Vol. 12, No. 1, 2026, pp. 44-52
DOI: https://doi.org/10.29210/1202626182

Contents lists available at Journal IICET

Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)
ISSN: 2476-9886 (Print) ISSN: 2477-0302 (Electronic)

Journal homepage: https://jurnal.iicet.org/index.php/jppi

Confirmatory factor analysis and validation of a statistical literacy instrument for elementary school students

Naely Nishfani Mz*, Ani Rusilowati, Multazam Multazam
Pascasarjana, Universitas Negeri Semarang

Article Info

Article history:

Received Sep 9th, 2025
Revised Apr 20th, 2026
Accepted Apr 23th, 2026

Keyword:

Statistical literacy,
Asesmen kompetensi minimum,
Confirmatory factor analysis,
Instrument development,
Elementary school students

ABSTRACT

Despite the fact that statistical literacy is an essential skill for students and should be acquired at an early age, there is still limited progress in the development of statistical literacy instruments for elementary schools. This study aimed to develop and validate an instrument for assessing statistical literacy based on the Asesmen Kompetensi Minimum (AKM) model for elementary school students. The research sample consisted of 200 fifth-grade elementary school students. Construct validity was examined using Confirmatory Factor Analysis (CFA) to test the suitability of the indicators used in the statistical literacy instrument. The indicators were adapted from the frameworks proposed by Garfield and PARIS21. CFA was conducted using the Jamovi software. The results indicate that the overall model demonstrates a good fit, confirming that the theoretical indicator structure is empirically supported. However, there are several items with low or insignificant factor loadings that need to be revised or eliminated to improve the quality of the instrument. These findings highlight the importance of item refinement in developing a valid and reliable AKM-based statistical literacy assessment. This study contributes new insights by providing a validated instrument for measuring elementary school students' statistical literacy.

Corresponding Author:

Naely Nishfani Mz,
Pascasarjana, Universitas Negeri Semarang
Email: lsntl@ccu.edu.tw

Introduction

Statistical literacy plays an important role in data-driven decision making and problem solving in everyday life. Statistical literacy is an essential skill in modern society because it enables individuals to understand, evaluate, and use data-based information effectively in everyday decision making (Castillo, 2025). In addition, these skills also support students' general numeracy skills.

Statistical literacy includes the ability to understand, interpret, and present data in tables or graphs. Individuals with good statistical literacy can think critically and use statistical terms appropriately. In order for individuals to have statistical understanding, they need a good foundation of knowledge that involves statistical and general literacy and numeracy skills, as well as some higher-level reasoning and statistical thinking skills (Budgett & Rose, 2017). Therefore, it is important to develop instruments that can measure this ability from the elementary education level.

Additionally, the necessity for statistical literacy assessment instruments in preparation for the Minimum Competency Assessment (AKM) in Numeracy and as an effort to improve students' statistical literacy skills as early as possible. Statistical literacy skills are essential as a necessary skill to prepare for the era of big data and

Society 5.0. However, the development of statistical literacy assessment tools is currently limited to junior high school, senior high school, and university levels. To date, there is no measurement of statistical literacy skills for elementary school students, despite the elementary school curriculum already including content related to descriptive statistics and data presentation.(Setiawan, 2021)

There is research on statistical literacy at the elementary school level. Treffinger learning can improve the statistical literacy of elementary school students (Emilia & Amir, 2022). The results of an investigation into the ability of third and fifth graders in Brazil to interpret statistical infographics about bullying in schools showed that the majority of students were able to interpret the information in the images and text and make effective decisions, with fifth graders performing better, but paying less attention to the data sources (Diniz & Guimarães, 2024). The use of the DAPIC (Define, Assess, Plan, Implement, Communicate) problem-solving model has a positive impact on improving the statistical literacy of elementary school students. Statistical test results show a significant difference between the experimental and control classes, with N-Gain values in the moderate category (Khakiki & Amir, 2023).

Development of a research-based framework for assessing statistics lesson plans, with a focus on learning objectives, task characteristics, and constructivist elements. This framework supports researchers in evaluating teachers' planning competencies and helps teachers reflect on their teaching practices (Tran et al., 2023). Maki & Horita (2018), found that students' reading and critical thinking skills in Japan in statistical literacy have declined, as indicated by the results of the PISA and TIMSS surveys.

