

## Research Article

# Dynamics of Content Knowledge: Learned from Indonesian Arabic Teachers

Mohammad Samsul Ulum <sup>1</sup>, Wahyu Indah Mala Rohmana <sup>2</sup>, and Melly Elvira <sup>3</sup>

<sup>1</sup>Teacher Professional Education Programs, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

<sup>2</sup>English Language Education Department, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

<sup>3</sup>Islamic Early Childhood Education Department, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

Correspondence should be addressed to Mohammad Samsul Ulum; [samsul@pai.uin-malang.ac.id](mailto:samsul@pai.uin-malang.ac.id)

Received 16 December 2022; Revised 27 April 2023; Accepted 28 April 2023; Published 15 May 2023

Academic Editor: Yuqing Geng

Copyright © 2023 Mohammad Samsul Ulum et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The dynamics of content knowledge among Indonesian Arabic teachers is a critical area of research due to the increasing demand for Arabic language education. This study aims to investigate the content knowledge of Arabic teachers in Teacher Professional Education Programs at UIN Malang, Indonesia. The research utilized factor analysis and exploratory factor analysis to explore the structure of the developed instrument measuring teachers' content knowledge. The findings revealed that the reliability of the test items in the instrument was in the moderate category, with variations in difficulty levels between two test packages. The components of *Nahwu* and *Shorof* were found to have the strongest influence on Package 1, while *Balaghah* had the strongest effect on Package 2. The results of Rasch model analysis indicated three categories of proficiency levels among Arabic language teachers (low, moderate, and high), providing insights for developing more accurate assessment instruments and training programs. The limitations of the study include the focus on content knowledge only and the need for further analysis with relevant methods in future studies. The implications of this research contribute to the understanding of Arabic language teachers' content knowledge dynamics and can inform the development of effective training programs and assessment instruments for Arabic language education.

## 1. Introduction

Teachers play a crucial role in achieving educational goals by formulating learning goals, choosing appropriate materials, employing effective teaching methods, and evaluating learning outcomes [1–4]. Moreover, research and scholarly consensus indicate that the quality of teachers directly impacts students' academic success [5–7]. To ensure teacher quality, educators must possess extensive and deep knowledge of the subjects they teach, including eloquent communication skills in the language of instruction, understanding of basic concepts and subject structure, and awareness of how subject knowledge is constructed, organized, and interconnected. Thus, giving more input to teaching leads to better performances

of students and higher awareness to participate in learning activities [8].

One significant factor that influences teacher quality is their professional knowledge, specifically their content-related knowledge. This refers to the teacher's understanding of facts, concepts, principles, methodologies, and generalizations related to the subject matter, which shapes their pedagogical thinking and decision-making. Studies conducted in Angola by Gunasekaran et al. [9] and in Nigeria by Ayeni [10] have shown that professional knowledge, including mastery of content-specific skills, significantly impacts the way teachers teach in the classroom. This supports the assertion by Widodo [11] that content knowledge is a crucial factor for teachers. Mazlan et al. [12] in their study discovered that there

is a positive and strong relationship between teachers' content knowledge of Arabic language and teaching effectiveness. Furthermore, Zakaria et al. [13] found that teachers with stronger content knowledge are more likely to use effective practices that help students construct and internalize knowledge, corroborating previous research that highlights the positive association between teacher content knowledge and student learning.

Currently, there is a pressing issue with the Arabic language competence of teachers, as their mastery of the language is often limited, resulting in suboptimal Arabic language teaching [14]. Kamarul Shukri and Mohd Hazli [15] stated that the requirements for precise language competence for Arabic language teachers have not been accurately identified, and there is a lack of instruments to measure teachers' competence in Arabic language teaching. Studies have also indicated that a limited and shallow understanding of the Arabic language is a contributing factor to the limited content knowledge among Arabic language teachers, as highlighted by Sirait [7] in discussing challenges in Arabic language teaching. Therefore, recognizing the competence of teachers in the Arabic language is essential for mapping purposes and providing the necessary teaching materials to improve the quality of Arabic language education.

In efforts to enhance teacher professional and content knowledge, many teachers participate in teacher professional development programs, such as the Teacher Professional Education Program (TPEP). For Arabic language teachers, this program provides learning materials in the form of Arabic modules that cover content knowledge. However, a current challenge is the limited availability of valid and reliable measurement scales for assessing Arabic content knowledge among teachers.

In language learning, particularly Arabic, the teacher must have adequate content knowledge, which includes mastery of several language aspects such as *Nahwu* (syntax), *Shorof* (morphology), *Balaghah* (semantics), and *Mu'jamiyah* (lexical). Previous studies have examined the difficulty levels of these language competencies. The easier competences that were mentioned were *Nahwu* and *Shorof* because *Shorof* only learns the formation, and words change according to the desired meaning, the origin of the word, and the change in word class. Meanwhile, *Nahwu* only study the relationship between words in sentences and grammar expression. A study by Mariyam [16] revealed that the content knowledge in *Nahwu* contributed 76.1% to the ability to read Arabic books. Apparently, *Shorof* is also considered a problem for Arabic students, specifically for beginners who want to learn Arabic speech, writing, or translating the L1 into Arabic [17]. Then *Mu'jamiyah* (lexical) knowledge also significantly influenced Arabic skills since it studied about vocabularies [18]. From these studies, *Balaghah* was identified as the most challenging aspect in Arabic, followed by *Nahwu* and *Shorof* [19]. *Balaghah*, which encompasses Arabic semantics and rhetoric, involves the understanding of the meanings of words, their usages, and the ability to express ideas eloquently and persuasively. Mastery of *Balaghah* is crucial for teachers to effectively teach Arabic language learners how to use language in various contexts, such as reading, writing, speaking, and listening.

Assessing the content knowledge of Arabic language teachers, particularly in *Balaghah*, is vital for identifying their strengths and weaknesses and providing targeted professional development opportunities. However, there is currently a lack of valid and reliable measurement scales specifically designed for assessing *Balaghah* content knowledge among Arabic language teachers. Existing measurement scales tend to focus on general language proficiency or do not comprehensively cover the specific aspects of *Balaghah*, making them inadequate for accurately assessing teachers' content knowledge in this area.

To address this gap, there is a need to develop and validate a measurement scale that specifically assesses *Balaghah* content knowledge among Arabic language teachers. This measurement scale should be based on robust theoretical frameworks and undergo rigorous validation processes to ensure its validity and reliability. It should also take into consideration the contextual factors of the Arabic language teaching environment, such as the level of proficiency of Arabic language learners, the curriculum, and the instructional practices commonly used in Arabic language classrooms.

Thus, this study aimed to find out the dynamics of content knowledge of Arabic teachers seen from the test instruments and the result of TPEP examination in UIN Malang. It also seeks to measure the validity and reliability of the test instruments in TPEP. Once a valid and reliable measurement scale for *Balaghah* content knowledge among Arabic language teachers is available, it can be utilized to assess teachers' strengths and weaknesses in this area, provide targeted professional development opportunities, and ultimately enhance the quality of Arabic language education. Additionally, the findings from the assessment can inform curriculum development, instructional practices, and policy decisions related to Arabic language education.

In conclusion, teacher content knowledge, particularly in the area of *Balaghah*, is a critical factor that impacts the quality of Arabic language education. However, there is a current gap in valid and reliable measurement scales for assessing *Balaghah* content knowledge among Arabic language teachers. Developing and validating a measurement scale specifically for this purpose is essential for accurately assessing teachers' content knowledge, providing targeted professional development opportunities, and ultimately improving the quality of Arabic language education.

## 2. Research Method

The objective of this study is to conduct an exploratory analysis on the dynamics of content knowledge among Arabic language teachers who are enrolled in the TPEP at UIN Malang. The study focuses on Arabic language teachers who participated in the TPEP and joined the professionalism exam. This study aimed to enhance the understanding of the content knowledge of Arabic language teachers who participated in TPEP, with a particular emphasis on their content knowledge dynamics.

Convenience sampling technique was used to select participants who met the research criteria, including geographical proximity to the research location, availability of time,

TABLE 1: Demographic profile of respondents based on gender, years of teaching experience, and teaching institution.

Variable	Criteria	Test Package 1	Test Package 2
Gender	Male (M)	103 (46.61%)	104 (47.06%)
	Female (F)	118 (53.39%)	121 (54.75%)
Teaching experience	<10 years	47 (21.27%)	37 (16.74%)
	$10 \leq x \leq 15$ years	45 (20.36%)	54 (24.43%)
	>15 years	129 (58.37%)	134 (60.63%)
Teaching institution	Elementary school (SD)	96 (43.44%)	100 (45.25%)
	Junior high school (SMP)	88 (39.82%)	78 (35.29%)
	Senior high school (SMA)	37 (16.74%)	47 (21.27%)

commitment to participate, and ease of access [20]. The sample comprised of 446 Arabic language teachers, which was considered sufficient as it exceeded the minimum sample size of 384. It was calculated based on the proportionate sampling formula from the population ( $N$ ) of 6675.

