

Developing an Instrument to Assess Writing Skills in Arabic as a Second Language: A Confirmatory Factor Analysis Approach

Lalu Agus Pujiartha¹, Apri Wardana Ritonga², Ayu Desrani³, Agus Supriatna⁴

¹ Institut Muslim Cendekia Sukabumi, Jawa Barat, Indonesia.

^{2&4} Sekolah Tinggi Ilmu Qur'an As-Syifa Lembang, Jawa Barat, Indonesia.

³ Universitas Negeri Yogyakarta, Yogyakarta, Indonesia.

*Correspondence to: pujiarthalaluagus@gmail.com

Abstract: This study aims to develop an assessment instrument for Arabic writing skills as a second language using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The instrument was developed based on the Oriondo and Antonio model, with a sample of 248 students who completed a writing test. Data were analyzed using R-Studio software for EFA, CFA, and descriptive statistics. The EFA results revealed one main factor, with eigenvalues and factor loadings above 0.6, indicating a consistent factor structure. In CFA, the four dimensions of writing skills Ideas and Content, Structure and Organization, Style and Rhetoric, and Language and Grammar demonstrated strong factor loadings and good model fit, confirming the robustness of the instrument. The instrument also showed high internal consistency and strong construct validity. This instrument can be practically used by educators and language program administrators to evaluate and enhance students' writing proficiency in Arabic as a second language. This study produces a valid and reliable instrument for assessing Arabic writing skills as a second language in higher education.

Keywords: Arabic learning; EFA, CFA; instrument validation; second language; writing skills

How To Cite: Pujiartha, L., Ritonga, A., Desrani, A., & Supriatna, A. (2025). Developing an instrument to assess writing skills in Arabic as a second language: A confirmatory factor analysis approach. *Al-Ta'lim Journal*, 32(2). doi:<https://doi.org/10.15548/jt.v32i2.937>

Article info: Submitted: 21st March 2024 | Revised: 30th May 2025 | Accepted: 29th Juli 2025

INTRODUCTION

Writing skills are an essential aspect that demands serious attention in the process of learning Arabic as a second language (L2) in Indonesia. These skills not only reflect mastery of grammar and vocabulary but also demonstrate critical thinking, creativity, and the systematic articulation of ideas (D'Souza, 2021). Writing plays a strategic role in strengthening overall Arabic language proficiency (Zaki, 2021), as it involves various linguistic components, including morphology, syntax, and semantics (Al-Janabi et al., 2020). Consequently, the development of valid and reliable instruments to measure writing skills has become an urgent necessity to support the evaluation of Arabic language learning.

The assessment of writing skills should be based on a comprehensive understanding of students' writing abilities in Arabic. Brown (2007) identifies five key aspects of writing: content, organization, vocabulary, grammar, and mechanics (such as spelling and punctuation). In the context of Arabic, this assessment must take into account specific elements such as Arabic sentence structure, the use of diacritical marks (*harakat*), and the precision in applying vocabulary (*mufradat*) (Rakhlin et al., 2021). Instruments designed to evaluate these skills must adequately reflect the diversity of these aspects in

a fair and accurate manner to ensure that the evaluation results genuinely represent the students' writing abilities in Arabic.

However, the availability of valid and reliable instruments for assessing Arabic writing skills as a second language remains limited, particularly those based on modern theoretical frameworks such as Confirmatory Factor Analysis (CFA). CFA provides a systematic approach to developing instruments with well-defined factor structures that are statistically validated (Akomah & Ramani, 2024), offering a robust foundation for the measurement process. The use of CFA is particularly crucial as it ensures that each indicator within the instrument truly represents the dimensions of writing skills being assessed (Sureshchandar, 2023a). Additionally, CFA enhances the construct validity of the instrument, leading to more accurate and scientifically credible evaluation results (Baharum et al., 2023; Black & Wiliam, 1998). The lack of CFA-based instruments often poses challenges to conducting objective and credible evaluations, which ultimately impacts the assessment of Arabic language learning processes (Brown, 2015).

