Jurnal Ilmiah Profesi Pendidikan
Volume 10, Nomor 2, Mei 2025
ISSN (Print): 2502-7069; ISSN (*Online*): 2620-8326

Validation of Pre-Service Science Teacher Artificial Intelligence Competence Self-Efficacy (AICS): Rasch Model Analysis

Hilman Qudratuddarsi[1]*, Wahyuni Adam[1], Dyah Puspitasari Ningthias[2], Aulia Rahmadhani[1], Evy Noviana[3]
[1]Program Studi Pendidikan IPA, Jurusan Pendidikan MIPA, FKIP, Universitas Sulawesi Barat, Prof. Dr. Baharuddin Lopa street, Majene, West Sulawesi, 91412. Indonesia
[2] Program Studi Pendidikan Kimia, Jurusan Pendidikan MIPA, FKIP, Universitas Sulawesi Barat, Majapahit Street No. 62, Mataram West Nusa Tenggara, 83125. Indonesia
[3]Program Studi Pendidikan Biologi, Jurusan Pendidikan MIPA, FKIP, Universitas Sulawesi Barat, Prof. Dr. Baharuddin Lopa street, Majene, West Sulawesi, 91412. Indonesia
*Corresponding Author: hilman.qudratuddarsi@unsulbar.ac.id

Article History
Received : April 06th, 2025
Revised : April 27th, 2025
Accepted : May 15th, 2025

Abstract: Artificial Intelligence (AI) is transforming science education through virtual labs, intelligent tutoring, and adaptive assessments. However, pre-service teachers often lack formal training in AI integration. This study aims to validated the Artificial Intelligence Competence Self-Efficacy (AICS) instrument using Rasch model, covering AI knowledge (AIK), AI Pedagogy (AIP), AI Assessment (AIA), AI Ethics (AIE), Human-Centred Education (HCE), and Professional Engagement (PEN). This study used a quantitative survey with 338 third-year pre-service science teachers selected through convenience sampling. Data were collected via Google Forms where ethical considerations and back-translation ensured data integrity. Data were analyzed through reliability, separation, item fit statistics, unidimensionality and Differential Item Functioning (DIF). The findings indicate that the AICS instrument is psychometrically sound, with high reliability (person reliability = 0.94, item reliability = 0.95) and excellent separation indices. The Wright Map showed that item difficulty was well-aligned with participant ability, effectively capturing various levels of AI self-efficacy. Item fit statistics confirmed all items functioned within acceptable ranges, and unidimensionality analysis supported the measurement of a single, coherent construct. DIF analysis showed minimal gender bias, though one item (AIP1) favored males. Overall, the instrument is valid and reliable for being used to assess AI competence self-efficacy among pre-service science teachers.

Keywords: Artificial Intelligence, Instrument Validation, Pre-service science teacher, Self-efficacy, Rasch Model

INTRODUCTION

Artificial Intelligence (AI) has rapidly emerged as a transformative force across various sectors, including education. With its ability to automate tasks, personalize learning, enhance assessment practices, and facilitate access to diverse resources, AI is becoming increasingly integrated into teaching and learning environments (Wang & Huang, 2025). Educational technologies powered by AI, such as adaptive learning systems, automated grading tools, intelligent tutoring systems, and content generation platforms (e.g., ChatGPT), have reshaped pedagogical approaches and expanded instructional possibilities (Ning, Zhang, Xu, Zhou & Wijaya, 2024). In particular, AI is playing a vital role in supporting teachers to improve efficiency and effectiveness while enabling students to engage with content in more meaningful and tailored ways (Qudratuddarsi, Fauziah, Agung & Yanti, 2025). As AI continues to permeate the educational landscape, it is essential to ensure that educators, especially pre-service teachers, are equipped with the necessary competence to navigate, implement, and critically assess AI-based tools in their future classrooms.

In the domain of science education, AI has demonstrated significant potential to transform the teaching and learning process. Applications such as AI-driven virtual laboratories, intelligent tutoring systems, automated assessment tools, and simulation-based learning environments are becoming increasingly prevalent (Jia, Sun & Looi, 2024). These technologies support students in developing scientific inquiry skills, conducting virtual experiments, and receiving immediate, personalized feedback. AI also facilitates the creation of adaptive learning pathways by analyzing individual student performance and tailoring scientific content to meet their unique needs (Kavitha & Joshith, 2024). For science educators, this presents new opportunities to implement differentiated instruction, promote inquiry-based learning, and foster higher student engagement in complex scientific concepts. However, the responsible and effective integration of AI in science education demands a strong foundation in AI literacy, pedagogical flexibility, and ethical considerations—critical competencies that must be developed during teacher training programs (Dewi, Qudratuddarsi, Ningthias & Cinthami, 2024).

