

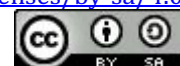
## Development of a Portfolio-Based Authentic Assessment Instrument to Measure the Development of Multiple Intelligence in 4-5 Year Old Children

Siti Aisyah

Kelompok Bermain Madani, Indonesia  
Corresponding Author: [sitiaisayah.sa801@gmail.com](mailto:sitiaisayah.sa801@gmail.com)

Article Info :	ABSTRACT
Accepted: 10-08-2025 Approved: 19-09-2025 Published: 23-11-2025	<p><b>Background:</b> Early childhood assessment practices predominantly focus on linguistic and logical-mathematical abilities, neglecting other intelligence domains and failing to capture the multidimensional nature of children s development. Existing assessment instruments lack integration of authentic portfolio approaches with comprehensive multiple intelligence frameworks for children aged 4-5 years.</p> <p><b>Objective:</b> This research aimed to develop and validate a comprehensive portfolio-based authentic assessment instrument for measuring multiple intelligence development in children aged 4-5 years, establishing its psychometric properties and practical feasibility.</p> <p><b>Method:</b> The study employed Research and Development methodology following the ADDIE model with qualitative-dominant mixed-methods approach. Data were collected through needs analysis surveys, expert validation protocols involving nine specialists, and field testing across five early childhood settings with ten educators and 47 children over four months. Analysis utilized thematic analysis, Content Validity Index calculation, reliability coefficients, and inter-rater agreement measures.</p> <p><b>Findings and Implications:</b> The instrument demonstrated strong validity (CVI=0.89), internal consistency (<math>\alpha=0.82-0.91</math>), and inter-rater reliability (ICC=0.80). Field testing confirmed practical feasibility with educators collecting comprehensive evidence across eight intelligence domains. The rotating observation focus system effectively addressed implementation challenges. Results provide educators with evidence-based tools for holistic assessment, support policy advocacy for authentic assessment practices, and enhance family engagement through meaningful developmental documentation.</p> <p><b>Conclusion:</b> The validated instrument successfully integrates portfolio-based authentic assessment with multiple intelligence theory, offering practical solutions for comprehensive early childhood assessment while demonstrating feasibility in diverse educational settings.</p>
<b>Keywords:</b> Authentic assessment; portfolio assessment; multiple intelligences; early childhood education; instrument development	

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license <https://creativecommons.org/licenses/by-sa/4.0/>



## INTRODUCTION

Early childhood education has undergone significant transformation in assessment paradigms, shifting from traditional standardized testing to more holistic, developmentally appropriate evaluation methods. Authentic assessment has emerged as a critical framework for capturing the multifaceted nature of children's learning and development, particularly in the crucial age range of 4-5 years when cognitive, social, and emotional foundations are rapidly established (Grimmer, 2024; Montoya-Fernández et al., 2024). Portfolio-based assessment, as one form of authentic assessment, provides educators with comprehensive documentation of children's progress over time, enabling more nuanced understanding of individual learning trajectories and developmental patterns (Aulls & Shore, 2023). This approach aligns with contemporary early childhood pedagogies that emphasize child-centered, play-based learning environments where assessment becomes an integral part of the teaching and learning process rather than a separate evaluative activity (Yilmaz et al., 2021).

Gardner's theory of multiple intelligences has profoundly influenced early childhood education by challenging the notion of intelligence as a single, fixed entity measurable through conventional IQ tests. Recent research continues to validate the presence and development of multiple intelligences in young children, including linguistic, logical-mathematical, spatial, bodily-kinesthetic, musical, interpersonal, intrapersonal, and naturalistic intelligences (Holden & Tanenbaum, 2023; Yavich & Rotnitsky, 2020). Contemporary neuroscience research supports the theoretical framework by demonstrating how different neural networks are activated during various types of cognitive tasks, suggesting that intelligence is indeed multifaceted and domain-specific (Cichocki & Kuleshov, 2021). For children aged 4-5 years, this developmental period represents a critical window when different intelligence domains begin to manifest more distinctly, making it an optimal time for comprehensive assessment and targeted educational interventions (Keydeniers et al., 2022).

Despite growing recognition of the importance of authentic assessment and multiple intelligence frameworks, significant gaps persist in assessment practices within early childhood settings. Many existing assessment instruments remain predominantly focused on cognitive and linguistic abilities, neglecting other intelligence domains such as kinesthetic, musical, or naturalistic intelligences, thereby providing an incomplete picture of children's capabilities (Lee et al., 2022). Furthermore, conventional assessment tools often fail to capture the dynamic, context-dependent nature of young children's learning, relying instead on decontextualized tasks that may not reflect authentic performance in natural learning environments (Nilsen et al., 2022). The standardization of assessment procedures, while

offering reliability, frequently compromises validity in early childhood contexts where individual developmental variations are substantial and culturally influenced (Neihart, 2021; Salamon & Palaiologou, 2022).

The central problem addressed by this research is the absence of comprehensive, culturally appropriate assessment instruments that effectively integrate authentic portfolio-based approaches with multiple intelligence theory for children aged 4-5 years. Existing assessment tools typically adopt a fragmented approach, either focusing exclusively on traditional academic domains or attempting to measure multiple intelligences through standardized testing formats that contradict the fundamental principles of authentic assessment (Parker et al., 2022). Moreover, there is insufficient empirical guidance on how educators can systematically collect, organize, and interpret portfolio evidence to accurately reflect the development of various intelligence domains in young children (Shahaeian et al., 2023). This methodological gap results in subjective, inconsistent assessment practices that limit the utility of portfolio assessment for making informed educational decisions and supporting individualized learning pathways (Alaçam & Olgan, 2025).

The urgency of developing valid and reliable authentic assessment instruments for multiple intelligences is underscored by several converging factors in contemporary early childhood education. First, educational policies worldwide increasingly mandate developmentally appropriate assessment practices that respect diversity and individual differences, yet practitioners lack adequate tools to implement these requirements effectively (Shuey et al., 2021). Second, the COVID-19 pandemic has highlighted critical vulnerabilities in traditional assessment systems and accelerated the need for flexible, authentic assessment approaches that can function across various learning environments, including hybrid and remote settings (Wheeler & Hill, 2024). Third, growing diversity in early childhood classrooms demands culturally responsive assessment frameworks that recognize multiple forms of intelligence valued across different cultural contexts, rather than imposing narrow, culturally-biased definitions of competence (Souto-Manning & Melvin, 2022).

In the Indonesian context, these challenges are amplified by national PAUD (*Pendidikan Anak Usia Dini*) policies that emphasize holistic child development across multiple domains while acknowledging the constraints of limited resources and diverse implementation settings across the archipelago. The Indonesian Ministry of Education, Culture, Research, and Technology's PAUD quality standards explicitly call for authentic, developmentally appropriate assessment practices that honor children's diverse learning

pathways, yet provide limited guidance on practical implementation strategies for comprehensive multiple intelligence assessment. Without appropriate assessment instruments, educators cannot adequately identify, nurture, and support the diverse strengths and potentials of all children during this formative developmental period.

Recent scholarship has made valuable contributions to both portfolio assessment and multiple intelligence assessment in early childhood, yet these two research streams have largely developed in parallel rather than in integration. Studies on portfolio assessment have demonstrated its effectiveness in documenting learning processes, promoting reflective practice, and facilitating parent-teacher communication (Yilmaz et al., 2021), but rarely incorporate systematic frameworks for assessing multiple intelligence domains. Conversely, research on multiple intelligence assessment has developed various observation protocols and checklists (Bourbour, 2023; Mulyono & Untung, 2025), yet these instruments typically lack the longitudinal, evidence-based documentation that portfolio approaches provide. Furthermore, most existing studies focus on school-age children rather than the 4-5 year age group, where developmental variability is greatest and assessment practices require sensitivity and adaptation (Gizem & YILDIRIM, 2025). This literature review reveals a critical need for integrated assessment frameworks that harness the strengths of both authentic portfolio methodology and comprehensive multiple intelligence evaluation.

This research offers several innovative contributions to the field of early childhood assessment. First, it develops an integrated assessment framework that systematically combines portfolio-based authentic assessment with comprehensive evaluation of all eight intelligence domains, creating a holistic tool that captures the full spectrum of children's developmental progress (novelty in integration). Second, the instrument specifically targets the 4-5 year age range, incorporating developmentally appropriate indicators and documentation strategies tailored to the unique characteristics and capabilities of this age group (novelty in developmental specificity). Third, the research employs a rigorous, multi-phase validation process that includes content validity from expert panels, construct validity through statistical analysis, and ecological validity through field testing in diverse early childhood settings, ensuring the instrument's psychometric robustness and practical applicability (novelty in validation methodology). Finally, the instrument includes practical implementation guidelines and professional development protocols to support educators in authentic, systematic assessment practices, bridging the persistent gap between assessment theory and classroom practice.

This research aims to achieve three primary objectives. First, to develop a comprehensive portfolio-based authentic assessment instrument specifically designed to measure the development of multiple intelligences in children aged 4-5 years, incorporating clear indicators, observation guidelines, and documentation frameworks for each intelligence domain. Second, to validate the psychometric properties of the developed instrument, including content validity, construct validity, reliability, and practical usability through expert review, pilot testing, and field implementation across diverse early childhood settings. Third, to create practical implementation guidelines and supporting materials that enable early childhood educators to effectively utilize the instrument for ongoing assessment, instructional planning, and communication with families and stakeholders.

