Validity and Reliability of the E-Aishi Interactive Professional Quality Module: A Scoping Review Using Cohen's Kappa with Expert Evaluation

Badhrianawati Haji Bidin , Romarzila Omar*

Department of Early Childhood Education, Faculty of Human Development, 35900, Tanjong Malim, Perak, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.1026EDU0081

Received: 20 January 2026; Accepted: 26 January 2026; Published: 16 February 2026

ABSTRACT

This scoping review examines the validity and reliability of the e-AISHI Interactive Professional Quality Module, a digital intervention designed to enhance the character and professional competencies of preschool teachers. Guided by the PRISMA-ScR framework and Arksey & O'Malley's six-stage methodology. This study maps existing literature on module validation specifically focusing on expert evaluation using the Cohen's Kappa statistical technique. Content validity was established by three subject matter experts in module development, counseling and ICT. The findings indicate that the e-AISHI module achieved a perfect inter-rater agreement (Kappa=1) demonstrating excellent initial validity and reliability compared to international studies that reported moderate to substantial agreement. While this perfect consensus underscores the clarity of the module's constructs, the study acknowledges that the score may reflect the homogeneity of a limited expert panel. Consequently, this validation serves as a foundational step within a broader Design and Development Research (DDR) framework with future phases intended to assess internal consistency and construct validity through empirical field testing. This review synthesizes methodological trends and gaps providing a rigorous framework for future validation studies in educational module development. Ultimately, the e-AISHI module serves as a scientifically grounded tool to support teacher professionalism and the achievement of Sustainable Development Goal (SDG) 4.

Keywords: e-AISHI Module, Validity, Reliability, Preschool Teachers, Cohen's Kappa, Scoping Review.

INTRODUCTION

The professional development and character building of preschool teachers are pivotal in ensuring the quality of early childhood education. As the first formal educators in a child's life, preschool teachers require not only pedagogical skills but also high professional integrity and emotional intelligence. However, despite the availability of various training modules, many lack rigorous scientific validation, particularly in the context of digital-based professional quality development.

To address this gap, the e-AISHI Interactive Professional Quality Module was developed as a digital intervention to strengthen the character and professional competencies of educators. In module development, ensuring validity and reliability is non-negotiable, as it determines whether the instrument accurately measures the intended constructs. One of the most robust statistical methods for measuring expert agreement in such validation processes is Cohen's Kappa.

While Cohen's Kappa is widely utilized in medical and technological fields to assess inter-rater reliability, its application in validating early childhood education (ECE) modules remains under-explored. Therefore, this study employs a scoping review guided by the PRISMA-ScR framework to map the existing literature on module validation and to report the validation findings of the e-AISHI module. By synthesizing current

methodological trends, this research aims to provide a standardized validation framework that can be adopted by future educational researchers to ensure the development of high-quality, reliable instructional materials.

Objective

1. To map existing literature on the use of expert evaluation in assessing the validity and reliability of educational modules, particularly in early childhood education.
2. To identify methods and approaches used to measure validity and reliability of modules, with a specific focus on the application of Cohen's Kappa statistical technique.
3. To analyze the extent to which expert consensus contributes to establishing the validity and reliability of the e-AISHI module.
4. To provide insights and recommendations for future studies on module development and validation processes in the field of teacher professional quality and character building.

Research Questions

1. What are the existing approaches used to evaluate the validity and reliability of educational modules in early childhood education?
2. How has Cohen's Kappa statistical technique been applied in assessing expert agreement on the validity and reliability of modules?
3. To what extent does existing literature suggest that expert evaluation contributes to establishing the validity and reliability of modules similar to e-AISHI?
4. What insights and implications can be drawn from existing studies to strengthen the validation process of the e-AISHI module?

Significance Of The Study

The significance of this study lies in its contribution to the methodological and practical framework of module development in early childhood education. Specifically this study provides:

1. Methodological Advancement: By focusing on the application of Cohen's Kappa. This study offers a clear and rigorous benchmark for researchers to quantify expert agreement. This addresses the common issue of subjectivity in expert evaluations, ensuring that the validity and reliability of educational tools are mathematically grounded.
2. Bridging the Gap in Early Childhood Education (ECE). While many modules exist, few are systematically validated for professional quality and character building. This study maps the existing literature to provide a foundational reference for ECE researchers and practitioners to develop high-quality, validated instruments like the e-AISHI module.
3. Enhancing Teacher Professionalism. The insights gained from this review offer practical implications for educators and policy-makers. By establishing a validated framework for the e-AISHI module, this study directly supports the improvement of teacher quality and the effectiveness of character building programs in preschools.
4. Practical Tool for Future Research. This review serves as a roadmap for future studies on digital and interactive module validation. It provides a synthesis of best practices and common pitfalls, helping future researchers streamline their validation processes and improve the overall standards of professional development modules.

MATERIAL AND METHOD

This study employs a scoping review design to map the existing literature on the validity and reliability of educational modules, with a specific focus on the application of Cohen's Kappa. The review process is strictly guided by the PRISMA Extension for Scoping Reviews (PRISMA-ScR) and the six-stage framework proposed by Arksey and O'Malley (2005).

