Al-Athfal:
Jurnal Pendidikan Anak

ISSN (p): 2477-4189; ISSN (e): 2477-4715
Vol 12 No 1 June 2026
Pages 1-18

Assessing Scientific Thinking in Early Childhood: Cross-Sectional Evidence for a Six-Dimensional Hierarchical Structure in Indonesian Preschoolers

Subhan[1]✉, Moh. Lalu Abid Zainul Puad[2], Anies Listyowati[3]
[1]Universitas Islam Negeri Palopo, Palopo, Indonesia
[2]Maurick College, Vught, Netherlands
[3]Universitas PGRI Adi Buana, Surabaya, Indonesia

Abstract

Purpose – Scientific thinking in early childhood remains understudied in non-Western contexts, with many existing models derived from Western samples and limited to two or three dimensions. This study examines a six-dimensional hierarchical framework proposing that domain-general cognitive capacities (Attention & Focus, Working Memory, Problem Solving) support domain-specific scientific competencies (Observation Skills, Prediction & Reasoning, Experimentation) in Indonesian preschoolers aged 4–6 years. Age-related patterns, gender differences, and institutional-type differences are also investigated.
Design/methods/approach – Using a quantitative cross-sectional design, data were collected from 105 children (4–6 years) enrolled in secular and Islamic early childhood education institutions in South Sulawesi, Indonesia. Scientific thinking was assessed using the Scientific Thinking Assessment for Early Childhood (STAEC), a 25-item teacher-rated instrument developed through expert review and pilot testing with 30 teachers. Analyses included descriptive statistics, reliability analysis, Pearson correlations, ANOVAs, and t-tests to evaluate interdimensional relationships and group differences.
Findings – Results provided initial cross-sectional evidence consistent with a six-dimensional hierarchical organization of early scientific thinking. Domain-general capacities were strongly intercorrelated ($r = .796–.831$) and showed higher mean scores than domain-specific competencies, suggesting a foundational role. Working memory displayed the strongest associations with advanced competencies, particularly prediction & reasoning and experimentation. A significant age-related difference emerged only for observation skills, whereas other dimensions showed non-significant developmental trends. No gender differences were observed across any dimension, and no differences emerged across secular and Islamic institution types.
Research implications/limitations – The cross-sectional design limits developmental and causal inferences. Teacher ratings may introduce rater bias and do not capture moment-to-moment reasoning processes. The single-region sample constrains generalizability; future research should use longitudinal, larger, multi-region, and multi-method designs.
Practical implications – Early childhood programs should strengthen foundational cognitive capacities while providing explicit, developmentally appropriate support for prediction and experimentation, and maintain equal learning expectations across genders and educational settings.
Originality/value – This study offers initial empirical support for a multidimensional hierarchical model of early scientific thinking in a non-Western context, including secular and Islamic early childhood education settings.

Keywords Scientific thinking, Early childhood, Cognitive architecture, Preschool education, Scientific reasoning

Paper type Research paper

✉ Corresponding author:
Email Address: subhan@uinpalopo.ac.id.
Received: 14 December 2025; Revised: 8 February 2026; Accepted: 10 February 2026

DOI: http://dx.doi.org./10.14421/al-athfal.2026.121-01

1. Introduction

Scientific thinking represents a fundamental cognitive capability enabling children to explore, understand, and make sense of their world through systematic observation, questioning, and reasoning (Kuhn, 2010). Developing scientific thinking during early childhood establishes foundational competencies supporting later academic achievement and problem-solving abilities (Nayfeld et al., 2011). Despite widespread recognition of its importance, fundamental questions about scientific thinking's cognitive architecture remain unresolved: Does it comprise a unitary construct or multiple distinct dimensions? If multidimensional, how do these dimensions relate to one another? Understanding this structure during the critical preschool period has profound implications for both developmental theory and educational practice, yet empirical evidence remains limited.

Contemporary theoretical frameworks propose that scientific thinking integrates domain-general cognitive capacities with domain-specific scientific competencies, with executive functions shown to predict children's conceptual learning in science (Schäfer et al., 2024), suggesting a hierarchical architecture wherein foundational executive functions scaffold the acquisition of specialized practices. This hierarchical perspective resonates with organizational and management theories of dynamic capabilities, which emphasize how foundational capacities enable higher-order competencies within complex systems (Teece, 2007; Helfat & Peteraf, 2015). Developmental research indicates that executive functions, including attention regulation, working memory, and cognitive flexibility, develop substantially during early childhood and provide a cognitive foundation for complex learning processes, including scientific reasoning (Best & Miller, 2010; Garon et al., 2008; Diamond, 2013; Miyake & Friedman, 2012). Gomez (2025) argued that scientific thinking encompasses both general reasoning abilities and specialized practices including hypothesis generation, experimental design, and evidence evaluation. Yangüez (2025) demonstrated that scientific reasoning recruits executive functions including working memory, cognitive flexibility, and inhibitory control, supporting hierarchical organization. However, these frameworks largely derive from research with older children and adults, leaving their applicability to early childhood uncertain. Tzuriel et al. (2024) documented substantial developmental changes in cognitive flexibility and hypothesis search across the preschool to adolescent transition, suggesting that scientific thinking's structure may differ across developmental periods.

Existing research on early scientific thinking has predominantly examined individual competencies in isolation rather than testing comprehensive structural models. Studies document that preschoolers demonstrate emerging capabilities in observation, classification, prediction, and simple experimentation. Samarapungavan et al. (2011) and Reith (2024) found that evidence evaluation, experimentation skills, and hypothesis generation develop across early and middle childhood, while Delserieys and Kampeza (2025) demonstrated that kindergartners can learn controlled experimentation strategies. Öztürk (2025) provided the most comprehensive assessment to date, documenting scientific thinking development through a multidimensional inventory, yet their work focused on ages 6–12 rather than the preschool period. This leaves critical gaps in understanding whether the multidimensional structure observed in older children already characterizes early childhood and how dimensions relate hierarchically.

The predominance of Western, Educated, Industrialized, Rich, and Democratic (WEIRD) samples in developmental research limits understanding of scientific thinking's universality versus cultural specificity (Nielsen et al., 2017; Henrich et al., 2010). This sampling bias parallels concerns in management and organizational research regarding the cross-cultural generalizability of Western-derived theoretical frameworks (Hofstede, 2011; House et al., 2004). Goddu and Gopnik (2024) demonstrated that causal learning, a core component of scientific thinking operates through universal cognitive mechanisms yet is shaped by cultural practices and values. This context matters theoretically, not merely empirically, because testing the proposed hierarchical model in a non-Western setting allows examination of whether the architecture of scientific thinking is a universal feature of cognitive development or is culturally contingent, thereby establishing the boundary conditions of existing theoretical models. Southeast Asian contexts, and Indonesia specifically, remain understudied despite representing substantially different cultural values and educational philosophies. Indonesia's dual system of secular and Islamic early childhood education provides particularly valuable opportunities for examining whether scientific thinking capabilities develop similarly across different educational philosophies (Ainnin & Ismail, 2024).