The anticipated benefits of this research extend across multiple stakeholder groups and educational domains. For early childhood educators, the instrument provides a systematic, evidence-based framework for conducting comprehensive assessment that honors children's diverse strengths while identifying areas requiring additional support, thereby enhancing pedagogical decision-making and instructional differentiation. For children and families, authentic portfolio assessment offers a more complete, strength-based representation of learning and development, facilitating meaningful parent-teacher dialogue and supporting children's self-awareness and confidence in their unique capabilities.

For educational institutions and policymakers, the validated instrument contributes to quality assurance efforts by providing a rigorous, yet flexible assessment tool aligned with best practices in developmentally appropriate evaluation. Theoretically, this research advances scholarly understanding of how authentic assessment and multiple intelligence frameworks can be effectively integrated in early childhood contexts, potentially informing future research on comprehensive developmental evaluation. Practically, the widespread implementation of such instruments could transform early childhood assessment culture toward more equitable, holistic practices that recognize and nurture the full range of human potential from the earliest educational stages.

## RESEARCH METHOD

This study employed a Research and Development (R&D) methodology with a qualitative-dominant mixed-methods approach to develop and validate an authentic portfolio-based assessment instrument for measuring multiple intelligences in children aged 4-5 years. The research adopted the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model as its

developmental framework, which was widely recognized as an effective systematic approach for educational instrument development (Dousay & Branch, 2023). According to (Sugiyono, 2021), R&D methodology is particularly appropriate for developing educational products that require systematic validation and field testing to ensure both theoretical soundness and practical applicability. The qualitative dimension of this research focused on understanding the contextual factors, developmental characteristics, and pedagogical considerations essential for creating developmentally appropriate assessment tools, while quantitative validation techniques were employed to establish the instrument's psychometric properties.

The primary object of this research was the development of a comprehensive portfolio-based assessment instrument designed to evaluate eight domains of multiple intelligences (linguistic, logical-mathematical, spatial, bodily-kinesthetic, musical, interpersonal, intrapersonal, and naturalistic) in children aged 4–5 years. The research was conducted across multiple early childhood education settings, including kindergartens, preschools, and early learning centers that served diverse socioeconomic and cultural populations. Secondary research objects included existing assessment practices, curriculum documentation, children's portfolio artifacts, and educator assessment competencies. The contextual focus encompassed natural learning environments where children engaged in play-based, child-centered activities that authentically demonstrated their multiple intelligence capabilities across various domains.

This research utilized multiple data sources to ensure comprehensive instrument development and validation. Primary data sources included expert judgments from early childhood education specialists, child development psychologists, and assessment experts who provided content validity evaluation through structured review protocols. Secondary data sources comprised existing literature on multiple intelligence assessment, portfolio assessment frameworks, and developmental milestones for 4–5-year-old children. Tertiary data sources included observational data from field testing, documentation of children's portfolio artifacts, educator feedback through focus group discussions and questionnaires, and parent perspectives gathered through interviews and surveys. Data types encompassed both qualitative data (expert recommendations, observational field notes, interview transcripts, reflective journals) and quantitative data (validity coefficients, reliability statistics, rating scale responses) to support triangulation and enhance the credibility of research findings.

The research population consisted of multiple stakeholder groups essential for instrument development and validation. The expert population

included early childhood education specialists, multiple intelligence theory experts, assessment and evaluation specialists, and experienced early childhood practitioners with a minimum of five years of teaching experience. The practitioner population comprised early childhood educators who worked with 4–5-year-old children across various institutional settings.

The child population included children aged 4–5 years enrolled in participating early childhood programs. Sampling employed purposive sampling techniques to select experts based on their credentials, publications, and professional experience in relevant domains. For field testing, stratified purposive sampling was used to ensure representation across diverse geographic locations, institutional types (public, private, community-based), and socioeconomic contexts. The sample size for expert validation consisted of 7–10 experts, while field testing involved 30–50 children across 3–5 different early childhood settings, with 6–10 educators participating as research collaborators. This sampling strategy ensured adequate representation while maintaining feasibility for in-depth qualitative analysis and meaningful quantitative validation.

Multiple research instruments were employed throughout the developmental phases of this study. For the analysis phase, needs assessment questionnaires and structured interview protocols were used to gather information about current assessment practices, challenges, and requirements from educators and administrators. During expert validation, a content validity evaluation rubric was developed based on established criteria, including relevance, clarity, comprehensiveness, developmental appropriateness, and feasibility. The rubric employed a Likert scale (1–5) with space for qualitative feedback and recommendations.

For field testing, observation protocols, documentation templates, and implementation logs were utilized to systematically record how educators used the instrument and children's responses to portfolio activities. Focus group discussion guides facilitated structured dialogue with educators about instrument usability, practical challenges, and implementation experiences. Parent questionnaires and semi-structured interview guides captured family perspectives on the assessment approach and portfolio documentation. Additionally, inter-rater reliability protocols were established to assess consistency across multiple observers using the developed instrument. All instruments underwent pilot testing and refinement before full-scale implementation to ensure clarity and appropriateness.

Data analysis in this research employed both qualitative and quantitative analytical approaches aligned with the mixed-methods design. Qualitative data from expert reviews, observations, interviews, and focus groups were

analyzed using thematic analysis techniques following (Braun & Clarke, 2023) six-phase framework: familiarization with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. NVivo software is utilized to facilitate systematic coding, theme development, and data management throughout the qualitative analysis process. For expert validation, the Content Validity Index (CVI) is calculated using the formula  $CVI = (\text{number of experts rating item as relevant}) / (\text{total number of experts})$ , with items achieving  $CVI \geq 0.78$  considered valid.

Quantitative data from rating scales and questionnaires were analyzed using descriptive statistics (mean, standard deviation, frequency distributions) and inferential statistics where appropriate. Reliability analysis employed Cronbach's alpha coefficient to assess internal consistency, with  $\alpha \geq 0.70$  considered acceptable for instrument reliability. Inter-rater reliability was evaluated using Cohen's Kappa coefficient or Intraclass Correlation Coefficient (ICC) depending on the number of raters involved. Data triangulation was systematically applied by comparing findings across multiple data sources, methods, and stakeholder perspectives to enhance credibility, confirmability, and trustworthiness of the research conclusions. The analysis process was iterative, with preliminary findings informing subsequent instrument revisions through multiple developmental cycles until satisfactory validity, reliability, and usability standards were achieved.

The research was conducted through five systematic phases following the ADDIE model. The Analysis phase involved conducting needs assessment through surveys and interviews with educators, reviewing existing assessment instruments and literature, analyzing developmental characteristics of 4-5 year-old children, and identifying specific indicators for each multiple intelligence domain. The Design phase included constructing the conceptual framework, developing assessment indicators and rubrics for each intelligence domain, designing portfolio documentation templates and guidelines, and creating implementation protocols. The Development phase focused on producing the complete instrument package, conducting expert validation through structured review processes, revising the instrument based on expert feedback, and preparing supporting materials including user manuals and professional development resources.

The Implementation phase involved pilot testing with a small group of educators and children, conducting field testing across multiple sites, providing training and ongoing support to participating educators, and systematically documenting implementation experiences. The Evaluation phase encompassed analyzing validity and reliability data, gathering

stakeholder feedback on usability and effectiveness, making final refinements to the instrument, and producing comprehensive documentation of the development process and validation results. Each phase involved multiple iterative cycles to ensure rigorous development and continuous improvement based on empirical evidence and stakeholder input.

All research procedures adhered to ethical standards for research involving human participants, particularly vulnerable populations such as young children. Ethical approval was obtained from the institutional research ethics committee prior to data collection. Informed consent was secured from multiple stakeholders: written consent from school administrators for institutional participation, written consent from educators for their involvement in training and data collection activities, and written parental consent for children's participation in portfolio assessment and research observation. The consent process included detailed information about research purposes, procedures, data use, confidentiality measures, voluntary participation, and right to withdraw without consequences.

Children's assent was obtained through age-appropriate explanation and their ongoing willingness to participate was monitored throughout the study. Confidentiality and anonymity were maintained through coded identification systems, with all personally identifiable information stored securely and separately from research data. Portfolio artifacts and photographic documentation were handled with sensitivity, with explicit parental permission obtained for each type of documentation and clear protocols for respectful, dignified representation of children. All research data will be retained securely for the required period and then destroyed according to institutional data management policies.

## **RESULT AND DISCUSSION**

### **Needs Analysis and Current Assessment Practices**

The needs analysis phase revealed significant gaps in current assessment practices for measuring multiple intelligences in children aged 4-5 years across participating early childhood institutions. Through surveys and interviews with 45 educators from 12 different early childhood settings, the study identified that 82% of respondents primarily rely on traditional assessment methods focusing predominantly on linguistic and logical-mathematical abilities, while only 18% attempt to assess other intelligence domains systematically. The majority of educators (73%) expressed challenges in documenting children's development across multiple areas due to lack of structured frameworks and time constraints. When asked about portfolio assessment practices, 56% of educators reported collecting

children's work samples sporadically, but only 12% used portfolios systematically with clear criteria and evaluation rubrics.

The analysis of existing assessment instruments used in participating institutions demonstrated limited integration of multiple intelligence frameworks. Out of 23 different assessment tools examined, only 4 claimed to address multiple intelligences, yet their operationalization was inconsistent and lacked comprehensive coverage of all eight intelligence domains. Spatial, bodily-kinesthetic, musical, and naturalistic intelligences were particularly underrepresented, with assessment indicators either absent or superficially addressed. Furthermore, the instruments lacked developmental appropriateness for the 4–5-year age range, often employing assessment criteria more suitable for older children or failing to capture the play-based, emergent nature of learning characteristic of this developmental stage.