Previous studies indicate that the development of Arabic writing skill assessment instruments has rarely been conducted using structured approaches grounded in robust theoretical frameworks. For example, Erguvan (2021) found that many instruments rely on simplistic scoring rubrics without in-depth statistical analysis. Meanwhile, Mellinger & Hanson (2020) emphasized the importance of quantitative approaches in language instrument validation to enhance measurement quality. Rakhlin et al (2021) highlighted the necessity of standardized evaluation tools for writing skills in Arabic language learning. This study argued that without valid assessment tools, achieving fair and credible evaluations is challenging. Lastly, Rufaiqoh et al (2023) investigated the development of criteria-based instruments for evaluating Arabic writing skills; however, their approach was primarily descriptive-analytical and lacked CFA. While these studies provide valuable insights, research that specifically employs CFA to develop and validate Arabic writing skill instruments remains scarce. This highlights a research gap that must be addressed to produce more statistically validated and reliable instruments for supporting Arabic language learning.

From the review of previous studies, it is evident that the demand for valid evaluation instruments is becoming increasingly urgent, especially given the growing interest in Arabic language learning among university students. Indonesian student university come from diverse linguistic backgrounds, necessitating evaluation tools that can accommodate varying levels of proficiency and challenges they face (Zakaria et al., 2024). Additionally, such instruments should account for the unique characteristics of Arabic as a second language, including its phonological, morphological, and syntactic features, which differ significantly from the students' first language (Rakhlin et al., 2021). The availability of validated instruments is expected to make the evaluation process more effective, transparent, and aligned with international standards while providing a comprehensive overview of students' writing skill development.

Based on the above discussion, several critical issues emerge: 1) How can a valid and reliable instrument for assessing Arabic writing skills be developed? 2) Does the factor structure align with theoretical expectations? 3) How can students' Arabic writing be evaluated more systematically? This research addresses these questions by developing a CFA-based instrument tailored to Arabic language learners in Indonesia. Beyond its theoretical contributions, the study also provides a practical tool that educators and

curriculum designers can use to track students' writing progress, identify specific weaknesses, and inform targeted instructional strategies in Arabic language programs.

METHOD

Research Design

This study is a Research and Development (R&D) study aimed at developing an assessment instrument for Arabic writing skills as a second language. The approach used is quantitative, involving Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The development process adopts the Oriondo & Dallo-Antonio (1984) model, which includes planning, testing, instrument validation, reliability measurement, and score interpretation, based on the Model Writing Competency theory by Bereiter & Scardamalia (1987). EFA was conducted to explore the factor structure, followed by CFA to confirm the dimensional fit of the proposed four-factor model: Ideas and Content, Structure and Organization, Style and Rhetoric, and Language and Grammar. To strengthen methodological rigor, alternative CFA models were tested and compared, including a one-factor model and a second-order factor model. The four-factor model showed superior fit indices compared to the alternatives, providing strong justification for its selection as the final model used in this study.

Participants

The population in this study consists of undergraduate students who are studying Arabic at higher education institutions, either as a major program or as a required/elective course. These participants were chosen due to their direct engagement with Arabic as a second language, particularly in developing writing skills. The students came from various academic backgrounds, including Arabic education, Islamic studies, and language and literature programs, with at least three semesters of formal Arabic instruction. Most participants had prior experience in writing Arabic essays, short texts, or reports as part of their coursework, ensuring a foundational exposure to Arabic writing tasks.

The sample size of 248 students satisfies the recommended minimum for CFA, which requires at least 5-10 respondents per instrument item. Purposive sampling was used to select participants who met specific criteria relevant to the study's objectives namely, those actively learning Arabic as a second language with demonstrated experience in writing. While this sampling strategy enhances focus and relevance, it also presents limitations in terms of representativeness. The use of non-random sampling and the focus on Indonesian higher education institutions may restrict the generalizability of the findings to broader populations or different educational contexts.

Instrument

The instrument developed in this study was designed to directly assess students' Arabic writing performance based on the Writing Competency Model by Bereiter & Scardamalia (1987), which includes four main dimensions: Ideas and Content, Structure and Organization, Style and Rhetoric, and Language and Grammar. The instrument consists of 20 writing tasks five for each dimension constructed based on indicators derived from the writing model and adapted to the academic context of Arabic as a second language. For instance, in the Ideas and Content dimension, students were asked to write a paragraph expressing their opinion on a social topic with supporting arguments; the Structure and Organization dimension included tasks such as writing a narrative with appropriate

paragraphing and transitions; Style and Rhetoric involved persuasive or descriptive writing adapted to the audience; and Language and Grammar included tasks focusing on accurate use of Arabic verb patterns, morphology, and sentence construction.