Gender equity in early scientific thinking represents another critical gap. Substantial evidence documents gender disparities emerging during elementary school (Miller et al., 2020; Wang & Degol, 2017). Suggesting cultural rather than biological origins, Zuo and Tang (2024) confirmed through meta-analysis that gender differences in early STEM learning are minimal and context-dependent. Xu (2025) found that gender gaps in STEM vary substantially across nations, suggesting cultural rather than biological origins. However, whether these disparities exist from the outset when competencies first emerge, or develop later through differential socialization and stereotype exposure, remains unclear. If capabilities are equivalent during early childhood, this would support sociocultural explanations while identifying the preschool period as the critical intervention window (Hyde, 2014).

The current study addresses these gaps by testing a six-dimensional hierarchical model of scientific thinking in Indonesian preschool children ages 4–6 years. The model proposes that scientific thinking comprises three domain-general cognitive capacities (Attention & Focus, Working Memory, Problem Solving) and three domain-specific scientific competencies (Observation Skills, Prediction & Reasoning, Experimentation), organized hierarchically such that foundational cognitive capacities scaffold acquisition of specialized scientific practices. This hierarchical conceptualization draws on both cognitive-developmental theory and organizational capability frameworks emphasizing how lower-order capacities enable higher-order performance (Teece, 2007; Senge, 2006). Four primary objectives guide the investigation. First, we test whether the six-dimensional structure characterizes scientific thinking during early childhood. Second, we investigate hierarchical organization by examining correlational patterns and mean-level differences. Third, we examine developmental patterns across ages 4–6 years and gender differences across all dimensions. Fourth, we provide the first large-scale evidence from an Indonesian context, testing whether the proposed structure generalizes beyond Western samples while examining scientific thinking development within Islamic early childhood education programs.

The Indonesian context offers unique contributions for understanding scientific thinking development across diverse educational and cultural settings. Indonesia represents the world's fourth most populous nation and largest Muslim-majority country, with early childhood education delivered through three primary systems: TK (Taman Kanak-kanak, secular public kindergarten), RA (Raudhatul Athfal, Islamic kindergarten under Ministry of Religious Affairs), and TKIT (Taman Kanak-kanak Islam Terpadu, integrated Islamic kindergarten combining religious and academic curricula). Validating the hierarchical model in Indonesian contexts including Islamic education settings addresses persistent Western-centrism in developmental science while providing evidence about whether foundational scientific thinking capabilities emerge through similar processes across diverse cultural-educational contexts (Gopnik, 2012).

This investigation advances scientific thinking research through providing the first comprehensive test of a multidimensional hierarchical model during the critical preschool period, examining developmental and gender patterns across dimensions, and validating the model in a non-Western context. By assessing scientific thinking through naturalistic teacher observation across diverse early childhood settings, the study addresses methodological limitations of brief direct assessment while capturing children's capabilities as they emerge in everyday educational contexts. Practically, the study provides educators and policymakers with evidence-based guidance for supporting scientific thinking development during the period when foundational capabilities are established, potentially preventing rather than remediating disparities that emerge later.

2. Methods

2.1. Research Design

This study employed a quantitative, cross-sectional survey design positioned as a psychometric validation and structural model-testing investigation, aiming to provide initial empirical evidence for the proposed six-dimensional hierarchical framework of scientific thinking during the preschool period. Cross-sectional designs enable efficient assessment of age-related patterns and individual differences while providing data suitable for psychometric validation (Kline, 2023; Creswell & Creswell, 2018). However, it is acknowledged that cross-sectional comparisons cannot substitute for longitudinal evidence regarding individual developmental trajectories or causal relationships among dimensions. The study was conducted during July–October 2025 in early childhood education settings across South Sulawesi, Indonesia. Correlational analyses, ANOVAs, and t-tests were selected as analytical techniques specifically aligned with the objective of testing the proposed hierarchical structure through patterns of interrelationships, mean-level differences, and group comparisons across the six dimensions.

2.2. Population and Sample

The target population comprised preschool children ages 4–6 years enrolled in early childhood education programs in South Sulawesi. South Sulawesi was selected as the study site due to its well-established networks of both secular (TK) and Islamic (RA, TKIT) early childhood education institutions, its cultural and demographic diversity representative of eastern Indonesia, and logistical accessibility enabling systematic comparison across educational philosophies. Participants were 105 children (54 males, 51 females) ages 4–6 years (M = 5.04, SD = 0.77) recruited through purposive cluster sampling from 12 institutions representing diverse educational settings: secular kindergartens (Taman Kanak-kanak, n = 38), Islamic kindergartens (Raudhatul Athfal, n = 41), and integrated Islamic kindergartens (TKIT, n = 26). Age distribution was: 29 four-year-olds (27.6%), 43 five-year-olds (41.0%), and 33 six-year-olds (31.4%). Sample size was determined through G*Power analysis (Faul et al., 2007) ensuring adequate power (.80, α = .05) for planned correlation analyses, t-tests, and ANOVAs, based on an assumed medium effect size (f = .25 for ANOVAs; r = .30 for correlations), consistent with conventions in early childhood developmental research (Cohen, 1988). Inclusion criteria required: (a) age 4.0–6.9 years, (b) regular enrollment, (c) teacher familiarity (≥3 months observation), and (d) parental consent. Children with identified developmental disabilities were excluded. Participating teachers (N = 28) had observed target children for an average of 7.3 months (SD = 3.5), with 76.2% having ≥6 months familiarity. To mitigate potential rater variability, all teachers received standardized written instructions detailing behavioral anchors for each rating point, and were instructed to base ratings on typical observed classroom behavior over extended periods rather than single observations. Participant characteristics are summarized in Table 1.

2.3. Data Collection Techniques and Instrument Development

Instrument Specifications. The Scientific Thinking Assessment for Early Childhood (STAEC) is a 25-item teacher-rating instrument assessing six theoretically derived dimensions organized hierarchically into domain-general cognitive capacities and domain-specific scientific competencies.

Domain-general cognitive capacities (15 items) include: (1) Attention & Focus (5 items; α = .750; e.g., "Maintains focus during extended observation activities"), (2) Working Memory (5 items; α = .865; e.g., "Remembers and integrates information during problem-solving"), and (3) Problem Solving (5 items; α = .875; e.g., "Develops and tests different strategies when encountering problems").

