confirmatory factor analysis was then conducted with a separate sample, and the three-factor structure demonstrated acceptable fit indices ($\chi^2/df = 1.134$, CFI = .955, GFI = .956, AGFI = .924, RMSEA = .030). In the reliability analyses, the KR-20 coefficient was .62 and Cronbach's alpha was .61. These values were considered acceptable for the purposes of the original scale-development study.

The NSSS was originally developed and psychometrically evaluated for preschool children. Therefore, its use in the present study with first-grade children requires careful interpretation. No new validity or reliability study was conducted specifically for the first-grade administration in the present study. Accordingly, first-grade scores are not treated as psychometrically equivalent to preschool scores obtained under the original validation conditions. Rather, repeated administration of the NSSS was used as an exploratory basis for examining patterns of response accuracy and strategy use across two developmental and assessment contexts.

This measurement consideration is particularly relevant to the interpretation of changes in total or subscale scores. Accordingly, quantitative comparisons are treated as descriptive evidence of change within the present sample rather than as evidence of formally equivalent measurement across developmental levels. Greater interpretive emphasis is placed on the pattern of children's responses and strategies across the two occasions than on direct comparison of total scores as if they represented identical measurement conditions.

In addition to the numerical responses recorded for each NSSS item, the children's verbal responses and observable solution strategies constituted the qualitative evidence for the strategy analysis. These responses were considered alongside the item-level accuracy data to determine the strategy category represented in each response. Thus, the NSSS functioned not only as a source of quantitative response data but also as a structured context for eliciting qualitative evidence about children's numerical reasoning and strategy use.

2.4. Data Collection Process

Data were collected from the same group of children at two assessment occasions approximately one year apart. The first assessment was conducted during the 2023-2024 academic year, when the children were attending preschool. Because the children were not yet expected to independently read the NSSS items, the scale was administered individually by the researcher. Each item was read aloud to the child, and the child's response and solution strategy were documented in the response form. The assessments were conducted individually in a quiet classroom environment under the supervision of the preschool teacher, and each administration lasted approximately 10 minutes. When a child indicated that an item had not been understood, the item was reread.

The second assessment was conducted at the end of the 2024-2025 academic year, after the same children had completed first grade. At this stage, the NSSS was administered individually, and the children independently read and responded to the items. Each administration lasted approximately 20 minutes. The children wrote their responses and observable solution strategies on the questionnaire themselves.

The same NSSS was therefore administered at both assessment occasions, while the mode of administration differed according to the children's developmental and educational context. The preschool assessment involved researcher-read items and oral responses, whereas the first-grade assessment involved independent reading and responding. This difference in administration modality was treated as an important contextual feature of the longitudinal comparison rather than as a controlled experimental condition. Consequently, changes in response accuracy and strategy use were interpreted in conjunction with the developmental transition and the differences in assessment conditions.

2.5. Data Analysis

Data analysis incorporated both quantitative and qualitative procedures. Quantitative analyses were used to examine students' NSSS scores and response accuracy across the two assessment occasions, whereas qualitative analysis focused on the strategies and explanations underlying students' responses. The two forms of analysis were considered complementary: quantitative analyses described patterns of performance and change, while qualitative analyses provided information about how students approached the number-sense tasks.

For the quantitative component, NSSS total scores and sub-dimension scores were calculated for each assessment occasion. Scores obtained during the preschool assessment were compared descriptively with those obtained at the end of first grade. For each child, the difference between the first-grade and preschool scores (first grade minus preschool) was also calculated to identify the direction and magnitude of individual change. These individual difference scores were used to describe within-case patterns of performance change rather than to infer developmental effects at the population level.

For the qualitative component, students' responses to the NSSS items and their accompanying explanations were examined to identify the strategies used to solve the tasks. To distinguish strategy choice from response outcome, each response was coded along two analytically distinct dimensions: (a) Response Accuracy, classified as correct or incorrect, and (b) Strategy Type, classified as flexible number-sense reasoning or rule-procedure-based reasoning. This distinction was made because response accuracy alone does not necessarily indicate the nature or quality of the reasoning used to obtain an answer. Thus, a correct answer could be produced through either flexible number-sense reasoning or a procedural approach, while an incorrect answer could nevertheless provide evidence of meaningful number-sense reasoning.

The qualitative explanations were independently coded by two researchers with expertise in early childhood mathematics education. Inter-coder agreement was assessed using Cohen's kappa, which yielded $\kappa = .88$, indicating a high level of agreement between the two coders. Following independent coding, disagreements were reviewed and resolved through discussion and consensus. The 0-1-2 scoring rubric used for composite scoring is summarized in Table 1.