Search Strategy and Data Sources

To identify relevant literature, a comprehensive search was conducted across three primary academic databases: Web of Science (WoS), Scopus, and ERIC. The search strategy utilized a combination of keywords, including *"Validity," "Reliability," "Module,"* and *"Cohen's Kappa."* The selection was restricted to articles published between 2024 and 2025 to capture the most recent methodological advancements.

Inclusion and Exclusion Criteria

Articles were screened based on the following criteria: (i) studies focusing on module validation or instrument reliability, (ii) research utilizing Cohen's Kappa as a statistical measure, and (iii) articles published in English or Malay. Studies lacking empirical data on reliability or those unrelated to education and professional development were excluded.

The e-AISHI Module Validation Framework

In addition to the literature synthesis, this study reports the empirical validation of the e-AISHI Interactive Professional Quality Module. The development and validation process followed a rigorous multi-methodological approach:

- Needs Analysis: Conducted using the Statistical Package for the Social Sciences (SPSS) Version 27 to identify the developmental requirements of the module.
- Content Validity: Established through the Fuzzy Delphi Method (FDM) to achieve a high level of expert consensus.
- Reliability Assessment: Quantified via Cohen's Kappa to evaluate inter-rater agreement and item consistency among three subject matter experts (SMEs) in module development, counseling, and ICT.
- Usability Testing: Evaluated using the Nominal Group Technique (NGT) to capture the perspectives and practical needs of end-users.

This integrated methodology ensures that the e-AISHI module is not only scientifically robust but also practically applicable for preschool teachers. Beyond the content validity and inter-rater reliability reported in this study, the e-AISHI module is designed within a comprehensive Design and Development Research (DDR) framework. Expert validation through Cohen's Kappa is a fundamental prerequisite before proceeding to test other psychometric properties. Future stages of this research will include assessing internal consistency (e.g., Cronbach's Alpha) and construct validity through pilot testing with the target end-users (preschool teachers). This ensures that the module is not only theoretically sound but also empirically consistent when implemented in real-world settings.

This research article employs the six-step methodology developed by Arksey and O'Malley (2005) as explained in the below section:

Step 1: Identifying the Research Question. This study investigates the validity and reliability of the e-AISHI Interactive Professional Quality Module. By employing a scoping review and expert evaluation via Cohen's Kappa, the research aims to establish a rigorous methodological foundation for the module which is specifically designed to enhance the professional quality and character building of preschool teachers. To guide the literature search, the following research questions were formulated:

1. To what extent does the e-AISHI module demonstrate content validity based on expert evaluation using Cohen's Kappa?
2. What is the level of inter-rater reliability among experts for the module as measured by the Cohen's Kappa index?

3. What does the existing scientific literature reveal regarding the methodologies used for validating interactive professional modules in Early Childhood Education (ECE)?

Step 2: Identifying Relevant Research. To ensure a comprehensive review, a systematic search was conducted across three primary databases: Web of Science (WoS), Scopus and ERIC. These selected for their extensive indexing of high-quality professional development and educational research. The search strategy employed specific keywords and Boolean operators across three thematic clusters:

1. Instructional Context: Interactive module, online professional development, online PD.
2. Target Population: Preschool teachers, early childhood education.
3. Methodological Focus: Validity, reliability and Cohen's Kappa. By synthesizing these terms, relevant literature concerning the e-AISHI module and teacher competency was identified. The selection process is visually summarized in Figure 1.

Step 3: Selecting Relevant Studies. To ensure the quality and relevance of the literature, specific inclusion and exclusion criteria were established for this scoping review. First, the scope was restricted to publications within the educational context. Second, only original research articles were eligible for inclusion; consequently, conference papers, book chapters, conference reviews, and books were excluded to maintain the rigor of the analysis. Third, the selection was limited to English-language papers published between 2024 and 2025. This selection process, which ensures the review remains current and focused, is integrated into the workflow illustrated in Figure 1.

Step 4: Charting the Data. To facilitate thematic and comparative analysis, the selected data were managed and organized using Microsoft Excel. A comprehensive charting process was undertaken to extract key information, which is presented in a structured table. This table includes essential details such as the author's name, year of publication, studied variables, research implications, and identified major and sub-themes. These organized data points provide the necessary evidence to address the research questions formulated in Step 1 and offer a graphical representation of the findings where applicable.

Step 5: Collating, Summarizing, and Reporting the Results. The gathered data from 9 selected articles were reorganized into specific topics, major themes and sub-themes identified through the thematic and comparative analysis conducted in Step 4. Each article is systematically classified and discussed based on these themes as detailed in Table 1.

Step 6: Discussion and Conclusion. In this final stage, the research outcomes are discussed in relation to the study's primary objectives. Given that the development of the e-AISHI Interactive Professional Quality Module is a relatively new area for educators, the discussion emphasizes how the scoping review and expert evaluation via Cohen's Kappa contribute to its validity and reliability. Furthermore, the practical limitations of the study are addressed, alongside strategic recommendations for future research to further refine the module's effectiveness. The process concludes with a summary of the study's contributions to the field of Early Childhood Education in the conclusion section.