Domain-specific scientific competencies (10 items) include: (4) Observation Skills (3 items; α = .865; e.g., "Notices and describes details others might overlook"), (5) Prediction & Reasoning (3 items; α = .902; e.g., "Makes logical predictions based on prior observations"), and (6) Experimentation (4 items; α = .832; e.g., "Tries different approaches to see different results," "Explains observation results," "Revises opinions after seeing different outcomes"). The

Experimentation dimension integrates behavioral experimentation with metacognitive explanation, reflecting contemporary frameworks emphasizing that scientific thinking encompasses both empirical testing and reflective reasoning (Gomez, 2025). This integration is developmentally appropriate for early childhood when experimentation and explanation are typically intertwined rather than differentiated processes.

Table 1. Sample Characteristics

Characteristic	n	%
Gender		
Male	54	51.4
Female	51	48.6
Age Group		
4 years	29	27.6
5 years	43	41.0
6 years	33	31.4
Type of Institution		
TK (Secular Kindergarten)	38	36.2
RA (Islamic Kindergarten)	41	39.0
TKIT (Integrated Islamic Kindergarten)	26	24.8
Teacher Familiarity		
3-6 months	25	23.8
≥6 months	80	76.2

Note. N = 105. TK = Taman Kanak-kanak (secular public kindergarten); RA = Raudhatul Athfal (Islamic kindergarten under Ministry of Religious Affairs); TKIT = Taman Kanak-kanak Islam Terpadu (integrated Islamic kindergarten combining religious and academic curricula). Participating teachers (N = 28) had observed target children for an average of 7.3 months (SD = 3.5).

The overall scale achieved excellent reliability (α = .954). Content validity was supported through expert judgment, and item wording was adapted to ensure cultural appropriateness for Indonesian preschool contexts. Inter-item correlations within each dimension ranged from .35 to .72, and corrected item–total correlations ranged from .48 to .81, indicating that all items contributed meaningfully to their respective dimensions without redundancy. The structure, item distribution, and internal consistency of the STAEC are summarized in Table 2.

Table 2. Structure and Reliability of the Scientific Thinking Assessment for Early Childhood (STAEC)

Dimension	Domain	Items	Example Indicator	α
Attention & Focus	Domain-general	5	Sustains attention during observation activities; resists distractions	.750
Working Memory	Domain-general	5	Follows sequential instructions; retains information to complete tasks	.865
Problem Solving	Domain-general	5	Develops alternative strategies when encountering difficulties	.875
Observation Skills	Domain-specific	3	Observes objects carefully; notices changes during activities	.865
Prediction & Reasoning	Domain-specific	3	Makes logical predictions; provides reasons for predictions	.902
Experimentation	Domain-specific	4	Tests different approaches; explains and revises based on evidence	.832
Overall Scale	—	25	—	.954

Note. Teachers rated children on a 5-point Likert scale (1 = Almost Never to 5 = Almost Always). The Experimentation dimension integrates active experimentation (trying different approaches, conducting simple experiments) and metacognitive explanation (explaining observations, revising predictions based on evidence), consistent with frameworks emphasizing that scientific thinking encompasses both empirical testing and reflective reasoning (Gomez, 2025). One Attention & Focus item was reverse-coded. All dimensions demonstrate acceptable to excellent reliability (α ≥ .75).

Material Specifications. Assessment materials consisted of: (a) the STAEC rating form administered via a secure Google Forms platform (25 items on 5-point Likert scales: 1 = Almost

Never to 5 = Almost Always), (b) demographic information forms requesting child age, gender, and enrollment duration, and (c) detailed administration instructions emphasizing ratings based on typical observed classroom behavior. The STAEC was developed following established scale development procedures for survey-based measurement instruments (DeVellis & Thorpe, 2021; Hinkin, 2023). Initial item generation produced 45 candidate items reviewed by five experts (two developmental psychologists, two early childhood educators, one science education researcher) for content validity, developmental appropriateness, and cultural relevance. Twenty items were eliminated based on expert feedback. The remaining 25 items underwent pilot testing with 30 teachers rating 60 children not included in the final sample. One item was reverse-coded. This reverse-coded item was included to reduce acquiescence bias and improve response quality in teacher ratings (Weijters & Baumgartner, 2012). Dimensional scores were computed as means of constituent items (range 1–5); overall scientific thinking scores were means across all 25 items.

Data Collection Procedure. Following institutional approval and parental consent, teachers received assessment packets with detailed instructions. Teachers completed assessments over a two-week period during non-instructional time, rating children based on typical observed behavior across regular classroom contexts. Completed assessments were submitted electronically via a secure Google Forms platform, with access restricted to the research team to ensure confidentiality and data integrity.

2.4. Data Analysis Techniques

Data analysis followed an iterative thematic approach, informed by interpretative phenomenological principles and structured coding procedures, and unfolded in three phases. First, transcripts and field notes were repeatedly read to develop a deep familiarity with the data, and initial codes were generated inductively by identifying recurring phrases, practices, concepts, and themes, without imposing predetermined categories. Second, relationships among initial codes were examined through axial coding, in which related codes were grouped thematically and constant comparison across data sources was conducted to identify patterns and variations. Third, core themes were identified that connected categories into a coherent narrative addressing the research questions, with three primary themes emerging: mediation strategies for digital technology, technology presence within households, and observed behavioural changes in children. These themes were refined through iterative cycles of coding and interpretation. NVivo 12 was used to organise data, manage codes, and track code development across the dataset; manual coding was performed first to maintain close engagement with the data, with NVivo serving as an organisational tool rather than for automated analysis. A reflexive journal was maintained throughout the analysis to document interpretive decisions, emerging insights, and potential biases, with regular entries reflecting on the researcher's positionality as a non-pesantren-affiliated outsider and how this might influence interpretations of religious practices.

Data analysis employed SPSS Version 28.0 with significance level $\alpha = .05$. Preliminary analyses examined data quality through missing data analysis (Little's MCAR test), distributional properties assessment (skewness and kurtosis within ±2.0 acceptable range), and outlier detection (Mahalanobis distance, $p < .001$). Given minimal missing data (<2%), listwise deletion was employed for each analysis. Procedures for handling and evaluating missing data followed established methodological recommendations for applied research (Enders, 2022). Descriptive statistics (means, standard deviations, ranges, skewness, kurtosis) characterized sample performance (Hair et al., 2019). Internal consistency reliability was evaluated using Cronbach's alpha with benchmarks: $\alpha \geq .70$ acceptable, $\alpha \geq .80$ good, $\alpha \geq .90$ excellent (Nunnally & Bernstein, 1994). Structural validity was examined through Pearson correlations among the six dimensions. Correlational analysis was employed as an initial exploratory approach appropriate for early childhood research where sample sizes typically preclude confirmatory techniques, providing preliminary evidence for the proposed hierarchical structure through patterns of interrelationships among dimensions (Tabachnick & Fidell, 2019). Developmental patterns were tested using one-way ANOVAs comparing three age groups (4, 5, 6 years) with effect sizes reported as partial eta-squared (η^2p: .01 small, .06 medium, .14 large), and independent samples t-tests comparing males and females with Cohen's d effect sizes (.20 small, .50 medium, .80 large).