Students' written responses were evaluated using structured assessment criteria aligned with the four dimensions. While a detailed rubric is not included in this article, scoring was conducted systematically by assessing the relevance, organization, linguistic accuracy, and rhetorical effectiveness of each response. The total writing score for each student was then used for further psychometric analysis. This scoring allowed for item-level data to be subjected to EFA and CFA, validating the internal structure of the instrument. The approach ensures that writing performance is assessed holistically and that the instrument meets empirical standards of validity and reliability, making it a robust tool for evaluating Arabic writing skills in a higher education context.

Data Analysis

The data analysis in this study involved two stages: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), aimed at testing the construct validity of the Arabic writing skills assessment instrument. The first stage, EFA, was conducted to explore the factor structure without prior assumptions to determine the number of significant factors based on eigenvalues and parallel analysis. This process included oblimin rotation to identify relationships between factors, with a loading factor ≥ 0.40 as the significance criterion. After the initial factor structure was identified, the next stage was CFA, which aimed to test the fit of the theoretical model based on the *Model Writing Competency* theory by Bereiter & Scardamalia (1987). The initial model comprised four dimensions: Ideas and Content, Structure and Organization, Style and Rhetoric, and Language and Grammar, each represented by five items.

The model was evaluated using model fit indices such as Chi-Square, CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), and RMSEA (Root Mean Square Error of Approximation) to ensure alignment with the data. Construct validity was assessed based on factor loading values, with items showing a factor loading ≥ 0.70 considered valid. Construct reliability was measured using Composite Reliability (CR) and Average Variance Extracted (AVE), with thresholds of $CR \geq 0.70$ and $AVE \geq 0.50$ indicating good reliability and consistency. Following the validation and reliability tests via CFA, the analysis proceeded with calculating the average writing skills scores of the sample using descriptive statistics. All analyses, including EFA and CFA, were conducted using R Studio software, employing the *psych* package for EFA and the *lavaan* package for CFA, ensuring precise and reliable analytical results.

RESULT AND DISCUSSION

Instrument Development

The development process of the Arabic writing skills instrument for students in this study follows the instrument development model by Oriondo & Dallo-Antonio (1984), which includes systematic steps to ensure the quality of the instrument. This instrument is designed based on the Writing Competency Model by Bereiter & Scardamalia (1987), which emphasizes four key dimensions of writing skills: ideas and content, structure and organization, style and rhetoric, and language and grammar. The first step is to define the measurement objectives, which aim to assess students' skills in writing Arabic as a second language based on these four dimensions. Then, test specifications are developed by

considering the type of text, difficulty level, and assessment components that align with the context of Arabic language learning at the university level.

Next, each dimension is broken down into specific indicators, which are then designed into instrument items in the form of writing tasks. The researcher uses a test to measure students' writing skills. The test, based on these four dimensions, is designed to comprehensively assess students' writing skills. Each dimension ideas and content, structure and organization, style and rhetoric, and language and grammar represents a key aspect of writing ability, so evaluating these four dimensions provides a holistic view of students' writing competence. At Table 1 below is an explanation of why these four dimensions are effective in assessing writing skills:

Table 1. Instrument Dimension Blueprint

Dimension	Blueprint
Ideas and Content	This dimension helps assess students' ability to generate relevant ideas and develop those ideas in depth. This test indicates whether students can understand the given topic and communicate their thoughts clearly and logically.
Structure and Organization	This dimension allows for the evaluation of students' ability to organize their writing systematically. By assessing the structure of the writing (introduction, body, and conclusion) and the coherence between paragraphs, instructors can gauge how well students can organize their ideas in a logical and understandable manner.
Style and Rhetoric	This dimension measures students' creative writing abilities, including how they vary sentence types and adjust their writing style to suit the communication purpose. The test provides insights into students' ability to engage readers and deliver messages in a way that is appropriate to the context.
Language and Grammar	This dimension enables the measurement of the technical aspects of writing skills, such as the correct use of grammar, vocabulary, spelling, and punctuation. This is important to ensure that students have a solid linguistic foundation for writing.

The instrument was then validated by experts to assess the relevance and clarity of each item, and revised based on their feedback. Following this, a pilot test was conducted with a sample of students to collect empirical data. The analysis of the pilot test included Exploratory Factor Analysis (EFA) and construct validity using Confirmatory Factor Analysis (CFA) to ensure that each item reflects the four dimensions of writing skills, as well as reliability testing. Items that did not meet the validity or reliability criteria were revised or removed. The final instrument includes a variety of writing tasks and a scoring rubric that reflects the indicators of each dimension. This process ensures that the instrument is valid, reliable, and appropriately measures students' writing skills in Arabic comprehensively, making it an effective evaluation tool for Arabic language learning in higher education.