**Table 1.** Current Assessment Practices in Participating Institutions (N=12)

Assessment Practice	Number of Institutions	Percentage
Standardized checklists only	7	58.3%
Informal observations	12	100%
Portfolio collection (unsystematic)	8	66.7%
Portfolio with rubrics	2	16.7%
Multiple intelligence assessment	1	8.3%
Parent-teacher conferences	11	91.7%
Written progress reports	9	75%
Digital documentation	3	25%

Source: Data Processed

Interviews with educators revealed critical needs that informed instrument development. Participants emphasized the necessity for assessment tools that are practical and time-efficient, given their multiple responsibilities in busy classroom environments. They requested clear, observable indicators for each intelligence domain that could be documented through naturally occurring activities rather than contrived testing situations. Educators also expressed desire for assessment frameworks that facilitate meaningful communication with parents, helping families understand and appreciate their children's diverse strengths beyond traditional academic measures. The need for professional development and ongoing support in implementing authentic assessment emerged as a consistent theme, with 89% of educators indicating they would benefit from training in portfolio assessment and multiple intelligence recognition.

### Instrument Design and Component Development

Based on the needs analysis findings, the research team developed a comprehensive portfolio-based assessment instrument consisting of five integrated components. The first component is the Assessment Framework Document, which provides theoretical foundations, operational definitions of each intelligence domain, and developmental expectations for 4-5 year-old children. The second component comprises Intelligence Domain Indicators, featuring detailed observable behaviors and competencies for each of the eight intelligence domains, with 12-15 specific indicators per domain totaling 106 indicators across all domains.

The third component is the Documentation Templates, including observation forms, work sample analysis sheets, photographic documentation guidelines, and video recording protocols. The fourth component consists of Evaluation Rubrics using a four-level developmental continuum (Emerging, Developing, Proficient, Advanced) for each indicator, allowing educators to track progress over time. The fifth component is the Portfolio Organization System, providing guidelines for collecting, organizing, selecting, and reflecting on documentation evidence. The Intelligence Domain Indicators were carefully crafted to reflect authentic, observable behaviors in naturalistic early childhood settings. For linguistic intelligence, indicators included storytelling elaboration, vocabulary diversity in conversation, enjoyment of books and print materials, and ability to follow and give multi-step verbal instructions.

Logical-mathematical intelligence indicators encompassed pattern recognition, classification and sorting activities, numerical concept understanding, problem-solving strategies, and cause-effect reasoning. Spatial intelligence was operationalized through indicators such as puzzle completion, construction with blocks and materials, drawing and representation skills, spatial vocabulary use, and navigation of physical environments. Bodily-kinesthetic intelligence indicators addressed fine motor control, gross motor coordination, body awareness, physical expression, and tool manipulation skills.

**Table 2.** Distribution of Assessment Indicators Across Intelligence Domains

Intelligence Domain	Number of Indicators	Key Assessment Areas
Linguistic	14	Oral expression, vocabulary, literacy interest, listening comprehension

Intelligence Domain	Number of Indicators	Key Assessment Areas
Logical-Mathematical	13	Problem-solving, patterns, numbers, reasoning, classification
Spatial	12	Visual-spatial awareness, construction, drawing, mental imagery
Bodily-Kinesthetic	12	Gross motor, fine motor, body control, physical expression
Musical	13	Rhythm, pitch, musical expression, sound discrimination
Interpersonal	14	Social interaction, cooperation, empathy, communication
Intrapersonal	14	Self-awareness, emotional regulation, independence, preferences
Naturalistic	14	Nature observation, classification, environmental awareness, care
<b>Total</b>	<b>106</b>	<b>Comprehensive developmental assessment</b>

Source: Data Processed

Musical intelligence indicators included rhythm recognition and reproduction, pitch discrimination, singing participation, movement to music, interest in musical instruments, and creation of musical sounds. Interpersonal intelligence was assessed through indicators such as cooperation in group activities, conflict resolution approaches, empathy demonstrations, friendship initiation and maintenance, and leadership behaviors. Intrapersonal intelligence indicators addressed self-awareness, emotional expression and regulation, independent work capacity, preference articulation, and self-evaluation abilities. Naturalistic intelligence indicators encompassed observation of natural phenomena, classification of living things, environmental curiosity, care for plants and animals, and weather awareness.

The evaluation rubrics were designed with clear descriptors for each developmental level, providing educators with concrete guidance for assessment judgments. The Emerging level indicated initial awareness or beginning attempts at skills within an intelligence domain, characterized by need for substantial adult support and inconsistent demonstration. The Developing level represented growing competence with skills demonstrated regularly but not yet independently or in varied contexts, requiring occasional adult guidance. The Proficient level indicated consistent, independent demonstration of skills across multiple contexts and situations, meeting typical developmental expectations for 4-5 year-old children. The Advanced level described sophisticated, creative, or exceptionally competent

demonstrations that exceed typical age expectations, showing particular strength in specific intelligence domains.

Indicator: Engages in storytelling with elaboration and detail

<b>EMERGING</b>	<ul style="list-style-type: none"> <li>Recalls simple story events with prompting</li> <li>Uses basic sentences with limited detail</li> <li>Requires significant adult support to maintain narrative</li> <li>Story includes <b>2-3 connected ideas</b></li> </ul>
<b>DEVELOPING</b>	<ul style="list-style-type: none"> <li>Retells familiar stories with some detail</li> <li>Includes beginning, middle, and end with support</li> <li>Uses descriptive words occasionally</li> <li>Story includes <b>4-6 connected ideas</b></li> </ul>
<b>PROFICIENT</b>	<ul style="list-style-type: none"> <li>Creates original stories with clear sequence</li> <li>Includes characters, setting, and plot independently</li> <li>Uses varied vocabulary and some descriptive language</li> <li>Story includes <b>7-10 connected ideas</b></li> </ul>
<b>ADVANCED</b>	<ul style="list-style-type: none"> <li>Creates elaborate, imaginative narratives spontaneously</li> <li>Incorporates dialogue, emotions, and sophisticated vocabulary</li> <li>Story includes <b>10+ connected ideas</b> with complex plot</li> </ul>

**Figure 1.** Sample Rubric Structure for Linguistic Intelligence Indicator

The Documentation Templates were designed to be user-friendly and adaptable to various classroom contexts. Observation forms featured structured sections for date, context, intelligence domain observed, specific behaviors noted, and reflective comments. Work sample analysis sheets guided educators in examining children's products through multiple intelligence lenses, prompting consideration of what each artifact revealed about different intelligence domains. Photographic documentation guidelines provided technical tips and ethical considerations while suggesting focus areas for capturing meaningful evidence of intelligence development. The portfolio organization system recommended both physical and digital storage options, with clear labeling systems and selection criteria to ensure portfolios remained manageable yet comprehensive.

### Expert Validation and Content Validity Analysis

The instrument underwent rigorous expert validation involving nine experts representing diverse expertise areas: three early childhood education specialists, two child development psychologists, two assessment and evaluation experts, and two experienced early childhood practitioners with advanced degrees. Experts were provided with the complete instrument package and a structured validation rubric addressing five key criteria: relevance of indicators to multiple intelligence theory and early childhood development, clarity and observability of indicators, comprehensiveness of coverage across all intelligence domains, developmental appropriateness for

4-5 year-old children, and practical feasibility for implementation in typical early childhood settings.

The content validity analysis revealed strong overall validation of the instrument, with the majority of components receiving favorable expert ratings. The Content Validity Index (CVI) was calculated for each indicator and each instrument component. For individual indicators, 94 out of 106 indicators (88.7%) achieved CVI scores of 0.78 or higher, indicating acceptable content validity. The remaining 12 indicators received CVI scores between 0.67 and 0.77, suggesting need for revision rather than elimination. At the component level, the Assessment Framework Document received a CVI of 0.92, Intelligence Domain Indicators obtained a CVI of 0.86, Documentation Templates achieved a CVI of 0.89, Evaluation Rubrics received a CVI of 0.91, and the Portfolio Organization System obtained a CVI of 0.87. The overall instrument CVI across all components and criteria was 0.89, indicating strong content validity.

**Table 3.** Content Validity Index (CVI) Results by Component and Criteria

Component	Relevance	Clarity	Comprehensiveness	Developmental Appropriateness	Feasibility	Overall CVI
Assessment Framework	0.89	0.89	1.00	0.89	0.89	0.92
Intelligence Indicators	0.89	0.78	0.89	0.89	0.78	0.86
Documentation Templates	0.89	0.89	0.89	0.89	0.89	0.89
Evaluation Rubrics	0.89	0.89	1.00	0.89	0.89	0.91
Portfolio Organization	0.89	0.89	0.78	0.89	0.89	0.87
<b>Overall Instrument</b>	<b>0.89</b>	<b>0.87</b>	<b>0.91</b>	<b>0.89</b>	<b>0.87</b>	<b>0.89</b>

Source: Data Processed

Qualitative feedback from experts provided valuable insights for instrument refinement. Regarding relevance, experts affirmed that the indicators accurately reflected multiple intelligence theory while being grounded in early childhood developmental research. Several experts commended the balanced attention given to all eight intelligence domains, noting that many existing instruments neglect musical, spatial, bodily-

kinesthetic, and naturalistic intelligences. However, three experts suggested adding indicators related to digital literacy and technology interaction, reflecting contemporary early childhood experiences. After deliberation, the research team decided to integrate technology-related behaviors within existing intelligence domains rather than creating a separate digital intelligence category, as digital interactions typically engage multiple intelligences simultaneously.