Two-way ANOVAs tested Age × Gender interactions. Levene's test verified homogeneity of variance assumptions. Confirmatory factor analysis was not conducted due to sample size limitations; the recommended minimum of 200 cases for CFA (Kline, 2023; Creswell & Creswell, 2018) exceeds the current sample, and the exploratory nature of validating a newly proposed structure in early childhood further supports the use of correlational and comparative approaches as an appropriate initial step. The proposed hierarchical structure of scientific thinking guiding the analyses is summarized in Table 3. It should be noted that this framework is theoretically proposed and empirically examined in the present study through patterns of intercorrelations and mean-level differences among dimensions, rather than through formal structural equation modeling. Level 1 represents the foundational domain-general cognitive capacities theorized to scaffold the development of Level 2 domain-specific scientific competencies.

Table 3. Conceptual Framework of the Six-Dimensional Model

Level	Dimension Type	Dimension	Functional Description
Level 2	Domain-specific scientific competencies	Experimentation	Systematically tests ideas through manipulation of variables.
Level 2	Domain-specific scientific competencies	Prediction & Reasoning	Generates explanations and predictions based on observed evidence.
Level 2	Domain-specific scientific competencies	Observation Skills	Identifies and describes relevant features of phenomena.
Level 1	Domain-general cognitive capacities	Problem Solving	Develops strategies to address challenges and reach solutions.
Level 1	Domain-general cognitive capacities	Working Memory	Maintains and integrates information during cognitive tasks.
Level 1	Domain-general cognitive capacities	Attention & Focus	Sustains concentration during observation and inquiry activities.

3. Result

3.1. Preliminary Analyses

Initial data screening revealed minimal missing data (1.8% of total data points), distributed randomly across items and participants (Little’s MCAR test: $\chi^2 = 142.35$, df = 156, p = .762). One participant with >20% missing data was excluded, yielding N = 105 for analyses. All variables showed acceptable distributional properties (skewness range: −0.84 to 0.92; kurtosis range: −0.65 to 1.12), well within the ±2.0 criteria for approximate normality (Kim, 2013). No multivariate outliers were detected (all Mahalanobis distances p > .001).

3.2. Descriptive Statistics and Reliability

Table 4 presents descriptive statistics for all six dimensions and overall scientific thinking.

Table 4. Descriptive Statistics and Internal Consistency Reliability for Six Dimensions and Overall Scientific Thinking (N = 105)

Dimension	Items	M	SD	Min	Max	Skew	Kurt	α
Attention & Focus	5	3.45	0.86	1.20	5.00	-0.12	0.34	.750
Working Memory	5	3.78	0.81	1.40	5.00	-0.32	0.28	.865
Problem Solving	5	3.63	0.84	1.60	5.00	-0.22	0.15	.875
Observation Skills	3	3.63	0.92	1.33	5.00	-0.28	0.42	.865
Prediction & Reasoning	3	3.10	0.95	1.00	5.00	0.18	-0.45	.902
Experimentation	4	3.48	0.88	1.25	5.00	-0.14	0.22	.832
Overall Scientific Thinking	25	3.51	0.73	1.56	4.96	-0.08	0.35	.954

Mean scores ranged from 3.10 (Prediction & Reasoning) to 3.78 (Working Memory), indicating capabilities moderately above the scale midpoint (2.5). The overall scientific thinking mean was 3.51 (SD = 0.73), with scores ranging from 1.56 to 4.96, demonstrating substantial individual variation. As illustrated in Figure 1, domain-general cognitive capacities tended to show higher mean scores than domain-specific scientific competencies, suggesting that foundational capacities may be more firmly established during the preschool period. Figure 2 shows overall scores approximating a normal distribution with substantial individual variability.

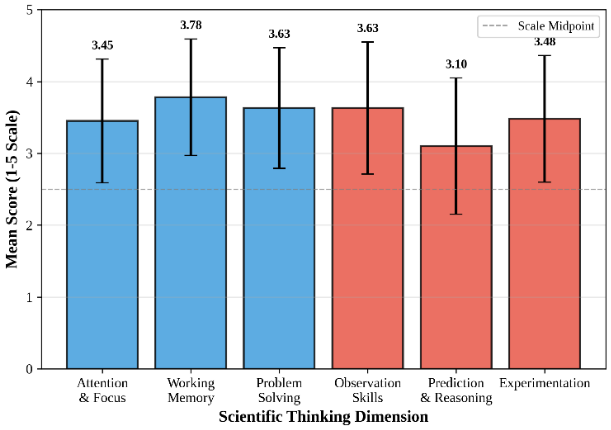

Figure 1. Mean Scores Across Six Scientific Thinking Dimensions (N = 105)

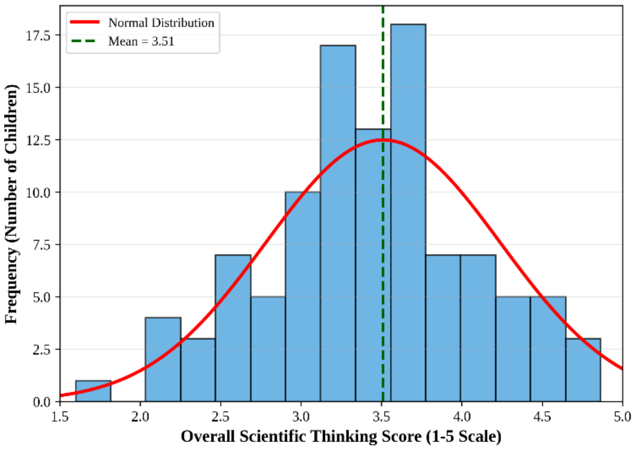

Figure 2. Distribution of Overall Scientific Thinking Scores (N = 105)

Internal consistency reliability was excellent. The overall STAEC achieved α = .954, with dimensional reliabilities ranging from .750 (Attention & Focus) to .902 (Prediction & Reasoning). All dimensions exceeded the .70 acceptable threshold, with five of six achieving the .80 benchmark (Figure 3).