Validity And Reliability

Validity and reliability are fundamental processes used to ensure that each item construct in an instrument is appropriate for addressing the research questions. In this study, validity is defined as the instrument's capacity to measure intended constructs accurately, thereby ensuring its defensibility and utility. Meanwhile, reliability refers to the consistency and stability of the e-AISHI module across different expert evaluations. Content validity was established through rigorous expert judgment while inter-rater reliability was quantified using the Cohen's Kappa statistical technique.

Cohen's Kappa Statistical Analysis

The Cohen's Kappa coefficient (K) is employed to measure the level of agreement between raters, accounting

for the possibility of agreement occurring by chance. The calculation for the e-AISHI module comprising 19 items evaluated by three experts are based on the following formula:

Cohen's Kappa Calculation Formula For E-Aishi Module.

The Kappa coefficient is calculated using the following formula:

$$K = \frac{Po - Pe}{1 - Pe}$$

Where:

(Po) : Observed agreement value between raters.

(Pe) : Expected agreement value by chance

N : Number of experts (3)

Total Items: 19 items.

Calculation of Observed Agreement (Po):

$$Po = \frac{\text{Number of items agreed upon by all raters}}{\text{Total Numbers Of Items}}$$

$$Po = \frac{19}{19}$$

$$= 1$$

The calculation of Observed Agreement (Po) above represents the actual level of consensus among the raters for the items evaluated. A value of 1 indicates perfect agreement among the experts across all items.

Calculation of expected agreement (Pe)

There are two response categories:

1. Agree (A)
2. Disagree (D)

In this study, the researchers observed the following distribution:

- Rater 1: 17 Agree and 2 Disagree.
- Rater 2: 18 Agree and 1 Disagree.
- Rater 3: 18 Agree and 1 Disagree.

Probability Calculation for Expected Agreement (Pe):

Agree (A)

Rater 1: $\frac{17}{19} = 0.89$

Rater 2: $\frac{18}{19} = 0.95$

Rater 3: $\frac{18}{19} = 0.95$

19

Disagree (D)

Rater 1: 2 = 0.11

19

Rater 2: 1 = 0.05

19

Rater 3: 1 = 0.05

19

In this context, the value of (Pe) is recorded as 0.8003.

Final Kappa Calculation:

$$K = \frac{Po - Pe}{1 - Pe}$$

$$= \frac{1 - 0.8003}{1 - 0.8003}$$

$$= \frac{0.1997}{0.1997}$$

$$= 1$$

Table 1. Kappa Value Table and Interpretation (Level)

	Kappa Value	Interpretation
Kurang daripada 0		Very weak
0.00 – 0.20		Weak
0.21 – 0.40		Moderately weak
0.41 – 0.60		Medium
0.61 – 0.80		Good
0.81 – 1.00		Very good

Source: (Cohen, 1960; Merriam, 2009; Mokhtar, 2011 Othman et al., 2018)

Table 2. Expert Kappa Values for e-AISHI Module

	Expert	Kappa Value
Expert 1		1
Expert 2		1
Expert 3		1

The calculation yielded a perfect Kappa score (K=1), indicating absolute consensus among the three experts regarding the module's item clarity and relevance. While this number of raters (N=3) meets the minimum requirements for initial content validation, the researchers acknowledge that such a result may reflect a homogeneity of perspectives. Consequently, these findings are treated as a foundational validation phase providing a basis for further reinforcement in subsequent development stages to minimize potential subjective bias.

FINDINGS

Explain about the PRISMA table.