Given the growing presence of AI in classrooms, a pertinent question arises: are pre-service teachers ready to apply AI in their teaching practices? This question is particularly significant as today's teacher candidates, primarily belonging to Generation Z, are digital natives with high exposure to technology, yet often lack formal training in AI integration (Karataş & Ataç, 2024). Many teacher education programs are still in the early stages of embedding AI-related content into their curricula. (Hava & Babayiğit, 2025). Consequently, pre-service teachers may have limited opportunities to explore the pedagogical, ethical, and practical implications of AI in education. To bridge this gap, it is essential to assess their readiness, competence, and confidence in using AI tools for instructional purposes. This assessment can provide valuable insights into the support systems, training modules, and professional development efforts needed to prepare future educators for the AI-enhanced classroom environment (Tram, 2024).

In response to this emerging need, the development and validation of a reliable and comprehensive instrument to measure pre-service teachers' AI competence self-efficacy is both timely and vital (Oved & Alt, 2025). Self-efficacy, defined as an individual's belief in their ability to perform a specific task or activity, plays a crucial role in shaping teachers' intentions, behaviors, and persistence in adopting new technologies (Wang & Chuang, 2024). Measuring AI competence self-efficacy allows researchers and teacher educators to identify strengths and areas for growth, tailor training interventions, and monitor progress over time. Moreover, a validated instrument ensures that the data collected is accurate, meaningful, and aligned with theoretical frameworks of technology acceptance and educational innovation (Chou, Shen, Shen & Shen, 2024).

The instrument validated in this study, the Artificial Intelligence Competence Self-Efficacy (AICS) scale, encompasses six key dimensions reflecting the multifaceted nature of AI integration in education. These dimensions are AI Knowledge (AIK), AI Pedagogy (AIP), AI Assessment (AIA), AI Ethics (AIE), Human-Centred Education (HCE), and Professional Engagement (PEN). Each construct is grounded in contemporary research and designed to capture distinct yet interconnected aspects of AI competence among pre-service science teachers. AI Knowledge (AIK) focuses on foundational understanding and practical exploration of AI tools. It includes the ability to identify AI-driven applications, experiment with AI-generated content, and articulate basic AI concepts. This dimension is critical at the early stages of teacher preparation, as it sets the groundwork for more advanced applications of AI in pedagogy and assessment. AI Pedagogy (AIP) addresses the integration of AI tools into teaching and learning scenarios. Items in this dimension assess participants' capacity to envision and plan AI-supported instruction, identify subject-relevant tools, and engage in collaborative discussions about pedagogical strategies involving AI. This reflects the growing emphasis on hypothetical and practice-based applications of technology in teacher education. AI Assessment (AIA) captures pre-service teachers' understanding of how AI can support assessment for and of learning. It includes designing AI-assisted assessments, leveraging AI for feedback and self-evaluation, and reflecting on the role of AI in monitoring student progress. This is particularly important given the evolving nature of assessment practices in digital learning environments. AI Ethics (AIE) emphasizes awareness of ethical, privacy, and well-being considerations related to AI use. Participants are expected to demonstrate an understanding of responsible data handling, potential biases in AI systems, and the importance of promoting ethical digital behavior among students. This construct addresses growing concerns about the ethical implications of technology use in education. Human-Centred Education (HCE) highlights reflective and societal aspects of AI integration. Items in this dimension encourage participants to consider the benefits and risks of AI in schools, the role of human judgment in AI applications, and the broader impact of AI on society. This fosters a critical perspective and promotes informed decision-making among future educators. Professional Engagement (PEN) focuses on pre-service teachers' motivation to explore, share, and continuously learn about AI in education. It includes seeking professional development opportunities, collaborating with peers, and contributing to knowledge-sharing practices. This dimension aligns with the concept of lifelong learning and the evolving nature of teaching in the digital age (Chiu, Ahmad & Çoban, 2024).

Given the scope and complexity of AI integration in education, it is imperative to validate the AICS instrument through rigorous methodological approaches. Validation ensures that the scale reliably measures the intended constructs, is applicable across diverse educational contexts, and yields results that can inform teacher education practices. In this study, the Rasch model was employed to analyze the instrument's psychometric properties (Guo, Shi & Zhai, 2025). The Rasch analysis provides insights into item reliability, fit statistics, unidimensionality, rating scale functionality, and potential differential item functioning (DIF) across subgroups. By employing this model, researchers can ensure that the instrument maintains consistency, fairness, and accuracy in measuring AI competence self-efficacy (Qudratuddarsi, Ramadhana, Indriyanti & Ismail, 2024). Research question of this study: Do the Artificial Intelligence Competence Self-Efficacy (AICS) valid and reliable?