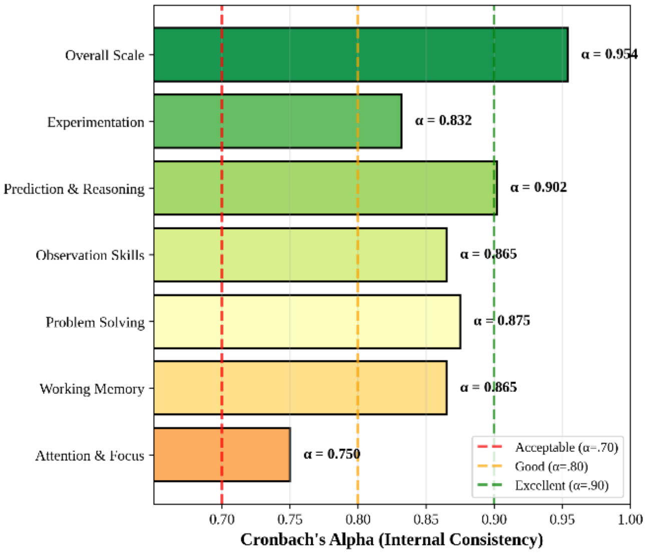

Figure 3. Internal Consistency Reliability (Cronbach's Alpha) by Dimension (N = 105)

3.3. Hierarchical Structure: Dimension Intercorrelations

Intercorrelations among the six dimensions were examined using Pearson correlation coefficients as an initial exploratory approach to evaluate the proposed hierarchical structure. Given the sample size (N = 105), correlational analysis provides preliminary evidence for hierarchical organization through patterns of interrelationships, rather than serving as a substitute for formal structural modeling such as confirmatory factor analysis or structural equation modeling. Table 5 presents the correlations among all dimensions.

Table 5. Pearson Correlations Among Six Scientific Thinking Dimensions (N = 105)

Dimension	1	2	3	4	5	6
1. Attention & Focus	—					
2. Working Memory	.796	—				
3. Problem Solving	.812	.831	—			
4. Observation Skills	.682	.738	.721	—		
5. Prediction & Reasoning	.581	.798	.672	.692	—	
6. Experimentation	.647	.773	.711	.728	.748	—

Note. N = 105. All correlations are significant at p < .001.

All correlations were positive and significant (ps < .001), ranging from r = .581 to r = .831. Domain-general capacities showed strong intercorrelations (rs = .796–.831), while correlations between domain-general and domain-specific dimensions were moderate to strong (mean rs = .647–.738), consistent with hierarchical organization. Working Memory emerged as the strongest correlate of advanced competencies (r = .798 with Prediction & Reasoning; r = .773 with Experimentation). Although several interdimensional correlations exceed .80, the six dimensions are retained as distinct constructs based on their differentiated developmental functions and theoretical foundations; these high correlations reflect developmental interrelatedness during early childhood rather than construct redundancy (Figure 4).

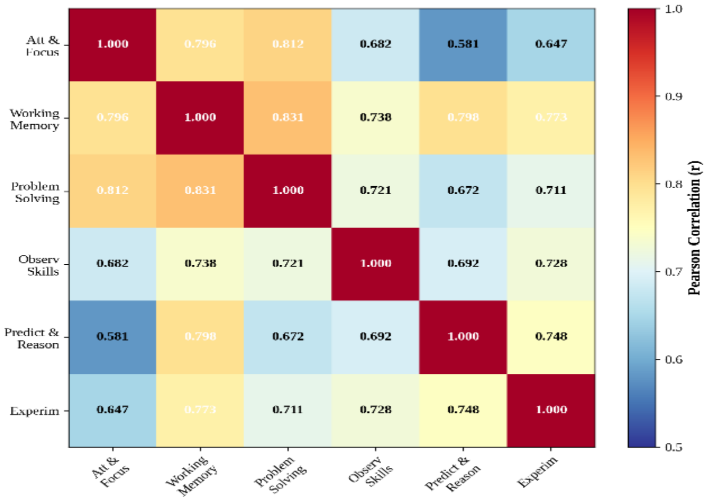

Figure 4. Correlation Matrix Heatmap for Six Scientific Thinking Dimensions (N = 105)

Mean-level analyses revealed a hierarchical pattern. Domain-general capacities (M = 3.62, SD = 0.69) significantly outperformed domain-specific competencies (M = 3.35, SD = 0.78), t(104) = 4.82, p < .001, d = 0.37. Within domain-specific skills, Observation (M = 3.63) exceeded both Prediction & Reasoning (M = 3.10) and Experimentation (M = 3.48), suggesting a developmental progression from information gathering to hypothesis generation and testing.

3.4. Age-Related Patterns

Table 6 presents mean scores by age group with ANOVA results. A significant age effect emerged for Observation Skills, F(2, 102) = 4.15, p = .018, η^2p = .075, with 6-year-olds significantly outperforming 4-year-olds (p = .015, d = 0.64). The remaining dimensions showed consistent positive trends across age groups without reaching statistical significance (all ps > .34, η^2ps =

.007–.021). In the context of early childhood research, these non-significant trends accompanied by small effect sizes remain developmentally informative, as substantial within-group variability during the preschool period can obscure meaningful age-related patterns. It should be noted that the cross-sectional design limits inferences to between-group comparisons and cannot establish intra-individual developmental trajectories. Figure 5 displays developmental patterns across age groups.

Table 6. Mean Scores by Age Group with One-Way ANOVA Results (N = 105)

Dimension	Age 4 years (n = 29) M (SD)	Age 5 years (n = 43) M (SD)	Age 6 years (n = 33) M (SD)	F (2,102)	p	η^2p
Attention & Focus	3.31 (0.88)	3.45 (0.85)	3.58 (0.87)	1.08	.343	.021
Working Memory	3.68 (0.84)	3.78 (0.79)	3.88 (0.82)	0.58	.561	.011
Problem Solving	3.51 (0.66)	3.55 (0.83)	3.67 (0.85)	0.34	.711	.007
Observation Skills	3.35^{a} (0.95)	3.64^{b} (0.91)	3.89^{b} (0.88)	4.15	.018*	.075
Prediction & Reasoning	2.98 (0.98)	3.08 (0.94)	3.25 (0.93)	0.98	.379	.019
Experimentation	3.42 (0.86)	3.41 (0.86)	3.61 (0.74)	0.65	.523	.012
Overall Scientific Thinking	3.38 (0.75)	3.51 (0.72)	3.68 (0.71)	1.98	.143	.037

*Note. N = 105. M = mean; SD = standard deviation; η^2p = partial eta-squared effect size. Within rows, means sharing subscript letters do not differ significantly at p < .05 (Tukey HSD post-hoc test). * p < .05.*

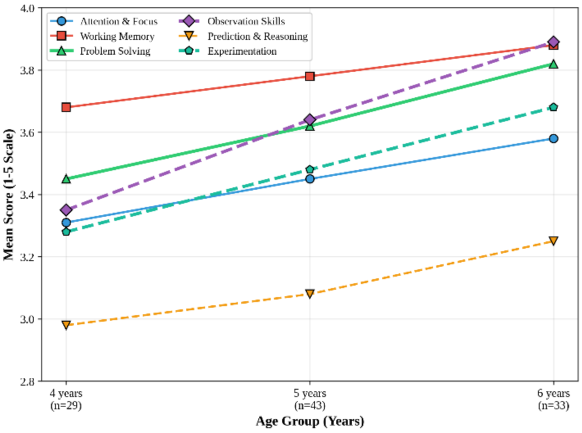

Figure 5. Mean Scores by Age Group Across Six Scientific Thinking Dimensions (N = 105)

Overall scientific thinking scores showed a positive but non-significant trend (F(2, 102) = 1.98, p = .143, η^2p = .037), suggesting that developmental changes during the preschool years are dimension-specific rather than uniform (Figure 6).