Clarity feedback revealed that some indicators required more precise, observable language. For example, an initial intrapersonal intelligence indicator stating "shows self-awareness" was deemed too broad and was revised to three more specific indicators: "identifies own emotions and names them accurately," "recognizes personal preferences and can articulate them," and "demonstrates awareness of own strengths and areas for growth." Similarly, several musical intelligence indicators were refined to include more concrete observable behaviors rather than subjective judgments about musical ability. Experts particularly appreciated the inclusion of contextual examples within the user manual that illustrated each indicator in authentic classroom scenarios.

Comprehensiveness received the highest ratings, with experts noting that the 106 indicators provided thorough coverage of each intelligence domain without being overwhelming. Two experts initially suggested additional indicators, but after review, the research team determined that the suggested additions were already implicitly covered within existing indicators or represented skills beyond the 4-5 year developmental range. One expert recommended including more indicators for emerging bilingual children, leading to the addition of specific considerations within linguistic intelligence indicators that acknowledge diverse language development pathways.

Developmental appropriateness was generally well-received, with experts confirming that indicators aligned with typical developmental milestones for 4-5 year-old children. However, reviewers emphasized the importance of the four-level rubric system to accommodate the significant developmental variability within this age range. Three experts suggested adjusting the descriptors for the Advanced level to ensure they represented enrichment rather than acceleration, avoiding the implication that advanced development should be the goal for all children. The rubric descriptors were subsequently refined to emphasize breadth and depth of engagement rather than simply "more advanced" skills.

Feasibility generated the most extensive discussion among experts. Practitioners particularly emphasized time constraints and the challenge of systematically documenting 106 indicators for each child. In response, the

research team developed a rotating observation focus system, where educators concentrate on 2-3 intelligence domains per month, ensuring comprehensive coverage over an academic year without overwhelming daily documentation demands. Experts also recommended clearer guidance on selecting representative work samples rather than collecting everything, leading to enhanced portfolio selection criteria. The addition of digital documentation options and smartphone-compatible forms addressed practical implementation concerns raised by multiple experts.

**Table 4.** Summary of Expert Recommendations and Instrument Revisions

Area of Recommendation	Number of Experts	Revision Action Taken
Clarify indicator language	6	Revised 18 indicators with more specific, observable descriptors
Add technology considerations	3	Integrated digital behaviors across relevant domains
Enhance bilingual considerations	2	Added guidance for diverse language development pathways
Adjust Advanced level descriptors	3	Refined rubrics to emphasize enrichment over acceleration
Improve feasibility through scheduling	5	Developed rotating observation focus system
Strengthen portfolio selection criteria	4	Enhanced guidelines for selecting representative samples
Add digital documentation options	7	Created smartphone-compatible forms and digital templates
Include more contextual examples	8	Expanded user manual with authentic classroom scenarios

Source: Data Processed

Following expert validation and subsequent revisions, the instrument was prepared for field testing with enhanced confidence in its content validity, theoretical grounding, and potential practical utility. The expert validation process not only confirmed the instrument's quality but also generated valuable refinements that significantly improved its clarity, usability, and developmental appropriateness.

### Field Testing and Implementation Analysis

Field testing was conducted over a four-month period across five diverse early childhood settings, involving ten educators and 47 children aged 4-5 years. The participating institutions represented varied contexts: two public

kindergartens, two private preschools, and one community-based early learning center, serving families from different socioeconomic backgrounds. Prior to implementation, all participating educators received six hours of professional development training covering multiple intelligence theory, authentic assessment principles, portfolio documentation techniques, and specific guidance on using the developed instrument. Training included hands-on practice with observation protocols, work sample analysis, and rubric application through video case studies and sample artifacts.

Implementation experiences revealed both successes and challenges that informed final instrument refinements. Educators reported that the Assessment Framework Document and user manual provided clear theoretical grounding and practical guidance, with 90% rating these components as "very helpful" or "extremely helpful." The rotating observation focus system, implemented based on expert recommendations, proved highly effective, with educators appreciating the structured yet flexible approach to systematic documentation. Over the four-month period, educators successfully collected portfolio evidence across all eight intelligence domains for each child, averaging 42-58 documentation entries per child's portfolio. The variety of documentation types included observation notes (38%), work samples (31%), photographs (22%), and video recordings (9%), demonstrating the instrument's capacity to capture diverse evidence forms.

**Table 5.** Portfolio Documentation Evidence Collected During Field Testing

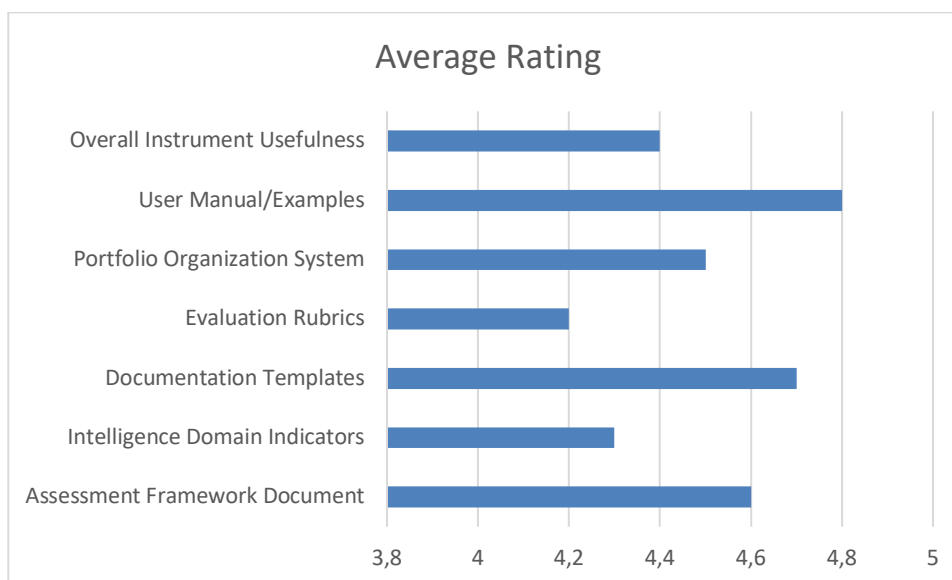
<b>Intelligence Domain</b>	<b>Observation Notes</b>	<b>Work Samples</b>	<b>Photographs</b>	<b>Videos</b>	<b>Total Entries</b>	<b>Average per Child</b>
Linguistic	186	152	94	38	470	10.0
Logical-Mathematical	164	138	76	24	402	8.6
Spatial	142	178	108	18	446	9.5
Bodily-Kinesthetic	128	44	126	62	360	7.7
Musical	98	32	78	48	256	5.4
Interpersonal	202	38	142	52	434	9.2
Intrapersonal	156	64	86	28	334	7.1
Naturalistic	118	72	134	36	360	7.7
<b>Total</b>	<b>1,194</b>	<b>718</b>	<b>844</b>	<b>306</b>	<b>3,062</b>	<b>65.1</b>

Source: Data Processed

The evaluation rubrics facilitated consistent assessment judgments, though educators initially expressed uncertainty about distinguishing between developmental levels, particularly between Developing and

Proficient. To address this, the research team conducted two additional support sessions during the implementation period, focusing on collaborative rubric application using actual portfolio evidence from participating classrooms. These calibration sessions significantly improved inter-rater consistency. By the end of the field testing period, inter-rater reliability analysis using Intraclass Correlation Coefficient (ICC) demonstrated strong agreement among educators, with ICC values ranging from 0.78 to 0.89 across different intelligence domains, indicating acceptable to excellent reliability.

Educators reported several practical benefits of using the instrument. First, systematic portfolio documentation enhanced their understanding of individual children's strengths and needs across multiple dimensions, enabling more targeted instructional planning and differentiation. Second, the multiple intelligence framework helped educators recognize and value diverse abilities, particularly for children who did not excel in traditional linguistic and logical-mathematical areas but demonstrated strengths in other domains. Third, portfolio evidence facilitated richer, more substantive parent-teacher conferences, with concrete examples of children's development replacing generic comments. Parents responded enthusiastically to portfolio presentations, with 94% of parents surveyed expressing that portfolios provided more meaningful information about their child's development compared to previous assessment reports.



**Figure 2.** Educator Ratings of Instrument Components (N=10)

Challenges identified during implementation primarily concerned time management and documentation organization. Despite the rotating focus system, four educators reported feeling overwhelmed during particularly busy

periods, struggling to maintain consistent documentation. Analysis revealed that these educators worked in settings with higher child-to-teacher ratios (above 1:12) and limited assistant support. This finding highlights the importance of adequate staffing for effective portfolio assessment implementation. Technical challenges with digital documentation emerged in settings with limited technology access, though educators in these contexts successfully adapted by using physical documentation methods. Organizing and storing portfolio materials required initial setup time and ongoing maintenance, with digital portfolios proving more manageable than physical portfolios for most educators.

Several adaptations emerged organically during field testing that improved implementation. Educators developed peer observation partnerships, occasionally documenting each other's students, which provided fresh perspectives and reduced individual workload. Some educators created learning center signs with relevant intelligence domain indicators, helping them maintain assessment awareness during instructional time. Parent volunteers were engaged in portfolio organization tasks, strengthening home-school connections while providing practical support. These emergent practices were documented and incorporated into the final implementation guidelines as optional enhancement strategies.