METHOD

This research adopted a quantitative survey approach, emphasizing the collection and analysis of numerical data from structured participant responses. As a survey-based study, it aimed to assess pre-service science teachers' Artificial Intelligence Competence Self-Efficacy (AICS) at a specific point in time, without introducing any intervention or manipulation to the participant group (Ramadhana & Qudratuddarsi, 2024). Utilizing a cross-sectional design enabled the researchers to obtain a snapshot of participants' self-efficacy, thereby avoiding challenges commonly associated with longitudinal studies—such as dropouts or shifting external influences that might affect perceptions over time (Wang, & Cheng, 2020). The use of a quantitative methodology ensured the objectivity and reliability of the data, allowing for statistical evaluation of trends and patterns in the responses. This approach was well-suited to the study's aim of producing generalizable findings that can inform the broader integration of virtual labs into science teacher education programs.

Participants

This study involved 338 Generation Z pre-service science teachers, selected using convenience sampling, which allowed for easy access to participants and efficient data collection. Although this sampling method may limit the generalizability of the findings, the selected participants were highly relevant to the study. All participants were in their third year of study, making them an appropriate group to provide informed responses regarding Artificial Intelligence Competence Self-Efficacy (AICS), as they have gained sufficient exposure to both pedagogical and technological aspects of their teacher education program. As shown in Table 1, the sample consisted of 28.99% male and 71.01% female participants. In terms of specialization, the group included 23.67% chemistry, 21.30% physics, 20.71% biology, and 34.32% general science pre-service teachers. This distribution highlights a diverse representation of science disciplines, providing valuable insights into the varying perspectives on AI competence across different fields within science education.

Table 1. Distribution of sample

Sample	N	Percentage
Gender		
Male	98	28.99%
Female	240	71.01%
Specialization		
Chemistry	80	23.67%
Physics	72	21.30%
Biology	70	20.71%
Science	116	34.32%
Total		100 %

Instrument

The instrument used in this study was carefully adapted from the work of Chiu, Ahmad, and Çoban (2024), who previously developed and validated a scale to measure Artificial Intelligence Competence Self-Efficacy (AICS). The original instrument was published in a reputable, high-impact journal and was originally constructed in English. As the target participants in the current study were native speakers of Bahasa Indonesia, a rigorous back-translation process was conducted to ensure both linguistic accuracy and cultural appropriateness. This process involved translating the instrument from English to Bahasa Indonesia and then independently translating it back to English by a bilingual expert. The back-translation technique was essential for preserving the semantic integrity, contextual relevance, and conceptual equivalence of the measurement items, thus minimizing potential misinterpretations or cultural biases (Behr, 2017). The decision to adapt this specific instrument was guided by its alignment with the objectives of the present study, which sought to explore pre-service science teachers' self-efficacy in integrating AI within their educational practices. Furthermore, the use of a previously validated and psychometrically robust instrument streamlined the research process, allowing the researchers to focus on contextual adaptation and Rasch model-based validation for the new participant group. This ensured that the adapted scale retained its reliability and validity within the unique educational and cultural setting of the current study.

The final instrument was designed to comprehensively capture the diverse dimensions of Artificial Intelligence Competence Self-Efficacy (AICS) among pre-service science teachers. It consisted of six key constructs: AI Knowledge (AIK), which measures confidence in understanding basic AI concepts relevant to science education; AI Pedagogy (AIP), which assesses the ability to design and implement AI-supported teaching strategies; AI Assessment (AIA), which evaluates confidence in using AI tools for student assessment and feedback; AI Ethics (AIE), which gauges awareness of ethical issues such as data privacy and fairness in AI use; Human-Centred Education (HCE), which focuses on maintaining student-centered teaching while integrating AI; and Professional Engagement (PEN), which reflects the commitment to continuous professional development in AI. Together, these constructs provide a solid framework for understanding the self-efficacy of pre-service science teachers in integrating AI responsibly and effectively into their future classrooms.

Data Collection

Data were collected using Google Forms, a digital platform that supports sustainable, paperless research practices. This method aligns with environmental responsibility initiatives while also enhancing efficiency and accuracy in data handling. The digital format enabled real-time access to participant responses and minimized the risk of data entry errors (Hidayat, Imami, Liu, Qudratuddarsi, & Saad, 2024). To ensure clarity and improve data reliability, the researcher was physically present during the data collection process. This allowed for immediate clarification of any confusing items and fostered a supportive environment, which encouraged participants to respond sincerely and attentively. Participation in the study was entirely voluntary, and participants were assured that their involvement would have no impact on their academic grades (Ahmad et. Al., 2019). Furthermore, all responses were treated as confidential, thereby protecting participant privacy and minimizing potential response bias. These ethical measures were essential for ensuring the integrity and authenticity of the data collected.