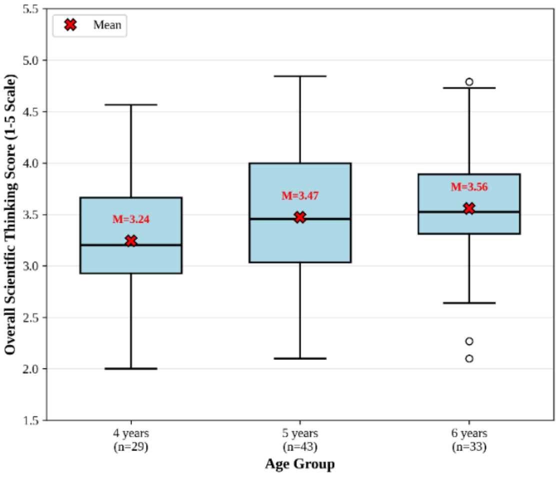

Figure 6. Distribution of Overall Scientific Thinking Scores by Age Group (N = 105)

3.5. Gender Comparisons

Table 7 presents gender comparisons across all six dimensions and overall scientific thinking. No significant gender differences emerged on any dimension or overall scientific thinking (all ps > .41, all |ds| < .16). As shown in Table 7 and Figure 7, males and females demonstrated equivalent capabilities across all six dimensions. Two-way ANOVAs testing Age × Gender interactions revealed no significant effects (all ps > .28).

Table 7. Gender Comparisons with Independent Samples t-test Results (N = 105)

Dimension	Males (n = 54) M (SD)	Females (n = 51) M (SD)	t (103)	p	d
Attention & Focus	3.42 (0.88)	3.49 (0.85)	-0.41	.682	0.08
Working Memory	3.77 (0.82)	3.79 (0.81)	-0.12	.906	0.02
Problem Solving	3.57 (0.86)	3.69 (0.83)	-0.73	.468	0.14
Observation Skills	3.60 (0.94)	3.66 (0.91)	-0.33	.741	0.07
Prediction & Reasoning	3.03 (0.97)	3.18 (0.93)	-0.82	.415	0.16
Experimentation	3.46 (0.90)	3.50 (0.87)	-0.23	.818	0.05
Overall	3.50 (0.71)	3.52 (0.76)	-0.14	.889	0.03

Note. N = 105 (54 males, 51 females). M = mean; SD = standard deviation; d = Cohen's d effect size. No statistically significant gender differences were observed (all ps > .05).

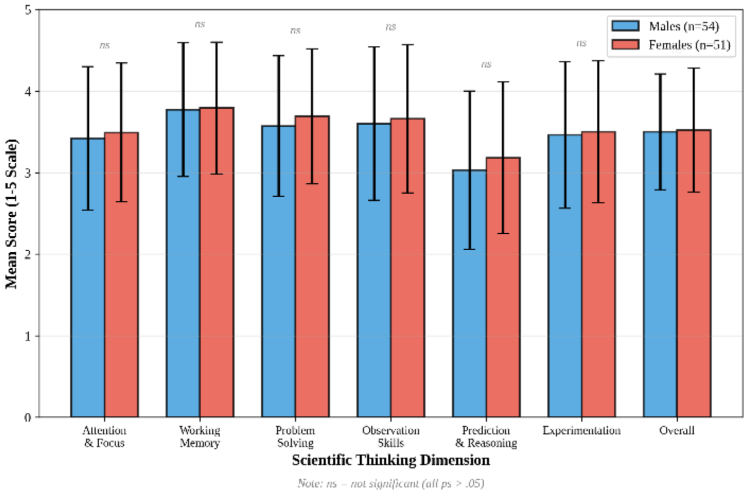

Figure 7. Gender Comparisons Across Six Scientific Thinking Dimensions (N = 105)

3.5. Institutional Type Comparisons

Table 8 presents comparisons across educational settings. No significant differences emerged across institution types for any dimension or overall scientific thinking (all Fs < 0.45, all ps > .64, all η^2ps < .009), indicating that scientific thinking capabilities develop equivalently across secular and Islamic early childhood education settings.

Table 8. Mean Scores by Institution Type with One-Way ANOVA Results (N = 105)

Dimension	TK (n = 38) M (SD)	RA (n = 41) M (SD)	TKIT (n = 26) M (SD)	F (2,102)	p	η^2p
Attention & Focus	3.48 (0.84)	3.42 (0.89)	3.47 (0.87)	0.08	.925	.001
Working Memory	3.82 (0.78)	3.75 (0.83)	3.77 (0.84)	0.11	.899	.002
Problem Solving	3.67 (0.81)	3.61 (0.86)	3.60 (0.87)	0.09	.916	.002
Observation Skills	3.69 (0.89)	3.60 (0.93)	3.58 (0.96)	0.19	.827	.004
Prediction & Reasoning	3.15 (0.92)	3.08 (0.96)	3.05 (0.99)	0.14	.873	.003
Experimentation	3.52 (0.85)	3.46 (0.90)	3.44 (0.91)	0.11	.900	.002
Overall Scientific Thinking	3.56 (0.70)	3.49 (0.75)	3.47 (0.76)	0.19	.830	.004

Note. N = 105. TK = Taman Kanak-kanak (secular kindergarten); RA = Raudhatul Athfal (Islamic kindergarten); TKIT = Taman Kanak-kanak Islam Terpadu (integrated Islamic kindergarten). M = mean; SD = standard deviation; η^2p = partial eta-squared effect size. No significant differences emerged across institution types (all ps > .82).

4. Discussion

This study provides the first comprehensive evidence for a six-dimensional hierarchical structure of scientific thinking in early childhood. Testing the model with 105 Indonesian preschool children aged 4–6 years yielded compelling support for the proposed architecture integrating domain-general cognitive capacities (Attention & Focus, Working Memory, Problem Solving) with domain-specific scientific competencies (Observation Skills, Prediction & Reasoning, Experimentation). The six dimensions demonstrated excellent reliability and meaningful intercorrelations, supporting scientific thinking as a coherent multidimensional construct. Hierarchical organization received robust support through correlational patterns, mean-level differences, and dimensional progression. Complete gender equity emerged across all dimensions, and the structure generalized across secular and Islamic educational settings.