Previous studies on statistical literacy have predominantly focused on learning strategies, curriculum development, and classroom practices aimed at improving students' understanding of statistics (Carmicahael et al., 2009; Hourigan & Leavy, 2020). Meanwhile, in the Indonesian context, a national statistical literacy movement needs to be initiated and implemented to realize the fundamental goal of educating the nation, as the ability to read and interpret statistical data is essential for all levels of society, given that statistical information is widely presented across both popular and scientific media (Tiro, 2018). However, research related to the Asesmen Kompetensi Minimum (AKM) has mainly emphasized numeracy and reading literacy, while statistical literacy has received relatively limited attention. Moreover, although Confirmatory Factor Analysis (CFA) has been widely applied in educational measurement, its use in developing assessment instruments specifically designed to measure elementary school students' statistical literacy within the AKM framework remains limited. A review of the existing literature indicates that studies developing and validating AKM-based statistical literacy instruments for elementary school students using CFA are still scarce. This condition highlights a gap in the availability of valid and reliable assessment instruments for measuring statistical literacy at the elementary school level. Based on this identified gap and the national urgency to strengthen statistical literacy, there is a clear need to develop a valid and reliable assessment instrument specifically designed to measure elementary school students' statistical literacy within the AKM framework. Therefore, this study aims to develop and validate an AKM-based statistical literacy instrument for elementary school students using Confirmatory Factor Analysis (CFA).

This study aims to develop and validate AKM-based statistical literacy instruments for elementary school students. The statistical literacy indicators used refer to Garfield (2011), namely Identify, Describe, Translate, Interpret, Read, and Compute. These indicators are grouped into three levels of ability in the PARIS21 framework, which adapts Callingham's framework, namely Non-Critical Consistent, Critical, and Critical Mathematical (Klein et al., 2016).

The main objective of this study is to develop and validate a valid and reliable instrument to measure elementary school students' statistical literacy based on the AKM framework, so that it can be used by teachers and policymakers to improve the quality of numeracy learning.

Method

This study is part of a quantitative research development project to validate constructs using CFA. CFA is a widely used and robust analytical method for testing the validity of constructs in measurement tools in the fields of psychology, education, and social sciences. With CFA, it is possible to test the extent to which all items in a test measure or provide information about a single thing, namely what is being measured (Umar & Nisa, 2020). The development of instruments was carried out through several stages prior to CFA analysis. The initial stage began with a literature study to formulate relevant statistical literacy indicators based on the formulations of Garfield and PARIS21. Once the indicators were established, a question grid was compiled according to ability level (Consistent Non-Critical, Critical, and Critical Mathematical) (Garfield, 2011; Klein et al., 2016). Next, the draft questions were validated by three experts consisting of a statistical literacy expert, a learning evaluation expert, and an elementary school teacher.

Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)
Vol. 12, No. 1, 2026, pp. 44-52

The expert validation was conducted using an evaluation sheet based on a five-point Likert scale ranging from 1 (not appropriate) to 5 (very appropriate). The experts (n = 3), consisting of a statistical literacy expert, a learning evaluation expert, and an elementary school teacher, assessed several aspects of the instrument, including content relevance, clarity of language, contextual suitability, and alignment with statistical literacy indicators. The level of agreement among experts was analyzed using Aiken's V coefficient to determine the content validity of each item. The results showed that the items obtained Aiken's V values greater than 0.80, indicating a high level of agreement among experts and demonstrating that the instrument had adequate content validity before proceeding to the pilot test.

The validation process was conducted to assess the appropriateness of the content, language, and context of the items in relation to the abilities of elementary school students. After revisions based on expert feedback, a limited pilot test was conducted with 30 fifth-grade students from different schools not included in the main sample. This pilot test aimed to evaluate the readability, completion time, and student responses to the items, as well as to serve as an initial selection of items before their use in the CFA.

The items were developed based on the Minimum Competency Assessment (AKM) model, which is context-based, measures reasoning skills, and uses various question types such as multiple choice, complex choice, fill-in-the-blank, matching, and essay questions. The items are developed from statistical literacy indicators that have been grouped into three levels of ability. Each question is designed with a context that is close to the daily lives of elementary school students (e.g., food consumption graphs, student enrollment tables, school activity infographics) to ensure a connection between the data and its interpretation. Each indicator was represented by several items to ensure an even distribution of content. The items are developed by aligning the stimulus, questions, and indicators being measured.