Participants voluntarily agreed to participate without coercion. Prior to the research, participants were provided with complete information about the research objectives, procedures, risks, benefits, and their rights as participants and were requested to provide written informed consent. Participant confidentiality was maintained through the use of codes or unidentifiable identification, and only authorized research team members had access to the data. Participant identities remained anonymous, and no personal information was disclosed in the research findings. Participant data privacy was ensured through secure storage and data deletion after the research was completed. Bias control measures were implemented to minimize potential biases in data collection and analysis.

The data obtained from the participants consisted of the results of the professionalism exam, which comprised of two test packages, namely Test Package 1 and Test Package 2. The sample of 446 Arabic language teachers was divided into two groups, with 221 teachers answering Test Package 1 and 225 teachers answering Test Package 2. Table 1 presents the demographic profile of the respondents based on gender, years of teaching experience, and teaching institution, as reported in Test Package 1 and Test Package 2. The table provides information on the number of male and female respondents, teaching experience categorized into three categories, and the distribution of respondents based on their teaching institutions.

Test Package 1 and Test Package 2 are different instruments, but they were derived from the same indicators and dimensions. Test Package 1 originally comprised of 110 items but underwent data reduction analysis resulting in the retention of only 49 items. Similarly, Test Package 2 initially contained 105 items, but after data reduction analysis, only 61 items were retained. The reduction of test items was meticulously conducted with a rational objective of selecting items that would optimally contribute to the test reliability in each package. This reduction was undertaken to enhance the quality of the tests and ensure that the tests utilized in this research possess high reliability, thus improving the validity of the research results and minimizing any

TABLE 2: KMO and Bartlett's test.

Kaiser–Meyer–Olkin measure of sampling adequacy	0.809
Bartlett's test of sphericity	
Approx. $\chi^2$	7,027.667
$df$	1,176
Sig.	0.000

confusion or misconceptions among readers regarding the number of test items employed.

The instruments were analyzed using exploratory factor analysis (EFA), a component of structural equation modeling, to validate the measurement model of the latent constructs used in this study. EFA was employed to identify the factors that predominantly influence the Arabic language competency of teachers. Furthermore, through EFA, the underlying factors of the data structure, specifically the four linguistic aspects of  $N$  = syntax (*Nahwu*),  $S$  = morphology (*Shorof*),  $B$  = *Balaghah*,  $M$  = lexical (*Mu'jamiyah*), were confirmed for the appropriateness of the instruments in measuring the Arabic language proficiency of teachers. EFA entails the process of reducing variables into main factors that are interrelated based on variable correlations, which represent the underlying constructs of the data. The data were subsequently analyzed to assess the test quality, including reliability, validity, and item characteristics. Additionally, this analysis facilitated the determination of teachers' pedagogical abilities in teaching based on the aforementioned factors in preparing competent Arabic teachers.

### 3. Results

**3.1. EFA for Package 1.** EFA is a statistical technique employed in psychometrics to identify and group correlated variables into smaller, interpretable factors. This method aids researchers in simplifying complex data into organized dimensions, facilitating data interpretation. In this study, EFA was utilized to analyze data collected from tests or questionnaires in order to uncover underlying patterns or structures. The EFA conducted in this study involved several stages, including the following.

**3.1.1. Sample Adequacy Test.** This stage assesses the adequacy of the sample by examining the Kaiser–Meyer–Olkin (KMO) measure, with a minimum threshold of 0.5 [21]. Additionally, Bartlett's test of sphericity is conducted to confirm that the

TABLE 3: Total variance explained.

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative (%)	Total	% of Variance	Cumulative (%)
1	27.060	55.225	55.225	27.060	55.225	55.225
2	5.979	12.202	67.427	5.979	12.202	67.427
3	4.481	9.145	76.572	4.481	9.145	76.572
4	3.548	7.241	83.813	3.548	7.241	83.813
5	0.627	4.953	88.766			

correlation matrix of the items is not an identity matrix, with a significant  $\chi^2$   $p$ -value ( $p$ -value < 0.05) [22]. The results of the analysis in Table 2 reveal that KMO = 0.809, indicating that the sample meets the minimum threshold. Bartlett's test yielded  $\chi^2(67) = 7,027.6$ ,  $p = 0.001$ , suggesting that the sample used in the analysis is sufficiently adequate and the data are suitable for factor analysis, as it is not random or structured.

The subsequent step involves testing the adequacy of the sample by examining the anti-image diagonal correlation or measures of sampling adequacy (MSA) in order to evaluate the strength of the correlation between one component and other components in the correlation matrix. The minimum correlation criterion is set at 0.5, as recommended by Neill [23], indicating that the data used in the factor analysis are not random or unstructured and thus suitable for factor analysis. Based on the feasibility test conducted to assess the data suitability and sample adequacy, the data in this study meet the criteria for adequacy for factor analysis, as evidenced by the results of KMO, Bartlett's test, and MSA correlation that meet the criteria.

**3.1.2. Component Feasibility Test.** Feasibility test for components is conducted to ensure that the data do not deviate in the factor analysis by evaluating the communality of each component. The acceptable range for communality generally falls within 0.4–0.7, categorized as low to moderate [22]. It reveals that the communality of the components ranges from 0.7 to 0.9, indicating that these variables are strongly related to each other and significantly contribute to the extracted factors. Communality is an estimate of the amount of variance in each variable that can be explained by the extracted factors in the factor analysis.

A communality range between 0.7 and 0.9 can be considered an indicator of good convergent validity, implying that the variables used in the factor analysis have a high degree of similarity in measuring the same construct. Thus, the results of the factor analysis can be relied upon in identifying the underlying factors of the analyzed construct.

**3.1.3. Determining the Number of Factors.** Determining the number of factors is done by examining the cumulative percentage of variance explained by the extracted factors. Table 3 shows that out of 49 factors, four factors have eigenvalues >1, indicating that these factors are able to explain a significant amount of variance in the data. Therefore, it can be concluded that the four extracted factors are significant in explaining the linguistic aspects being analyzed. The total variance explained by these four factors is 83.81%, indicating that the extracted factors account for a large portion of the

variation in the data. However, there is still ~16.19% of unexplained variation by these factors, which may be attributed to other factors not included in the analysis or natural variability in the data. The consensus on the cumulative percentage of variance in factor analysis varies depending on the field of research conducted [22]. For instance, in the field of science, the minimum cumulative variance is usually set at 95%, while in the social sciences, the cumulative variance is generally around 50%–60% [19].

The preliminary analysis of the proficiency of Arabic language teachers revealed the presence of four factors. However, the initial results of the factor analysis showed interrelated factors that were challenging to interpret clearly. Hence, varimax rotation was employed to enhance the interpretation of these factors. The varimax rotation involved rotating the factor loadings identified, resulting in simplified factors with high loadings on one or two variables. This aided in reducing the complexity of the factors and facilitated a more comprehensible interpretation of the results in the context of prior research. The rotated results of the four components, along with instrument communality and reliability, are concurrently presented in Table 4.

The factor analysis results revealed the presence of four factors identified after rotation using the varimax method, as shown in Table 4 provided. The table includes information on communalities, Cronbach's  $\alpha$  with item deleted, and the contribution of each item to the identified factors. Communalities, which indicate the extent to which the variation of an item can be explained by the identified factors, ranged from 0 to 1, with higher communalities indicating a greater contribution of the item to the identified factors. In Table 4, it is observed that several items have high communalities above 0.8, such as N1, N2, N4, N5, N7, N22, N23, N27, N29, N30, N31, N32, S2, S5, S6, S7, S8, S9, S10, B12, B17, B20, B24, B28, B30, B31, and B37, indicating their significant contribution to the identified factors.

Internal reliability, measured by Cronbach's  $\alpha$ , is a reflection of the internal reliability of a factor or construct, with values ranging from 0 to 1. Higher values indicate higher internal reliability of the factor. The contribution of each item to the identified factors is also presented in Table 4, as seen from the communalities values. Items with higher communalities have a greater contribution to the identified factors, and items with high communalities (above 0.8) significantly contribute to the identified factors.

Furthermore, the EFA conducted on the Arabic language proficiency of teachers, which includes four linguistic aspects (*Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*), using the component matrix output from SPSS, concludes that:

TABLE 4: Component, comuality, and reliability statistics Package 1.