The hierarchical cognitive architecture received compelling empirical support across multiple complementary analyses. Stronger correlations among domain-general capacities (rs = .796–.831) than between these capacities and domain-specific competencies (mean rs = .647–.738) provide clear evidence for hierarchical organization, supporting Gomez's (2025) framework that foundational cognitive processes scaffold specialized scientific practices. It should be noted that this support constitutes preliminary empirical evidence for the proposed hierarchical structure; formal confirmation through latent variable modeling awaits future research with larger samples. While these high intercorrelations might raise questions about construct redundancy, the six dimensions are theoretically differentiated based on distinct developmental functions and cognitive mechanisms rather than purely statistical criteria, consistent with developmental models in which capacities are functionally interconnected yet conceptually distinct (Miyake & Friedman, 2012). Working memory emerged as especially central, showing the strongest correlations with advanced reasoning (r = .798 with Prediction & Reasoning; r = .773 with Experimentation), consistent with developmental models emphasizing its function in integrating information and supporting complex reasoning (Cowan, 2014; Blair & Razza, 2007) and with Tzuriel et al. (2024) evidence that working memory capacity constrains hypothesis search and cognitive flexibility. Mean-level analyses further substantiated hierarchical structure, with domain-general capacities significantly outperforming domain-specific competencies (M = 3.62 vs. 3.35, d = 0.37), suggesting foundational skills develop earlier or more robustly. Within domain-specific dimensions, the progression from Observation (M = 3.63) to Prediction & Reasoning (M = 3.10) to Experimentation (M = 3.48) aligns with frameworks wherein scientific thinking advances from descriptive observation toward hypothesis generation and empirical testing (Klahr & Dunbar, 1988). The substantial gap between observation and prediction capabilities (d = 0.68) indicates that generating testable predictions represents a more challenging developmental achievement. Together, these findings provide converging evidence for hierarchical architecture during early childhood, with domain-general capacities providing cognitive foundations enabling domain-specific competencies (Zimmerman, 2007).

Developmental analyses revealed a significant age effect for Observation Skills (η^2p = .075), with 6-year-olds outperforming 4-year-olds (d = 0.64). Problem Solving and Experimentation showed positive but non-significant trends (η^2ps = .007–.012), with substantial within-group variability (SDs = 0.66–0.86) suggesting that experiential factors play important roles alongside maturation. These findings align with Reith (2024) and Tzuriel et al. (2024), while intervention research demonstrating that kindergartners can learn experimentation through targeted instruction (Van der Graaf et al., 2020; García-Rodeja, 2024) suggests that the observed individual variation reflects differential learning opportunities rather than solely maturational constraints. This interpretation indicates that appropriate educational experiences during the preschool period could meaningfully advance scientific thinking capabilities.

Complete gender equity across all six dimensions (all ps > .41, all |ds| < .16) represents a critical finding for understanding STEM gender disparities' developmental origins. Overall scientific thinking scores were virtually identical between boys and girls (3.50 vs. 3.52, d = 0.03). This pattern contradicts biological explanations for STEM gender gaps, which predict capability differences should manifest when competencies first emerge. Consistent with Hyde's (2014) gender similarities hypothesis and meta-analytic evidence that gender gaps vary across cultural contexts (Xu, 2025; Miller et al., 2020), these findings support sociocultural explanations and identify the preschool period as a critical intervention window before stereotype internalization creates disparities.

Cross-cultural validation in Indonesian contexts addresses persistent Western-centrism in developmental research while demonstrating the model's generalizability. The absence of significant differences among TK, RA, and TKIT settings suggests scientific thinking capabilities develop similarly regardless of educational philosophy, consistent with Goddu and Gopnik's (2024) evidence that fundamental cognitive processes operate universally while being shaped by cultural practices. This finding demonstrates that foundational scientific capabilities emerge through similar processes whether education emphasizes secular or religious frameworks (Ainnin & Ismail, 2024), with implications for educational policy across culturally diverse nations. The naturalistic teacher assessment approach also demonstrated excellent psychometric properties (α = .954 overall; .750–.902 dimensional), addressing Guarrella et al. (2023) and Brenneman (2011) documentation that teachers lack confidence in science assessment and suggesting that with appropriate tools, teachers can reliably differentiate scientific thinking dimensions.

These findings yield four actionable implications for educational practice. First, supporting foundational cognitive capacities—particularly working memory and sustained attention—may facilitate scientific thinking broadly, as activities promoting executive functions could yield benefits extending to scientific competencies (French, 2004). Second, the progression from observation to prediction to experimentation suggests instructional sequences wherein children receive extensive guided observation before explicit prediction instruction, consistent with Hsin et al. (2025). Third, the substantial individual variation alongside age-related trends suggests that developmentally appropriate expectations should accommodate wide capability ranges while providing targeted support. Fourth, complete gender equity demonstrates that preschool programs should maintain high expectations for all children while actively counteracting subtle biases, with policymakers prioritizing the preschool period for preventing STEM gender gaps rather than remediating them later.

4.1. Research Contribution

This study makes three primary contributions. Theoretically, it provides the first systematic test of a multidimensional hierarchical model during early childhood, extending evidence from older samples (Öztürk, 2025; Reith, 2024) to the preschool period and identifying working memory as an especially central capacity for educational intervention. Empirically, the comprehensive documentation of complete gender equity (all ps > .41, all |ds| < .16) across both foundational capacities and specialized competencies at the point of capability emergence provides compelling evidence supporting sociocultural rather than biological explanations for later STEM disparities, with profound implications for educational policy. Cross-culturally, the first large-scale validation in an Indonesian context encompassing secular and Islamic educational philosophies demonstrates that fundamental aspects of scientific thinking's structure may be universal, providing evidence that foundational scientific capabilities and religious education are compatible.

4.2. Limitations

Several limitations warrant consideration. The cross-sectional design precludes conclusions about individual developmental trajectories or causal mechanisms, as between-group age comparisons cannot establish how individual children change over time or disentangle maturation from experience and cohort effects. Reliance on teacher ratings, while offering ecological validity, introduces potential biases including halo effects and differential familiarity across children; multi-method approaches combining direct assessment, parent reports, and systematic observation would strengthen validity through triangulation. The sample size (N = 105), while adequate for planned analyses, limited power for detecting small effects and precluded confirmatory factor analysis, measurement invariance testing, or structural equation modeling at the latent level. Geographic scope limited to South Sulawesi constrains generalizability across Indonesia's diverse regions, and the correlational design cannot determine which instructional approaches effectively promote scientific thinking development.

4.3. Suggestions

Several research priorities emerge from these findings. Longitudinal studies following children from preschool through elementary school are critically needed to map developmental trajectories, examine whether hierarchical structure strengthens over time, and investigate whether early capabilities predict later outcomes across diverse educational contexts (TK, RA, TKIT). Multi-method approaches combining teacher reports, direct assessment, and systematic observation would strengthen validity, while larger samples would enable confirmatory factor analysis, measurement invariance testing, and investigation of moderators including socioeconomic status and home learning environment. Geographic expansion across Indonesian regions and cross-national comparisons would further test universality versus cultural specificity. Intervention research represents an equally critical priority: randomized trials comparing pedagogical models would identify effective practices, test whether supporting executive functions enhances scientific thinking, and examine whether targeted instruction can address the substantial gap between observation and prediction capabilities. Finally, research on the preschool-to-primary transition would identify factors preserving gender equity versus creating disparities as children encounter differentiated opportunities and stereotype exposure.