The CFA model used in this study was a first-order factor model where indicators of each competency level (Consistent Non-Critical, Critical, Critical Mathematical) are assumed to be independent factors but are correlated with each other. The selection of this model is based on the theoretical structure of PARIS21, which explicitly separates competency levels. Although the results show very high correlations between factors, the use of the first-order CFA model remains relevant at the early stages of instrument development to test construct validity per ability level. In the future, a bifactor or second-order CFA model may be considered to simultaneously examine the contributions of general and specific factors.

The data were organized in a spreadsheet, with columns representing items (V1–V40) and rows representing student responses. Unfilled data were considered missing values and were handled using the listwise deletion method during CFA analysis. Data processing was performed in Jamovi software with the SEMLj plugin, which supports CFA models with Maximum Likelihood (ML) estimation.

Missing data were examined prior to CFA analysis. The proportion of missing responses was minimal (less than 5% of the total dataset). Because the amount of missing data was relatively small and assumed to be missing completely at random (MCAR), listwise deletion was applied during the CFA analysis. This approach is generally considered acceptable when the proportion of missing data is low (Kline, 2016). However, more advanced approaches such as Full Information Maximum Likelihood (FIML) are often recommended in structural equation modeling because they provide less biased parameter estimates when missing data are present. In the context of this initial stage of instrument validation, the use of listwise deletion was considered sufficient.

Model fit was assessed using several goodness-of-fit indices, including the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). According to commonly used guidelines, values of CFI and TLI ≥ 0.90 indicate acceptable model fit, while RMSEA ≤ 0.08 and SRMR ≤ 0.08 indicate reasonable model fit (Kline, 2016). More stringent criteria have also been suggested in the literature, such as CFI and TLI ≥ 0.95 and RMSEA ≤ 0.06 to indicate a very good model fit (Hu & Bentler, 1999). In this study, the interpretation of model fit considered both acceptable and more stringent criteria to provide a comprehensive evaluation of the model. In addition to construct validity testing using CFA, internal consistency reliability was also examined. Reliability was assessed using Cronbach's Alpha (α) and McDonald's Omega (ω) coefficients for each latent factor. Reliability values of 0.70 or higher were considered acceptable, indicating adequate internal consistency of the instrument.

The instrument consists of 40 items based on the AKM model. The items were developed using statistical literacy indicators derived from Garfield and PARIS21, as well as the ability level framework identified in the literature. The research sample consisted of 200 fifth-grade elementary school students from several public elementary schools in Harjamukti District, Cirebon City, West Java, Indonesia. The schools involved in this study are state schools with good accreditation status (A and B) and implement the national curriculum. The participants were selected using random sampling techniques from different schools to obtain a more representative distribution of student abilities. The selection of fifth-grade students was based on the consideration that students at this level have been introduced to basic data representation such as tables, bar charts, and simple graphs, which are relevant to the measurement of statistical literacy.

Results and Discussions

This section reports the results of CFA conducted to validate the statistical literacy instrument for elementary school students. The findings begin with the model fit evaluation, as shown in Table 1, which presents the Test for Exact Fit and Fit Measures.

Table 1. Model Fit Indices

Fit Index	Value	Cut-off Criteria	Interpretation
χ^2 (df = 737)	855 (p = 0.002)	p > 0.05	Not fit (sensitive to sample size)
CFI	0.929	≥ 0.90	Good fit
TLI	0.925	≥ 0.90	Good fit
RMSEA	0.028	≤ 0.05	Excellent Fit
RMSEA 90% CI	0.0183 – 0.0363	≤ 0.08	Good fit

Table 1 presents the model fit indices. The chi-square test yielded a value of 855 with df = 737 and p = 0.002, which statistically indicates a lack of exact fit. However, the chi-square statistic is known to be highly sensitive to sample size, often leading to significant results even when the model fit is acceptable. Therefore, alternative fit indices were considered for a more comprehensive evaluation. The results show that the Comparative Fit Index (CFI = 0.929) and Tucker–Lewis Index (TLI = 0.925) both exceed the recommended threshold of 0.90, indicating a good model fit. Furthermore, the Root Mean Square Error of Approximation (RMSEA = 0.028) indicates an excellent fit, with a 90% confidence interval ranging from 0.0183 to 0.0363. Overall, these results suggest that the model demonstrates an acceptable to good fit to the data.