Exploratory Factor Analysis (EFA)

Kaiser-Meyer-Olkin (KMO) Test and Bartlett's Test of Sphericity

Before conducting the model, fit test using CFA, Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity were performed to ensure that the data met the requirements for factor analysis. The analysis results revealed that the Kaiser-Meyer-Olkin (KMO) value was 0.96, indicating a very high level of sample adequacy for conducting factor analysis. A KMO value close to 1, as in this study, suggests that the correlations between the instrument items are sufficient to form a clear factor structure. Additionally, the results of

Bartlett's Test of Sphericity, with a significance value of 0.00 ($p < 0.05$), indicated that the correlation matrix among the items was significantly different from an identity matrix, thus supporting the assumption that the data can be analyzed using factor analysis techniques. With a very high KMO value and a significant Bartlett's Test, it can be concluded that the data is of excellent quality for factor analysis, both for the exploratory and confirmatory stages of the factor model.

Eigen values and Parallel Analysis

In this study, an Exploratory Factor Analysis (EFA) has been conducted to identify the underlying factor structure of the variables under investigation. The results of the EFA will be presented through several indicators, including the eigenvalue, scree plot, and parallel analysis. These three results provide an overview of the number of significant factors as well as the contribution of each variable to the main factors identified in the data. The results of the eigen value are as follows:

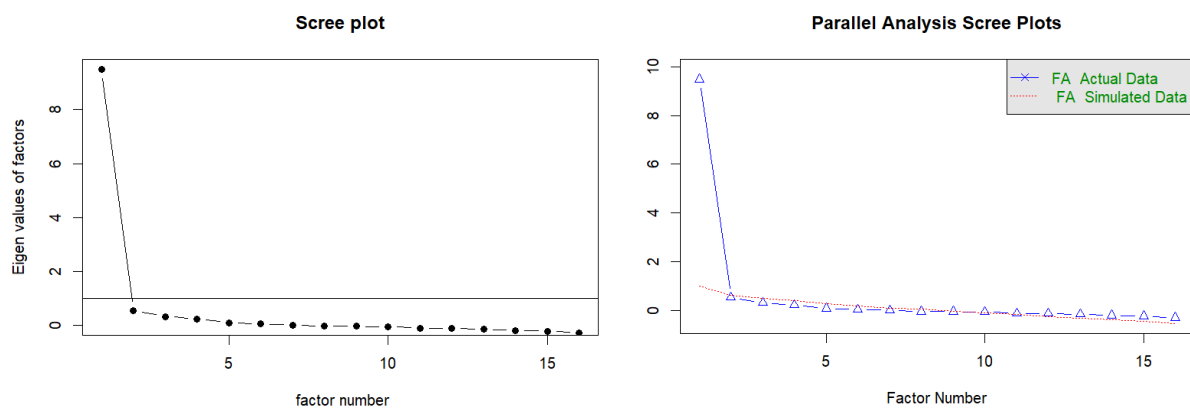


Figure 1. Scree plots Eigen values and parallel analysis

Based on the Scree Plots analysis, the results indicate that the first factor has a much higher eigenvalue compared to the other factors, with a significant decrease from the first to the second factor, forming an elbow point. This suggests that the first factor contributes the most in explaining the variance of the data. According to the Kaiser criterion, only factors with an eigenvalue greater than 1 are considered significant, and in this scree plot, only one factor meets this criterion, while the other factors have eigenvalues close to or below 1. These findings suggest that the data structure is predominantly unidimensional, where one main factor dominantly explains the relationships among the variables in the data.

Loading Factor EFA

Table 2. Loading Factor EFA

Factor Analysis, Rotation = PA			
	Item	Loading Factor	Results
Principal Component (PA1)	X2	0.841	Accepted
	X3	0.768	Accepted
	X4	0.788	Accepted
	X5	0.756	Accepted
	X6	0.777	Accepted
	X7	0.800	Accepted
	X8	0.816	Accepted
	X9	0.678	Accepted
	X10	0.763	Accepted
	X11	0.766	Accepted

Factor Analysis, Rotation = PA		
X12	0.790	Accepted
X13	0.726	Accepted
X14	0.709	Accepted
X15	0.757	Accepted
X16	0.809	Accepted