Data Analysis

Following the data collection phase, the responses were systematically organized and tabulated using Microsoft Excel 2019 to streamline further analysis with Winsteps version 3.7.3. The Rasch measurement model was employed as the primary analytical framework to evaluate several psychometric properties of the instrument, including reliability, separation indices, item fit statistics, unidimensionality, and rating scale functioning. Each sub-construct related to Artificial Intelligence Competence Self-Efficacy (AICS) was analyzed independently to ensure precision in the interpretation of results. Reliability analysis was conducted to assess the internal consistency and stability of the measurement instrument across items and participants. Meanwhile, separation statistics were used to determine the instrument's capacity to differentiate between respondents with varying levels of self-efficacy or perceived competence regarding technology use. These indices are vital for understanding the precision and discriminative power of the scale (Revelle &

Condon, 2019). Item fit statistics played a central role in identifying whether each item adequately aligned with the latent construct being measured. Misfitting items could indicate ambiguity or misinterpretation and therefore needed to be evaluated to ensure the overall validity of the scale. The analysis of unidimensionality verified that each subscale measured a single, coherent latent trait—an essential assumption of the Rasch model to maintain construct integrity (Hagquist & Andrich, 2017). In addition, rating scale calibration was performed to confirm the proper functioning of response categories. This process ensures that the Likert-type scale options provided clear and meaningful distinctions across different levels of the measured construct. By employing this comprehensive Rasch-based analytical approach, the study ensured a robust validation process for the instrument, ultimately reinforcing the credibility and generalizability of the findings related to pre-service science Artificial Intelligence Competence Self-Efficacy.

RESULT AND DISCUSSION

Wright Map

The Wright Map (Item-Person Map) is a key diagnostic tool in Rasch analysis, providing a simultaneous representation of item difficulty and person ability on the same logit scale. In this study, the map indicates that the AICS instrument is well-aligned with the measured abilities of the pre-service science teachers. On the left side of the map, the distribution of items ranges from approximately –5 to +6 logits, showing a broad spectrum of difficulty levels. This ensures that the instrument can effectively differentiate between low, moderate, and high levels of AI self-efficacy. Easier items, such as those related to ethical awareness (e.g., AIE1, PEN1), were generally endorsed by most participants, while more challenging items, such as those requiring application and integration of AI in teaching (e.g., AIK1, AIP3), were located at higher logit levels, requiring greater perceived competence. This vertical spread of item difficulty supports the content and construct validity of the scale by adequately covering the underlying trait continuum of AI competence self-efficacy.