The composite scoring rubric was designed to capture the nature of students' solution strategies rather than response accuracy alone. Accordingly, a correct response based on flexible number-sense reasoning received 2 points, an incorrect response accompanied by evidence of number-sense reasoning received 1 point, and a response based on mechanical counting, rigid procedures, or a blank/non-informative explanation received 0 points. This scoring system intentionally distinguishes the outcome of a response from the reasoning evident in the response. Therefore, a correct procedural response may receive fewer points than an incorrect response that demonstrates number-sense reasoning. The composite score was consequently not interpreted as a conventional measure of

mathematical accuracy. Response Accuracy and Strategy Type were also retained as separate analytic variables in order to avoid conflating performance outcome with strategy use.

Given the small sample size ($n = 7$) and the paired nature of the longitudinal data, non-parametric procedures were used for the quantitative comparisons. Descriptive statistics, including medians and interquartile ranges (IQRs), were calculated for NSSS scores at the two assessment occasions. For each child, a difference score (first-grade score minus preschool score) was calculated to describe the direction and magnitude of individual change. The numbers of positive ranks, negative ranks, and ties were also examined. The Wilcoxon signed-rank test was used to evaluate whether the distribution of paired scores differed between the preschool and first-grade assessments. Given the exploratory nature of the study and the very small sample, the inferential results were interpreted cautiously and were considered alongside the individual-level and strategy-level findings. The effect size (r) was calculated to provide an estimate of the magnitude of the observed difference and was interpreted using Cohen's (1988) guidelines. An individual change plot was also used to visually examine score changes between the two assessment occasions. All quantitative analyses were conducted using IBM SPSS Statistics 22.

Table 1

Coding and Scoring Students' Scale Responses

Response Accuracy	Strategy Type	Description	SPSS Code	Scoring
Correct	Flexible number-sense reasoning	Correct response accompanied by flexible reasoning, benchmark use, or mental decomposition.	1	2
Incorrect	Flexible number-sense reasoning	Incorrect numerical answer, but the explanation demonstrates valid number-sense logic or benchmark reasoning.	2	1
Correct or incorrect	Answering Based on Process-Rule Approach/Meaningless Explanations/Blank	Response based primarily on mechanical counting, rigid procedures, or a blank/non-informative explanation.	3	0

3. Findings

This exploratory longitudinal multi-case study examined changes in students' number sense performance and strategy use across two assessment occasions: preschool and the end of first grade. Quantitative analyses focused on changes in NSSS total and component scores, whereas qualitative analyses examined item-level response accuracy and strategy use across the two occasions. The findings are presented in three stages. First, changes in NSSS scores are reported descriptively and inferentially. Second, item-level strategy transitions are examined to identify patterns of strategy maintenance and change. Finally, selected individual cases are presented to illustrate how changes in strategy use and response accuracy occurred across the two assessment occasions.

Table 2 indicates substantial individual variation in NSSS total scores across the two assessment occasions. Three students (S2, S5, and S6) obtained higher scores at the end of first grade, whereas four students (S1, S3, S4, and S7) obtained lower scores. The magnitude of change also varied across cases, ranging from a 6-point increase for S5 to a 5-point decrease for S1. Thus, the score patterns do not indicate a uniform direction of change across the cohort. When the students' sub-dimension scores were examined, different change patterns were observed in the three components of number sense. Changes in component scores were likewise heterogeneous. Scores on the three components did not show a consistent pattern of increase or decrease across students. The largest individual variations were observed in the first two components, whereas scores on the third component were comparatively more stable for several students. These patterns further indicate that changes in total NSSS scores were not uniform across the different dimensions of number sense.

Table 2*Students' NSSS Total Scores at the Preschool and First-Grade Assessments*

Student	Test	Component 1	Component 2	Component 3	Total Score
S1	Preschool assessment	3	4	5	12
	First-grade assessment	3	2	2	7
S2	Preschool assessment	2	4	6	12
	First-grade assessment	6	4	6	16
S3	Preschool assessment	4	4	6	14
	First-grade assessment	3	2	6	11
S4	Preschool assessment	6	6	6	18
	First-grade assessment	5	7	4	16
S5	Preschool assessment	2	2	4	8
	First-grade assessment	4	6	4	14
S6	Preschool assessment	3	4	4	11
	First-grade assessment	6	4	6	16
S7	Preschool assessment	3	5	3	11
	First-grade assessment	2	2	6	10

To provide a visual representation of individual score changes across the two assessment occasions, Figure 1 presents students' NSSS scores at the preschool and first-grade assessments.