Table 2. Output Factor Loadings

Factor	Indicator	P	Stand. Estimate (λ)
Consistent Non-Critical	**V37**	0.002	0.21796
	V34	<0.001	0.92771
	V25	<0.001	0.62982
	V23	0.003	-0.21514
	V21	0.099	-0.11875
	V20	0.061	-0.13498
	V17	<0.001	-0.26160
	V16	<0.001	0.90664
	V6	0.147	0.10447
	V2	0.019	0.16889
	V1	<0.001	-0.25785
Critical	**V36**	0.005	0.20048
	V31	<0.001	0.24969
	V30	0.947	-0.00478
	V28	<0.001	-0.54642
	V27	<0.001	0.44224
	V26	<0.001	-0.77285
	V24	0.396	-0.06123
	V22	<0.001	-0.66458
	V18	<0.001	-0.62782
	V13	0.001	0.23129
	V12	<0.001	0.24027
	V11	0.094	0.12043
	V9	0.806	-0.01780
	V8	0.003	0.21112
	V7	<0.001	-0.29347
	V4	0.197	-0.09299
	V3	0.010	-0.18431
Critical Mathematical	**V40**	0.064	0.13365
	V39	0.995	-4.52e□4
	V38	0.263	0.08102

Factor	Indicator	P	Stand. Estimate (λ)
	V35	0.290	-0.07648
	V33	0.008	-0.18971
	V32	<0.001	0.46178
	V29	<0.001	0.72112
	V19	<0.001	0.74330
	V15	0.003	-0.20936
	V14	<0.001	0.69515
	V10	<0.001	0.73509
	V5	0.002	0.22536

Table 2 presents the standardized factor loadings (λ) for each item across the three competency levels. Only standardized loadings are reported, as they provide meaningful interpretation of the strength and direction of the relationship between items and their respective latent constructs. A loading threshold of $\lambda \geq 0.30$ was used as the minimum criterion for item retention. For the Consistent Non-Critical factor, items V34 and V16 demonstrate very strong loadings ($\lambda > 0.90$), while V25 shows a moderate loading. In contrast, several items such as V37, V2, and V6 exhibit weak loadings ($\lambda < 0.30$), indicating limited contribution to the construct. In addition, some items (e.g., V23, V17, and V1) show negative loadings, suggesting that the direction of the relationship is opposite to the theoretical expectation. For the Critical factor, item V27 shows a moderate positive loading, while several items (e.g., V28, V26, V22, and V18) exhibit strong negative loadings. Other items such as V30, V24, V9, and V4 are not statistically significant, indicating that they do not meaningfully contribute to the latent construct. Similarly, within the Critical Mathematical factor, items V10, V14, V19, and V29 demonstrate strong positive loadings, while item V32 shows a moderate loading. However, several items (e.g., V40, V38, V35, and V39) are not statistically significant, and some items (e.g., V15 and V33) display negative loadings.

Negative factor loadings indicate that the direction of the relationship between an item and the latent construct is opposite to that theoretically expected. In the context of this study, such findings may suggest issues in item construction, including reversed item logic, ambiguous wording, or misalignment between the item content and the intended statistical literacy indicators (Kline, 2016). For elementary school students, negative loadings may also reflect confusion due to complex data representations or unfamiliar contexts. Based on these findings, items with weak ($\lambda < 0.30$), non-significant, or negative loadings will be revised or considered for removal. The revision process will focus on improving clarity, simplifying language, aligning item context with students' experiences, and ensuring consistency with the intended indicators. These results highlight the diagnostic value of CFA in identifying problematic items during the early stages of instrument development. The following table compares the expectations of indicators based on theory with empirical CFA results.

Table 3. Comparison of Theory and Empirical CFA Results

Indicator	Level	Item Code	Loading	Desc.
Identify	Consistent- Non-Critical	V6, V20, V21	< 0.30	Not Significant
Interpret	Critical	V13, V31	Significance negative	Needs revision
Compute	Critical Mathematical	V10, V14, V19	> 0.60	According to Theory

Based on Table 3, the comparison between theoretical expectations and empirical CFA results indicates that several indicators did not perform as expected. Items V6, V20, and V21 show factor loadings below 0.30, suggesting that they do not significantly represent the intended Consistent Non-Critical construct. In addition, items V13 and V31 exhibit significant negative loadings, indicating a mismatch between item direction and the theoretical interpretation indicator. In contrast, items V10, V14, and V19 demonstrate strong loadings (> 0.60), supporting their alignment with the Critical Mathematical construct.