On the right side of the Wright Map, the participants' abilities are distributed across a relatively wide range, though the majority fall between 0 and +2 logits. This concentration suggests that most pre-service science teachers perceive themselves as having moderate to high AI competence self-efficacy. Importantly, the alignment between the person and item distributions suggests that the test is well-targeted: most items are situated at levels appropriate to the abilities of the participants. This good targeting enhances measurement precision, as it allows the instrument to reliably capture differences in self-efficacy across the cohort. Additionally, the absence of floor or ceiling effects indicates that neither the items were too easy nor too difficult for the sample as a whole, further affirming the instrument's suitability. In conclusion, the Wright Map substantiates the psychometric robustness of the AICS instrument, providing strong evidence for its validity in assessing the AI-related self-efficacy of pre-service science teachers.

```
         Item - MAP - Person
           <rare>|<more>
 6               + 010F  082F  112F  116F
                 | 081M
 5               +
                 | 193F  194F
                 | 009F  234F
 4               +
                 | 150F
                 | 198M  207F  215F  259F  323M
                 | 201F
                 |T
 3               + 199F  214F  239F
                 | 090M  249F  251M
                 | 137M  202M  230F  232F  258M  274M  289F
                 | 143M  306F
                 | 040F  152M  213F  328F
 2               + 073F  088F  108M  142F  154F  191F  196F  226F  228F
                 | 027F  145F  212F  225F  261F  266F  282M  337F
                 |S 018F  141F  183M  192F  199M  204M  208F  211F  241M  281M
                 | 002F  031F  037F  091F  096M  187F  188F  203F  216F  220F
                   278F
             X T| 093M  180F  206M  231M  242F  277F  294M  322F  326F
 1               + 012F  045F  063F  085F  095F  107M  125F  167F  195F  200F
                   229M  238F  285F  303F
            XX   | 006F  008F  034M  127M  153F  161F  221F  272F  275F  329F
           XXX S| 001F  015F  068F  080M  098F  155F  178M  190M  222F  248F
                   273M  284F  292M  304F  309M
           XXX   | 014M  016M  048M  055F  083F  084F  087F  102F  103F  106F
                   134F  140F  144F  148F  159F  210F  233F  247F  287M  297F
                   325F  333F
            XX   |M 032F  043F  151M  166M  174M  244F  286M  330M
 0           X M+ 003F  004F  020M  021F  024F  025F  028F  030F  044F  051F
                   053F  054M  061M  071M  076F  114F  118M  119M  121F  124F
                   126F  129F  130F  131F  136F  139F  156F  160F  170F  171F
                   173M  176F  179F  185F  197M  205F  224F  243F  245M  250M
                   252F  253F  256F  262M  267M  291F  321F  324F  335M
            XX   | 013F  022F  036F  065M  066M  077F  111F  113F  175F  217M
                   219F  237F  336F
          XXXX   | 005M  007F  011F  019F  029F  059F  060F  075F  092F  101F
                   157M  165F  169F  227F  236F  255F  300M  305F  311F
           XXX S| 035F  072F  086M  089F  115F  146M  158M  182F  249F  264F
                   283F  298F  313F
             X   | 017M  023M  033M  041F  062F  070M  100F  104F  132F  133F
                   135F  147M  186F  263M  265F  271F  290F  299F  301M  316F
                   318F  331M  338M
-1          XX   + 026F  038F  039F  042F  058M  110F  120M  138M  163M  235F
                   240M  254F  308M  314F  315M
                T| 046F  047M  050F  056F  097M  269F  293M
                 |S 052F  064F  067F  078F  149F  172F  181F  189F  218F  246M
                   276F  310F  317F  319M
                 | 128M  164F  184M  288F
                 | 074F  105F  260F  280M  307F
-2               + 117F  123F  177M  257M  270F  279F  312M  320M
                 | 295F  332F
                 | 122F  168M  296F
                 | 268M
                 |T 057F  223F
-3               + 049F  334M
                 | 069F
                 | 302F
                 | 079F
-4               + 094F
                 | 327M
-5               + 099F  162M
            <frequ>|<less>
```

Figure 1. Wright Map

DOI: https://doi.org/10.29303/jipp.v10i2.3662

Reliability and Separation

The results of the Rasch analysis presented in Table 2 indicate that the AICS (Artificial Intelligence Competence Self-Efficacy) instrument possesses strong psychometric properties in terms of reliability and item-person separation. The person reliability index of 0.94 and item reliability of 0.95 demonstrate excellent internal consistency, indicating that the instrument is highly effective in distinguishing individuals with varying levels of AI self-efficacy and that the items are stable across different samples. This is further supported by a high Cronbach's alpha coefficient of 0.94, which confirms the internal consistency of the scale and suggests that the items are measuring the same underlying construct. Additionally, the person separation index of 4.09 implies that the instrument can categorize respondents into at least four distinct levels of ability, which is well above the minimum threshold of 2.0 for reliable separation. Similarly, the item separation index of 7.23 indicates a wide range of item difficulties and confirms that the sample size is sufficient to validate the item hierarchy. The statistically significant chi-square value (χ^2 = 16,684.85, df = 7,610, $p < .01$) further supports the presence of real differences among item difficulties, rather than random variation. Collectively, these findings provide strong evidence that the AICS instrument is both reliable and valid for assessing the self-efficacy of pre-service science teachers in relation to artificial intelligence competence.

Table 2. Reliability and Separation of NATAI

Indicator	Value
Person Reliability	0.94
Item Reliability	0.95
Cronbach Alpha	0.94
Person Separation	4.09
Item Separation	7.23
Chi-square	16684.85** (d.f. 7610)

Fit Statistics

The item fit statistics from the Rasch model analysis indicate that all items in the AICS instrument function acceptably within the expected range of the model. Using the commonly accepted threshold of 0.5 to 1.5 for Mean Square (MNSQ) values, all items—both for Infit and Outfit—fall within this acceptable range. This suggests that each item contributes meaningfully to measuring the latent trait of artificial intelligence competence self-efficacy among pre-service science teachers. Notably, items such as AIK1 (Infit MNSQ = 1.46; Outfit MNSQ = 1.48) and AIP3 (Infit MNSQ = 1.36; Outfit MNSQ = 1.26) are on the higher end but still remain within the permissible bounds, suggesting that while they are somewhat unpredictable, they do not significantly distort measurement. Additionally, the standardized ZSTD values for these items (e.g., AIP3 ZSTD Outfit = 9.6) are elevated, likely due to the large sample size, which can cause ZSTD to become overly sensitive; therefore, interpretation of ZSTD values should be made cautiously. Importantly, all items have positive Point-Measure Correlation (Pt. Mea Corr) values ranging from 0.55 to 0.81, indicating that each item aligns positively with the overall construct and contributes effectively to measuring the intended trait. These findings collectively support the internal structural validity of the instrument and affirm that the AICS scale is well-targeted and reliable for assessing AI competence self-efficacy in pre-service science teachers.