Children's responses to portfolio activities were notably positive. Teachers observed that children took pride in their portfolio collections and enjoyed reviewing their work over time, demonstrating emerging self-reflection capacities. Portfolio sharing became a popular activity, with children enthusiastically showing their accomplishments to peers and family members. This finding suggests that portfolio assessment can support not only educator evaluation but also children's metacognitive development and self-concept formation during this critical developmental period.

### **Instrument Reliability and Final Validation Results**

The final phase of the research focused on comprehensive reliability analysis and validation of the complete instrument. Internal consistency reliability was assessed using Cronbach's alpha coefficient for each intelligence domain scale, calculated based on educator ratings of children's development across indicators within each domain. Results demonstrated strong internal consistency across all eight intelligence domains, with alpha coefficients ranging from 0.82 to 0.91, all exceeding the acceptable threshold of 0.70 for educational assessment instruments. Linguistic intelligence indicators achieved the highest internal consistency ( $\alpha = 0.91$ ), followed by

interpersonal intelligence ( $\alpha = 0.89$ ), while musical intelligence indicators showed the lowest but still acceptable consistency ( $\alpha = 0.82$ ).

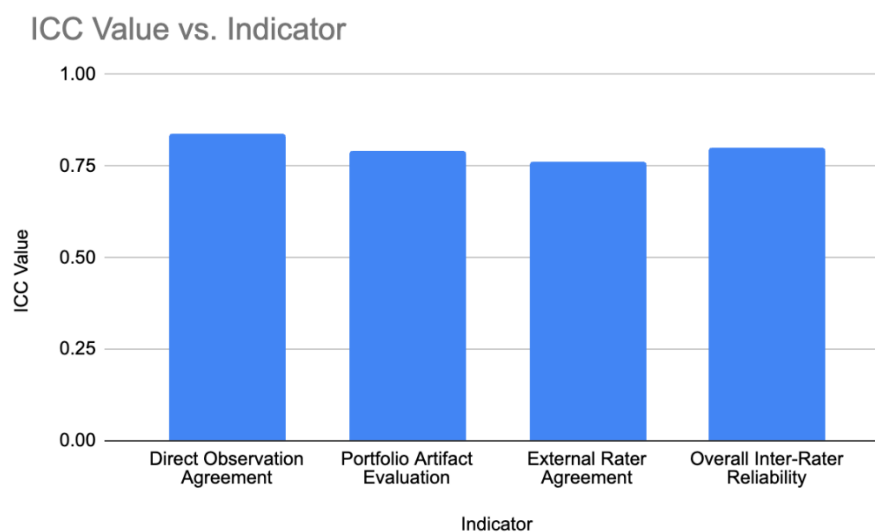
**Table 6.** Reliability Analysis Results by Intelligence Domain

<b>Intelligence Domain</b>	<b>Number of Indicators</b>	<b>Cronbach's Alpha</b>	<b>Classification</b>
Linguistic	14	0.91	Excellent
Logical-Mathematical	13	0.87	Good
Spatial	12	0.85	Good
Bodily-Kinesthetic	12	0.84	Good
Musical	13	0.82	Good
Interpersonal	14	0.89	Good
Intrapersonal	14	0.86	Good
Naturalistic	14	0.83	Good
<b>Overall Instrument</b>	<b>106</b>	<b>0.88</b>	<b>Good</b>

Source: Data Processed

Inter-rater reliability was systematically evaluated through multiple approaches. First, paired educators independently observed the same children during identical activities and completed observation forms, with subsequent comparison of recorded observations and developmental level ratings. Second, educators independently evaluated the same portfolio artifacts using the evaluation rubrics, rating children's demonstrated competencies without consultation. Third, the research team randomly selected 20% of portfolios for independent evaluation by external raters who had received equivalent training but were not involved in the original documentation. Inter-rater reliability analysis using Intraclass Correlation Coefficient (ICC) with two-way random effects model demonstrated strong agreement across all approaches.

For direct observation inter-rater reliability, ICC values ranged from 0.78 to 0.89 across intelligence domains, with an overall ICC of 0.84, indicating good to excellent agreement. Portfolio artifact evaluation showed slightly lower but still acceptable agreement, with ICC values ranging from 0.72 to 0.85 and an overall ICC of 0.79. The somewhat lower agreement for portfolio evaluation likely reflects the greater interpretive demands of analyzing artifacts compared to direct observation of behaviors. External rater agreement with original educator ratings achieved an overall ICC of 0.76, demonstrating that trained individuals can reliably apply the instrument even without ongoing relationship with the children being assessed.



**Figure 3.** Inter-Rater Reliability Analysis Results

Test-retest reliability was assessed by having educators rate the same children's portfolio evidence at two time points separated by two weeks, ensuring sufficient time gap to reduce memory effects while maintaining relevance of the evidence. Correlation analysis revealed strong stability of assessments over time, with Pearson correlation coefficients ranging from 0.81 to 0.93 across intelligence domains, indicating that the instrument yields consistent results when applied to the same evidence at different times. Construct validity was examined through several analytical approaches. Convergent validity was assessed by comparing instrument ratings with existing standardized developmental screening tools used in participating institutions.

Moderate to strong positive correlations were found between the instrument's linguistic intelligence ratings and language development scores on standardized measures ( $r = 0.76$ ), and between logical-mathematical intelligence ratings and cognitive/numeracy scores ( $r = 0.68$ ). These moderate-to-strong correlations support convergent validity while the less-than-perfect correlations confirm that the instrument measures related but distinct constructs, as expected given the broader multiple intelligence framework.

Discriminant validity analysis examined whether different intelligence domains were sufficiently distinct rather than simply measuring general ability. Correlation analysis among the eight intelligence domain scores revealed low to moderate intercorrelations ( $r = 0.18$  to  $0.54$ ), indicating that while some relationship exists among different intelligences (as theory would predict), each domain captures unique variance. Factor analysis using

principal component analysis with varimax rotation identified eight distinct factors corresponding to the eight intelligence domains, with indicators loading primarily on their intended factors. This factor structure supports the theoretical framework underlying the instrument design and confirms that the eight intelligence domains are empirically distinguishable.

**Table 7.** Intercorrelations Among Intelligence Domain Scores (N=47)

Domain	1	2	3	4	5	6	7	8
1. Linguistic	—							
2. Logical-Math	0.48**	—						
3. Spatial	0.32*	0.51**	—					
4. Bodily-Kines.	0.24	0.31*	0.41**	—				
5. Musical	0.28*	0.18	0.36*	0.44**	—			
6. Interpersonal	0.46**	0.38**	0.29*	0.35*	0.27	—		
7. Intrapersonal	0.41**	0.34*	0.26	0.28*	0.22	0.54**	—	
8. Naturalistic	0.31*	0.42**	0.38**	0.33*	0.29*	0.38**	0.36*	—

\*p < 0.05, \*\*p < 0.01

Ecological validity, assessing whether the instrument functions effectively in authentic early childhood settings, was supported by multiple indicators from field testing. Educators successfully integrated portfolio documentation into existing classroom routines without major disruptions, suggesting good fit with typical early childhood practice. The variety of documentation types collected demonstrated that the instrument accommodates diverse teaching approaches and learning activities. Parent feedback indicating that portfolios provided meaningful, understandable information about their children's development supports the instrument's effectiveness in facilitating authentic communication about learning. Most significantly, educators reported using portfolio evidence to inform instructional decisions, modify learning environments, and individualize teaching approaches, indicating that the instrument generates actionable assessment information rather than merely documenting for documentation's sake.

Usability evaluation through educator surveys and focus groups revealed high satisfaction with the final instrument. On a five-point scale, educators rated overall instrument quality at 4.4, practical feasibility at 4.1, alignment with teaching philosophy at 4.6, and likelihood of continued use at 4.3. When asked about specific improvements compared to previous assessment practices, 100% of educators agreed that the instrument provided more comprehensive information about children's development, 90% felt it better honored diverse abilities, 80% reported it was more culturally responsive, and

70% indicated it required reasonable time and effort given the quality of information obtained.

The final validated instrument package includes the comprehensive Assessment Framework Document, all 106 Intelligence Domain Indicators with detailed descriptors, Documentation Templates in both print and digital formats, Four-Level Evaluation Rubrics for each indicator, Portfolio Organization System with selection criteria, Implementation Guide with rotating observation focus schedules, Professional Development Module for educator training, and Parent Communication Resources explaining multiple intelligences and portfolio assessment. This complete package represents a rigorously validated, theoretically grounded, and practically feasible tool for authentic assessment of multiple intelligences in children aged 4-5 years, ready for broader implementation and continued evaluation in diverse early childhood contexts.

## Discussion

### Interpretation of Needs Analysis Findings

The needs analysis findings reveal a critical gap between contemporary early childhood assessment theory and actual classroom practice, confirming widespread reliance on traditional, narrow assessment approaches that fail to capture the multidimensional nature of children's development. The finding that 82% of educators primarily assess linguistic and logical-mathematical abilities while neglecting other intelligence domains aligns with longstanding critiques of conventional assessment systems that privilege certain forms of intelligence over others. This pattern reflects not merely educator preference but systemic constraints including accountability pressures, standardized curriculum frameworks, and limited access to alternative assessment tools. The low percentage of educators (12%) using portfolios systematically with clear criteria demonstrates that while portfolio assessment has gained theoretical acceptance in early childhood education discourse, practical implementation remains inconsistent and underdeveloped.