Item	Component				Komunalaitas	Cronbach's $\alpha$ if item deleted
	1	2	3	4		
N1	0.901				0.913	0.979
N2	0.898				0.906	0.979
N4	0.903				0.924	0.979
N5	0.876				0.906	0.979
N6	0.828				0.832	0.979
N7	0.886				0.881	0.979
N8	0.814				0.839	0.979
N10	0.878				0.876	0.979
N18	0.829				0.769	0.979
N22	0.905				0.901	0.979
N23	0.901				0.907	0.979
N24	0.844				0.809	0.979
N25	0.889				0.911	0.979
N27	0.901				0.923	0.979
N29	0.893				0.919	0.979
N30	0.859				0.836	0.979
N31	0.861				0.899	0.979
N32	0.899				0.928	0.979
S2			0.873		0.893	0.980
S5			0.899		0.929	0.980
S6			0.891		0.905	0.980
S7			0.877		0.898	0.980
S8			0.895		0.915	0.980
S9			0.870		0.866	0.980
S10			0.852		0.853	0.980
B4		0.828			0.822	0.979
B12		0.896			0.911	0.979
B17		0.893			0.918	0.979
B19		0.868			0.859	0.979
B20		0.891			0.868	0.979
B24		0.877			0.907	0.979
B25		0.838			0.819	0.979
B27		0.832			0.817	0.979
B28		0.929			0.917	0.979
B30		0.885			0.893	0.979
B31		0.904			0.898	0.979
B36		0.871			0.887	0.979
B37		0.910			0.910	0.979
M4				0.897	0.884	0.981
M7				0.949	0.916	0.981
M9				0.891	0.845	0.980
M6				0.902	0.837	0.980
M12			0.673		0.953	0.979
M15			0.645		0.951	0.979
M17			0.604		0.917	0.979
M18			0.569		0.827	0.979
M19			0.672		0.941	0.979
M20			0.649		0.945	0.979
M21			0.604		0.917	0.979
Cronbach's $\alpha$ (49 Items)						0.980



- (1) The syntactic aspect (N) has a significantly strong positive correlation with component 1 (0.901), and a moderate correlation with component 2 (0.17), component 3 (0.218), and component 4 (0.253). This indicates that the syntactic aspect has a significant influence on component 1, and a relatively weaker influence on components 2, 3, and 4.
- (2) The morphological aspect (S) has a significantly strong positive correlation with component 3 (0.899), and a moderate correlation with component 1 (0.128), component 2 (0.229), and component 4 (0.131). This suggests that the morphological aspect has a significant influence on component 3, and a relatively weaker influence on components 1, 2, and 4.
- (3) The *Balaghah* aspect (B) has a significantly strong positive correlation with component 2 (0.929), and a moderate correlation with component 1 (0.211), component 3 (0.285), and component 4 (0.151). This indicates that the *Balaghah* aspect has a significant influence on component 2, and a relatively weaker influence on components 1, 3, and 4.
- (4) The lexical aspect (M) has a significantly strong positive correlation with component 4 (0.897), and a moderate correlation with component 1 (0.297), component 2 (0.255), and component 3 (0.202). This suggests that the lexical aspect has a significant influence on component 4, and a relatively weaker influence on components 1, 2, and 3.

**3.1.4. Validity and Reliability.** In the context of EFA, various types of validity and reliability measures are commonly employed. These include determinant validity, which assesses the validity of factors identified in the EFA analysis; convergent validity, which evaluates the similarity of relationships among measurement variables that are purported to measure the same construct as the identified factors; and reliability, which gauges the consistency of results produced by the measurement tool. The utilization of validity and reliability measures in EFA analysis is essential to ensure the accuracy, consistency, and validity of the identified factor analysis results. Based on the findings presented in Table 4, the following explanations can be provided.

Construct validity can be ascertained from the communality of each item in the instrument, which consistently exceeds an average of 0.80. This indicates that the items in the instrument possess good construct validity, as they exhibit high communality, and most of the variance in these items can be accounted for by the common factor or construct being measured.

Convergent validity can be inferred from the corrected item-total correlation, which denotes the correlation between the item score and the total score of the instrument after excluding the score of that particular item. The table reveals that the corrected item-total correlation for each item in the instrument is relatively high, averaging above 0.30. This suggests that the items in the instrument exhibit good convergent validity, as they display a sufficiently high correlation

with the total score of the instrument and are capable of measuring the same construct as other validated measurement instruments.

Discriminant validity can be observed from the intercorrelation among item scores in the instrument. The table indicates that the intercorrelation among item scores in the instrument is relatively low, averaging below 0.80. This implies that the instrument possesses good discriminant validity, as the items in the instrument are not highly correlated with each other and can differentiate between different constructs.

Reliability: reliability refers to the extent to which a measurement instrument can produce consistent and stable results in repeated measurements. The reliability of the instrument can be assessed using Cronbach's  $\alpha$  if item deleted, which is the reliability coefficient of the instrument after excluding one item. According to Table 4, Cronbach's  $\alpha$  values for the identified factors are sufficiently high, ranging from 0.979 to 0.981, indicating good reliability for these factors. This signifies that the instrument exhibits very good reliability, as it demonstrates very high-reliability coefficients and is capable of producing consistent and stable results in repeated measurements.

### 3.2. The Characteristics of Item Questions with Rasch Model

**3.2.1. Difficulty Index.** Based on the Rasch model, the characteristics of item questions can provide valuable insights into the difficulty level of questions. These insights can help teachers in administering tests on Arabic language content knowledge. The Rasch model can be used to determine the distribution of item difficulty ( $\delta$ ) and teacher ability ( $\theta$ ). The difficulty index serves as an indicator of the likelihood of teachers successfully answering a question based on their ability. If the teacher's ability ( $\theta$ ) is greater than the item difficulty ( $\delta$ ), it can be inferred that the teacher has a higher chance of correctly answering the question, and vice versa. According to Sainuddin [24], the acceptable range for question difficulty falls within  $-2 \leq \delta \leq 2$  logits. Difficulty categories on the logit scale include difficult for  $\delta > 0.5$ , moderate for  $-0.5 \leq \delta \leq 0.5$ , and easy for  $\delta < -0.5$ . The table above presents the difficulty level categories for Arabic language content knowledge questions in Package 1.

Table 5 illustrates that 70% of the questions in Package 1 are categorized as having moderate difficulty, while the remaining 30% are classified as difficult. There are no questions categorized as easy in Package 1. From a psychometric perspective, all the items analyzed exhibit measurement outcomes that closely align with the expected values in the Rasch model, as evident from the infit mean-square values presented in Figure 1.

Question M21 ( $\delta = 1.493$ ) has the highest difficulty level in Package 1, while questions B5, B4, and B20 ( $\delta = -0.206$ ) are the easiest. In general, the *Mu'jamiyah* category has the highest number of questions with high difficulty levels, totaling six questions, while *Balaghah* has two questions categorized as having moderate difficulty (easier compared to other indicators). An interesting finding is that there are no questions categorized as easy in this package, whereas typically,

TABLE 5: Difficulty level categories for question Package 1.

Dimension	Difficulty level		
	Difficult	Moderate	Easy
<i>Nahwu</i>	N1, N8, N23, and N24	N2, N4, N5, N6, N7, N10, N18, N22, N25, N27, N29, N30, N31, and N32	–
<i>Shorof</i>	S2 and S7	S5, S6, S8, S9, and S10	–
<i>Balaghah</i>	B24, B25, and B31	B4, B12, B17, B19, B20, B27, B28, B30, B36, and B37	–
<i>Mu'jamiyah</i>	M9, M12, M15, M18, M19, and M21	M4, M6, M7, M17, and M20	–
Total	15	34	0