Tabel 3. Item Fit Statistics

Item	**MNSQ**		**ZSTD**		**Pt Mea Corr**
	Infit	Outfit	Infit	Outfit	
AIK1	1.46	1.48	5.3	5.2	0.64
AIK2	0.93	0.92	-0.9	-1.6	0.72
AIK3	0.98	0.89	-1.3	-1.4	0.72
AIK4	0.97	0.97	-0.4	-0.3	0.73
AIP1	1.37	1.37	4.4	4.3	0.63
AIP2	1.25	1.21	3.1	2.6	0.67
AIP3	1.36	1.26	7.7	9.6	0.55

Item	MNSQ		ZSTD		Pt Mea Corr
	Infit	Outfit	Infit	Outfit	
AIP4	1.24	1.25	3.6	2.9	0.66
AIA1	0.8	0.83	-2.8	-2.3	0.74
AIA2	0.95	0.94	-0.7	-0.7	0.72
AIA3	1.27	1.33	3.2	3.8	0.62
AIA4	0.79	0.83	-3.6	-2.2	0.75
AIE1	0.67	0.66	-4.8	-4.8	0.79
AIE2	0.75	0.79	-3.5	-2.9	0.76
AIE3	0.99	1.0	-0.1	-1.0	0.75
AIE4	1.16	1.19	2.6	2.3	0.76
HCE1	0.8	0.81	-2.8	-2.6	0.77
HCE2	0.87	0.85	-1.8	-2.0	0.76
HCE3	0.93	0.93	-1.6	-0.9	0.76
HCE4	0.76	0.76	-3.5	-3.4	0.77
PEN1	0.61	0.66	-5.9	-5.9	0.81
PEN2	0.98	0.99	-0.3	-0.2	0.75
PEN3	0.81	0.81	-2.6	-2.7	0.77
PEN4	0.87	0.92	-1.8	-1.1	0.76

Unidimensionality

The unidimensionality of the AICS (Artificial Intelligence Competence Self-Efficacy) instrument was assessed using Rasch model analysis, and the results support the scale's validity as a measure of a single latent trait. The analysis showed that 56.0% of the total raw variance was explained by the Rasch measures, with 32.5% attributed to persons and 23.5% to items. This level of explained variance indicates that the model effectively captures the construct being measured—pre-service science teachers' self-efficacy in AI competence. Furthermore, the unexplained variance in the first contrast was relatively low, with an eigenvalue of 1.4 and a percentage of 7.8%. These values fall well below the commonly accepted thresholds (eigenvalue < 2.0; < 10% variance), suggesting the absence of any dominant secondary dimension. Such findings provide strong evidence for the unidimensionality of the scale, confirming that the AICS instrument is appropriately structured to assess a single, coherent construct. This supports the instrument's use in educational research and teacher training contexts where valid and reliable measurement of AI competence self-efficacy is essential.

Table 4. Unidimensionality of AICS

	Value
Raw variance explained by persons	32.5%
Raw variance explained by items	23.5%
Raw variance explained by measures	56.0%
Unexplained variance in 1st contrast (eigenvalue)	1.4
Unexplained variance in 1st contrast (percentage)	7.8%

DIF Analysis

The Differential Item Functioning (DIF) analysis was conducted to examine whether male and female pre-service science teachers responded differently to specific items within the instrument, despite having comparable overall ability levels. The graph illustrates the DIF measures across selected items, with values on the y-axis indicating the magnitude and direction of DIF. A positive DIF value suggests the item was more favorable or easier for males, while a negative value implies it was more accessible to females. Overall, the instrument demonstrated minimal DIF across most items, indicating that it generally functions equitably for both genders. However, a notable exception was Item 5 (AIP1), which displayed a DIF value exceeding +1.0, suggesting that male participants found this item significantly easier. This item pertains to envisioning how AI tools could support teaching and learning, which may reflect gender-based differences in technological confidence or familiarity with AI applications in education. Conversely, Items 1 (AIK1) and 3 (AIK3), which relate to fundamental AI knowledge, showed moderate negative DIF values, indicating that female participants responded more favorably to these items. Other items, such as AIA3, PEN1, and PEN3, exhibited negligible DIF, with overlapping response patterns between male and female participants. These findings suggest that while the instrument is largely gender-neutral, a few items—particularly AIP1—may require further review or revision to ensure consistent interpretation and fairness across gender groups. Addressing these discrepancies will enhance the instrument's validity and ensure more reliable measurement of AI competence self-efficacy among diverse populations.