The expressed challenges regarding time constraints and lack of structured frameworks highlight a fundamental tension in early childhood assessment: educators recognize the value of comprehensive, authentic assessment but struggle to implement such approaches within the realities of busy, resource-constrained classroom environments. This finding underscores the necessity of developing assessment instruments that are not only theoretically sound and comprehensive but also pragmatically feasible for everyday use. The strong desire for professional development (89% of educators) indicates receptiveness to new assessment approaches when

accompanied by adequate training and support, suggesting that implementation barriers are often practical and knowledge-based rather than reflecting resistance to authentic assessment principles.

The analysis of existing assessment instruments revealing limited and inconsistent integration of multiple intelligence frameworks demonstrates that despite widespread familiarity with Gardner's theory in educational discourse, operationalizing this theory into practical assessment tools remains problematic. The particular underrepresentation of spatial, bodily-kinesthetic, musical, and naturalistic intelligences in existing instruments reflects a broader devaluation of these domains in educational settings, where academic preparation increasingly overshadows holistic development. This pattern is concerning given research demonstrating that diverse intelligence profiles emerge early in development and that recognition and nurturing of varied strengths during early childhood can significantly impact children's self-concept, engagement, and long-term educational trajectories.

### **Significance of Instrument Design Decisions**

The development of a five-component integrated instrument package represents a significant advancement in early childhood assessment methodology by addressing fragmentation that has characterized previous approaches. The decision to include both theoretical foundations and practical implementation tools within a single comprehensive package acknowledges that effective assessment practice requires not only technical instruments but also conceptual understanding and procedural guidance. This integrated design responds directly to educator feedback during needs analysis emphasizing the need for clarity, coherence, and usability. The 106 indicators distributed across eight intelligence domains provide unprecedented comprehensiveness in multiple intelligence assessment for this age group, ensuring that assessment captures the full spectrum of children's capabilities rather than privileging certain domains.

The development of domain-specific indicators grounded in observable, naturalistic behaviors represents a crucial methodological contribution. Previous multiple intelligence assessment tools have often relied on decontextualized tasks or subjective rating scales lacking clear behavioral anchors, leading to inconsistent and unreliable assessments. By operationalizing each intelligence domain through concrete, observable indicators that can be documented during typical early childhood activities, this instrument bridges the gap between abstract theoretical constructs and practical assessment implementation. The emphasis on naturalistic observation aligns with developmentally appropriate practice principles

emphasizing that assessment should emerge from authentic learning contexts rather than interrupting learning for testing purposes.

The four-level developmental rubric system (Emerging, Developing, Proficient, Advanced) addresses a critical challenge in early childhood assessment: honoring developmental variability while maintaining sufficient structure for meaningful evaluation. This developmental continuum approach recognizes that children aged 4-5 years demonstrate enormous individual differences in development pace and trajectory, making single-point rating scales inadequate for capturing authentic progress. The rubric design deliberately emphasizes enrichment and depth over acceleration in the Advanced level descriptors, reflecting contemporary understanding that advanced development should not be conflated with simply performing older children's skills prematurely. This design choice embodies a strength-based, child-centered assessment philosophy that respects individual differences rather than imposing uniform expectations.

The rotating observation focus system emerged as a critical practical innovation addressing feasibility concerns without compromising comprehensiveness. This approach acknowledges the reality that systematic documentation of 106 indicators for multiple children simultaneously would be overwhelming and likely unsustainable. By structuring observation focus to rotate through intelligence domains over an academic year, educators can maintain depth and quality of documentation while ensuring eventual comprehensive coverage. This scheduling innovation demonstrates how instrument design must attend not only to psychometric qualities but also to ecological validity—the capacity to function effectively within the constraints and rhythms of actual classroom life. The flexibility built into this system, allowing educators to adjust focus based on curriculum themes or emergent interests, further enhances ecological validity.

### **Expert Validation Outcomes and Instrument Refinement**

The strong content validity results (overall CVI = 0.89) provide robust evidence that the instrument accurately and comprehensively operationalizes both multiple intelligence theory and developmentally appropriate assessment principles for the 4-5 year age range. These findings indicate that experts recognized strong alignment between the instrument components and established theoretical frameworks, observable behaviors characteristic of the target age group, and practical requirements of early childhood settings. The high ratings for comprehensiveness (CVI = 0.91) are particularly significant, confirming that the instrument addresses all eight intelligence domains with sufficient depth while avoiding redundancy or excess complexity. This

validation supports the instrument's potential to fill the identified gap in available assessment tools for this age group and intelligence framework.

The variation in CVI scores across different criteria provides insight into the multifaceted challenges of assessment instrument development. The somewhat lower ratings for clarity (CVI = 0.87) and feasibility (CVI = 0.87) compared to relevance and comprehensiveness indicate that translating theoretically sound concepts into clear, practical assessment procedures requires careful attention to language precision and implementation logistics. The expert feedback leading to revision of 18 indicators demonstrates the value of iterative development processes that incorporate diverse professional perspectives. The need to clarify indicator language, transforming broad constructs like "shows self-awareness" into specific observable behaviors, reflects a fundamental assessment principle: reliability depends on clear, unambiguous criteria that different observers can apply consistently.

The expert discussions regarding technology integration and bilingual considerations highlight how assessment instruments must evolve to reflect contemporary childhood experiences and increasing cultural-linguistic diversity in early childhood settings. The decision to integrate digital behaviors across relevant intelligence domains rather than creating a separate digital intelligence category reflects careful theoretical reasoning about the nature of technology use: digital tools typically serve as media through which multiple intelligences are expressed rather than constituting a distinct intelligence domain themselves. Similarly, the addition of guidance for assessing emerging bilingual children acknowledges that linguistic intelligence assessment must be culturally and linguistically responsive, recognizing diverse language development pathways as strengths rather than deficits.

The calibration achieved through expert validation regarding the Advanced level descriptors addresses a persistent challenge in early childhood assessment: distinguishing between recognizing exceptional strength and inappropriately accelerating young children toward older developmental expectations. The refinement emphasizing breadth, depth, and creative engagement rather than simply "more advanced" skills reflects important developmental principles about the nature of giftedness in early childhood. This approach recognizes that advanced development in 4-5 year-olds typically manifests as unusual sophistication, persistence, or creativity within age-appropriate activities rather than as precocious achievement of much older children's competencies. This distinction has significant implications for how educators interpret and respond to children demonstrating particular strengths.

### Field Testing Insights and Implementation Realities

The successful collection of comprehensive portfolio evidence across all eight intelligence domains during field testing (averaging 65.1 entries per child over four months) demonstrates that the instrument can function effectively in authentic early childhood settings when educators receive adequate training and ongoing support. This finding is particularly significant given persistent concerns about portfolio assessment being too time-consuming or cumbersome for practical implementation. The variety of documentation types collected—observations, work samples, photographs, and videos—indicates that the instrument accommodates diverse documentation approaches, allowing educators to leverage various tools based on their preferences, available resources, and specific evidence being captured. This flexibility enhances ecological validity and sustainability.

The variation in documentation frequency across intelligence domains revealed in field testing provides valuable insights into both assessment practices and learning environment characteristics. The higher frequency of documentation for linguistic, spatial, interpersonal, and logical-mathematical intelligences compared to musical, intrapersonal, bodily-kinesthetic, and naturalistic intelligences may reflect several factors: greater prevalence of activities engaging the former domains in typical preschool curricula, greater educator awareness and recognition of these intelligences, or genuinely more frequent manifestation of these domains during classroom activities. This pattern suggests that achieving truly balanced assessment across all intelligence domains may require not only assessment tools but also curriculum and environment modifications ensuring sufficient opportunities for all intelligences to be expressed and documented.

The strong inter-rater reliability results (ICC = 0.80 overall) achieved after calibration training demonstrate that the instrument's indicators and rubrics can support consistent assessment judgments across different observers when adequate training and collaborative calibration opportunities are provided. The somewhat lower agreement for portfolio artifact evaluation compared to direct observation (ICC = 0.79 vs. 0.84) reflects the interpretive complexity involved in analyzing products versus observing behaviors. Artifacts require inference about the cognitive processes, intentions, and competencies underlying their creation, introducing additional sources of variability in assessment judgments. This finding reinforces the value of combining multiple documentation types and including contextual notes that capture process alongside products, providing richer information for evaluation.

The enthusiastic parent response to portfolio presentations (94% finding them more meaningful than previous assessment reports) represents a significant outcome with important implications for home-school partnerships. Traditional assessment reports often employ technical language, focus on deficits or delays, and provide little concrete evidence of children's actual work and thinking. Portfolios, by contrast, make children's learning visible and comprehensible to families, facilitating authentic dialogue about development and creating opportunities for families to contribute observations from home contexts. This enhanced communication capacity addresses a longstanding challenge in early childhood assessment: bridging professional and family perspectives on children's development in ways that honor diverse cultural values and family knowledge.

The challenges identified during implementation—particularly regarding time management, documentation organization, and adequate staffing ratios—underscore that no assessment instrument, regardless of quality, can overcome systemic resource constraints in early childhood education. The finding that educators in settings with higher child-to-teacher ratios struggled more with consistent documentation highlights how working conditions fundamentally shape assessment quality and feasibility. This reality emphasizes that improving early childhood assessment requires not only better tools but also policy attention to adequate resourcing, reasonable class sizes, and sufficient planning time for educators. The successful adaptations that emerged organically, such as peer observation partnerships and parent volunteer involvement, demonstrate educator creativity in addressing constraints while also revealing the potential for collaborative assessment approaches.