The results of the factor loading analysis (Table 2) show the contribution of each variable to the main factor (PA1) in the dataset. Based on these results, the factor loading values range from 0.679 to 0.841, indicating a strong correlation between each variable and the main factor. In general, factor loading values above 0.60 are considered significant, meaning that all the variables analyzed have a valid contribution to the formation of the main factor. The variable with the highest factor loading is x2 (0.841), which shows that this variable has the strongest correlation with the main factor, while the variable with the lowest factor loading is x9 (0.679), which, although lower compared to the other variables, still falls within the significance threshold. Overall, the relatively homogeneous factor loading values suggest that all variables consistently represent the same construct without any outliers or irrelevant variables. These findings indicate that the first factor (PA1) is a sufficient single dimension to explain the relationships among the variables in the dataset, and can be interpreted as the main construct underlying all the variables tested. In the context of instrument development, these results confirm that each item or variable contributes evenly to the same dimension, which supports the overall validity of the instrument.

Confirmatory Factor Analysis (CFA)

Loading Factor CFA

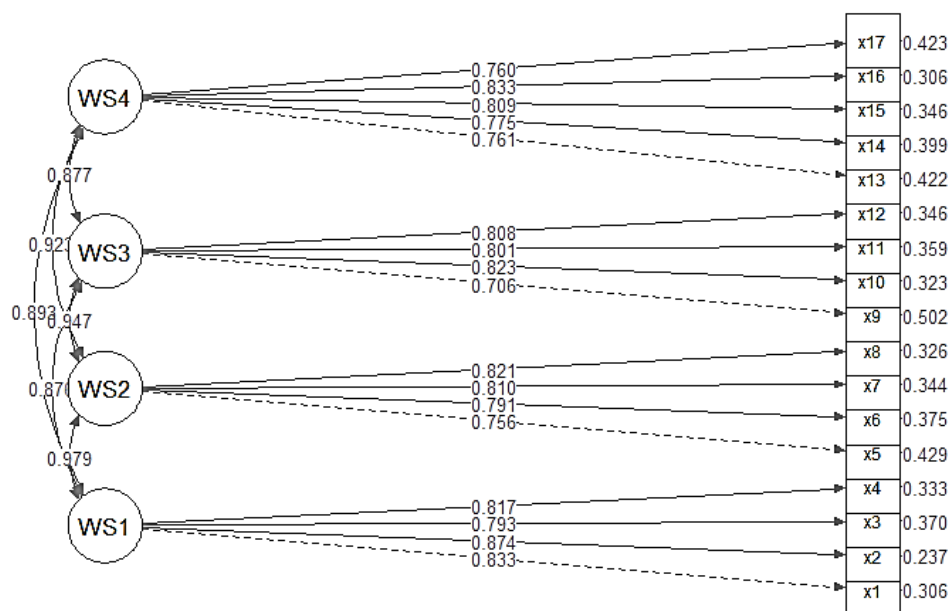


Figure 2. Loading Factor Writing skills

Based on the Writing Competency Model by Bereiter & Scardamalia (1987), writing skills consist of several key dimensions: Ideas and Content, Structure and Organization, Style and Rhetoric, and Language and Grammar. In the results of the CFA model analysis, these four dimensions are mapped into the latent factors WS1, WS2, WS3, and WS4. The high factor loading values on most of the indicators suggest that this instrument generally represents these four dimensions well.

Dimension WS1 (*Ideas and Content*): This dimension relates to the writer's ability to generate relevant, deep, and clear ideas in the writing. The factor loading values for the indicators in WS1 range from 0.817 to 0.833, indicating that these indicators strongly contribute to measuring students' ability to generate ideas and content in their writing. This is consistent with the Bereiter & Scardamalia (1987) theory, which emphasizes the importance of organizing ideas as a core element of writing competency.

Dimension WS2 (*Structure and Organization*): This dimension reflects the writer's ability to arrange ideas logically and systematically in a clear structure. The factor loading values for the indicators in WS2 range from 0.756 to 0.821, showing a strong relationship between the indicators and the latent factor. This is consistent with the Bereiter & Scardamalia (1987) theory, which highlights that the organization of writing is a key component of effective writing processes.

Dimension WS3 (*Style and Rhetoric*): This dimension is concerned with the ability to use engaging language styles and effective rhetorical strategies to communicate messages to the audience. In WS3, the factor loading values range from 0.706 to 0.923, with specific indicators like x9 and x11 showing very high contributions (above 0.90). This indicates that the indicators in this dimension are generally effective in measuring aspects of style and rhetoric, as explained in the Bereiter & Scardamalia (1987) model.