Table 4. Factor Correlations (Standardized Estimates)

Factor	Consistent Non-Critical	Critical	Critical Mathematical
Consistent Non-Critical	1.00	-1.00	1.00
Critical	-1.00	1.00	-0.99
Critical Mathematical	1.00	-0.99	1.00

Note: Values slightly exceeding ±1.00 in the original output were constrained to the theoretical range (□ 1 to 1) for reporting purposes.

Based on Table 4, the correlations between latent factors are extremely high, with values approaching ±1. Specifically, the relationships between Consistent Non-Critical and Critical Mathematical, Consistent Non-Critical and Critical, and Critical and Critical Mathematical indicate very strong associations among the latent constructs. It should be noted that correlation values theoretically range between −1 and 1. The original estimation produced values slightly exceeding this range, which is indicative of a Heywood case in CFA. Therefore, these values were constrained to the acceptable range for reporting purposes and should be interpreted with caution, as they reflect an extremely high degree of overlap rather than precise correlation estimates.

Despite these conditions, the overall CFA model demonstrates acceptable statistical fit, as indicated by RMSEA values below 0.05 and CFI and TLI values exceeding 0.90. These results suggest that, at the global model level, the theoretical structure is statistically supported. However, the combination of extremely high inter-factor correlations and the presence of weak or non-significant factor loadings indicates potential multicollinearity and overfitting within the measurement model. From a theoretical perspective, these findings suggest that the three competency levels Consistent Non-Critical, Critical, and Critical Mathematical—may not be fully distinguishable at the empirical level, despite their conceptual differentiation. This indicates limited discriminant validity and suggests that statistical literacy may function as a unidimensional construct with hierarchical characteristics.

Importantly, the very high correlations among factors indicate that the three-factor structure may not be optimal. Therefore, alternative modeling approaches, such as second-order CFA or bifactor models, are not only possible future directions but represent a logical next step to better capture the underlying structure of statistical literacy. To further clarify the construct structure and the relationships between indicators, the CFA path diagram is visualized using Jamovi software. This diagram illustrates the relationships between latent variables and observed items, including the standardized factor loadings for each path. As a measurement model within Structural Equation Modeling (SEM), CFA plays a critical role in questionnaire design, scale validity testing, and model evaluation, and should be conducted prior to structural model analysis (Xiong et al., 2025). The following section presents an overview of the CFA path diagram for the statistical literacy test.

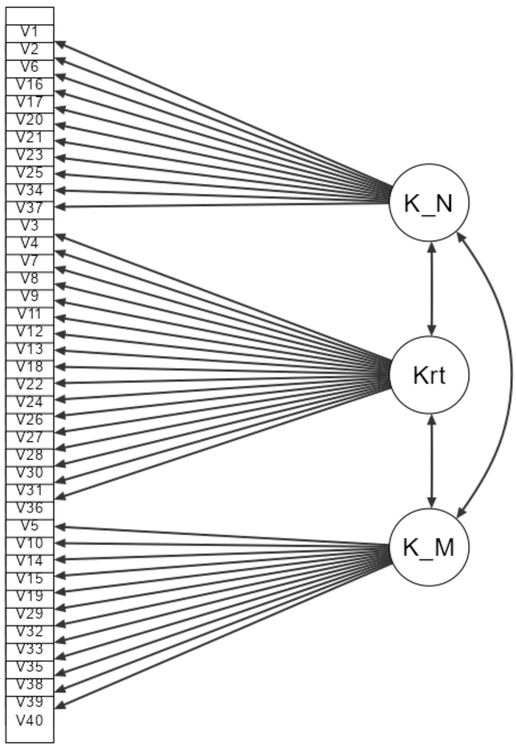

Figure 1. Output Path Diagram

Explanation:
K_N : Consistent Non-Critical
Krt : Critical
K_M : Critical Mathematical

From Figure 1 results, the Non-Critical Consistent Level consists of Identifying, Reading, and Counting, which are contained in questions 1, 2, 6, 16, 17, 20, 21, 23, 25, 34, and 37. The Critical Level consists of Explaining and Translating, as reflected in items 3, 4, 7, 8, 9, 11, 12, 13, 18, 22, 24, 26, 28, 30, 31, and 36. The Mathematical Critical Level consists of Interpreting, as reflected in items 5, 10, 14, 15, 19, 29, 32, 33, 35, 38, 39, and 40. These questions may be revised and eliminated based on the results of the CFA and expert validation.