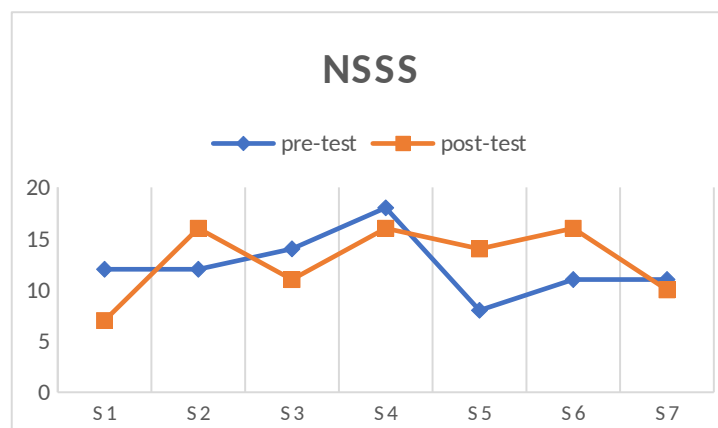
Figure 1*Comparison of Students' NSSS Scores Across the Two Measurement Occasions*

Figure 1 presents students' NSSS total scores obtained during the preschool period (preschool assessment) and at the end of first grade (first-grade assessment). As illustrated in the figure, score changes were not homogeneous across students. While three students (S2, S5, and S6) demonstrated increases in their scores, four students (S1, S3, S4, and S7) showed decreases. These findings indicate individual variation in the direction and magnitude of score changes over time.

Table 3*Descriptive Statistics for NSSS Scores*

Measurement	Median	Q1	Q3	IQR
Preschool assessment	12	11	14	3
First-grade assessment	14	10	16	6
Difference Score (First g.-Pre)	-1	-3	5	8

Note. Positive ranks = 3; Negative ranks = 4; Ties = 0.

The Wilcoxon signed-rank test was administered to determine the significance of the differences between preschool assessment and first-grade assessment scores. The analysis results are presented in Table 4.

Table 4*Wilcoxon Signed-Rank Test Results*

Variable	Median preschool assessment	Median first-grade assessment	W ⁺	W ⁻	Z	Exact two-sided p	Effect size (r)
Total NSSS Score	12	14	15	11	-0.170	.813	.064

Note. W⁺ and W⁻ represent the sums of positive and negative ranks, respectively. The exact two-sided p-value was calculated based on all possible sign permutations for the seven paired observations. Effect size was calculated using $r = Z/\sqrt{N}$.

The Wilcoxon signed-rank test indicated that the difference between the preschool assessment scores (Median = 12, IQR = 3) and first-grade assessment scores (Median = 14, IQR = 6) was not statistically significant, $Z = -0.170$, $p = .813$. Examination of individual changes showed that three students demonstrated higher scores at first-grade assessment, whereas four students obtained lower scores. No tied observations were identified between the two measurement occasions. The median difference score (first-grade assessment - preschool assessment) was -1. The calculated effect size was negligible ($r = .064$), indicating that the magnitude of the observed change was extremely small. The non-significant result should not be interpreted as evidence that individual students showed no change. Rather, the analysis did not provide evidence of a consistent group-level difference across the two assessment occasions. Given the small sample and the differences in administration modality, the quantitative results are interpreted descriptively and in conjunction with the item-level qualitative analysis.

To examine changes in strategy use across the two assessment occasions, each student's response was compared at the item level. Response accuracy and strategy type were examined as separate dimensions, allowing strategy transitions to be identified independently of changes in correctness. Table 5 therefore shows whether students maintained or changed their strategy type and whether these transitions were accompanied by improved, unchanged, or decreased response accuracy.

Table 5*Item-Level Transitions in Response Accuracy and Strategy Type Across the Two Assessment Occasions*

Stud.	Item	Preschool Response	First-Grade Response	Transition	Stud.	Item	Preschool Response	First-Grade Response	Transition
S1	1	Incorrect + NS	Incorrect + NS	NS maintained; accuracy unchanged (incorrect)	S5	1	Incorrect + RP	Correct + NS	RP → NS; accuracy improved
S1	2	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S5	2	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S1	3	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)	S5	3	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)
S1	4	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)	S5	4	Correct + RP	Correct + NS	RP → NS; accuracy unchanged (correct)
S1	5	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)	S5	5	Correct + RP	Incorrect + NNS	RP → NNS; accuracy decreased

Stud.	Item	Preschool Response	First-Grade Response	Transition	Stud.	Item	Preschool Response	First-Grade Response	Transition
S1	6	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S5	6	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S1	7	Correct + NS	Correct + RP	NS → RP; accuracy unchanged (correct)	S5	7	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S1	8	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S5	8	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S1	9	Incorrect + NS	Correct + RP	NS → RP; accuracy improved	S5	9	Incorrect + NNS	Correct + RP	NNS → RP; accuracy improved
S1	10	Correct + NS	Incorrect + NNS	NS → NNS; accuracy decreased	S5	10	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S2	1	Incorrect + NS	Correct + NS	NS maintained; accuracy improved	S6	1	Incorrect + NS	Correct + NS	NS maintained; accuracy improved
S2	2	Correct + NNS	Correct + NS	NNS → NS; accuracy unchanged (correct)	S6	2	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S2	3	Incorrect + NS	Correct + NS	NS maintained; accuracy improved	S6	3	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S2	4	Correct + NS	Correct + RP	NS → RP; accuracy unchanged (correct)	S6	4	Correct + NS	Correct + RP	NS → RP; accuracy unchanged (correct)
S2	5	Correct + RP	Correct + NS	RP → NS; accuracy unchanged (correct)	S6	5	Incorrect + NS	Correct + RP	NS → RP; accuracy improved
S2	6	Incorrect + NNS	Correct + RP	NNS → RP; accuracy improved	S6	6	Incorrect + NS	Correct + NS	NS maintained; accuracy improved
S2	7	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S6	7	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S2	8	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S6	8	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S2	9	Correct + RP	Correct + NS	RP → NS; accuracy unchanged (correct)	S6	9	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S2	10	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S6	10	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S3	1	Incorrect + NS	Incorrect + NS	NS maintained; accuracy unchanged (incorrect)	S7	1	Correct + NS	Correct + NNS	NS → NNS; accuracy unchanged (correct)