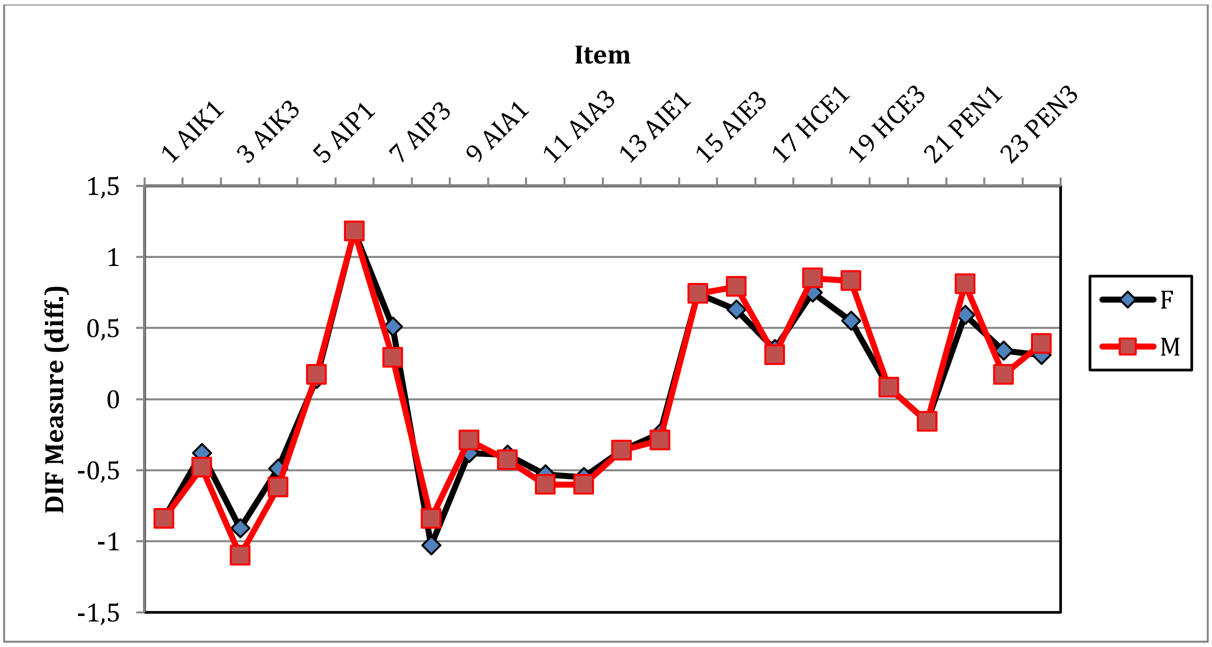

Figure 2. DIF analysis based on gender

Despite its valuable contributions, this study has several limitations. First, the use of a convenience sampling method limits the generalizability of the findings, as the participants were drawn from courses taught by the researchers and may not fully represent the broader population of pre-service science teachers. Additionally, the cross-sectional design captures participants' perceptions at a single point in time, which does not account for potential changes in attitudes or technology use over longer periods. The gender imbalance in the sample, with a predominance of female participants, may also introduce bias and limit the applicability of the results across more balanced populations. Furthermore, reliance on self-reported data may be subject to social desirability bias, where participants could provide favorable responses rather than fully accurate ones.

CONCLUSION

The results of this study provide strong evidence for the validity, reliability, and fairness of the AICS (Artificial Intelligence Competence Self-Efficacy) instrument in assessing AI-related self-efficacy among pre-service science teachers. The Wright Map analysis demonstrated that the instrument is well-targeted, with a broad and balanced distribution of item difficulties that align appropriately with the participants' ability levels. This indicates that the instrument can effectively differentiate between varying degrees of AI competence self-efficacy without presenting items that are too easy or too difficult. The reliability and separation indices further affirm the psychometric strength of the AICS instrument. High person and item reliability values, along with strong person and item separation indices, confirm the internal consistency and the instrument's ability to classify respondents into distinct levels of ability. The fit statistics also showed that all items perform within acceptable ranges, supporting the instrument's structural validity and ensuring that each item contributes meaningfully to the measurement of AI self-efficacy. Additionally, the unidimensionality analysis confirmed that the AICS instrument accurately measures a single underlying construct, validating its theoretical framework. Minimal Differential Item Functioning (DIF) across gender groups suggests that the instrument is largely free from gender bias.

