INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 13, Issue 1, (2024) E-ISSN: 2226-6348

# Developing E-Muhadathat Kit Instrument (I-KEM) for Non-Arabic Speakers: A Rasch Model Analysis

Siti Rahmah Borham, Saipolbarin Ramli, Mohammad Taufiq Abdul Ghani

Department of Modern Languages, Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, Malaysia

Abd Hadi Borham, Miftachul Huda

Department of Islamic Studies, Faculty of Human Sciences, Universiti Pendidikan Sultan Idris, Malaysia

Nurziaty Nordin

Department of Computer and Information Sciences, Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, Malaysia

**To Link this Article:** http://dx.doi.org/10.6007/IJARPED/v13-i1/20640 DOI:10.6007/IJARPED/v13-i1/20640

***Published Online:*** 18 January 2024

**Abstract**

More attention should be given to determining the precise instrument validity and reliability measurement tools in order to reduce error and produce a quality instrument. Hence, this study aims to identify the validity and reliability of the E-Muhadathat kit or (I-KEM) instrument, which has 21 items, via the Rasch Measurement Model. This model is believed to provide accurate estimates of construct validity and instrument reliability. This instrument aims to determine the development requirements for the E-Muhadathat kit targeted at non-Arabic speakers in Malaysian public higher learning institutions. Three experts in Malay, Arabic, and general knowledge were consulted to determine the instrument's face and content validity, which contains 21 items. I-KEM pilot questionnaires were distributed online using Google Forms to 50 respondents from three public Malaysian universities: MARA University of Technology (UiTM), University of Malaya (UM), and University Putra Malaysia (UPM). Subsequently, the validity and reliability of I-KEM were measured based on the Rasch Model using Ministeps software version 5.1.4.0. The findings show that I-KEM needs to be improved in terms of face validity and content validity on the recommendation of expert evaluators. In contrast, the measurement of the Rasch Model shows that I-KEM has good construct validity with a noise level of 10% and is controlled by a maximum value of 15%. In addition, the item reliability index of 0.96 and the person reliability index of 0.90 indicate a high degree of instrument reliability. This study presents significant findings in determining the developmental needs for the E-Muhadathat kit for non-Arabic speakers in Malaysian public higher learning institutions.

**Keywords:** Rasch Measurement, Validity, Reliability, Arabic Teaching And Learning, E-Muhadathat , Non-Arabic Speakers

## Introduction

Interactive multimedia plays an important role in revolutionizing education with technology and innovation, cultivating active learning practices to boost the quality of teaching and learning. In Arabic education, adopting Information and Communication Technology (ICT) has helped empower Arabic learning and produced professionals skilled in Arabic communication. As Marpuah (2015) described, Muhadathat entails communication or speech in Arabic and reflects an individual's capability to speak and practice Arabic fluently and spontaneously. In addition, these skills also require a psychological readiness in a person to speak a second or foreign language spontaneously and fluently (Hashimoto, 2002).

The concept of Muhadathat requires an emphasis on Arabic language communication activities among non-Arabic speakers (Tu'aimat, 2006). This is supported by the opinion of Ashour and al-Hawamidat (2014), who stated that speaking skills reflect the individual mind because information, ideas, and emotions can be conveyed through those skills. According to Ramli et al (2017), among the alternatives that Arabic language teachers need to do in creating interactive and communicative classrooms is through a teaching approach with media kits. Furthermore, as asserted in previous studies (Hat et al., 2013; Mahmuda, 2018), multimedia technologies can provide an interactive and student-centered learning environment. As a result, students will have a higher interest and motivation in Arabic learning. This, ultimately, improves the effectiveness of Arabic teaching and learning activities.

Moreover, the traditional learning method is predominantly one-way and teacher-centered, which can negatively impact students' conceptual understanding and learning motivation (Hassan et al., 2017). In addition, conventional learning also prevents students from adopting a student-centered, interactive learning environment, which multimedia technologies can greatly influence. This can ensure the effectiveness of teaching and learning through higher students' motivation and engagement in learning. Self-directed learning or heutagogy has become synonymous with today's modern era (Samin et al., 2020). Hence, in line with today's technology, this study plans to develop an interactive multimedia kit or the E-Muhadathat kit, which can simulate dialogues with native Arabic speakers in real-life settings. This kit is an Android smartphone app based on Natural Language Processing (NLP). It uses Artificial Intelligence (AI) technology to recognize the voices of native Arabic speakers and can be used to help students practice Arabic correctly in a real context.

## Problem Statement

Conventional teaching methods that are one-sided and passive lead to Arabic language teaching being seen as difficult, boring, and uninteresting (Zaini et al., 2017). According to Mazlan (2018), in universities, one-way teaching is still predominantly used in Arabic language classrooms, and lecturers tend to use traditional teaching approaches like lectures and discussions supplemented with less interactive visual and auditory teaching aids. The internet allows students to access online resources regardless of time and place and adopt self-directed learning or heutagogy (Blaschke, 2012). However, the reliance on conventional learning has prevented students from adopting a culture of heutagogy.