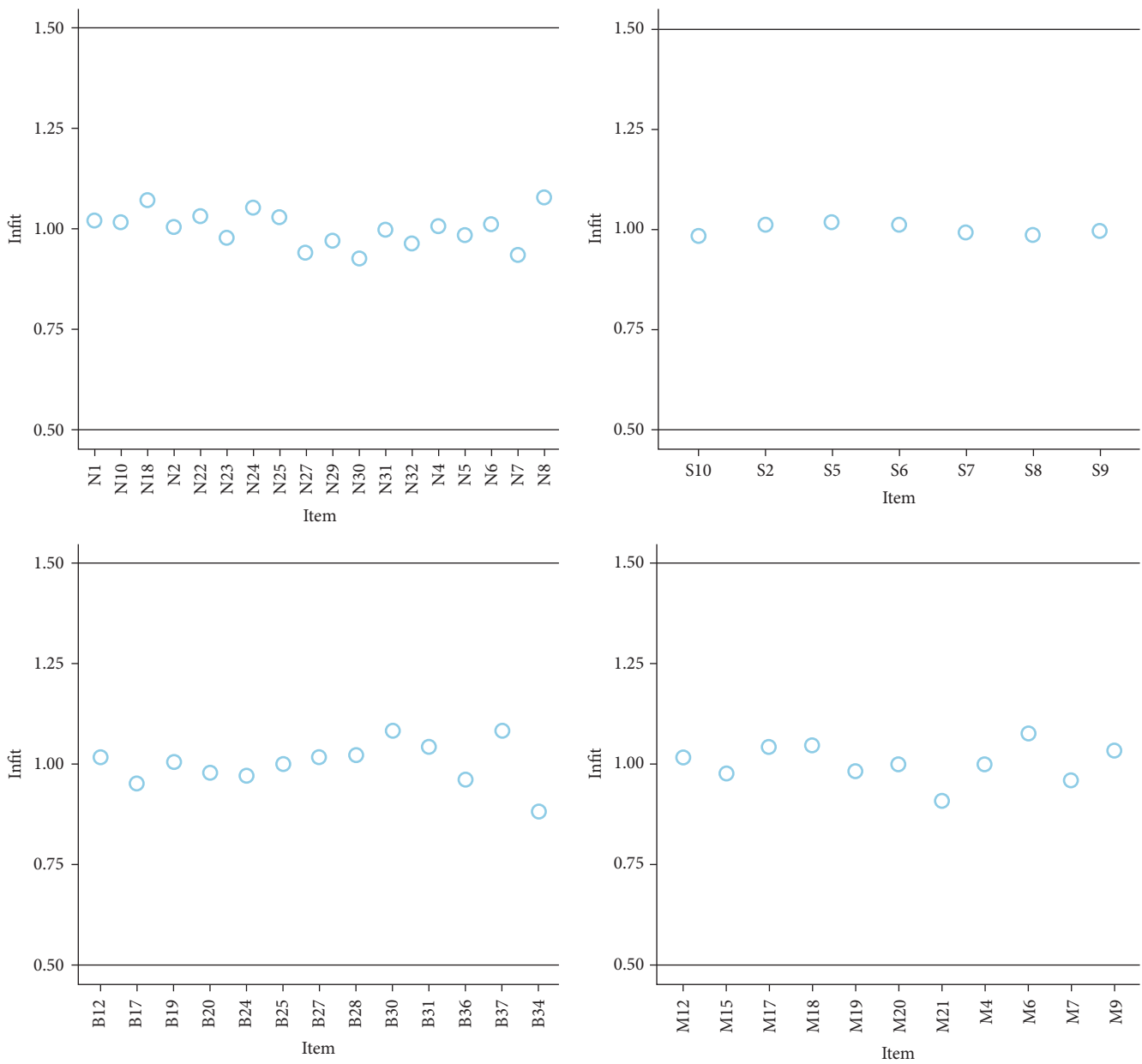


FIGURE 1: Infit mean square of questions in Package 1.

TABLE 6: Proficiency of arabic language teachers in answering questions in Package 1.

Category	$\theta$ (logit scale)	Frequency
Low	$<-1$	18 (8.16%)
Moderate	$-1 \leq x \leq 1$	180 (81.63%)
High	$>1$	23 (10.20%)

for proficiency tests, there should be around 10%–15% of questions categorized as easy. This may be due to the reduction of items during the initial analysis, as easy questions did not significantly contribute to measuring the Arabic language content knowledge of teachers.

3.2.2. *Arabic Language Teachers' Proficiency.* Table 6 presents information on the proficiency of Arabic language teachers in answering questions in Package 1, categorized into three levels: low, moderate, and high, based on the measured  $\theta$  (theta) values in the logit scale.

Table 6 presents the findings of the psychometric analysis of Arabic language teachers' abilities in answering Package 1 questions, as measured using a logit scale. The results indicate three categories of abilities: low, moderate, and high. A total of 8.16% of teachers were categorized as low, with  $\theta$  values  $<-1$ , while the majority of teachers, 81.63%, fell into the moderate category, with  $\theta$  values ranging from  $-1$  to  $1$ . Additionally, 10.2% of teachers were categorized as high, with  $\theta$  values  $>1$ . These findings offer valuable insights into the distribution of Arabic language teachers' abilities based on the logit scale, with the majority falling in the moderate category. This information can serve as a reference for the development of targeted training or teaching programs to enhance Arabic language teachers' abilities.

3.2.3. *Total Information Function (TIF) and Standard Error of Measurement (SEM).* Figure 2 displays the TIF and SEM for Package 1. The figure shows that the questions used in Package 1 provide optimal information in the ability range of  $-3.85$  to  $3.92$  logit (intersection of TIF and SEM). This suggests that teachers' abilities in answering Package 1 questions fall within the range of approximately  $-2.2$  to  $1.5$  logit. These measurement results offer valuable information regarding Arabic language content knowledge of teachers.

3.3. *EFA for Package 2.* Similar to the factor analysis conducted in the previous section, the analysis for Package 2 also went through several stages, including the following.

3.3.1. *Sample Adequacy Test.* The results of EFA output in Table 7 indicate that the data used have an adequate sample size, as evidenced by the KMO MSA value of 0.883, which is above the minimum threshold of 0.5 [17]. Additionally, Bartlett's test of sphericity shows that the correlations between variables in the data are statistically significant, with a  $\chi^2$  (1,830) = 54,253.143,  $p = 0.000$ . This indicates that the data used in the factor analysis are sufficiently representative, and the correlations between variables in the data are strong enough to conduct factor analysis.

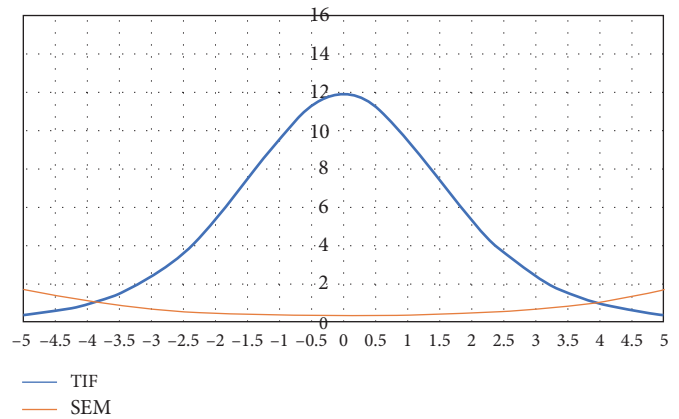


FIGURE 2: TIF and SEM Package 1.

TABLE 7: KMO and Bartlett's test.

Kaiser–Meyer–Olkin measure of sampling adequacy	0.883
Bartlett's test of sphericity	
Approx. $\chi^2$	54,253,143
df	1,830
Sig.	0.000

The next step is to test the adequacy of the sample by examining the anti-image diagonal correlations or MSA. This test is conducted to evaluate the strength of the correlation between one component and other components in the correlation matrix. The minimum correlation criterion is 0.5 [23], indicating that the data used in the factor analysis are not random or unstructured, making it suitable for factor analysis. Based on the feasibility test conducted to examine the data suitability and sample adequacy, the data in this study meet the criteria for sample adequacy for factor analysis. This is indicated by the results of the KMO, Bartlett's test, and MSA correlation, which meet the criteria.

3.3.2. *Component Feasibility Test.* The feasibility test is conducted to ensure that the data are not deviant in the factor analysis by evaluating the communality of each component. The minimum acceptable communality interval generally ranges from 0.4 to 0.7, which is categorized as low to moderate [22]. Table 4 shows that the communality of the components has a range of values between 0.7 and 0.9, which can be interpreted as indicating a strong correlation and significant contribution of these variables to the extracted factors. Communality is an estimation of the amount of variance in each variable that can be explained by the extracted factors in the factor analysis. A communality range between 0.7 and 0.9 can be considered as an indicator that the variables used in the factor analysis have good convergent validity, meaning that these variables have a high level of similarity in measuring the same construct. Thus, the results of the factor analysis can be relied upon in identifying the underlying factors of the analyzed construct.



TABLE 8: Total variance explained.

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative (%)	Total	% of Variance	Cumulative (%)
1	17.565	28.795	28.795	17.565	28.795	28.795
2	16.191	26.542	55.337	16.191	26.542	55.337
3	5.325	8.730	64.067	5.325	8.730	64.067
4	5.149	8.442	72.509	5.149	8.442	72.509
5	.770	6.770	79.279			

3.3.3. *Determining the Number of Factors.* The determination of the number of factors was carried out by examining the cumulative percentage of variance based on the formed factors. Table 8 shows that out of the 61 factors, four factors have eigenvalues  $>1$ . Eigenvalues  $>1$  indicate that these factors are able to explain a significant amount of variance in the data. Therefore, it can be concluded that the four extracted factors can be considered significant factors in explaining the linguistic aspects being analyzed. The total variance explained by these four factors is 72.51%. This indicates that the extracted factors are able to explain a large portion of the variation in the data. However, there is still  $\sim 27.93\%$  of unexplained variance by these factors, which may be caused by other factors not included in the analysis or natural variability in the data. There is no consensus on the cumulative percentage of variance in factor analysis, as it depends on the field of research being conducted [22]. For example, in the field of sciences, the minimum cumulative variance is typically set at 95%, while in the social sciences, it is generally around 50%–60% [19].