Stud.	Item	Preschool Response	First-Grade Response	Transition	Stud.	Item	Preschool Response	First-Grade Response	Transition
S3	2	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S7	2	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S3	3	Incorrect + NNS	Correct + RP	NNS → RP; accuracy improved	S7	3	Incorrect + NS	Correct + RP	NS → RP; accuracy improved
S3	4	Incorrect + NS	Correct + RP	NS → RP; accuracy improved	S7	4	Correct + NS	Correct + RP	NS → RP; accuracy unchanged (correct)
S3	5	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)	S7	5	Correct + RP	Correct + RP	RP maintained; accuracy unchanged (correct)
S3	6	Incorrect + NS	Correct + RP	NS → RP; accuracy improved	S7	6	Incorrect + NS	Correct + RP	NS → RP; accuracy improved
S3	7	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S7	7	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S3	8	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S7	8	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)
S3	9	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S7	9	Incorrect + NS	Correct + NS	NS maintained; accuracy improved
S3	10	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)	S7	10	Incorrect + NNS	Correct + NS	NNS → NS; accuracy improved
S4	1	Correct + NS	Incorrect + NS	NS maintained; accuracy decreased					
S4	2	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)					
S4	3	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)					
S4	4	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)					
S4	5	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)					
S4	6	Incorrect + NS	Correct + NS	NS maintained; accuracy improved					
S4	7	Incorrect + NNS	Incorrect + NS	NNS → NS; accuracy unchanged (incorrect)					

Stud.	Item	Preschool Response	First-Grade Response	Transition	Stud.	Item	Preschool Response	First-Grade Response	Transition
S4	8	Correct + NS	Incorrect + NS	NS maintained; accuracy decreased					
S4	9	Correct + NS	Incorrect + NS	NS maintained; accuracy decreased					
S4	10	Correct + NS	Correct + NS	NS maintained; accuracy unchanged (correct)					

Note. NS = number-sense-based strategy; RP = rule/procedural strategy; NNS = non-number-sense-based strategy. “Maintained” indicates that the same strategy type was observed across the two assessment occasions. Accuracy changes were coded separately from strategy transitions.

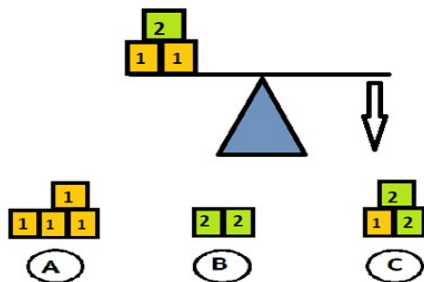
The item-level analysis revealed both stability and change in students’ strategy use. Across the cases, number-sense strategy maintenance (NS → NS) was the most frequently observed pattern, although its prevalence varied considerably among students. S4 demonstrated particularly strong strategy stability, with NS → NS observed on nine of the ten items. Other students showed more heterogeneous profiles, including shifts from number-sense to rule-/procedure-based strategies (NS → RP), shifts toward number-sense strategies from rule-/procedure-based or non-number-sense strategies, and, in some cases, shifts from number-sense to non-number-sense strategies. Thus, the transition from preschool to first grade was not characterized by a single, uniform trajectory of strategy change.

Some students shifted from number-sense-based strategies to rule-/procedure-based strategies at the second assessment, whereas others maintained number-sense strategies or shifted toward number-sense-based reasoning. These heterogeneous patterns suggest that changes in strategy use cannot be characterized as a uniform movement from flexible reasoning toward procedural approaches.

Figure 2

Question on Knowing the Meaning and Magnitude of Numbers – S4’s Answers

1- The scale below has cubes with numbers on them placed on one arm. What do you think the numbers on these cubes add up to? And which group of numbered cubes should be placed on the other arm of the scale so that it moves down in the direction of the arrow?