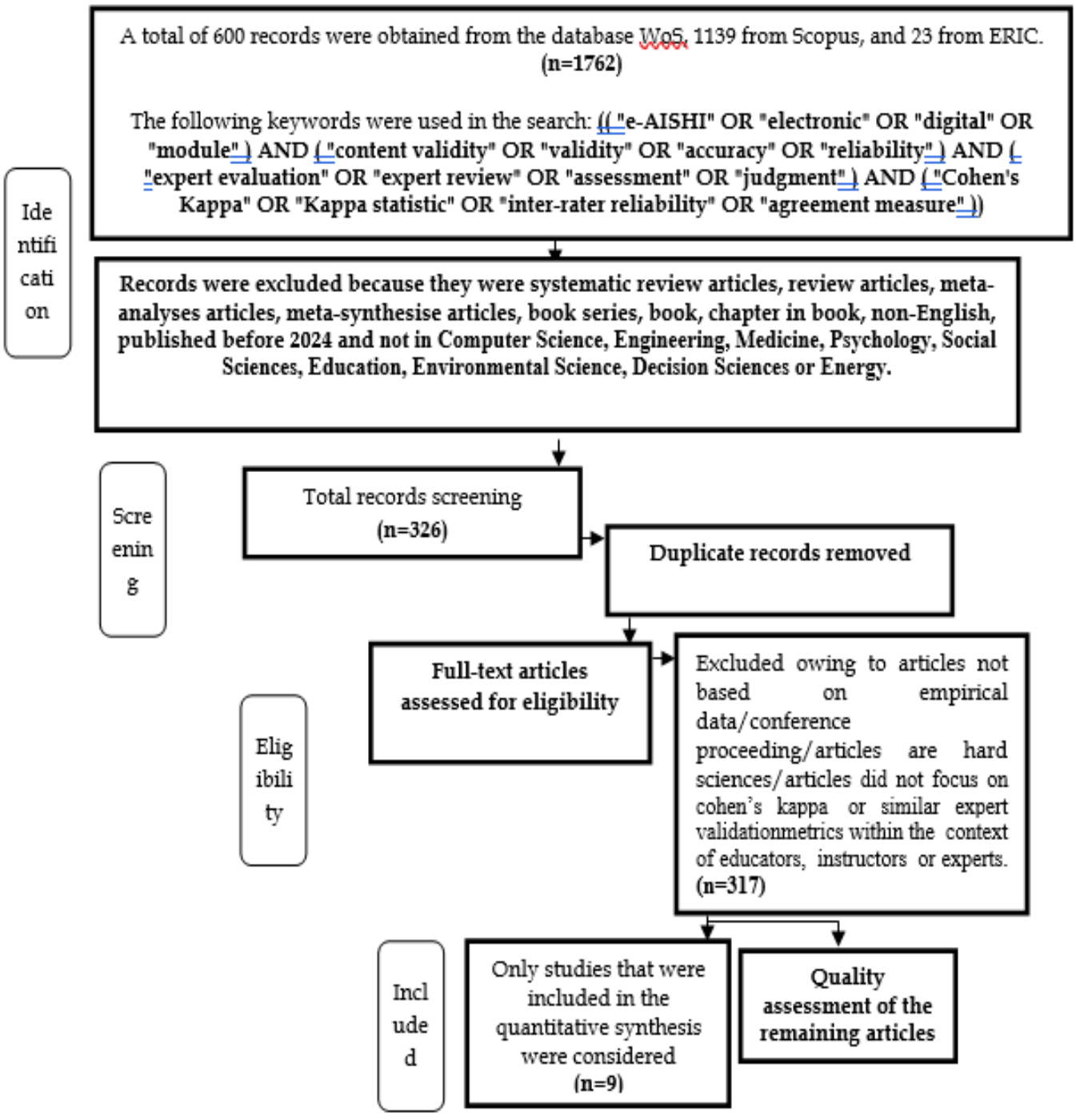


Figure 1: Flow diagram of research selection process using Preferred Reporting Items for Systematic Reviews (PRISMA) adapted from a study by Moher et al. (2015)

MAIN FINDINGS

A total of 9 articles met the inclusion criteria for this scoping review. These studies were sourced from Scopus, Web of Science (WoS), and ERIC, providing a global perspective on the application of Cohen's Kappa in module validation. The perfect agreement reached by the e-AISHI module (Kappa = 1) stands as a significant finding when compared to the varying levels of agreement (moderate to excellent) reported in the other 8 international studies.

Table 3 Charting the data

Publication Location	Variables/ Construct	Validity and Reliability of Module	Sub-Theme	Theme
Scopus Johar R et al. (2025)	Independent: E-module based on Realistic Mathematics Education (RME) approach. Mediating: Media and material experts' evaluation (Plomp's Model). Dependent: Validity, practicality and effectiveness (learning outcomes) of the e-module.	The e-module demonstrated high validity based on expert evaluations, achieving mean scores of 4.68 from media experts and 4.8 from material experts. In terms of reliability, the inter-rater agreement was classified as substantial, with a Cohen's Kappa value of kappa=0.654. Furthermore, the effectiveness of the module was statistically significant as evidenced by a notable difference in learning outcomes (p=0.015) and a moderately strong effect size (Cohen's d=0.65).	RME-Based Digital Resource Validation	Comprehensive Validation of Interactive e-Modules
Scopus Moon Flip et al. (2025)	Independent: Checkbox grading Mediating: Multiple Assessors Dependent: Large-scale mathematics exams	This paper presents 'checkbox grading,' a digital method where exam designers have predefined checkboxes with both feedback and associated partial grades. Assessors then tick the checkboxes relevant to a student solution. Dependencies between checkboxes ensure consistency among assessors in following the grading scheme. Moreover, the approach supports 'blind grading' by hiding the grades associated with the checkboxes, thus focusing assessors on the criteria rather than the scores. Results show that assessors perceived	Digital Assessment Consistency and Feedback Mechanisms	Innovative Digital Evaluation Methodologies