The results obtained from the initial analysis of Arabic language proficiency of teachers revealed four factors. However, the initial results of the factor analysis showed that these factors were interrelated and difficult to interpret clearly. Therefore, varimax rotation was used to improve the interpretation of these factors. Varimax rotation was performed by rotating the factor loadings found, resulting in simpler factors with high factor loadings on one or two variables. This helped reduce the complexity of the factors and made the interpretation of the factor analysis results more easily understood in the context of previous research. The rotated results of the four components, along with instrument communality and reliability, are presented collectively in Table 9.

The results of factor analysis indicate that four factors were identified after rotation using the varimax method. The provided table contains information regarding communality, Cronbach's  $\alpha$  if items were removed, and the contribution of each item to the identified factors. Based on Table 9, items with high communality (below 0.5) are items N11, N13, N12, S4, S7, S8, S9, S10, B53, B54, B55, B56, B37, B46, B48, B51, B52, B53, B54, and B55. Communalities below 0.5 indicate that these variables have a relatively low contribution to the common factors or other latent factors in the data table; thus, their communality is considered poor. However, these items still have theoretical or conceptual relevance to the common factors being analyzed. Despite their low communality, the

concepts or variables represented by these items are still relevant and important in the conceptual framework or theory used in the research or analysis.

Internal reliability (Cronbach's  $\alpha$ ): Cronbach's  $\alpha$  is a measure of the internal reliability of a factor or construct. Cronbach's  $\alpha$  values range from 0 to 1, and higher Cronbach's  $\alpha$  values indicate higher internal reliability of the factor. Contribution of each item to the identified factors: the table also presents the contribution of each item to the identified factors, which can be seen from the communality values. The higher the communality value of an item, the greater its contribution to the identified factors. Therefore, items with high communality values (above 0.8) have a significant contribution to the identified factors.

Furthermore, the results of the EFA conducted on the Arabic language proficiency of teachers, which includes four linguistic aspects (*Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*) using the component matrix output from SPSS, concluded that some items with high factor loadings on specific components could be identified as the strongest indicators representing certain components or aspects of teachers' Arabic language proficiency, as follows:

- (1) The first factor has the highest loading for items N11, N12, and N13, representing the syntactic (*Nahwu*) aspect of Arabic language, indicating that these items represent the syntactic (*Nahwu*) aspect of teachers' Arabic language proficiency.
- (2) The second factor has the highest loading for items S4, S7, S8, S9, and S10, representing the morphological (*Shorof*) aspect of Arabic language.
- (3) The third factor has the highest loading for items B2, B3, B4, B6, B8, B9, B10, B11, B12, B13, B15, B17, B19, B21, B22, B23, B25, B26, B27, B28, B30, B32, B33, B34, B35, B36, B37, B38, B40, B43, B45, B46, B47, B48, B49, B50, B51, B52, B53, B54, B55, and B56, representing the *Balaghah* aspect of Arabic language. This indicates that these items represent the morphological *Balaghah* (rhetoric) aspect of teachers' Arabic language proficiency.
- (4) The fourth factor has the highest loading for items M4, M9, M11, M12, M14, M16, and M19, representing the lexical (*Mu'jamiyah*) aspect of Arabic language, indicating that this component represents the lexical (*Mu'jamiyah*) aspect of teachers' Arabic language.

TABLE 9: Component, comuality, and reliability statistics Package 2.

Items	Component				Communalities	Cronbach's $\alpha$ if item deleted
	1	2	3	4		
N11	0.518				0.430	0.927
N13	0.444				0.398	0.928
N12	0.447				0.404	0.927
S4			0.652		0.430	0.927
S7			0.630		0.410	0.927
S8			0.664		0.448	0.927
S9			0.687		0.482	0.927
S10			0.672		0.465	0.928
B2		0.961			0.924	0.923
B3		0.955			0.914	0.923
B4		0.954			0.916	0.923
B6		0.972			0.945	0.923
B8		0.921			0.853	0.923
B9		0.944			0.895	0.923
B10		0.905			0.824	0.924
B11		0.922			0.852	0.924
B12		0.900			0.816	0.924
B13		0.946			0.898	0.923
B15		0.961			0.924	0.923
B17		0.906			0.830	0.924
B19		0.943			0.899	0.923
B21		0.939			0.884	0.924
B22		0.913			0.836	0.924
B23		0.904			0.820	0.924
B25		0.937			0.887	0.923
B26		0.963			0.932	0.923
B27		0.624			0.455	0.925
B28		0.663			0.493	0.926
B30		0.947			0.910	0.925
B32		0.953			0.914	0.924
B33		0.941			0.899	0.924
B34		0.956			0.929	0.925
B35		0.907			0.833	0.924
B36		0.944			0.898	0.925
B37		0.899			0.810	0.925
B38		0.936			0.885	0.924
B40		0.909			0.830	0.925
B43		0.943			0.892	0.925
B45		0.957			0.921	0.924
B46		0.886			0.791	0.925
B47		0.934			0.876	0.925
B48		0.957			0.920	0.924
B49		0.944			0.905	0.924
B50		0.912			0.839	0.925
B51		0.949			0.907	0.925
B52		0.957			0.921	0.924
B53		0.587			0.367	0.926
B54		0.563			0.338	0.926
B55		0.791			0.015	0.929
B56		0.876			0.031	0.929
M4			0.932		0.870	0.927

(continued)

TABLE 9: Continued.

Items	Component				Communalities	Cronbach's $\alpha$ if item deleted
	1	2	3	4		
M9			0.950		0.916	0.928
M11			0.922		0.872	0.929
M12			0.956		0.920	0.928
M14			0.945		0.902	0.928
M16			0.955		0.920	0.928
M19				0.698	0.495	0.928
M20				0.761	0.586	0.928
M22				0.743	0.563	0.928
M26				0.788	0.624	0.928
M28				0.771	0.597	0.928
Cronbach's $\alpha$ (61 Items)						0.927

**3.3.4. Validity and Reliability.** There are several types of validity and reliability commonly used in EFA analysis, namely: determinant validity, which measures the validity of the factors identified in the EFA analysis. Convergent validity measures the appropriateness of the relationships between the measurement variables considered to measure the same construct as the identified factors. Reliability, which measures the consistency of measurement results. It is important to use validity and reliability in EFA analysis to ensure the accuracy, consistency, and validity of the identified factor analysis results. Based on the analysis results presented in Table 9, the following can be explained.

Construct validity indicates that the communality for each instrument item is very high, with an average above 0.70. This indicates that the items in the instrument have good construct validity, as they have high communality, and most of the variation in these items can be explained by the common factor or construct being measured. Convergent validity can be seen from the corrected item-total correlation, which is the correlation between the item scores and the total instrument scores after removing the item scores. Based on the table, it can be seen that the corrected item-total correlation for each instrument item is sufficiently high, with an average above 0.30. This indicates that the items in the instrument have good convergent validity, as they have a sufficiently high correlation with the total instrument score and can measure the same construct as other tested and validated measurement instruments. Discriminant validity can be seen from the intercorrelation between item scores in the instrument. Based on the table, it can be seen that the intercorrelation between item scores in the instrument is relatively low, with an average below 0.80. This indicates that the instrument has good discriminant validity, as the items in the instrument are not highly correlated with each other and can distinguish between different constructs. Reliability: reliability refers to the extent to which a measurement instrument can produce consistent and stable results in repeated measurements. The reliability of an instrument can be seen from Cronbach's  $\alpha$  if item deleted, which is the reliability coefficient of the instrument after removing one item. Based on Table 9, it can be seen that Cronbach's  $\alpha$  values for the identified factors are

quite high, ranging from 0.923 to 0.929, indicating good reliability for these factors. This indicates that the instrument has very good reliability, as it has very high-reliability coefficients and is capable of producing consistent and stable results in repeated measurements.

#### 3.4. The Characteristics of Item Questions with Rasch Model

**3.4.1. Difficulty Index.** The characteristics of Rasch model items indicate the level of difficulty of questions. These characteristics show how difficult or easy the questions are for teachers to answer when implementing a test on Arabic language content knowledge. Studies using the Rasch model can be chosen to determine the distribution of item difficulties ( $\delta$ ) and teacher's abilities ( $\theta$ ). The difficulty index indicates the likelihood of a teacher successfully answering a question with their current ability. If the teacher's ability  $\theta > \delta$ , it can be said that the teacher has the opportunity to answer the question correctly, and vice versa. According to Sainuddin [24], the acceptable range of item difficulties is within the interval of  $-2 \leq \delta \leq 2$  logit scale. Difficulty categories on the logit scale are as follows:  $\delta > 0.5$  indicates difficult,  $-0.5 \leq \delta \leq 0.5$  indicates moderate, and  $\delta < -0.5$  indicates easy. Table 10 below informs that generally, 70% of the questions in Package 2 fall into the moderate difficulty level category, with only 30% of questions categorized as easy difficulty. In contrast to the previous Package 1, which did not have any easy difficulty questions, Package 2, on the other hand, does not have any high-difficulty questions. The table below presents the distribution of difficulty categories for Arabic language content knowledge questions in Package 2.