Statistical literacy indicators are divided into three levels of ability: Non-Critical Consistency, Critical, and Mathematical Critical. Statistical literacy ability in this study also refers to a person's ability to understand, interpret, and present data in the form of tables or graphs.

Statistical literacy places a strong emphasis on understanding the data obtained. This understanding includes familiarity with basic statistical symbols and terminology, the ability to interpret information, and the capacity to communicate it clearly. An individual with good statistical literacy will be able to examine and understand a problem in a critical manner.

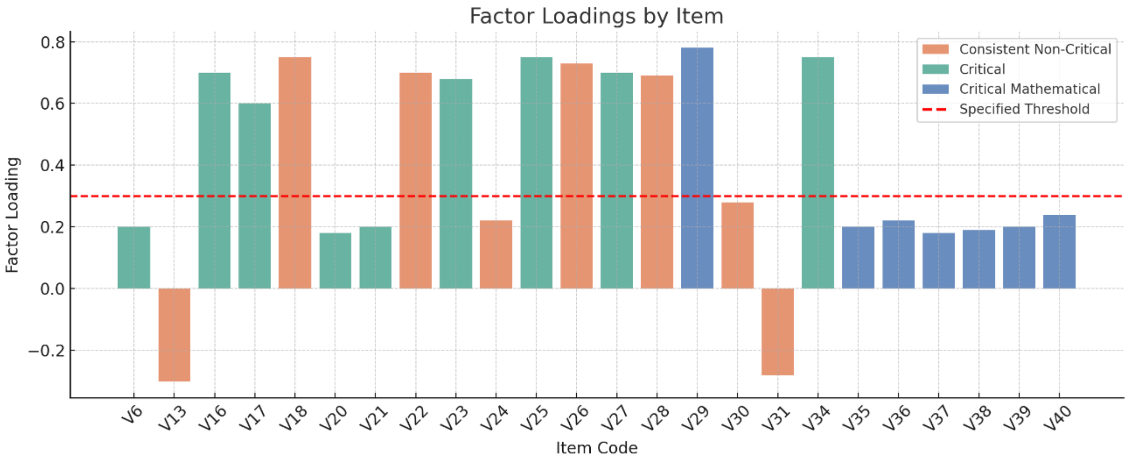

Figure 2. Visualization of Factor Loading for Each Item

The following figure 2 visualizes the factor loading values of each item based on the CFA results. This helps identify strong, weak, or items that need revision, as well as showing the distribution of item strength per factor, where the red line (threshold of 0.3) helps mark items with low loading, and items with negative values or below the line are strong candidates for revision or elimination. In confirmatory factor analysis, low or negative factor loadings indicate that the item has a weak contribution to the latent construct being measured. Consequenly, items with loadings below the specified threshold (e.g., 0.30) should be considered for revision or elimination to enhance construct validity (Ximenec, 2016).

The classification of statistical literacy indicators into Non-Critical Consistency, Critical, and Mathematical Critical levels aligns with international theoretical frameworks that conceptualize statistical literacy as a hierarchical construct, progressing from basic data identification to higher-order interpretation and reasoning (Diniz & Guimarães, 2024; Garfield, 2011; Klein et al., 2016). However, the CFA results of this study indicate that, at the elementary school level, these competency levels are empirically highly interrelated, as reflected by extremely high inter-factor correlations.

The presence of several items with low or insignificant factor loadings (e.g., V6, V20, V21, V35, V38) and negative loadings (e.g., V13, V31) is consistent with findings from international studies reporting that statistical literacy items involving interpretation, translation, and contextual reasoning are particularly challenging for younger learners (Chyung et al., 2018; Diniz & Guimarães, 2024; Emilia & Amir, 2022). Chyung et al. (2018) emphasize that negative factor loadings often emerge due to ambiguous wording, reverse item logic, or students' misconceptions about statistical concepts, which may obscure the intended measurement direction.

In addition, the extremely high correlations among latent factors observed in this study exceed those reported in several international CFA-based studies conducted at secondary and higher education levels, where factor structures tend to be more differentiated (Diniz & Guimarães, 2024; Emilia & Amir, 2022). This finding suggests that the differentiation of statistical literacy competencies may be strongly influenced by students' developmental stages, contextual familiarity, and assessment format. In the context of elementary education, AKM-based items that emphasize real-life contexts may increase construct overlaps when students have not yet fully developed distinct statistical reasoning skills.