		checkbox grading as very useful.		
Scopus Geara et al. (2025)	Independent: Key clinical variables in the GynOp Register (e.g., uterine weight, delivery date, and ASA classification). Mediating: Single-blinded multicenter audit process comparing register entries against source medical records. Dependent: Internal validity and data quality of the national gynecological surgery registry.	The study found high overall validity, with exact agreement reaching 100% for pathological diagnosis and 97% for perineal rupture. For numerical data, the correlation was perfect at 1.00. However, ordinal variables such as ASA-classification and genital prolapse stage showed only moderate agreement through Cohen's Kappa calculation, indicating specific areas in the registration module that require further standardization and user education.	Audit-Based Validation of National Registries.	Methodological Rigor in Digital Assessment.
Web of Science Lin, Y et al. (2025)	Independent: Assessment methods: Digital palpation (Modified Oxford Scale) and Transperineal Ultrasound (TPUS). **Mediating:** Variable Independent blinded raters (two clinicians and two ultrasound raters) and the process of quantifying structural displacement (e.g., proximal urethral axis angle change).	The results indicated that TPUS demonstrated high inter-rater reliability (ICC=0.92), whereas digital palpation achieved only a moderate level of agreement (Cohen's Kappa, kappa=0.55). Despite this difference, a strong and statistically significant correlation was observed between the two assessment methods ($r=0.727$, $p<0.001$).	Comparative Reliability of Objective and Subjective Measures	Methodological Rigor in Digital Assessment

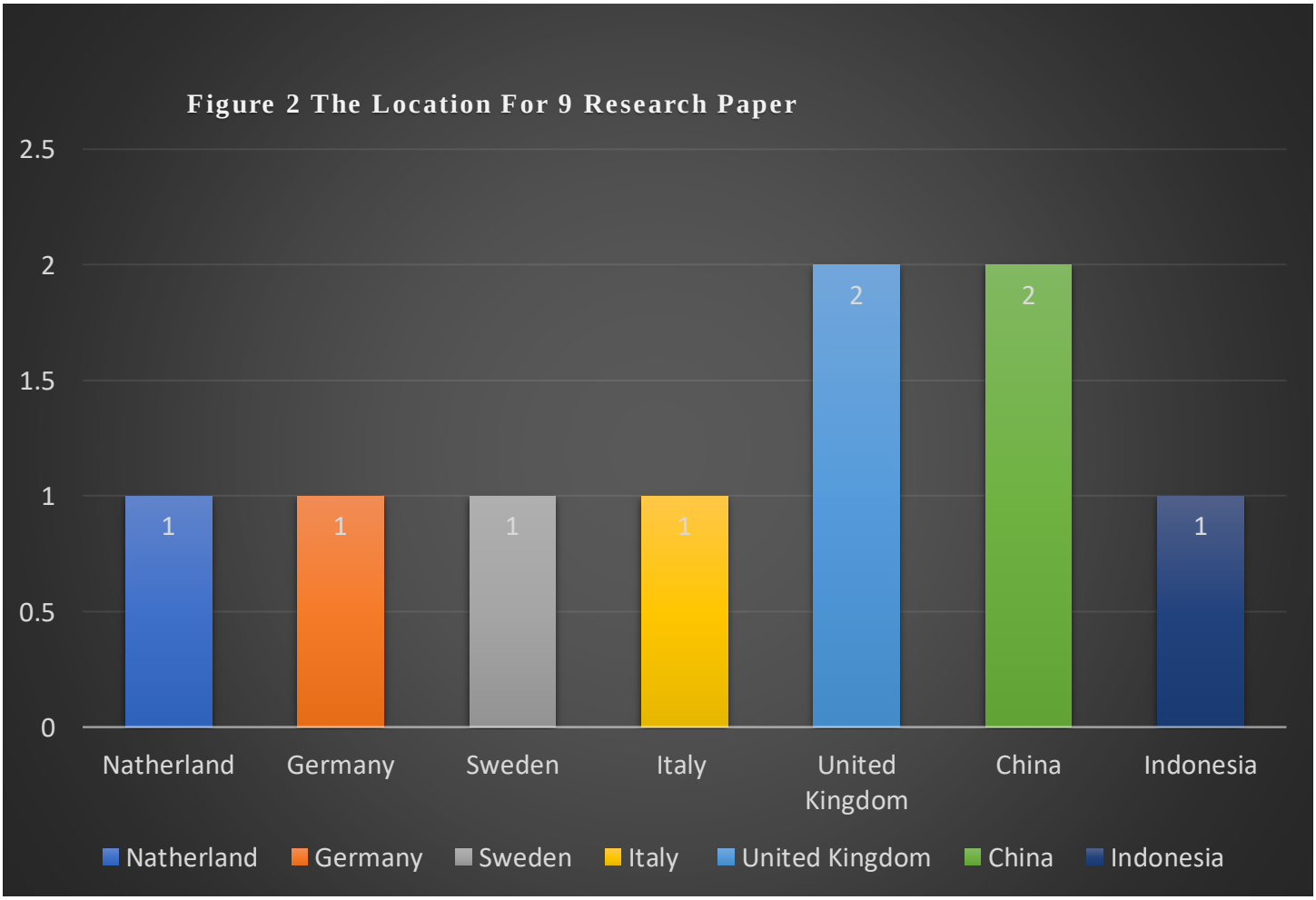
	Dependent: Accuracy and reliability of pelvic floor muscle contractility assessment.			
Web of Science D'Abbronzo et.al (2025)	Independent: Reticulin Fibrosis Assessment Mediating: Digital pathology system & Expert Annotation Process Dependent: An Automated Artificial Intelligence-Based Tool	This study utilized 86 bone marrow biopsy specimens from patients diagnosed with Philadelphia chromosome-negative myeloproliferative neoplasms, collected between 2018 and 2023. A fully convolutional network based on the InceptionV3 architecture was trained to assess fibrosis grades (MF0–MF3) from whole slide images of reticulin-stained sections. The model was trained using 3814 annotated images and validated using a separate set of 40 BMBs. The algorithm's performance was evaluated by comparing its fibrosis grading to expert hematopathologists' assessments, yielding a Cohen's kappa coefficient of 0.831, indicating excellent agreement.	AI-Driven Agreement and Expert Validation	Automated Evaluation Systems •
ERIC Messer M. et al. (2025)	Independent: Human grading with shared rubrics. Mediating: Multiple assessors (Teaching Assistants) and assignment batches. Dependent: Grading consistency (Inter-rater & Intra-rater reliability) in programming assignments.	Measured inter-rater reliability using Krippendorff's alpha- an [alpha] = 0.2 for correctness. Intra-rater reliability was tested via duplicated assignments, revealing high individual inconsistency (average grade difference > 1.5).	Assessor Inconsistency and Rubric Limitations	Human Factors in Assessment Reliability