Studies have found that heutagogy is only moderately practiced in Arabic lessons. This highlights the need to extend this practice, specifically in higher education institutions in Malaysia (Yahya et al., 2013; Hamzah et al., 2019; Rahman & Ahmad, 2020; Hai et al., 2020; Yahaya et al., 2021; Ghani et al., 2022; Fuad & Al-Yahya, 2022). The development needs of this media kit are also due to the issues of speech, communication, and willingness to communicate (WTC) faced by Arabic language students in Malaysian universities (Haron, 2011; Ibrahim, 2013; Mohamad, 2009; Arshad & Bakar, 2012; Noor et al., 2014; Daud & Pisal, 2014; Tohlong, 2015; Borham, 2018; Rahman & Ahmad, 2020). Thus, before conducting the E-Muhadathat kit development needs analysis, the I-KEM instrument should be tested to ensure its validity and reliability with Rasch Model measurements are at a high level.

**Research Problem**

This study's main aim is to use the Rasch-based Model to determine the validity and reliability of the I-KEM instrument. The measurement encompasses the reliability item index, person index, item polarity, unidimensionality, standardized residual correlation, and item fit.

**Methodology**

This pilot study involved 50 Arabic language students in three Malaysian public universities (UiTM, UM, and UPM). A pilot study is administered to a small portion of the original sample to identify deficiencies and weaknesses of the instrument (Fraenkel and Wallen, 1996). This study used the E-Muhadathat kit development questionnaire (I-KEM) for non-Arabic speakers in higher education in Malaysia. This instrument was adapted from (Sahrir et al., 2017; Rahman et al., 2015; MacIntyre et al., 1998; Tohlong, 2015). Once the I-KEM instrument was developed, it was distributed to three experts in Malay, Arabic, and quantitative studies to test the face and content validity. Then, respondents were asked for their consent to participate in data collection.

Several enumerators helped to distribute informed consent forms and questionnaires online through the Google Form platform. Measurements based on the Rasch Model were used to identify the construct validity and instrument reliability. This measurement model formulates dichotomy data (in the form of right and wrong forms such as tests) and polytomy data (in the form of Likert scales such as questionnaires), which correlate respondents' ability with item difficulty (Rasch, 1980). According to Wright and Linacre (1992), the Rasch Model meets the criteria of scientific measurement principles such as determining linear measurements, overcoming the issue of missing data, providing accuracy estimates, detecting misfit or outlier items, and the measurement instrument is not dependent on the observed object parameters. This model is believed to provide accurate estimates of construct validity and instrument reliability (Aziz et al., 2015). Therefore, the data of this pilot study were analyzed using Ministeps software version 5.1.4.0.

**Validity of Instrument**

Punch (1998) described validity as how much the measurement accurately represents a concept. Therefore, face validity and content validity should be conducted by field experts to check aspects of language, structure, and sentence order used in the questionnaire items and to see the suitability of the items with the components in the measurement (Darusalam & Hussin, 2016). I-KEM instrument went through a process of face validity and content validity by three experts in the field of Arabic linguistics, Malay education, and quantitative research

from the Arabic department and the Malay department, Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris (UPSI).

Then, the researcher conducted a face validity and content validity measurement using the Content Validity Index (CVI) value to determine the average validity value of the instrument administered by experts. According to Davis (1992), the CVI value that meets the requirements is ≥ 0.80 for new or modified instruments. The CVI value can be measured for each item (I-CVI) or the CVI value for the entire instrument (S-CVI). For the I-CVI value, the number of experts assigned a scale of 3 or 4 is divided by the total number of experts who completed the assessment. In addition, the S-CVI value is determined by dividing the sum of the I-CVI for each item by the total number of items. A summary of the formula for measuring CVI is as follows (Polit et al., 2007):

I-CVI = *Expert number giving a scale of 3 or 4*
*Total number of experts*
S-CVI = *Number of I-CVI for each item*

*Number of Items*

Based on the formula, the researcher carries out the face validity of the I-KEM instrument, and the CVI value obtained is 1.00, which indicates high face validity and is accepted by the expert panel (Davis, 1992). Table 1 below shows the face validity by measuring the CVI value obtained for the I-KEM instrument.

Table 1
*Face Validity of I-KEM Instrument*

| No. | Item | E1 | E2 | E3 | I-CVI | Total Expert Agreement | Interpretation |
|---|---|---|---|---|---|---|---|
| 1. | The size of the font used is appropriate. | √ | √ | √ | 1.00 | 3 | Accepted |
| 2. | The sentence structure used is correct. | √ | √ | √ | 1.00 | 3 | Accepted |
| 3. | The spelling used is correct. | √ | √ | √ | 1.00 | 3 | Accepted |
| 4. | The content used is appropriate. | √ | √ | √ | 1.00 | 3 | Accepted |
| 5. | The format used is appropriate. | √ | √ | √ | 1.00 | 3 | Accepted |
| | CVI Average | | | | 1.00 | | |

Table 2 shows the content validity with the CVI value obtained for the I-KEM instrument. Based on the CVI formula, the CVI value is 1.00, which indicates that the content validity of the I-KEM instrument is high and accepted by the expert panel.