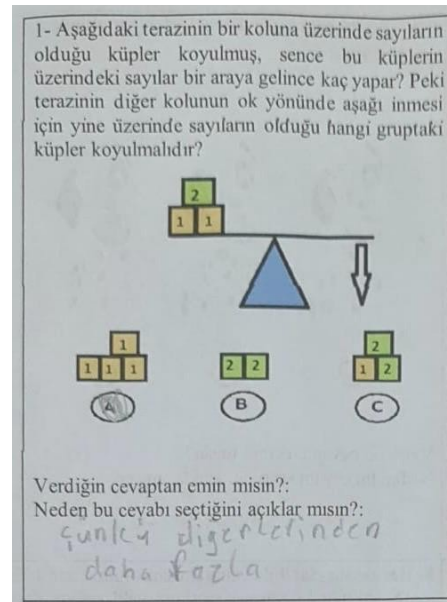


Are you sure about your answer?:

Could you explain why you chose this answer?:

Preschool assessment

C: 1 is heavier because it's less than the other, and 2 are more.



First-grade assessment

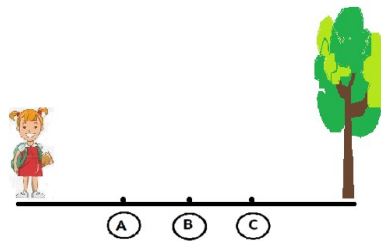
A: Because there are more cubes than the others.

An illustrative example can be seen in S4's response to Item 1 (Figure 2). In the preschool assessment, when the item was presented orally, S4's explanation focused primarily on the number of cubes rather than on the numerical magnitude represented on the cubes. In the first-grade assessment, when the student independently read the item, the explanation focused on comparing the numerical quantities represented by the cubes. Thus, although the response changed across occasions, the example illustrates a difference in the reasoning expressed under the two assessment conditions.

Figure 3

Questions on Using Reference Points and Question-Making Inferences – S1's Answers

9- Ekin walks 10 steps from where she is to the tree. At what point will Ekin be after 5 steps from where she started walking?

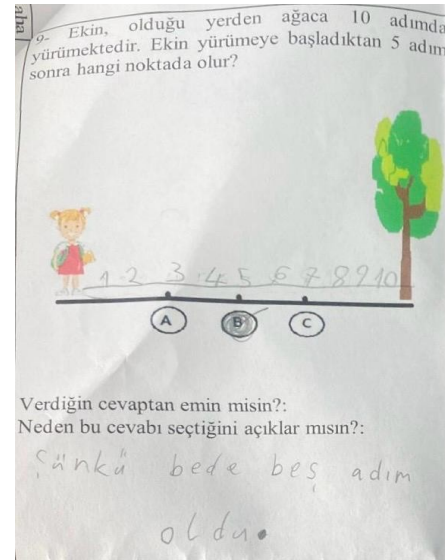


Are you sure about your answer?:

Could you explain why you chose this answer?:

Preschool assessment

C: If she walks a lot, she will reach point C.



First-grade assessment

B: Because B has taken five steps.

A different pattern was observed for S1 on Item 9 (Figure 3). In the preschool assessment, S1 provided an incorrect response accompanied by a number-sense-based explanation. In the first-grade assessment, the student provided a correct response using a rule-procedure-based explanation. Thus, the change involved both improved response accuracy and a shift in strategy type from NS to RP. This example illustrates why accuracy and strategy use were examined as separate dimensions in the present study.

4. Discussion, Conclusion, and Recommendations

This study examined individual and group-level changes in number sense performance and strategy use among the same students from preschool to the end of first grade. Designed as an exploratory longitudinal multi-case study, the study focused on continuity and change in students' solution strategies across two assessment occasions. An important methodological characteristic was that the assessment procedures differed between the two occasions. During the preschool assessment, items were read aloud by the researcher and students responded orally, whereas during the first-grade assessment, students independently read and responded to the items. Therefore, the observed differences should be interpreted as patterns occurring across a developmental and educational transition accompanied by changes in assessment conditions, rather than as direct evidence of the effects of literacy acquisition, formal mathematics instruction, or P4C practices. The central contribution of the study lies in showing how examining individual strategy trajectories alongside overall performance can provide a more differentiated account of children's number sense across this transition.

The quantitative findings provide limited evidence of systematic change in overall number sense performance across the two assessment occasions. Although the median NSSS score increased from 12 (IQR = 3) in the preschool assessment to 14 (IQR = 6) in the first-grade assessment, the median

individual difference score (first-grade assessment - preschool assessment) was -1. Three students showed positive changes, whereas four showed negative changes. Consistent with these individual patterns, the Wilcoxon signed-rank test indicated no statistically significant difference between scores at the preschool and first-grade assessments ($Z = -0.170$, $p = .813$), and the effect size was negligible ($r = .064$). Thus, the higher median score observed at the first-grade assessment should not be interpreted as evidence of general improvement in number sense performance. Rather, the wider interquartile range and divergent individual trajectories point to substantial heterogeneity in students' performance across the two occasions. Given the small sample size, the non-significant result should also not be interpreted as evidence that no developmental change occurred. Instead, the quantitative findings are best understood as descriptive evidence of varied individual trajectories.