ERIC Radovic, Niels Seidel (2024)	Independent: Self-Regulated Learning Support (SRL-S) rubric. Mediating: International and interdisciplinary experts' reviews. Dependent: Assessment of SRL support in technology-enhanced learning environments.	Content, criterion-related, and construct validation were performed through expert reviews. Reliability was established using Intraclass Correlation Coefficients and Cohen's Kappa tests (inter-rater and intra-rater).	Rubric Development and Multi-Dimensional Validation	Theoretical-Based Assessment Tools
Scopus Naomi Dale et al. (2024)	Independent: An Autism Observational Assessment Including ADOS-2® Mediating: Modification and Validation Dependent: Children with Visual Impairment	The primary objective of the study is to validate a modified version of the Autism Diagnostic Observation Schedule (ADOS-2®, Module 3), for children with VI. A cross-sectional observational study was undertaken with 100 (mean 5½ years, SD 10.44 months, range 4–7 years; 59 (59%) males) children with congenital disorders of the peripheral visual system with moderate/severe-profound VI. As the primary objective, 83 (83%) who were 'verbally fluent' were assessed with the modified ADOS-2® (Module 3). Their scores were investigated for reliability, construct and criterion validity against expert clinician formulation and parent-rated social and communication questionnaires (Social Responsiveness Scale-2, SRS-2; Children's Communication Checklist-2).	Modified Observational Tools for Vulnerable Populations	Inclusive Clinical Assessment
Web of Sciences Wang, J. X., Zhang, Q. T., et al. (2024).	Independent: Enhanced dual-channel model (EfficientNetB7, SAM, Unet++). Mediating:	The model's reliability was validated through multiple metrics: Cohen's Kappa coefficient (0.9185 and 0.8835) indicating almost perfect agreement, Overall Accuracy (91.53% - 98.92%) and IoU	Deep Learning-Based Automated Validation	Technological Integration in High-Precision Assessment

	Transformers Module, Deep separable CNN, and Variational Autoencoder (VAE). <u>Dependent:</u> Landslide detection accuracy and region extraction (segmentation) precision.	(0.8631).		

Background Of The Research

Figure 2 depicts the location for the research on Cohen Kappa in fixing the validation and reliability the validation and reliability to highlight the context of the research.



DISCUSSION

The mapping of previous studies as synthesized in this review, reveals that Cohen's Kappa is a highly versatile and rigorous statistical tool across a wide range of disciplinary domains. Analysis across fields such as Computer Science, Engineering, Medicine, Psychology and Education demonstrates that this technique remains the gold standard for validating the consistency and reliability of instruments, modules and developmental models. Conceptually, the role of Cohen's Kappa in strengthening the integrity of educational modules can be categorized into three primary dimensions: the process dimension, the outcome dimension and the enabler dimension.

Table 4 Shows how Cohen Kappa determine the validation and reliability of module as a concept.

	Independent	**Mediating**	**Dependent**
Process	Traditional vs. Digital Assessment Methods (e.g., Manual palpation vs. Ultrasound or Human grading vs. AI).	Blinded Expert Review and Expert Annotation Procedures (e.g., Using independent raters to mitigate bias).	Objective levels of Inter-rater Agreement.
Outcome	Draft Instructional Modules or Diagnostic Tools (e.g., e-AISHI module, RME e-module, or SRL-S Rubric).	Cohen's Kappa Statistical Analysis (Measuring agreement beyond the element of chance).	Scientifically proven Content Validity and instrument reliability.
Enabler	Technological Integration and Digital Platforms (e.g., InceptionV3 AI system or digital diagnostic platforms).	Standardization of Evaluation Criteria (e.g., Use of shared rubrics or digital grading checkboxes).	Assessment consistency and reduction of human subjectivity in the evaluation process.