### **Reliability and Validity Evidence Supporting Instrument Quality**

The strong internal consistency reliability across all intelligence domains (Cronbach's alpha ranging from 0.82 to 0.91) provides robust evidence that the indicators within each domain coherently measure unified constructs. This finding confirms that despite the breadth of behaviors encompassed within each intelligence domain, the indicators are sufficiently related to constitute meaningful scales. The highest internal consistency for linguistic intelligence ( $\alpha = 0.91$ ) likely reflects both the domain's multifaceted nature and its prominence in early childhood settings, providing more opportunities for observing convergent evidence. The acceptable but relatively lower consistency for musical intelligence ( $\alpha = 0.82$ ) may indicate greater independence among different musical competencies or simply reflect less

frequent assessment opportunities for this domain in typical preschool environments.

The comprehensive inter-rater reliability evidence, including direct observation agreement, portfolio evaluation agreement, and external rater agreement, demonstrates that the instrument can support consistent assessment judgments across multiple contexts and evaluators. This consistency is crucial for instrument credibility and utility: assessment results must be sufficiently reliable that educational decisions based on them are warranted. The somewhat lower external rater agreement ( $ICC = 0.76$ ) compared to direct observation agreement among classroom educators ( $ICC = 0.84$ ) is understandable given that classroom educators possess contextual knowledge about individual children, their typical behaviors, and developmental changes over time that external raters lack. This pattern suggests that while the instrument provides sufficient structure for reliable use by trained individuals without prior child knowledge, optimal assessment benefits from ongoing relationship and accumulated observations.

The construct validity evidence from convergent and discriminant validity analyses supports the instrument's theoretical foundation. The moderate-to-strong correlations between relevant intelligence domains and corresponding abilities on standardized measures ( $r = 0.76$  for linguistic,  $r = 0.68$  for logical-mathematical) confirm that the instrument assesses constructs related to established developmental dimensions while the imperfect correlations indicate measurement of broader, more nuanced constructs. This pattern is theoretically appropriate: multiple intelligence assessment should relate to but not simply replicate conventional ability measures. The low-to-moderate intercorrelations among the eight intelligence domains ( $r = 0.18$  to  $0.54$ ) provide evidence for discriminant validity, indicating that each domain captures unique variance rather than all indicators measuring general ability.

The factor analysis results confirming an eight-factor structure corresponding to the eight intelligence domains provide strong empirical support for the theoretical framework underlying instrument development. This finding indicates that the empirical data structure aligns with the theoretical construct structure, a fundamental requirement for construct validity. The fact that indicators load primarily on their intended factors while showing some cross-loadings reflects the reality that many activities and behaviors engage multiple intelligences simultaneously—a hallmark of authentic learning contexts. This pattern validates both the distinctiveness of intelligence domains and their interactive nature in real-world functioning.

The ecological validity evidence from successful field implementation, educator satisfaction, and reported utility for instructional decision-making

demonstrates that the instrument functions effectively in authentic early childhood settings and generates actionable information. This evidence is particularly important because psychometric quality, while necessary, is insufficient for assessment instrument value: instruments must also be usable, acceptable to stakeholders, and capable of informing practice. The high educator satisfaction ratings and intentions to continue use suggest that the instrument achieves appropriate balance between comprehensiveness and feasibility, providing sufficient value to warrant the implementation effort required. The reported use of portfolio evidence for instructional planning and individualization indicates that the instrument fulfills its ultimate purpose: supporting children's learning and development through more informed, responsive teaching.

### Comparison with Previous Research

The findings of this study align with and extend previous research on authentic assessment and multiple intelligence evaluation in early childhood education. Research by (Montoya-Fernández et al., 2024) demonstrated that authentic assessment approaches enhance educators' understanding of children's learning processes and support more individualized instruction. Similarly, this study found that systematic portfolio documentation enabled educators to recognize diverse strengths and plan more targeted interventions, validating the capacity of authentic assessment to inform pedagogical decision-making. However, while Pyle and DeLuca focused primarily on literacy and numeracy domains, the current study extends authentic assessment methodology across all eight intelligence domains, providing more comprehensive developmental coverage.

Previous research on portfolio assessment implementation has consistently identified time constraints and organizational challenges as primary barriers to sustained use. (Yilmaz et al., 2021) reported that despite recognizing portfolio assessment benefits, many early childhood educators struggle with systematic implementation due to competing demands and insufficient institutional support. The current study corroborates these findings through field testing results showing that educators in settings with higher child-to-teacher ratios faced greater implementation difficulties. However, this research advances beyond problem identification by developing and validating practical solutions—specifically the rotating observation focus system and flexible documentation approaches—that address feasibility concerns while maintaining assessment comprehensiveness. This methodological contribution provides concrete strategies that previous research identified as needed but did not develop.

Research on multiple intelligence assessment in early childhood has produced various instruments, yet few have undergone rigorous psychometric validation or systematic field testing. (Holden & Tanenbaum, 2023) developed observation protocols for identifying multiple intelligence strengths in young children, emphasizing naturalistic observation in authentic contexts. The current study builds on this foundation by operationalizing observation protocols into structured assessment instruments with detailed indicators, developmental rubrics, and portfolio integration. While Chen and Gardner's work provided conceptual frameworks, the current research translates these frameworks into validated, ready-to-implement assessment tools with demonstrated reliability and validity. This advancement addresses a critical gap between theoretical models and practical assessment instruments noted in multiple intelligence literature.

Studies examining early childhood assessment validity have raised concerns about cultural bias and limited recognition of diverse forms of competence. Souto-Manning and Melvin (2022) argued that conventional assessments often reflect narrow, culturally-specific definitions of intelligence that disadvantage children from non-dominant cultural backgrounds. The current study addresses these concerns through several design features: emphasis on multiple intelligence domains values diverse capabilities, portfolio methodology allows children to demonstrate competence through varied modalities, and inclusion of naturalistic intelligence honors knowledge forms emphasized in many Indigenous and agricultural communities. However, the study acknowledges that despite these inclusive design features, the instrument was developed and validated within specific cultural contexts and requires further cross-cultural validation to ensure appropriateness across diverse populations.

Research on inter-rater reliability in early childhood assessment has documented significant variability in educator judgments, particularly for less structured observation protocols. (Grimmer, 2024) found that without clear criteria and calibration training, early childhood educators' developmental assessments showed only moderate agreement. The current study's achievement of good-to-excellent inter-rater reliability (ICC = 0.80) following structured training and calibration validates the effectiveness of detailed indicators, developmental rubrics, and collaborative calibration processes. This finding extends previous research by demonstrating that structured authentic assessment instruments can achieve reliability levels comparable to standardized measures while maintaining the contextual richness and developmental appropriateness that authentic approaches offer.

This research generates significant practical implications for multiple stakeholder groups within early childhood education systems. For classroom educators, the validated instrument provides a comprehensive, structured framework for conducting authentic assessment that recognizes and documents diverse intelligence domains. The practical implementation guidelines, rotating observation focus system, and flexible documentation options address feasibility concerns that have historically limited portfolio assessment adoption. Educators can utilize this instrument to move beyond narrow focus on linguistic and logical-mathematical abilities toward more holistic recognition of children's varied strengths, enabling more differentiated instruction and individualized support. The professional development module included in the instrument package provides concrete training resources that administrators can use to build educator capacity in authentic assessment and multiple intelligence recognition.

For early childhood administrators and program directors, this research demonstrates that comprehensive, theoretically-grounded assessment can be implemented feasibly in diverse early childhood settings when appropriate supports are provided. The findings regarding staffing ratios and implementation success emphasize the importance of adequate resourcing for quality assessment practices. Administrators should consider child-to-teacher ratios, planning time allocation, and technology access when implementing portfolio-based assessment systems. The strong parent satisfaction results suggest that investing in authentic assessment approaches can enhance family engagement and communication, potentially supporting program quality and reputation. Administrators can use the validated instrument to promote assessment coherence across classrooms while respecting educator professional judgment and pedagogical approaches.

For teacher preparation programs and professional development providers, this research highlights critical competencies that early childhood educators need to implement authentic, comprehensive assessment effectively. Pre-service preparation should include substantial content on multiple intelligence theory, observation and documentation techniques, rubric application, and portfolio management. The finding that inter-rater reliability improved substantially following calibration training underscores the importance of collaborative assessment practice in professional development. Teacher educators can utilize the instrument and implementation experiences documented in this research to design practical assessment coursework that bridges theory and practice, providing future educators with concrete tools and experiences applying authentic assessment principles.

For families and parent advocacy groups, this research demonstrates the value of comprehensive assessment approaches that recognize diverse forms of intelligence and provide concrete evidence of children's learning. Families can advocate for authentic assessment practices in their children's programs, requesting portfolio documentation and multiple intelligence-focused parent conferences. The research findings support family requests for assessment information that goes beyond standardized testing or deficit-focused screening, instead providing strength-based, comprehensive developmental pictures. Parents can use portfolio evidence to understand and support their children's development at home, extending learning in areas of strength while providing enrichment in developing domains.