This heterogeneity becomes particularly meaningful when performance is considered together with students' strategy use. The qualitative analysis suggests that number sense development across the preschool-to-primary-school transition cannot be adequately characterized as a simple progression toward either greater flexibility or greater proceduralization. Some students maintained number-sense-based strategies, whereas others changed the strategies they used across the two assessment occasions. Importantly, these changes did not always occur in the same direction. This pattern suggests that strategy use may be more dynamic and individually differentiated than a developmental account based solely on total performance scores would indicate. In this respect, examining how children reasoned through tasks provides information about mathematical thinking that is not fully captured by whether their answers were correct.

One particularly important implication concerns the distinction between response accuracy and strategy use. Some students produced correct answers while relying on counting-based, rule-based, or otherwise non-flexible approaches, whereas an incorrect response could, in some cases, be accompanied by reasoning that demonstrated meaningful number-sense understanding. This distinction is consistent with conceptualizations of number sense as involving not only correct numerical outcomes but also understanding numerical relationships, flexibility, estimation, mental calculation, and the use of meaningful reference points (McIntosh et al., 1992; Griffin, 2004). Accordingly, accuracy alone may provide an incomplete picture of the quality of young children's mathematical reasoning. For early mathematics assessment, attention to the reasoning underlying a response may therefore be particularly valuable when the aim is to understand children's developing number sense.

The observed shifts toward more rule- or counting-based explanations in some cases are descriptively noteworthy, but they should not be interpreted as evidence that formal schooling, literacy acquisition, or classroom instruction reduced students' number sense. Because the present study did not directly measure classroom instructional practices, students' reading proficiency, reading comprehension, or other learning experiences during the intervening period, the source of these changes cannot be isolated. Several processes may have contributed simultaneously, including developmental maturation, increasing familiarity with formal mathematical procedures, individual learning experiences, changes in task demands, and differences in assessment modality. The longitudinal design therefore allows these changes to be documented, but not attributed to a single developmental or instructional mechanism.

The difference in administration modality is particularly important in interpreting the findings. In the preschool assessment, the researcher read the items aloud, allowing students to focus primarily on the numerical and quantitative aspects of the tasks. In the first-grade assessment, students were required to read and respond to the items independently. Reading and mathematical problem solving involve multiple cognitive processes, including attention, working memory, prior knowledge, meaning-making, and inference (Akyol, 2013; Pardo, 2004; Snow, 2002). Consequently, differences in reading, comprehension, and response demands may have influenced how students accessed and expressed their mathematical reasoning. This possibility is especially relevant in a longitudinal study in which developmental change and measurement change occur simultaneously. The present findings therefore

provide a basis for further investigation of the relationship between literacy-related demands and number sense performance rather than evidence of a direct effect of literacy acquisition.

The finding that different dimensions of number sense may show different patterns across students also supports the view of number sense as a multidimensional construct. Number sense encompasses interrelated but distinguishable aspects, including understanding numerical magnitude, mental calculation and flexibility, reference-point use, and quantitative reasoning (Griffin, 2004; McIntosh et al., 1992). Previous longitudinal research has similarly suggested that different components of number sense may demonstrate different developmental patterns (Hojnoski et al., 2023). Taken together, these perspectives suggest that longitudinal assessment based solely on total scores may obscure meaningful changes within particular aspects of children's mathematical thinking. Examining strategy use at the item level may consequently provide a complementary perspective for understanding how different dimensions of number sense are maintained or transformed over time.

The present findings are broadly consistent with previous research indicating that students may rely substantially on rule-based or operation-oriented strategies when solving mathematical tasks. Palabıyık (2022), in a study covering preschool through fourth grade of primary school, reported frequent use of rule-based and operation-oriented strategies, while Can (2017) found relatively low number sense performance among primary school students alongside substantial use of operation-based solution methods. The present study extends this literature by providing an individual longitudinal perspective and showing that strategy trajectories are not necessarily uniform across children. Rather than positioning number-sense reasoning and procedural reasoning as mutually exclusive developmental stages, the findings suggest that children may maintain, adopt, or move between different strategy types depending on the task and assessment context. Nevertheless, because instructional experiences were not directly measured, these patterns cannot be interpreted as evidence that formal schooling causes a shift away from number-sense reasoning.

At the same time, the presence of positive changes in some students is important because it demonstrates that the transition to first grade was not uniformly associated with greater reliance on procedural strategies. Higher performance or more number-sense-oriented reasoning at the first-grade assessment may reflect developmental maturation, increased familiarity with mathematical tasks, mathematical learning experiences, individual differences, or changes in students' ability to independently interpret task statements. Reading skills may support students' ability to access and interpret written mathematical situations (Başaran, 2013; Fuchs et al., 2001), but the present study does not allow the contribution of reading development to be separated from these other factors. The contrasting trajectories therefore point toward a potentially complex relationship between literacy-related task demands and mathematical reasoning, one that warrants investigation through designs capable of measuring these processes simultaneously.