In the context of the process dimension, the application of Cohen's Kappa serves as a bridge connecting subjective human judgment with objective digital measurements. This is clearly illustrated in studies comparing traditional and digital assessment methods such as the comparison between manual palpation and ultrasound. In this process, procedures such as blinded expert reviews act as critical mediating factors to minimize rater bias. Through this systematic approach, the degree of inter-rater agreement can be measured objectively thereby ensuring that the internal validity of an instrument is preserved before it is widely implemented.

Furthermore, Cohen's Kappa acts as a primary determinant in the outcome dimension particularly in assessing a module's readiness for implementation. For the e-AISHI module, achieving a perfect Kappa score (K=1) categorizes its validity and reliability as 'Excellent.' This finding is significant when compared to other

instructional or diagnostic frameworks reviewed, such as the RME-based e-module (K = 0.654) or AI-driven diagnostic tools (K = 0.831). However, the researchers interpret this absolute agreement with caution. Given that the expert panel was limited to three individuals (N = 3), this score may be influenced by a homogeneity of perspectives. Consequently, these results are treated as a foundational validation phase that will be further reinforced in subsequent development stages by involving a broader and more multidisciplinary panel of experts.

Finally, the enabler dimension illustrates how digital integration facilitates higher levels of consensus. The use of automated tools and standardized rubrics as seen in high-precision landslide monitoring and AI-driven pathology models proves that technology effectively mitigates human inconsistency in the evaluation process. By integrating these digital enablers, the e-AISHI module not only enhances the methodological rigor of the validation process but also ensures that this character-building intervention for preschool teachers is consistent, impactful and scientifically sound.

Overall, the integration of Cohen's Kappa within the e-AISHI module is not merely a statistical procedure but a deliberate effort to ensure this digital module meets global standards for quality education, aligned with the objectives of Sustainable Development Goal (SDG) 4.

LIMITATIONS AND RECOMMENDATIONS

Despite the rigorous methodological approach, this study acknowledges certain limitations that provide avenues for future scholarly inquiry.

Limitations

Despite its contributions, this study acknowledges several limitations that should be considered when interpreting the findings. Primarily, the research relied exclusively on three major academic databases: Web of Science (WoS), Scopus and ERIC. While this deliberate focus ensured the inclusion of high-quality and peer-reviewed literature, it may have inadvertently excluded relevant studies indexed in supplementary, regional or grey literature databases that could provide additional perspectives on module validation.

Furthermore, the scoping review was restricted to a specific timeframe (2024–2025). This window was strategically selected to capture the most recent methodological trends in integrating digital platforms and interactive modules into educational assessment. However, the researchers recognize that such a narrow timeframe limits the synthesis of foundational longitudinal data and seminal theories that shaped the application of Cohen's Kappa in previous decades.

The reliance on a targeted expert panel also presents a constraint while the perfect Kappa score (K=1) indicates strong initial consensus, it reflects the perspectives of a specific group of three raters. These limitations however present a significant opportunity for future researchers to conduct broader systematic reviews or meta-analyses that compare the evolution of validation standards transitioning from traditional pedagogical methods to modern AI-driven interactive platforms.

Recommendations For Future Research

To build upon the findings of this study, the following recommendations are proposed:

- Expansion of Search Scope: Future research should integrate a broader array of data sources, including grey literature and alternative databases such as Google Scholar or ProQuest. This would capture a more diverse range of emerging validation methodologies and a more expansive panorama of inter-rater reliability applications.

- Longitudinal Empirical Assessment: Subsequent studies should move beyond the initial expert validation phase to investigate the long-term effectiveness of the e-AISHI module. Longitudinal

research is essential to assess the actual impact of the module on preschool teachers' professional quality and character development within real-classroom settings.

- Comparative Analysis: Further studies could compare the effectiveness of Cohen's Kappa with other statistical measures, such as the Content Validity Index (CVI) or Intra-class Correlation (ICC) to provide a more multi-dimensional perspective on module reliability.

- Empirical Field Testing: It is strongly recommended that subsequent studies move beyond expert evaluation to conduct empirical field tests with preschool teachers to measure the module's actual impact on professional behavior and classroom integrity.

- Methodological Triangulation: Future research should consider comparing the effectiveness of Cohen's Kappa with other metrics such as the Content Validity Index (CVI) or Intra-class Correlation (ICC) to provide a more multi-dimensional perspective on instrument reliability.

Implications

This research elucidates the critical nexus between methodological rigor and expert consensus in validating educational modules, particularly within the digital landscape. The findings of this scoping review offer several profound implications for teacher professional development and the standards of instrument validation in early childhood education (ECE).