Table 10 shows that there are no difficult questions in Package 2. This is likely due to question reduction, where high-difficulty questions were eliminated because they fell below the criteria. Statistically, all analyzed items had measurement outcomes that closely approached the expected values in the Rasch model. This is indicated by the infit mean square of all items falling within the acceptance range of 0.5–1.5, as presented in Figure 3.

The most difficult question in Package 2 is question B53 ( $\delta = 0.252$ ), while the easiest is question B22 ( $\delta = -0.916$ ).

TABLE 10: Difficulty level categories for question Package 2.

Dimension	Difficulty level		
	Difficult	Moderate	Easy
<i>Nahwu</i>	–	N11 and N12	N2, N10, and N13
<i>Shorof</i>	–	S4, S7, S8, S9, and S10	–
<i>Balaghah</i>	–	B4, B6, B9, B10, B11, B12, B13, B15, B17, B19, B21, B23, B27, B28, B30, B32, B36, B37, B40, B45, B46, B47, B48, B49, B50, B51, B52, B53, and B54	B2, B3, B8, B22, B25, B26, B33, B34, B35, B38, B43, B55, and B56
<i>Mu'jamiyah</i>	–	M11, M14, M19, M20, M22, and M26	M4, M9, M12, M16, and M28
Total	0	42	21

Overall, the category *Balaghah* has the highest number of questions categorized as moderate to easy compared to other dimensions. An interesting finding is that there are no questions categorized as difficult in this category, whereas a test instrument should ideally have 25%–30% of difficult questions to assess proficiency. This could be due to the initial item reduction analysis indicating that difficult questions did not significantly contribute to measuring the Arabic language content knowledge of teachers.

**3.4.2. Arabic Language Teachers' Proficiency.** Table 11 presents information on the proficiency of Arabic language teachers in answering Package 1 questions, categorized into three levels: low, moderate, and high, based on the  $\theta$  (theta) values measured in logit scale.

Based on the results of logit scale analysis of Arabic language teachers' abilities in answering Package 2 questions, it was found that 14 teachers (6.12%) were categorized as having low abilities, 193 teachers (85.71%) had moderate abilities, and 18 teachers (8.16%) had high abilities. Low ability in answering Package 2 questions may need to be improved through training or more effective learning approaches. Meanwhile, moderate ability is considered satisfactory, but there is still room for improvement in certain aspects. High ability, on the other hand, is considered excellent and can serve as a role model for other teachers in mastering Arabic language content. These findings are important to be further analyzed and reinforced with relevant statistical analysis methods to comprehensively understand the abilities of Arabic language teachers in answering Package 2 questions in the context of the research conducted.

**3.4.3. TIF and SEM.** Figure 4 shows that the TIF and SEM provide information that the questions used in Package 2 can provide optimal information in the ability range of  $-4.13$  to  $4.13$  logits (intersection of TIF and SEM). This result indicates that teachers' ability to answer Package 2 questions on a scale of  $-1.9$  to  $1.8$  is within the range where Package 2 provides optimal information. Thus, the information obtained from the measurement results using Package 2 provides optimal information about the Arabic language content knowledge of teachers.

## 4. Discussion

The results of the EFA revealed that four factors represent different aspects of Arabic language proficiency among teachers, as measured in this study. These factors are Arabic language teachers' knowledge of grammar (*Nahwu*), word forms (*Shorof*), rhetoric (*Balaghah*), and vocabulary (*Mu'jamiyah*), which are important components in assessing their knowledge [25, 26]. The factor analysis results indicated that their knowledge of word structure, grammar rules, root words, rhetorical figures, and appropriate word usage in context all play significant roles as main components of Arabic language teachers' knowledge content.

The EFA results also showed that the items developed in this study have strong relationships with the predetermined dimensions, and the number of factors formed aligns with the number of identified components. This indicates the validity and reliability of the instrument developed for measuring Arabic language teachers' knowledge content. These findings also reinforce the quality of the instrument in terms of construct validity, discriminant validity, convergent validity, and test reliability. Thus, the obtained measurement results can be considered accurate and dependable in assessing Arabic language teachers' proficiency in the studied knowledge content.

The research findings revealed interesting results regarding Package 1 and Package 2 tests in measuring Arabic language teachers' knowledge content. In the Package 1 test, the highest number of questions with a high difficulty level was found in the *Mu'jamiyah* component, while in the Package 2 test, the *Balaghah* component had the highest number of questions. These findings depict the characteristics of Package 1 and Package 2 tests in relation to *Mu'jamiyah* and *Balaghah* components in the context of this research and can serve as a basis for developing more accurate assessment instruments in future studies [27, 28].

The results of the Rasch model analysis revealed three categories of Arabic language teachers' abilities: low, moderate, and high. The majority of teachers fell into the moderate category, and this finding can serve as an important reference for the development of training or instructional programs. The analysis also indicated the need for improvement in

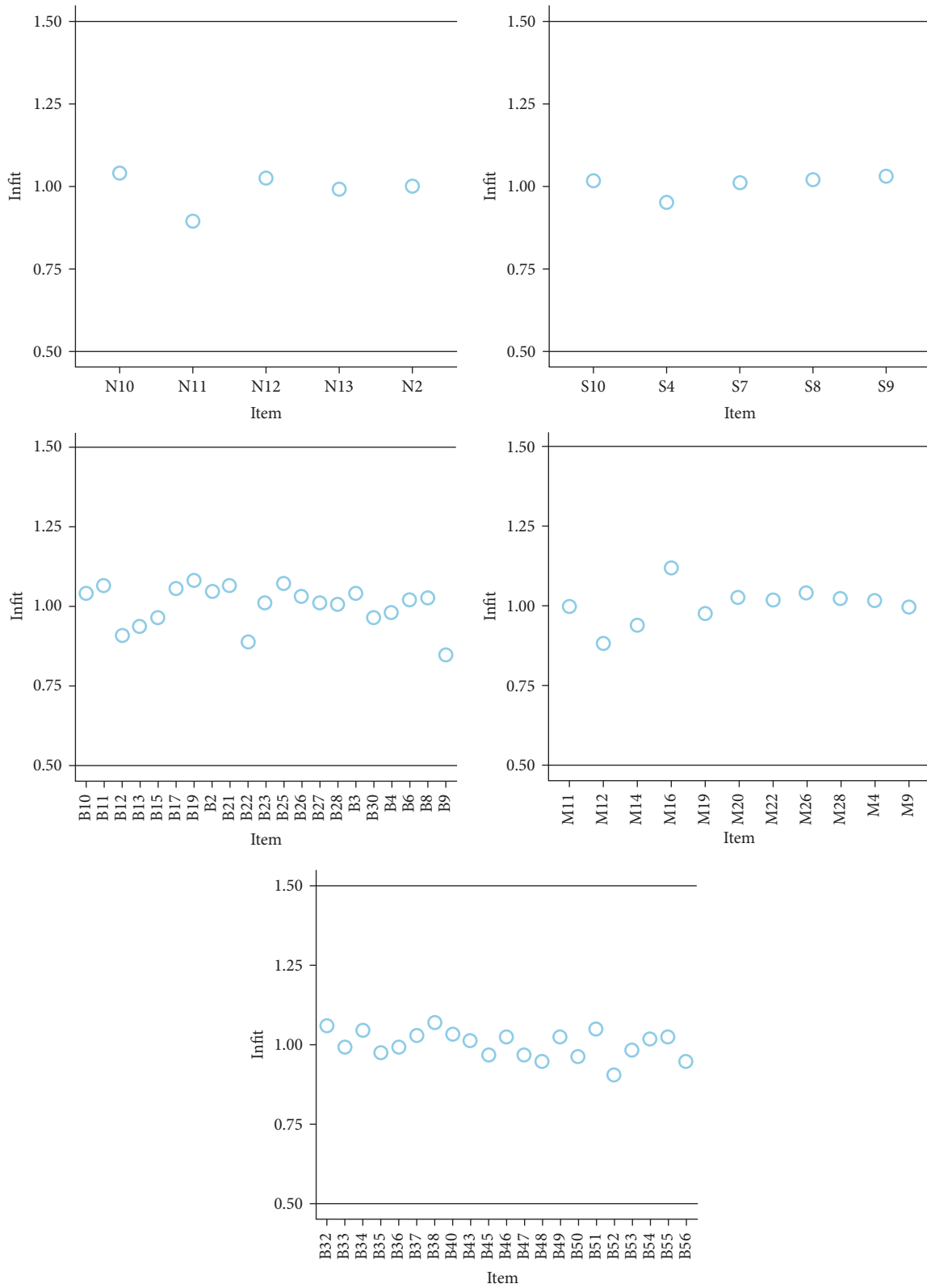


FIGURE 3: Infit mean square of Package 2 questions.