These findings have implications for early mathematics assessment and instruction. Rather than focusing exclusively on whether students obtain correct answers, assessment practices may benefit from examining how children arrive at their answers and how flexibly they use numerical relationships. Opportunities to explain, compare, justify, and revise different solution strategies may make children's number-sense reasoning more visible. Activities involving estimation, mental calculation, reference points, mathematical discussion, and flexible solution approaches may provide useful contexts for supporting the development of number sense. Previous research has emphasized the importance of early mathematical experiences and systematic support for number sense (Erdoğan & Baran, 2003; Jordan et al., 2010; NCTM, 2000; Şengül & Gülbağcı Dede, 2013). For the preschool-to-primary transition specifically, assessment procedures should also be sensitive to changes in literacy demands so that differences in children's mathematical performance are not automatically interpreted as changes in mathematical understanding. These implications should, however, be regarded as pedagogical considerations arising from the exploratory findings rather than as experimentally established effects.

Several limitations should be considered when interpreting the findings. First, the final sample consisted of only seven students, which substantially limits statistical power and the generalizability

of the findings. The longitudinal multi-case design nevertheless allowed detailed examination of individual trajectories that may be difficult to capture in larger group-level analyses. Second, the NSSS was originally developed and psychometrically evaluated for preschool children, and the present study did not provide separate validity or reliability evidence for its independent administration to first-grade students. Therefore, scores obtained at the two assessment occasions should not be interpreted as fully equivalent measures obtained under identical measurement conditions. The repeated administration was used primarily as an exploratory basis for examining patterns of response accuracy and strategy use across two developmental and assessment contexts. Third, the administration modality changed from researcher-administered oral presentation in preschool to independent reading and responding in first grade. Differences in literacy-related demands and response requirements may therefore have contributed to observed differences in performance. Finally, literacy development, reading comprehension, classroom mathematics instruction, teacher practices, and other learning experiences during the one-year period were not directly measured. The four students who did not participate in the follow-up assessment also represent a potential source of attrition bias.

These limitations point to several directions for future research. Studies with larger and more diverse samples should follow students across multiple assessment occasions and, where possible, examine mathematical performance under comparable and systematically varied administration conditions. Future research should directly assess reading proficiency and reading comprehension alongside number sense and document classroom mathematics practices and students' exposure to formal mathematical procedures. Such designs would make it possible to examine more precisely the relationships among literacy development, instructional experiences, assessment conditions, and number sense. In addition, think-aloud protocols, individual interviews, and systematic qualitative coding by multiple coders could provide richer evidence concerning the reasoning underlying students' responses. A particularly useful direction would be to combine longitudinal measurement with repeated assessment under both orally presented and independently read conditions. Such a design could help distinguish developmental changes in number sense from changes associated with the linguistic and cognitive demands of the assessment itself.

Overall, the findings suggest that changes in number sense across the transition from preschool to first grade are not necessarily uniform across students and cannot be adequately represented by total scores alone. Individual differences in strategy use, including the distinction between obtaining a correct answer and demonstrating flexible number-sense reasoning, provide an important dimension for understanding children's mathematical thinking. The findings further highlight the importance of considering assessment conditions when interpreting longitudinal changes in young children's mathematical performance. Given the small sample size, the differences in assessment modality, and the absence of direct measures of literacy and classroom instruction, the findings should be regarded as exploratory and hypothesis-generating rather than as evidence of causal effects. The main contribution of the study is therefore its longitudinal, individual-level description of continuity and change in number-sense strategy use across an important developmental and educational transition, while also drawing attention to the role of assessment conditions in how young children's mathematical reasoning becomes visible.

Declarations

Acknowledgements: Not applicable.

Authors' contributions: N.T. contributed to the conceptualization of the study, methodology, validation, data collection, formal analysis, research, and references, as well as writing, reviewing, and editing the manuscript. S.Ş. contributed to the conceptualization of the study, methodology, validation, and reviewing and editing of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Funding: The authors declare that this research received no external funding.

Ethics approval and consent to participate: The study received ethical approval from the Marmara University Institute of Educational Sciences Publication Ethics Committee, with approval number 651808 dated 01.11.2023.

Thesis and conference disclosure: The core dataset used in this study was derived from the first author's doctoral dissertation entitled *Çocuklar için felsefe uygulamalarının okul öncesi öğrencilerin sayı algısı becerileri üzerindeki etkisi* (Marmara University, 2025). This article extends that initial dataset as an independent longitudinal follow-up study that includes data collected at the end of first grade. This study has not been presented at any conference.

AI declaration: During the preparation of this article, the authors used generative AI tools solely for language editing, translation assistance, and spell checking to improve readability.