Methodological Standardization

The study underscores the urgent need for standardized validation procedures in educational research. Historically, the validation of instructional tools has frequently relied on qualitative expert feedback which is inherently susceptible to subjectivity. By establishing Cohen's Kappa as a robust metric for inter-rater reliability, this research provides a roadmap for scholars to adopt objective and mathematically grounded frameworks. Such precision ensures that developed modules are not merely collections of content but scientifically verified instruments capable of delivering consistent educational outcomes.

Advancement of Teacher Professional Quality

The successful validation of the e-AISHI module with perfect inter-rater agreement (kappa = 1) carries significant practical implications. In an era where preschool teachers are the primary architects of a child's character, the availability of a validated digital tool grounded in integrity and empathy is indispensable. This result implies that character building interventions can achieve a high degree of expert consensus when developed through systematic phases such as the Design and Development Research (DDR) framework employed in this study.

Technological Integration and Scalability

The review highlights the escalating importance of technological integration in pedagogy. The transition toward interactive digital modules suggests that digital platforms offer superior reproducibility and accessibility compared to traditional methods. For the e-AISHI module, this means professional development can be delivered reliably across diverse geographical locations, effectively mitigating the inconsistencies often associated with human-led training sessions.

Alignment with Global Educational Goals

On a broader scale, these implications align with the global pursuit of Sustainable Development Goal (SDG) 4 which advocates for quality education and the strengthening of teacher professionalism. By providing a rigorous methodological roadmap, this research empowers practitioners and policymakers to implement

validated, high-impact training resources that are essential for the holistic development of both educators and students.

CONCLUSION

The development of the e-AISHI Interactive Professional Quality Module represents a significant step forward in integrating rigorous scientific validation with digital professional development for preschool teachers. Through a systematic scoping review and a robust multi-methodological approach including FDM, NGT and Cohen's Kappa, this study has established that the e-AISHI module possesses exceptional validity and reliability. The perfect inter-rater agreement (Kappa = 1) attained in this study underscores the module's readiness for practical implementation.

The implications of this study extend beyond academic discourse. The e-AISHI module provides a proven empirical foundation that can be utilized by the Ministry of Education, teacher training institutions and other educational stakeholders as a benchmark for designing holistic and high-impact intervention programs. By focusing on character building and professional values, this module directly supports the achievement of Sustainable Development Goal (SDG) 4: Quality Education. It ensures that preschool educators are equipped with the integrity and competencies necessary to provide high-quality early childhood education, ultimately fostering a more ethical and professional teaching workforce.

Future research should focus on longitudinal studies to evaluate the long-term impact of the e-AISHI module on teacher performance and student outcomes in diverse educational settings.

REFERENCES

1. Badhrianawati, H. B., & Romarzila, O. (2025). The Design Of E-Aishi Interactive Professional Quality Module To Preschool Teacher Character Development Using Design And Development Research (DDR) In Early Childhood Education. Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia.
2. D'Abbronzo, G., et al. (2025). Development Of An Automated Artificial Intelligence-Based Tool For Reticulin Fibrosis Assessment In Bone Marrow Biopsies. Università Degli Studi Della Campania Luigi Vanvitelli, Via Luciano Armanni 5, 80138 Naples, Italy. Journal of Clinical Pathology. Web of Science.
3. Dale, N., et al. (2024). Modification and Validation of an Autism Observational Assessment Including ADOS-2® for Use with Children with Visual Impairment. Great Ormond Street Hospital NHS Trust, Great Ormond Street, London WC1 3JH, UK**.** Scopus.
4. Geara et al. (2025). Assessment of the internal validity of the Swedish National Quality Register in Gynecological Surgery (GynOp). Division of Obstetrics and Gynecology at Södersjukhuset, Stockholm, Sweden. Scopus.
5. Johar, R. (2025). Development Of Algebraic Expressions E-Modules Through Realistic Mathematics Education Approach. Department Of Mathematics Education, Universitas Syiah Kuala, Aceh, Indonesia Scopus.
6. Lin, Y., et al. (2025). Transperineal ultrasound versus digital palpation: Identifying key parameters for objective pelvic floor muscle contraction assessment. Fuzhou University Affiliated Provincial Hospital, Fuzhou, China. Web of Science.
7. Messer, M., et al. (2025). How Consistent Are Humans When Grading Programming Assignments?. King's College London, London, United Kingdom of Great Britain and Northern Ireland. *ERIC*.
8. Moon Flip, et al. (2025). Innovative Digital Evaluation Methodologies: Checkbox Grading For Large-Scale Mathematics Examinations. Freudenthal Institute, Utrecht University, the Netherlands. Scopus.
9. Radović, S., & Seidel, N. (2024). Development And Validation Of The Self-Regulated Learning Support (SRL-S) Rubric In Technology-Enhanced Environments. Center of Advanced Technology

for Assisted Learning and Predictive Analytics (CATALPA), Fern Universität in Hagen, Germany. *ERIC*.

10. Wang, J. X., Zhang, Q. T., et al. (2024). Enhanced Dual-Channel Model-Based With Improved Unet++ Network For Landslide Monitoring And Region Extraction.
Faculty of Arts and Sciences, Beijing Normal University, Zhuhai 519087, China. Web of Science.