TABLE 11: Arabic language teachers' proficiency in answering questions in Package 2.

Category	$\theta$ (logit scale)	Frequency
Low	$< -1$	14 (6.12%)
Moderate	$-1 \leq x \leq 1$	193 (85.71%)
High	$> 1$	18 (8.16%)

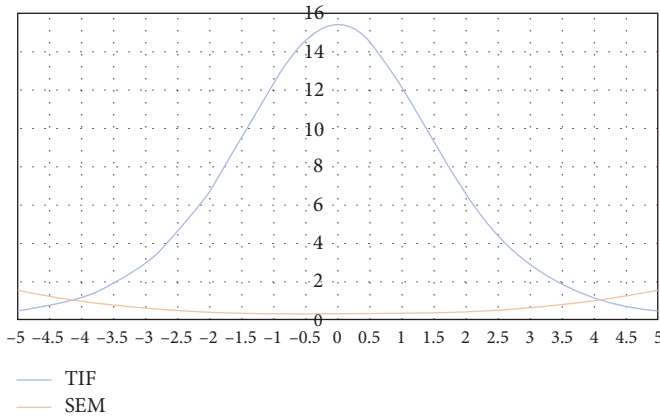


FIGURE 4: TIF and SEM Package 2.

teachers' abilities to answer questions in Package 1 and Package 2. These findings need to be further analyzed using relevant methods to comprehensively understand the Arabic language teachers' abilities in content knowledge within the TPEP at UIN Malang.

There are several theories that may be related to the findings of this study. For example, some Arabic linguistic theories may consider aspects of *Balaghah* or rhetoric as less relevant in everyday Arabic language proficiency and instead prioritize lexical aspects (*Mu'jamiyah*) as more important in daily communication [29, 30]. Additionally, there may be linguistic theories that propose that syntactic aspects (*Nahwu*) are not as significant in Arabic language proficiency and that understanding context and the use of phrases or expressions in everyday communication are far more important [31]. Furthermore, the results of the Rasch model analysis, which revealed three categories of Arabic language teachers' abilities (low, moderate, and high), can also serve as a basis for the development of training or instructional programs tailored to the needs of teachers in improving their knowledge in Arabic language content. Training or instructional programs that can enhance teachers' abilities to answer questions in Packages 1 and 2 can also be developed to improve their abilities in the components of *Mu'jamiyah* and *Balaghah*.

Moreover, the findings of this study can also be linked to the theory of professional development, which emphasizes the importance of continuous learning and improvement among teachers [32, 33]. Teachers who engage in ongoing professional development activities, such as workshops, seminars, and training programs, are more likely to enhance their knowledge and skills in their subject matter, which can positively impact their proficiency in teaching Arabic language

[34, 35]. Therefore, the results of this study can provide insights for the development of effective professional development programs for Arabic language teachers to improve their knowledge and abilities in different components of Arabic language proficiency.

Furthermore, the findings of this study have implications for curriculum development in Arabic language teacher education programs. The emphasis on different components of Arabic language proficiency, such as *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*, in this study, can inform the design of curriculum content and instructional strategies to better prepare Arabic language teachers for their teaching roles [36, 37]. For example, the curriculum can include specific modules or courses that focus on enhancing teachers' knowledge of word structure, grammar rules, root words, rhetorical figures, and appropriate word usage in context, which are identified as important components of Arabic language teachers' knowledge content in this study. Moreover, the curriculum can also incorporate opportunities for teachers to practice and apply their knowledge in real-world teaching contexts, such as through classroom observations, teaching practicum, and reflective activities, to further enhance their language proficiency and pedagogical skills.

In addition, the findings of this study can contribute to the improvement of Arabic language teacher assessment practices. The identification of different components of Arabic language proficiency and their relationship with test difficulty levels in this study can inform the development of more accurate and reliable assessment instruments for evaluating Arabic language teachers' knowledge and abilities [38, 39]. For instance, future assessment instruments can be designed to include a balanced representation of different components of Arabic language proficiency, such as *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*, to provide a comprehensive measure of teachers' language proficiency. Moreover, the findings of this study can also guide the development of assessment items that align with the identified components of Arabic language teachers' knowledge content to ensure that the assessment accurately reflects the specific knowledge and skills required for effective Arabic language teaching.

It is worth noting that this study has some limitations that should be acknowledged. First, the sample of Arabic language teachers in this study was limited to a specific TPEP at UIN Malang, which may not fully represent the diverse population of Arabic language teachers in different contexts. Therefore, caution should be exercised when generalizing the findings to other contexts. Future research could consider including a more diverse sample of Arabic language teachers from different regions, institutions, and levels of education to obtain a more comprehensive understanding of their knowledge and abilities. Second, the measurement instrument used in this study was developed based on the researchers' conceptualization of Arabic language proficiency and the identified components, which may not fully capture all aspects of Arabic language proficiency. Future research could consider using multiple measurement instruments or triangulation of data sources to strengthen the validity and reliability of the findings. Lastly, this study

focused on Arabic language teachers' knowledge and abilities but did not investigate other factors that may influence their proficiency, such as motivation, attitudes, and classroom practices. Future research could explore the interplay of these factors and their impact on Arabic language teachers' proficiency, such as their motivation to teach Arabic, attitudes toward the language, and classroom practices. Understanding the complex interplay between these factors and teachers' proficiency can provide a more holistic understanding of Arabic language teaching and can inform the development of targeted interventions to improve Arabic language teachers' proficiency.

In conclusion, this study contributes to the understanding of Arabic language teachers' knowledge and abilities in different components of Arabic language proficiency. The findings have implications for professional development, curriculum development, and assessment practices in Arabic language teacher education programs. However, it is important to acknowledge the limitations of this study, such as the limited sample size and measurement instrument used. Future research should consider addressing these limitations and further exploring other factors that may influence Arabic language teachers' proficiency. By addressing these gaps in the literature, we can continue to advance our understanding of Arabic language teaching and ultimately enhance the quality of Arabic language education.

## 5. Conclusion

In conclusion, this study investigated the components of Arabic language teachers' knowledge content, specifically *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*, and their relationship with test difficulty levels. The findings of this study contribute to the existing literature on Arabic language teaching by shedding light on the importance of these knowledge components for Arabic language teachers' proficiency. The results have implications for professional development programs, curriculum development, and assessment practices in Arabic language teacher education.

The study suggests that Arabic language teachers need to have a strong foundation in *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah* in order to effectively teach Arabic language to learners. These components are essential for teachers to understand the complex nature of Arabic language teaching and to provide quality education to their students. The findings also highlight the need for ongoing professional development programs to enhance Arabic language teachers' knowledge and abilities in these areas.

Furthermore, the study indicates that test difficulty levels are influenced by the proficiency of Arabic language teachers in *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah*. This implies that teachers with a higher level of proficiency in these components are likely to develop more challenging and appropriate assessments for Arabic language learners. Therefore, it is crucial to consider the proficiency of Arabic language teachers in these components when designing assessments for Arabic language learners.

Despite the limitations of this study, such as the small sample size and the focus on specific knowledge components,

the findings provide valuable insights for future research in the field of Arabic language teaching. Future research could explore the interplay of these knowledge components with other factors, such as teaching strategies, classroom practices, and student outcomes, to gain a more comprehensive understanding of Arabic language teachers' proficiency and its impact on Arabic language education.

This study contributes to the literature on Arabic language teaching by highlighting the importance of *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah* as essential components of Arabic language teachers' knowledge content. The findings have implications for teacher education, curriculum development, and assessment practices and call for further research to enhance our understanding of Arabic language teachers' proficiency and its implications for Arabic language education.

## Data Availability

The data that support the findings of this study are available upon reasonable request from the corresponding author, due to privacy or security restrictions. The data are not publicly available to protect the confidentiality of the study participants.

## Ethical Approval

This article followed all ethical standards for research without direct contact with human or animal subjects.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to thank for Universitas Islam Negeri Maulana Malik Ibrahim Malang all its facilities to support this research.