REFERENCES

Ahmad, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., & Farheen, Z. (2019). Qualitative v/s. quantitative research-a summarized review. *population*, *1*(2), 2828-2832.

Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, *20*(6), 573-584.

Chiu, T. K., Ahmad, Z., & Çoban, M. (2024). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Education and Information Technologies*, 1-19.

Chou, C. M., Shen, T. C., Shen, T. C., & Shen, C. H. (2024). Developing and validating an AI-supported teaching applications' self-efficacy scale. *Research & Practice in Technology Enhanced Learning*, *19*.

Dewi, H. R., Qudratuddarsi, H., Ningthias, D. P., & Cinthami, R. D. D. (2024). The Current Update of ChatGPT Roles in Science Experiment: A Systemic Literature Review. *Saqbe: Jurnal Sains dan Pembelajarannya*, *1*(2), 74-85.

Guo, S., Shi, L., & Zhai, X. (2025). Developing and validating an instrument for teachers' acceptance of artificial intelligence in education. *Education and Information Technologies*, 1-23.

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and quality of life outcomes*, *15*, 1-8.

Hava, K., & Babayiğit, Ö. (2025). Exploring the relationship between teachers' competencies in AI-TPACK and digital proficiency. *Education and information technologies*, *30*(3), 3491-3508.

Hidayat, R., Imami, M. K. W., Liu, S., Qudratuddarsi, H., & Saad, M. R. M. (2024). Validity of engagement instrument during online learning in mathematics education. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(2).

Jia, F., Sun, D., & Looi, C. K. (2024). Artificial intelligence in science education (2013–2023): Research trends in ten years. *Journal of Science Education and Technology*, *33*(1), 94-117. https://doi.org/10.1007/s10956-023-10077-6

Karataş, F., & Ataç, B. A. (2024). When TPACK meets artificial intelligence: Analyzing TPACK and AI-TPACK components through structural equation modelling. *Education and Information Technologies*, 1-26.

Kavitha, K., & Joshith, V. P. (2024). Pedagogical Incorporation of Artificial Intelligence in K-12 Science Education: A Decadal Bibliometric Mapping and Systematic Literature Review (2013-2023). *Journal of Pedagogical Research*, *8*(4), 437-465. https://doi.org/10.33902/JPR.202429218

Li, M., & Nugraha, M. G. (2025). Validating a context-specific TPACK scale for primary mathematics education in China. *International Electronic Journal of Mathematics Education*, *20*(2), em0819.

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978

Oved, O., & Alt, D. (2025). Teachers' technological pedagogical content knowledge (TPACK) as a precursor to their perceived adopting of educational AI tools for teaching purposes. *Education and Information Technologies*, 1-27.

Qudratuddarsi, H., Fauziah, A., Agung, A., & Yanti, M. (2025). "Status Quo" ChatGPT dalam pengajaran dan pembelajaran fisika: systematic literature review. *PHYDAGOGIC: Jurnal Fisika dan Pembelajarannya*, *7*(2), 110-118.

Qudratuddarsi, H., Ramadhana, N., Indriyanti, N., & Ismail, A. I. (2024). Using Item Option Characteristics Curve (IOCC) to unfold misconception on chemical reaction. *Journal of Tropical Chemistry Research and Education*, *6*(2), 105-118.

Ramadhana, N., & Qudratuddarsi, H. (2024). Analisis Self Efficacy Mahasiswa pada Mata Kuliah Biologi Sel. *Saqbe: Jurnal Sains Dan Pembelajarannya*, *1*(1), 33-38.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological assessment*, *31*(12), 1395.

Sun, F., Tian, P., Sun, D., Fan, Y., & Yang, Y. (2024). Pre-service teachers' inclination to integrate AI into STEM education: Analysis of influencing factors. *British Journal of Educational Technology*, *55*(6), 2574-2596. https://doi.org/10.1111/bjet.13469

Tram, N. H. M. (2024). Unveiling the Drivers of AI Integration Among Language Teachers: Integrating UTAUT and AI-TPACK. *Computers in the Schools*, 1-21.

Wang, D., & Huang, X. (2025). Transforming education through artificial intelligence and immersive technologies: enhancing learning experiences. *Interactive Learning Environments*, 1-20. https://doi.org/10.1080/10494820.2025.2465451

Wang, X., & Cheng, Z. (2020). Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*, *158*(1), S65-S71.

Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, *29*(4), 4785-4808.