References

- Akyol, H. (2013). *Programa uygun Türkçe öğretim yöntemleri*. (6th ed.). Pegem Akademi.
- Başaran, M. (2013). Okuduğunu anlamanın bir göstergesi olarak akıcı okuma, *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 13 (4), 2277-2290.
- Beliakoff, A., Jordan, N., Klein, A., Devlin, B., & Huang, C.W. (2025). Stability of early number sense competencies for predicting mathematics difficulties. *Learning and Individual Differences*, 118, 102633. <https://doi.org/10.1016/j.lindif.2025.102633>
- Can, D. (2017). *İlkokul dördüncü sınıf öğrencilerinin sayı duyularının bağlam temelli ve bağlam temelli olmayan problem durumlarında incelenmesi*. [Examination of fourth grade elementary school students' number sense in context-based and noncontext-based problems] [Doctoral dissertation, Hacettepe University]
- Clements, D. H., Dong, Y., Day-Hess, C. A., & Sarama, J. (2024). Longitudinal investigation of early mathematical achievement and classroom strategic diversity: A replication and extension study. *Learning and Individual Differences*, 114, 102516. <https://doi.org/10.1016/j.lindif.2024.102516>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2009). *Research methods in education* (6th ed.). Routledge.
- Devlin, D., Moeller, K., & Sella, F. (2022). The structure of early numeracy: Evidence from multi-factorial models. *Trends in Neuroscience and Education*, 26, 1-9. <https://doi.org/10.1016/j.tine.2022.100171>
- Erdoğan, S., & Baran, G. (2003). Erken çocukluk döneminde matematik. *Eğitim ve Bilim*, 28(130), 32-40.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256. https://doi.org/10.1207/S1532799XSSR0503_3
- Gözüm, A.İ.C., Özberk, E.H., Kaya, Ü.Ü., & Uyanık Aktulun, Ö. (2024). Number sense across the transition from preschool to elementary school: A latent profile analysis. *Early Childhood Education Journal*, 52(6), 1221-1243. <https://doi.org/10.1007/s10643-023-01511-w>
- Griffin, S. (2004). Building number sense with number worlds: A mathematics program for young children. *Early Childhood Research Quarterly*, 19(1), 173-180. <https://doi.org/10.1016/j.ecresq.2004.01.012>

- Hickendorff, M. (2022). Flexibility and adaptivity in arithmetic strategy use: What children know and what they show. *Journal of Numerical Cognition*, 8(3), 367-381. <https://doi.org/10.5964/jnc.7277>
- Hojnoski, R. L., Polignano, J. C., & Caskie, G. I. (2023). Number sense development during the preschool years: Relations within and between key skill indicators. *Early Education and Development*, 34(7), 1631-1646. <https://doi.org/10.1080/10409289.2022.2120327>
- Howden, H. (1989). Teaching number sense. *Arithmetic Teacher*, 36(6), 6-11.
- Jordan, N. C., Devlin, B. L., & Botello, M. (2022). Core foundations of early mathematics: Refining the number sense framework. *Current Opinion in Behavioral Sciences*, 46, 101181. <https://doi.org/10.1016/j.cobeha.2022.101181>
- Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, 39(2), 181-195. <https://doi.org/10.1080/02796015.2010.12087772>
- McIntosh, A., Reys, B.J., & Reys, R.E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2-8. <https://flm-journal.org/Articles/94F594EF72C03412F1760031075F2.pdf>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
- Palabıyık, E. (2022). *Okul öncesi ve ilkokul öğrencilerinin sayı hissi becerilerinin incelenmesi*. [Examination of pre-school and primary school students' number sense skills] [Doctoral dissertation, Gazi University].
- Pardo, L. S. (2004). What every teacher needs to know about comprehension. *The reading teacher*, 58(3), 272-280. <https://doi.org/10.1598/RT.58.3.5>
- Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464-510). Wiley.
- Snow, C. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. RAND Education. https://www.rand.org/pubs/monograph_reports/MR1465.html
- Şengül, S., & Gülbağcı, H. (2013). 7. ve 8. sınıf öğrencilerin sayı hissi ile matematik öz yeterlikleri arasındaki ilişkinin incelenmesi. *The Journal of Academic Social Science Studies*, 6(4), 1049-1060.
- Torres-Peña, R. C., Peña-González, D., & Ariza-Echeverri, E. A. (2025). Mathematical thinking in preschool: Strengthening seriation and counting through problem solving. *International Journal of Early Childhood*, 57(2), 379-401. <https://doi.org/10.1007/s13158-024-00402-4>
- Tutkun, N. (2025). *Çocuklar için felsefe uygulamalarının okul öncesi öğrencilerinin sayı hissi becerilerine etkisi*. [The effect of philosophy for children practices on number sense skills of preschool students] [Doctoral dissertation, Marmara University].
- Yıldırım, A., & Şimşek, H. (2016). *Sosyal Bilimlerde nitel araştırma yöntemleri* (10th ed.). Seçkin Yayıncılık.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.