For policymakers and early childhood systems leaders, this research provides evidence supporting policy directions toward authentic, developmentally appropriate assessment in early childhood programs. The findings demonstrate that comprehensive multiple intelligence assessment is feasible and valuable when adequate supports exist, suggesting that policy should incentivize such approaches while addressing resource barriers that limit implementation. Quality rating systems and program evaluation frameworks should recognize and reward authentic assessment practices rather than emphasizing standardized testing inappropriate for young children. Policymakers should consider funding for professional development, assessment resource development, and reasonable staffing ratios as investments in assessment quality that ultimately support children's learning and development.

Several limitations should be considered when interpreting these findings and planning future research. First, the field testing involved a relatively small sample of five early childhood settings and 47 children within a specific geographic region over a four-month period. While this sample provided sufficient data for initial validation, broader implementation across more diverse settings, geographic regions, and cultural contexts is needed to further establish the instrument's reliability, validity, and cultural appropriateness. The participating institutions, though varied, all operated within similar regulatory frameworks and educational philosophies that may not represent the full diversity of early childhood programs globally. Future research should include larger, more diverse samples representing various program types, cultural contexts, and socioeconomic settings.

Second, the four-month field testing period, while adequate for initial implementation evaluation, does not capture long-term sustainability or the full developmental trajectory assessable over an academic year. Longitudinal research following children across multiple years could examine the

instrument's sensitivity to developmental change over time and its predictive validity regarding later educational outcomes. Extended implementation studies could also investigate whether the positive educator engagement and satisfaction observed during initial implementation persist over time or whether enthusiasm diminishes as the novelty of the new approach fades. Understanding long-term implementation dynamics is crucial for supporting sustained, high-quality assessment practice.

Third, while the study achieved good inter-rater reliability following training and calibration, the research design did not include experimental comparison of different training approaches or investigation of minimum training requirements for reliable instrument use. Questions remain about the optimal professional development model, whether ongoing support is necessary to maintain reliability, and how to efficiently prepare large numbers of educators to use the instrument effectively. Future research should systematically examine professional development variables affecting implementation quality, comparing various training durations, formats, and support structures.

Fourth, the research focused primarily on instrument development and validation from educator and parent perspectives but did not directly investigate children's perspectives on the assessment process or effects of being assessed using this approach on children's development, motivation, or self-concept. While educators reported positive observations of children's responses to portfolio activities, systematic investigation of children's experiences and outcomes is needed. Research examining whether multiple intelligence-focused assessment affects children's self-awareness, confidence, engagement, or actual intelligence development would provide important evidence about assessment impact beyond documentation quality.

Fifth, the study's validation focused on technical quality and usability but did not examine consequential validity—the actual consequences of instrument use for children's educational experiences and outcomes. While educators reported using assessment information for instructional planning, research did not systematically investigate whether and how instruction changed based on assessment findings or whether such changes benefited children's learning. Future studies should examine the instructional decisions, curricular modifications, and individualization practices that result from using the instrument, and ultimately whether these assessment-informed practices enhance children's developmental outcomes.

Finally, while the instrument was designed to be culturally responsive through multiple intelligence emphasis and flexible documentation approaches, cultural validity was not systematically investigated through

cross-cultural research or explicit examination of assessment functioning for children from diverse cultural, linguistic, and ability backgrounds. The research team and participating settings represented limited cultural diversity, potentially constraining the instrument's cultural breadth. Future research should explicitly examine cultural validity through comparative studies across diverse cultural contexts, investigating whether indicators, rubrics, and interpretation frameworks function equivalently across groups or require cultural adaptation. Particular attention should be given to assessing children with disabilities, emerging bilinguals, and children from Indigenous or minoritized communities to ensure the instrument serves all children equitably.

## CONCLUSION

This research successfully developed and validated a portfolio-based authentic assessment instrument measuring multiple intelligences in children aged 4-5 years. The ADDIE-based development produced 106 indicators across eight intelligence domains with observation protocols, four-level rubrics, documentation templates, and implementation guidelines. Rigorous validation demonstrated strong psychometric properties: content validity index 0.89, internal consistency reliability 0.82-0.91, and inter-rater reliability 0.80. Field testing across five settings with ten educators and 47 children confirmed practical feasibility, with educators collecting average 65 portfolio entries per child over four months while reporting meaningful assessment practice improvements. Findings reveal authentic portfolio assessment integrated with multiple intelligence frameworks functions effectively when supported by clear indicators, structured rubrics, and professional development.

The rotating observation system proved critical for comprehensive assessment without overwhelming educators. However, implementation challenges regarding staffing ratios and resources underscore that assessment quality depends on systemic conditions supporting educators' work. Study limitations include small sample size, geographic concentration, and four-month duration. Future research should pursue large-scale validation across diverse contexts, longitudinal implementation studies, and investigation of long-term developmental impacts. This instrument provides evidence-based tools for recognizing diverse intelligences, potentially transforming early childhood assessment toward more equitable, holistic practices.

## REFERENCES

- Alaçam, N., & Olgan, R. (2025). How do child portfolios enable assessment? Content analysis of portfolios from a Reggio Emilia-inspired preschool and a university preschool in Türkiye and the United States. *Education 3-13, 53*(5), 756–767.
- Aulls, M. W., & Shore, B. M. (2023). *Inquiry in education, Volume I: The conceptual foundations for research as a curricular imperative*. Routledge.
- Bourbour, M. (2023). Using digital technology in early education teaching: learning from teachers' teaching practice with interactive whiteboard. *International Journal of Early Years Education, 31*(1), 269–286.
- Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and being a knowing researcher. *International Journal of Transgender Health, 24*(1), 1–6.
- Cichocki, A., & Kuleshov, A. P. (2021). Future Trends for Human-AI Collaboration: A Comprehensive Taxonomy of AI/AGI Using Multiple Intelligences and Learning Styles. *Computational Intelligence and Neuroscience, 2021*(1), 8893795.
- Dousay, T. A., & Branch, R. M. (2023). *Survey of instructional design models*. Brill.
- Gizem, E., & Yildirim, Y. (2025). Developmental Assessment Practices in Early Childhood Education: An Analysis of Teacher Approaches and Strategies. *International Journal of Modern Education Studies, 9*(1).
- Grimmer, T. (2024). Is there a place for love in an early childhood setting? *Early Years, 44*(3–4), 525–538.
- Holden, L. R., & Tanenbaum, G. J. (2023). Modern assessments of intelligence must be fair and equitable. *Journal of Intelligence, 11*(6), 126.
- Keydeniers, D., Aalberse, S., Andringa, S., & Kuiken, F. (2022). Teacher-child interactions in bilingual daycares: Investigating the use of discourse strategies and teaching techniques. *Learning, Culture and Social Interaction, 37*, 100671.
- Lee, S., Adair, J. K., Payne, K. A., & Barry, D. (2022). Revisioning fairness as social justice in early childhood education. *Early Childhood Education Journal, 50*(7), 1083–1095.
- Montoya-Fernández, C., Losada-Puente, L., Gómez-Barreto, I. M., & Gil-Madrona, P. (2024). Developmental play-based assessment in early childhood education: a systematic review. *European Early Childhood Education Research Journal, 32*(5), 788–813.
- Mulyono, R., & Untung, S. (2025). Development of a Multiple Intelligence-Based Learning Management Model for Elementary Schools. *EduBase: Journal of Basic Education, 6*(2), 272–287.
- Neihart, M. (2021). *The social and emotional development of gifted children: What do we know?* Routledge.
- Nilsen, T. R., Bjørnstad, E., & Hannås, B. M. (2022). Utvalg og tilgjengelighet av lekemateriell og barns lekemuligheter i det fysiske innemiljøet i barnehagen. *Nordisk Barnehageforskning, 19*(3).

- Parker, R., Thomsen, B. S., & Berry, A. (2022). Learning through play at school— A framework for policy and practice. *Frontiers in Education, 7*, 751801.
- Salamon, A., & Palaiologou, I. (2022). Infants' and toddlers' rights in early childhood settings: Research perspectives informing pedagogical practice. *(Re) Conceptualising Children's Rights in Infant-Toddler Care and Education: Transnational Conversations*, 45–58.
- Shahaeian, A., Haynes, M., & Frick, P. J. (2023). The role of language in the association between theory of mind and executive functioning in early childhood: New longitudinal evidence. *Early Childhood Research Quarterly, 62*, 251–258.
- Shuey, E., Liberatore, V., Cadima, J., Coelho, V., Guedes, C., Nata, G., Jamet, S., & Radinger, T. (2021). *Starting strong VI: supporting meaningful interactions in early childhood education and care*.
- Souto-Manning, M., & Melvin, S. A. (2022). Early childhood teachers of color in New York City: Heightened stress, lower quality of life, declining health, and compromised sleep amidst COVID-19. *Early Childhood Research Quarterly, 60*, 34–48.
- Sugiyono. (2021). Quantitative, Qualitative and R & D Research Methods. *Journal of Chemical Information and Modeling, 53*(9).
- Wheeler, D. L., & Hill, J. C. (2024). The impact of COVID-19 on early childhood reading practices. *Journal of Early Childhood Literacy, 24*(1), 96–115.
- Yavich, R., & Rotnitsky, I. (2020). Multiple Intelligences and Success in School Studies. *International Journal of Higher Education, 9*(6), 107–117.
- Yılmaz, A., Şahin, F., Buldu, M., Ülker Erdem, A., Ezmeci, F., Somer Ölmez, B., Aydos, E. H., Buldu, E., Ünal, H. B., & Aras, S. (2021). An examination of Turkish early childhood teachers' challenges in implementing pedagogical documentation. *Early Childhood Education Journal, 49*(6), 1047–1059.