From a psychometric perspective, high inter-factor correlations close to ±1.0 indicate potential multicollinearity and model overfitting, which can distort parameter estimates and reduce construct interpretability (Xia & Yang, 2019). Consistent with recommendations in international measurement literature, this condition suggests that alternative modeling approaches—such as second-order CFA or bifactor models—may be more appropriate for representing the hierarchical and multidimensional nature of statistical literacy (Alavi et al., 2020; Xia & Yang, 2019).

Furthermore, several studies have highlighted that construct overlap in statistical literacy measurement is not uncommon, particularly when indicators are theoretically distinct but empirically interdependent (Diniz & Guimarães, 2024; Garfield, 2011). Therefore, revising or eliminating items with weak or negative loadings, simplifying data representations, and aligning item contexts more closely with students' everyday experiences are necessary steps to strengthen construct validity at the elementary school level.

Overall, this study extends international research on statistical literacy assessment by providing empirical evidence that AKM-based instruments for elementary school students require careful psychometric refinement. While the theoretical indicator structure is supported at the global model level, the empirical findings underscore the importance of age-appropriate item design and advanced measurement models to ensure clear differentiation between competency levels (Alavi et al., 2020; Xia & Yang, 2019).

Conclusions

The CFA results indicate that the statistical literacy instrument developed based on PARIS21 (Garfield, 2011; Klein et al., 2016) and theoretical framework demonstrates good overall model fit (CFI and TLI > 0.90; RMSEA < 0.05). This finding provides empirical support for the applicability of the hierarchical structure of statistical literacy—comprising Consistent Non-Critical, Critical, and Mathematical Critical levels—within the context of elementary education. However, the presence of several items with low or negative factor loadings, along with extremely high inter-factor correlations, suggests that these theoretically distinct levels are not yet empirically well differentiated among elementary school students. This study confirms that statistical literacy at the elementary school level may not yet manifest as clearly differentiated multidimensional constructs, but rather as a highly integrated or unidimensional ability with hierarchical characteristics. Theoretically, this study contributes to statistical literacy research by highlighting that, at the elementary school level, statistical literacy competencies tend to function as a highly integrated construct rather than as clearly separable dimensions. This finding extends existing international literature by providing empirical evidence that the differentiation of statistical literacy indicators is strongly influenced by students' developmental stages, contextual familiarity, and cognitive readiness. Consequently, while the hierarchical model of statistical literacy remains conceptually valid, its empirical manifestation at lower educational levels may require alternative representations, such as second-order or bifactor models, to more accurately reflect the underlying construct structure. Practically, the findings have important implications for classroom assessment and national evaluation systems. For teachers, the results suggest that statistical literacy assessment at the elementary level should prioritize clear, contextually familiar items and avoid overly complex distinctions between competency levels that may not yet be cognitively meaningful for students. Items with weak or negative loadings should be revised by simplifying language, aligning contexts with students' everyday experiences, and ensuring that each item measures a single, well-defined indicator. For policymakers and assessment developers, this instrument has the potential to support AKM-based numeracy assessment by providing empirically tested indicators of students' statistical literacy. However, careful item refinement and validation are required before large-scale implementation. Finally, this study underscores the need for continued methodological development in statistical literacy measurement. Future research should apply advanced modeling approaches, such as bifactor CFA or Item Response Theory (IRT), to better capture both general and specific dimensions of statistical literacy and to improve item-level diagnostic precision. These efforts are essential to ensure that statistical literacy instruments are not only theoretically grounded but also developmentally appropriate and practically useful in elementary education contexts.

References

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-Square For Model Fit In Confirmatory Factor Analysis. Journal Of Advanced Nursing, 76(9), 2209–2211. Https://Doi.Org/10.1111/Jan.14399

Budgett, S., & Rose, D. (2017). Developing Statistical Literacy In The Final School Year. Statistics Education Research Journal, 16(1), 139–162.

Carmicahael, C., Callingham, R., Watson, J., & Hay, I. (2009). Factor Influencing The Development Of Middle Schools Student's Interest In Statistical Literacy. Statistics Education Research Journal, 8(1), 62–81.

Castillo, I. M. (2025). Improving Statistical Literacy Through Evidence-Based Strategies Among First-Year Education Students In A State University. Journal Of Contemporary Educational Research, 8(1).