## Supplementary Materials

The supplementary information presented in this manuscript is the result of analyzing the Dichotomous Rasch Model using JAMOVI software, which includes a summary table of the Item Statistics, the Wright Map graph, and the Infit Item Plot for each component of *Nahwu*, *Shorof*, *Balaghah*, and *Mu'jamiyah* from two packages of test items (Package 1 and Package 2). These figures have been included as supplementary material to provide additional information related to the analysis and results of the study, and they are intended to support the findings and interpretations presented in the main body of the manuscript. (*Supplementary Materials*)

## References

- [1] P. Dao, "Effects of task goal orientation on learner engagement in task performance," *International Review of Applied Linguistics in Language Teaching*, vol. 59, no. 3, pp. 315–334, 2021.
- [2] T. Dicke, J. Elling, A. Schmeck, and D. Leutner, "Reducing reality shock: the effects of classroom management skills training on

- beginning teachers," *Teaching and Teacher Education*, vol. 48, pp. 1–12, 2015.
- [3] P. Hiver, A. H. Al-Hoorie, J. P. Vitta, and J. Wu, "Engagement in language learning: a systematic review of 20 years of research methods and definitions," *Language Teaching Research*, 2021.
  - [4] S. Z. Bin Tahir and Y. Rinantanti, "Multilingual lecturers' competence in English teaching at the University of Iqra Buru, Indonesia," *Asian EFL Journal*, vol. 5, pp. 79–92, 2016.
  - [5] R. Seebruck, "Teacher quality and student achievement: a multilevel analysis of teacher credentialization and student test scores in California High Schools," *McGill Sociological Review*, vol. 5, pp. 1–18, 2015.
  - [6] F. Faize, "Impact of teacher quality on the academic achievement of students at secondary stage in Punjab (Pakistan)," *European Journal of Social Sciences*, vol. 19, no. 1, pp. 97–105, 2011.
  - [7] S. Sirait, "Does teacher quality affect student achievement? An empirical study in Indonesia," *Journal of Education and Practice*, vol. 7, no. 27, pp. 34–41, 2016.
  - [8] Y. Geng and F. Huang, "Coupling coordination between higher education and environmental governance: evidence of western China," *PLOS ONE*, vol. 17, no. 8, Article ID e0271994, 2022.
  - [9] A. Gunasekaran, C. Patel, and E. Tirtiroglu, "Performance measures and metrics in a supply chain environment," *International Journal of Operations & Production Management*, vol. 21, no. 1/2, pp. 71–87, 2001.
  - [10] V. Ayeni, *Empowering the Customer: The Citizen in Public Sector Reform*, Commonwealth Secretariat, 2001.
  - [11] A. Widodo, "Teacher pedagogical content knowledge (PCK) and students' reasoning and wellbeing," *Journal of Physics: Conference Series*, vol. 812, Article ID 012119, 2017.
  - [12] N. A. Mazlan, Z. Arifin, S. Zailani et al., "Indicators for Arabic language lecturers' competencies in Universiti Teknologi Mara on teaching effectiveness: a hybrid PLS-SEM modelling approach," *Central Asia & the Caucasus*, vol. 23, no. 1, pp. 1603–1617, 2022.
  - [13] Z. M. Zakaria, A. Atan, R. Yusuf, and S. Y. Mei, "Content knowledge competency of Arabic language teacher trainees during teaching practice," *International Journal of Academic Research in Business and Social Sciences*, vol. 9, no. 9, pp. 307–319, 2019.
  - [14] F. Caena and C. Redecker, "Aligning teacher competence frameworks to 21st century challenges: the case for the European Digital Competence Framework for Educators (*Digcompedu*)," *European Journal of Education*, vol. 54, no. 3, pp. 356–369, 2019.
  - [15] M. T. Kamarul Shukri and Y. A. Mohd Hazli, "Pendidikan bahasa Arab di Malaysia: tinjauan literatur mengenai tahap penguasaan bahasa," *Diges Pendidik*, vol. 8, no. 2, pp. 91–102, 2008.
  - [16] S. Mariyam, "Hubungan penguasaan nahwu sharaf dengan kemampuan membaca kitab kuning pesantren riyadhul huda," *Jurnal Pendidikan Bahasa Arab*, vol. 2, no. 1, pp. 71–81, 2021.
  - [17] M. A. Abduh, Mustaufir, and R. N. Indah, "Arabic language production: challenges for Indonesian adult and children learners," in *International Conference Recent Innovation*, pp. 767–772, SciTePress, 2018.
  - [18] C. E. Setyawan, "Paradigma Studi bahasa Arab di perguruan tinggi agama islam (PTAI)," *Jurnal Pendidikan Bahasa Arab*, vol. 1, no. 1, pp. 17–43, 2018.
  - [19] H. Q. M. Asaad, M. K. Hasan, N. Alidoust, A. A. Shabdin, A. Ounissi, and M. D. Hossain, "The impact of morphological awareness intervention in enhancing L2 postgraduate students' productive vocabulary breadth," *Education Research International*, vol. 2022, Article ID 5930822, 11 pages, 2022.
  - [20] S. J. Stratton, "Population research: convenience sampling strategies," *Prehospital and Disaster Medicine*, vol. 36, no. 4, pp. 373–374, 2021.
  - [21] H. Taherdoost, S. Sahibuddin, and N. Jalaliyoon, "Exploratory factor analysis: concepts and theory," *Advances in Applied and Pure Mathematics*, vol. 27, pp. 375–382, 2010.
  - [22] S. Sainuddin, B. Subali, Jailani, and M. Elvira, "The development and validation prospective mathematics teachers holistic assessment tools," *Ingénierie des Systèmes d'Informa- Informatio*, vol. 27, no. 2, pp. 171–184, 2022.
  - [23] J. Neill, *Sample Factor Analysis Write-Up: Exploratory Factor Analysis of the Short Version of the Adolescent Coping Scale*, Centre for Applied Psychology, University of Canberra, 2008.
  - [24] S. Sainuddin, "Analisis karakteristik butir tes matematika pada tes buatan MGMP matematika kota makassar berdasarkan teori moderen (Teori Respon Butir)," *Jurnal Penelitian Matematika Dan Pendidikan Matematika*, vol. 1, no. 1, pp. 1–12, 2018.
  - [25] N. Baya'a and W. Daher, "The development of college instructors' technological pedagogical and content knowledge," *Procedia-Social and Behavioral Sciences*, vol. 174, pp. 1166–1175, 2015.
  - [26] T. N. Ma'mun, A. Kosasih, Y. Rohmayani, E. K. Fimansyah, and M. H. Al-khresheh, "Foreign language teachers' technological and pedagogical content knowledge: a study with AFL teachers in Indonesia," *Journal of Language and Linguistic Studies*, vol. 17, no. 4, pp. 1998–2021, 2021.
  - [27] H. Sudani, *The Impact of Ferdinand de Saussure on Arabic Linguistic Research: Arabic Reception of Linguistics*, al-Markaz al-'Arabī lil-Abḥāth wa-Dirāsāt al-Siyāsāt, 2019.
  - [28] R. Laatar, C. Aloulou, and L. H. Belguith, "Towards a historical dictionary for Arabic language," *International Journal of Speech Technology*, vol. 25, pp. 29–41, 2022.
  - [29] A. H. Abdullah and M. C. Omar, "A prolonged dilemma: traditional curriculum of Arabic rhetoric (Balaghah) in Malaysia," *The Social Sciences*, vol. 11, no. 6, pp. 1008–1014, 2016.
  - [30] Y. Suleiman, *The Arabic Language and National Identity*, Edinburgh University Press, 2019.
  - [31] U. Bahruddin, A. Gunawan, and L. Budianto, "Grammar cube strategy: the newest alternative media to learn Arabic grammar for beginner in Indonesia," *The Educational Review, USA*, vol. 4, no. 4, pp. 93–100, 2020.
  - [32] L. Darling-Hammond, S. P. Newton, and R. C. Wei, "Developing and assessing beginning teacher effectiveness: the potential of performance assessments," *Educational Assessment, Evaluation and Accountability*, vol. 25, pp. 179–204, 2013.
  - [33] R. M. Ingersoll and M. Strong, "The impact of induction and mentoring programs for beginning teachers: a critical review of the research," *Review of Education Research*, vol. 81, no. 2, pp. 201–233, 2011.
  - [34] M. Alosh, H. M. El-Khafai, and S.-D. Hammoud, "Professional standards for teachers of Arabic," in *Handbook for Arabic Language Teaching Professionals in the 21st Century*, pp. 409–417, Routledge, 2006.
  - [35] A. A. S. Tabieh, M. M. Al-Hileh, H. M. J. A. Afifa, and H. Y. Abuzagha, "The effect of using digital storytelling on developing active listening and creative thinking skills," *European Journal of Educational Research*, vol. 10, no. 1, pp. 13–21, 2021.
  - [36] R. Nasser, "Qatar's educational reform past and future: challenges in teacher development," *Open Review of Educational Research*, vol. 4, no. 1, pp. 1–19, 2017.

- [37] K. Wahba, Z. A. Taha, and L. England, *Handbook for Arabic Language Teaching Professionals in the 21st Century*, Routledge, 2014.
- [38] M. Al-Muslim, M. F. Ismail, S. A. Ghani, Z. Nawawi, M. A. Rahman, and M. N. Rostam, "What are the features of quality for Arabic teachers agreed by students and teachers?" *Journal of Education and e-Learning Research*, vol. 7, no. 1, pp. 56–63, 2020.
- [39] Zurqoni, H. Retnawati, S. Rahmatullah, H. Djidu, and E. Apino, "Has Arabic language learning been successfully implemented?" *International Journal of Instruction*, vol. 13, no. 4, pp. 715–730, 2020.