Dimension WS4 (*Language and Grammar*): This dimension includes the ability to use grammar, spelling, and punctuation correctly. The factor loading values for WS4 range from 0.761 to 0.833, indicating that the indicators in this dimension consistently measure aspects of language and grammar. This dimension aligns with the Bereiter & Scardamalia (1987) theory, which states that technical aspects of writing form the foundation of good writing competency. Overall, all indicators have factor loading values above 0.70, indicating excellent validity.

Goodness of Fit Indexes for the Measurement Model of Arabic Writing Skills

Tabel 2. Model Fit Evaluation

Fit Indexes	Fit index value	Good fit
Root Mean Square error of Approximation (RSMEA)	0.076	≤ 0.08
Standardized Root Mean Square (SRMR)	0.045	≥ 0.50
Comparative Fit Index (CFI)	0.922	≥ 0.90
Turkey-Lewis Index (TLI)	0.906	≥ 0.90

The results of the model fit evaluation (Table 2) indicate that the hypothesized model has a fairly good fit with the analyzed data. The Root Mean Square Error of Approximation (RMSEA) has a value of 0.076, which is within an acceptable range (< 0.08), indicating that the model has a relatively low approximation error and can be considered a moderately good fit. Additionally, the Standardized Root Mean Square Residual (SRMR) is 0.045, which is below the ≤ 0.05 threshold, indicating an excellent model fit.

For the comparative-based indicators, the Comparative Fit Index (CFI) is 0.922, and the Tucker-Lewis Index (TLI) is 0.906, both exceeding the minimum threshold of 0.90. This suggests that the model has an adequate fit, although there is room for improvement to make it more optimal (for example, by increasing the CFI and TLI values to ≥ 0.95 for an excellent fit). Overall, the model can be considered a moderately to well-fitting model, with the results showing that the hypothesized factor structure aligns with the empirical data. However, minor improvements to the model may further enhance the fit.

Construct Validity and Reliability

Table 3. Construct Validity and reliability

Construct Validity and reliability	Fit index value	Good fit
Average Variance Extracted (AVE)	0.669	≥ 0.50
Composite Reliability (rho_a)	0.776	≥ 0.70
Cronbach's Alpha	0.961	≥ 0.70

The results of the validity and reliability analysis in Table 3 above, of the construct indicate that the model demonstrates excellent quality. The Average Variance Extracted (AVE) of 0.669 exceeds the minimum threshold of 0.50, meaning that the construct explains more than 66% of the variance in its indicators. This indicates strong convergent validity, where the indicators within the construct are highly correlated. Regarding reliability, the Composite Reliability (CR/rho_a) of 0.776 indicates good internal reliability, as the value is above the minimum threshold of 0.70. Additionally, the very high Cronbach's Alpha of 0.961 signifies outstanding internal consistency among the indicators within the construct, although such a high reliability value may also suggest redundancy among some of the indicators. Overall, the construct exhibits strong convergent validity and excellent reliability, making it a valid and reliable instrument for measuring the intended construct.

Descriptive Statistics Writing Skills

Table 4. Average Description of Writing Skills

SMT	N	Descriptive statistics		
		Mean	Std. Deviation	Range
3	87	65.4625	11.27779	34.00
5	81	68.1481	11.83333	46.00
7	80	73.5862	7.35567	44.00
Total	248	69.1895	10.79513	46.00

Based on the descriptive statistics on Table 4 above, there is a noticeable trend of increasing average scores (mean) as the semesters (SMT) progress. In the third semester, the average score was 65.462 from 87 respondents, with a standard deviation of 11.27, indicating considerable variation in scores among students in this semester. The fifth semester saw a higher average score of 68.14 from 81 respondents, with a standard deviation of 11.83, reflecting slightly greater variability compared to the third semester. The seventh semester recorded the highest average score of 73.58 from 80 respondents, with a lower standard deviation of 7.355, indicating that the scores were more concentrated around the mean.

Overall, the average score across all semesters was 69.18 from a total of 248 respondents, with a standard deviation of 10.79, reflecting a moderate level of variation. These results suggest that students tend to show improvement in academic performance in the later semesters, which may be attributed to the maturation of learning skills or increased adaptation to academic demands. The lower standard deviation in the seventh semester also indicates that student performance in the final semester is more consistent compared to earlier semesters.

The results of this study reveal a significant difference between the outcomes of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) in the development of a measurement instrument for writing skills. During the EFA stage, the analysis identified only one significant main factor, indicating that the data has a relatively simple structure, with most variables interrelated within a single overarching dimension.