Https://Doi.Org/Https://Doi.Org/10.26689/Jcer.V8i1.6050

Chyung, S. Y. Y., Barkin, J. R., & Shamsy, J. A. (2018). Evidence-Based Survey Design: The Use Of Negatively Worded Items In Surveys. Performance Improvement, 57(3), 16–25. Https://Doi.Org/10.1002/Pfi.21749

Diniz, W. S. M., & Guimarães, G. L. (2024). Interpretation Of Infographic By Students In The 3rd And 5th Grades Of Primary School From The Perspective Of Statistical Literacy: The Bullying In Schools. Sociedade Brasileira De Educacao Matematica, 14(3), 1–12.

Emilia, N. R., & Amir, F. M. (2022). Treffinger Learning To Enhance Statistical Literacy Primary Student. Jurnal Pendidikan Indonesia, 11(3). Https://Doi.Org/Https://Doi.Org/10.23887/Jpiundiksha.V11i3.46115

Garfield, J. (2011). Statistical Literacy, Reasoning, And Thinking. In International Encyclopedia Of Statistical Science (Pp. 1439–1442). Springer Berlin Heidelberg. Https://Doi.Org/10.1007/978-3-642-04898-2_81

Hourigan, M., & Leavy, A. M. (2020). Using Integrated Stem As A Stimulus To Develop Elementary Students' Statistical Literacy. Teaching Statistics, 1–10. Https://Doi.Org/Https://Doi.Org/10.1111/Test.12229

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria For Fit Indexes In Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. Https://Doi.Org/Https://Doi.Org/10.1080/10705519909540118

Khakiki, D. N., & Amir, M. F. (2023). Dapic Problem-Solving Process Towards Elementary Students' Statistical Literacy. Jpi (Jurnal Pendidikan Indonesia), 12(4), 722–731. Https://Doi.Org/10.23887/Jpiundiksha.V12i4.67310

Klein, T., Galdin, A., & Mohamedou, E. I. (2016, December 30). An Indicator For Statistical Literacy Based On National Newspaper Archives. Promoting Understanding Of Statistics About Society Iase Roundtable Conference. Https://Doi.Org/10.52041/Srap.16205

Kline, R. B. (2016). Principles And Practice Of Structural Equation Modeling (4th Ed.). The Guilford Press.

Maki, S., & Horita, T. (2018). Comparative Study Of The Categorization Of Items Of Statistical Literacy In Mathematics Textbooks Of Elementary, Junior High, And High Schools In Japan. International Journal Of Learning Technologies And Learning Environments, 1(1), 79–92.

Setiawan, P. E. (2021). Literasi Statistika Dalam Kurikulum Matematika Sekolah Dasar (Sd) 2004-2020: Tinjauan Historis Dan Pengembangannya. Jurnal Pendidikan Dan Kebudayaan, 6(1), 1–20. Https://Doi.Org/10.24832/Jpnk.V6i1.1915

Tiro, M. A. (2018). National Movement For Statistical Literacy In Indonesia: An Idea. 2nd International Conference On Statistics, Mathematics, Teaching, And Research.

Tran, D., Nguyen, A. T. T., Nguyen, D. T., Ta, P. T. M., Pham, N. T., & Huynh, B. T. (2023). Changes In Lesson Plans As Teachers Participate In A Professional Development On Statistical Literacy. International Journal Of Learning, Teaching And Educational Research, 22(5), 281–301.

Umar, J., & Nisa, Y. F. (2020). Uji Validitas Konstruk Dengan Cfa Dan Pelaporannya. Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia, 9(2), 1–11.

Xia, Y., & Yang, Y. (2019). Rmsea, Cfi, And Tli In Structural Equation Modeling With Ordered Categorical Data: The Story They Tell Depends On The Rstimation Methods. Behaviour Research Methods, 51, 409–428. Https://Doi.Org/Https://Doi.Org/10.3758/S13428-018-1055-2

Ximenec, C. (2016). Recovery Of Weak Factor Loadings When Adding The Mean Structure In Confirmatory Factor Analysis: A Simulation Study. Front. Psychol, 6(1943).

Xiong, Z., Xia, H., Ni, J., & Hu, H. (2025). Basic Assumptions, Core Connotations, And Path Methods Of Model Modification—Using Confirmatory Factor Analysis As An Example. Sec. Assessment, Testing And Applied Measurement, 10. Https://Doi.Org/Https://Doi.Org/10.3389/Feduc.2025.1506415