5. Conclusion

This study provides initial cross-sectional evidence consistent with a six-dimensional hierarchical organization of scientific thinking in early childhood among Indonesian preschoolers aged 4–6 years. Findings suggest that domain-general cognitive capacities (attention, working memory, and problem solving) are strongly interrelated and show higher mean levels than domain-specific scientific competencies (observation, prediction & reasoning, and experimentation), supporting the interpretation that foundational capacities may scaffold emerging scientific practices. Working memory displayed the strongest associations with higher-order competencies.

Across this sample, we found no evidence of gender differences in any dimension, indicating comparable performance between boys and girls at the preschool stage. While these results do not support early-emerging gender gaps, longitudinal and multi-method research is needed to examine how gender patterns may change with schooling and sociocultural exposure. Finally, the absence of differences across secular and Islamic early childhood education settings suggests that the proposed structure is robust across institutional types within the studied Indonesian context, though broader generalization requires larger, multi-region and cross-national replication. Practically, early childhood programs may benefit from strengthening foundational cognitive capacities while providing explicit, developmentally appropriate support for prediction and experimentation, alongside equitable learning opportunities for all children.

Declarations

Author contribution statement

Subhan: Conceptualization, Methodology, Visualisation, and Writing - Original Draft.
Moh. Lalu Abid Zainul Puad: Investigation, Data Curation, Formal Analysis.
Anies Listyowati: Validation, Resources, Supervision, and Writing - Review & Editing.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request. Due to ethical considerations involving young children, the data are not publicly available.

Declaration of interests statement

All authors declare that they have no financial or personal interests that could influence the work presented in this manuscript.

Additional information

Correspondence and material requests should be addressed to subhan@uinpalopo.ac.id.

ORCID

Subhan https://orcid.org/0000-0002-0215-2716
Moh. Lalu Abid Zainul Puad https://orcid.org/0000-0002-1434-8994
Anies Listyowati https://orcid.org/0009-0003-6077-6679

References

Ainnin, I., & Ismail. (2024). Integration of Islamic education into early childhood curriculum: Building character in the digital era. *Absorbent Mind: Journal of Early Childhood Education, 4*(2), 267–283. https://doi.org/10.37680/absorbent_mind.v4i2.6093

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Brenneman, K. (2011). Assessment for preschool science learning and learning environments. *Early Childhood Research & Practice, 13*(1), 1–9.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Delserieys, A., & Kampeza, M. (2025). Current research and learning in the field of early childhood science education. *Education Sciences, 15*(9), Article 1194. https://doi.org/10.3390/educsci15091194

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). Guilford Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly, 19*(1), 138–149. https://doi.org/10.1016/j.ecresq.2004.01.004

García-Rodeja, I. (2024). Inquiry-based activities with woodlice in early childhood: Implementation and observations. *Education Sciences, 14*(7), Article 710. https://doi.org/10.3390/educsci14070710

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Goddu, M. K., & Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology, 3*, 319–339. https://doi.org/10.1038/s44159-024-00300-5

Gomez, M. J. (2025). The impact of inquiry-based learning in science education: A systematic review of student engagement and achievement. *Journal of Education, Learning, and Management, 2*(2), 353–363. https://doi.org/10.69739/jelm.v2i2.1143

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*(6102), 1623–1627. https://doi.org/10.1126/science.1223416

Guarrella, C., van Driel, J., & Cohrssen, C. (2023). Toward assessment for playful learning in early childhood: Influences on teachers' science assessment practices. *Journal of Research in Science Teaching, 60*(3), 675–707. https://doi.org/10.1002/tea.21811

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Helfat, C. E., & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal, 36*(6), 831–850. https://doi.org/10.1002/smj.2247

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hinkin, T. R. (2023). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967–988. https://doi.org/10.1177/014920639502100509

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture, 2*(1), Article 8. https://doi.org/10.9707/2307-0919.1014

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. SAGE Publications.

Hsin, C.-T., Wu, H.-K., Luu, D. T., & Wei, M.-E. (2025). Fostering young children's scientific practices in urban and Indigenous areas: An investigation of instructional strategies. *International Journal of Science Education, 47*(4), 582–606. https://doi.org/10.1080/09500693.2024.2343437

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch19

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2020). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of U.S. draw-a-scientist studies. *Child Development, 91*(2), 368–398. https://doi.org/10.1111/cdev.13039

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education and Development, 22*(6), 970–988. https://doi.org/10.1080/10409289.2010.507496

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology, 162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Öztürk, E. (2025). Assessing scientific thinking in early childhood: Development and validation of the Scientific Thinking Skills Assessment Tool (STS-AT). *The Eurasia Proceedings of Educational and Social Sciences, 45*, 88–97. https://doi.org/10.55549/epess.949

Reith, M. (2024). Fostering scientific reasoning competencies: Experimental investigation of instructional sequences impacting skills development. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2024.2394708

Samarapungavan, A., Patrick, H., & Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction, 29*(4), 416–470. https://doi.org/10.1080/07370008.2011.608027

Schäfer, J., Reuter, T., Leuchter, M., & Karbach, J. (2024). Executive functions and problem-solving: The contribution of inhibition, working memory and cognitive flexibility to science problem-solving performance in elementary school students. *Journal of Experimental Child Psychology, 244*, Article 105962. https://doi.org/10.1016/j.jecp.2024.105962

Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization* (Rev. ed.). Doubleday.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640

Tzuriel, D., Weiss, T., & Kashy-Rosenbaum, G. (2024). The effects of working memory training on working memory, self-regulation, and analogical reasoning of preschool children. *British Journal of Educational Psychology, 94*(3), 695–714. https://doi.org/10.1111/bjep.12709

Van der Graaf, J., Segers, E., & Verhoeven, L. (2020). Scientific reasoning in kindergartners: Bridging the gap between skills and knowledge. *Learning and Instruction, 69*, Article 101367. https://doi.org/10.1016/j.learninstruc.2020.101367

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-x

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737–747. https://doi.org/10.1509/jmr.11.0368

Xu, L. (2025). A conceptual framework for fostering gender equity in early years STEM education. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-025-10553-y

Yangüez, M. (2025). Development and differentiation of executive function: Inhibition, working memory, and cognitive flexibility across early childhood. *Journal of Cognitive Development*. https://doi.org/10.1080/15248372.2025.2547621

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Zuo, H., & Tang, S. (2024). Gender differences in early childhood STEM learning: A meta-analysis. *Early Childhood Research Quarterly, 66*, 182–195. https://doi.org/10.1016/j.ecresq.2023.10.004