This finding aligns with the fundamental principle of EFA, which is exploratory in nature and often results in simpler factor structures, especially when the tested variables are highly correlated or similar. However, although the EFA results pointed to a single dominant factor, the CFA stage was based on the theoretical framework of writing skills, which posits the existence of four distinct dimensions, each reflecting a critical aspect of writing skills.

The CFA results demonstrated that the data fit better with this four-factor model, consistent with existing theories describing writing skills as a complex and multidimensional ability. The difference between the EFA and CFA results can be explained by the differing purposes and approaches of these two analytical methods. EFA provides a general and exploratory overview of the factor structure in the data without any prior assumptions (Akar, 2025), whereas CFA tests more specific hypotheses grounded in existing theory and literature (Bean & Bowen, 2021). Therefore, while EFA suggested a simpler factor structure, the CFA findings supporting the four-dimensional model provide stronger evidence of the complexity of writing skills, which encompass various stages and distinct abilities (Yilmaz & Jafarova, 2022).

This study emphasizes the importance of integrating both exploratory and confirmatory approaches in instrument development to ensure measurement validity and accuracy, particularly in the context of measuring multidimensional skills like writing. In psychometric literature and instrument development, the discrepancy between EFA and CFA outcomes is not uncommon, and in many cases, such differences are acceptable as long as they are supported by a robust theoretical foundation (Sureshchandar, 2023b). However, it is essential to recognize that EFA and CFA serve different purposes, and discrepancies between their results should not be viewed as problematic or universally undesirable.

Several researchers argue that differences between EFA and CFA do not necessarily indicate flaws in the instrument or data but instead reflect differences in their methodological approaches. Costello & Osborne (2005) suggest that EFA often yields simpler models, particularly when the tested variables are highly correlated. Conversely, in the context of CFA, more complex models are often appropriate as they better align with existing theoretical frameworks. Hair (2011), in their book *Multivariate Data Analysis*, argue that although EFA and CFA may produce divergent results, this does not imply that the instrument is invalid. Researchers must provide a strong theoretical rationale to justify CFA results, particularly when significant differences from EFA outcomes arise, especially if the CFA model aligns with well-established theories and literature.

Thus, it can be concluded that differences between EFA and CFA outcomes do not signify flaws in the instrument but rather reflect the distinct analytical approaches of exploratory and confirmatory methods. EFA offers an initial, simplified view of the data structure, whereas CFA is designed to test more complex and theory-driven models. In the context of my study, this difference offers significant novelty by demonstrating that the development of a theoretically grounded instrument for writing skills captures more complex dimensions compared to the initial exploratory insights from EFA. Accordingly, this study underscores the critical role of a robust theoretical foundation in instrument development, particularly for assessing multidimensional skills such as writing. These findings also contribute to the literature by highlighting that theory-based model validity can be better represented through CFA, even when the initial structure derived from EFA appears simpler.

CONCLUSION

Based on the results of the study, the development of an instrument to measure Arabic writing skills demonstrated excellent outcomes through both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). During the EFA stage, the instrument successfully met the established criteria, as evidenced by significant eigenvalues, support from parallel analysis, and strong loading factor values for all items. These findings indicate that each item contributes validly to the main factor formed, reflecting the suitability of the instrument's initial structure. Furthermore, the CFA analysis supported the validity of the theoretical model previously established. All statistical indicators, such as RMSEA, SRMR, CFI, TLI, AVE, Composite Reliability, and Cronbach's Alpha, showed excellent values, confirming that the four-dimensional model developed based on the theory of writing skills aligns well with the empirical data. Additionally, the loading factors in CFA demonstrated significant contributions of each item to their respective dimensions, reinforcing the consistency and reliability of the measurements.

REFERENCES

- Akar, S. G. M. (2025). Students' Disengagement in Online Courses: Validity and Reliability of an Instrument. *Journal of Education and Learning*, 19(1), 506–514. <https://doi.org/10.11591/edulearn.v19i1.21733>
- Akomah, B. B., & Ramani, P. V. (2024). Confirmatory Factor Analysis of the Positive Factors Influencing the Performance of Ghanaian Construction Projects. *Construction Innovation*. <https://doi.org/10.1108/CI-08-2022-0190>
- Al-Janabi, A., Al-Zubaidi, E. A., & Abdulzhrara Al-Sagheer, R. H. (2020). Encapsulation of Semantic Description with Syntactic Components for the Arabic Language. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 961–967.
- Baharum, H., Ismail, A., Awang, Z., Mckenna, L., Ibrahim, R., Mohamed, Z., & Hassan, N. H. (2023). The Study Adapted Instruments Based on Confirmatory Factor Analysis (CFA) to Validate Measurement Models of Latent Constructs. *International Journal of Environmental Research and Public Health*, 20(2860).
- Bean, G. J., & Bowen, N. K. (2021). Item Response Theory and Confirmatory Factor Analysis: Complementary Approaches for Scale Development. *Journal of Evidence-Based Social Work (United States)*, 18(6), 597–618. <https://doi.org/10.1080/26408066.2021.1906813>
- Bereiter, C., & Scardamalia, M. (1987). The Psychology of Written Composition. In *The psychology of written composition*. Lawrence Erlbaum Associates, Inc.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Brown, H. D. (2007). *Principles of Language Learning And Teaching* (5th Ed). Longman.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (Second Edi). Guilford Publications.
- Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four

- Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*, 10(7), 135–138. <https://doi.org/10.7275/jyj1-4868>
- D'Souza, R. (2021). What Characterises Creativity in Narrative Writing, and How do we Assess it? Research Findings from a Systematic Literature Search. *Thinking Skills and Creativity*, 42, 100949. <https://doi.org/10.1016/j.tsc.2021.100949>
- Erguvan, I. D. (2021). Gathering Evidence on E-rubrics: Perspectives and Many Facet Rasch Analysis of Rating Behavior. *International Journal of Assessment Tools in Education*, 8(2), 454–474. <https://doi.org/https://doi.org/10.21449/ijate.818151>
- Hair, J. F. (2011). Multivariate Data Analysis: An Overview. In *International Encyclopedia of Statistical Science* (pp. 904–907). Springer Berlin Heidelberg.
- Mellinger, C. D., & Hanson, T. A. (2020). Methodological Considerations for Survey Research: Validity, Reliability, and Quantitative Analysis. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 172–190. <https://doi.org/10.52034/lanstts.v19i0.549>.
- Oriondo, L. L., & Dallo-Antonio, E. M. (1984). *Evaluating Educational Outcome: Test, Measurement, Evaluation*. Rex Book Store, Inc.
- Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2021). Assessing language development in Arabic: The Arabic language: Evaluation of function (ALEF). *Applied Neuropsychology: Child*, 10(1), 37–52. <https://doi.org/10.1080/21622965.2019.1596113>
- Rufaiqoh, E., Sutaman, S., Arifa, Z., Ibrahim ElJack Ibrahim, N., & Asy'ari, H. (2023). Jacob's Analytical Assessment in the Evaluation of Arabic Writing Skills at Madrasah Aliyah Negeri Program Keagamaan. *Alsinatuna*, 9(1), 29–40.
- Sureshchandar, G. S. (2023a). Quality 4.0 – A Measurement Model using the Confirmatory Factor Analysis (CFA) Approach. *International Journal of Quality & Reliability Management*, 40(1), 280–303. <https://doi.org/10.1108/IJQRM-06-2021-0172>
- Sureshchandar, G. S. (2023b). Quality 4.0 – A Measurement Model Using the Confirmatory Factor Analysis (CFA) Approach. *International Journal of Quality & Reliability Management*, 40(1), 280–303. <https://doi.org/10.1108/IJQRM-06-2021-0172>
- Yilmaz, E., & Jafarova, G. (2022). Development of Data Driven Decision Making Scale: A Validity and Reliability Study. *Research on Education and Psychology*, 6(Special Issue), 69–91. <https://doi.org/10.54535/rep.1104114>
- Zakaria, M. K., Saputra, R. H., Embong, A. M., Khadim, K. A., Ab Wahab, N. A., Hailan, S., Abdul Kadir, F. K., Munawarah, J., Abdul-Kadir, N. A., Nasir, N. Y. M., & Kadir, K. A. (2024). Similar but Not Same: Language Barriers and the Facet of Life Faced by the Indonesians as International Students in Malaysia. *International Journal of Sustainable Development and Planning*, 19(4), 1347–1359. <https://doi.org/10.18280/ijstdp.190412>
- Zaki, M. (2021). Self-correction through Corpus-based Tasks: Improving Writing Skills of Arabic Learners. *International Journal of Applied Linguistics*, 31(2), 193–210. <https://doi.org/10.1111/ijal.12312>