Table 2
*Content Validity of I-KEM Instrument*

| No. | Item | E1 | E2 | E3 | I-CVI | Total Expert Agreement | Interpretation |
|---|---|---|---|---|---|---|---|
| 1. | The content of the instrument is appropriate for the target population. | √ | √ | √ | 1.00 | 3 | Accepted |
| 2. | The use of items is consistent with the objectives of the study. | √ | √ | √ | 1.00 | 3 | Accepted |
| 3. | The content of the instrument is appropriate for students to answer the questionnaire within the time allocated (two-week period). | √ | √ | √ | 1.00 | 3 | Accepted |
| 4. | The content of the instrument can improve the performance of students' language skills. | √ | √ | √ | 1.00 | 3 | Accepted |
| 5. | The content of the instrument can change students' attitudes toward excellence. | √ | √ | √ | 1.00 | 3 | Accepted |
| | CVI Average | | | | 1.00 | | |

The experts agreed that the I-KEM instrument can measure the aspects of content that need to be measured. In addition, some improvements need to be made, such as refining word spelling errors, refinement of verse expressions that carry the double meaning so as not to confuse the respondent, adding questionnaire items, and detailing the reasons why respondents learned Arabic to be used as a reference for the development of the kit. Therefore, the researcher has refined the I-KEM instrument that has been developed based on expert recommendations.

**Pilot Study**

This study conducted a pilot study to evaluate the findings' consistency, stability, and repeatability. This process entails administering the questionnaire to a small group of respondents before questionnaires are distributed to the respondents (Asbulah et al., 2018). Reliability means the extent to which an instrument is free from measurement errors (Fraenkel & Wallen, 2003). According to Linacre (1994), the minimum stable and adequate rate for the pilot study sample is 30 people with a confidence level (95%), so it can produce a stable and meaningful statistical analysis. The pilot study involved 50 respondents who were Bachelor of Arabic students from UiTM, UM, and UPM, similar to the target respondents for the actual questionnaire. For this purpose, the researcher has appointed several enumerators to distribute informed consent forms and questionnaires online through Google Forms to the respondents. In this study, research data collection from respondents is conducted based on the guidelines of the UPSI Research Ethics Committee (JKEPU, UPSI) (Ethical Reference Number: 2021-0249-01).

**Results and Discussion**

The results are aligned with the study's objectives-(1) identifying the reliability of the instrument, (2) identifying the polarity of the item measuring each construct, (3) examining the unidimensionality of items measuring a single construct, (4) measuring the standardized residual correlation values to avoid items overlapping or confusing each other, and (5) analyzing the suitability of the items measuring each construct. The instrument contained 21

items with 5 Likert scales, starting with 1 strongly disagree, disagree, neutral, agree, and 5 strongly agree.

1) Reliability Index

In the Rasch Model measurement, the ideal values for Cronbach's Alpha are between 0.71-0.99 (Bond & Fox, 2015). The interpretation of Cronbach's Alpha (α) values is shown in Table 3.

Table 3
*Content Validity of I-KEM Instrument*

| Cronbach's Alpha score | Interpretation |
|---|---|
| 0.9-1.0 | Very good, effective with a highly consistent |
| 0.7-0.8 | Good and acceptable |
| 0.6-0.7 | Acceptable |
| <0.6 | The item should be refined |
| <0.5 | The item should be removed |

This study determined the instrument's reliability using statistical analyses grounded on the Rasch Model, as shown in Figure 1.

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 87.4 | 21.0 | 2.80 | .45 | | | | |
| SEM | 1.6 | .0 | .26 | .04 | | | | |
| P.SD | 11.4 | .0 | 1.79 | .25 | | | | |
| S.SD | 11.5 | .0 | 1.81 | .25 | | | | |
| MAX. | 105.0 | 21.0 | 8.01 | 1.86 | | | | |
| MIN. | 65.0 | 21.0 | .06 | .31 | | | | |
| REAL RMSE | .55 | TRUE SD | 1.70 | SEPARATION | 3.07 | Person RELIABILITY | .90 | |
| MODEL RMSE | .51 | TRUE SD | 1.71 | SEPARATION | 3.36 | Person RELIABILITY | .92 | |
| S.E. OF Person MEAN = .26 | | | | | | | | |

Person RAW SCORE-TO-MEASURE CORRELATION = .95
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .94 SEM = 2.81
STANDARDIZED (50 ITEM) RELIABILITY = .96

Figure 1. Reliability and Person Separation Values towards the I-KEM Instrument

Based on Figure 1, the Cronbach's Alpha (α) value for the I-KEM instrument is 0.94. Furthermore, the study reported a person reliability index of 0.90; the person separation index was reported at 3.07. The results further indicate that if 21 items were administered to a group of respondents of different and similar abilities, there is a high expectation of repetition for the respondents' feedback in the questionnaire (Bond & Fox, 2015). In addition, the reliability and separation values for items are listed in Figure 2 below.

| | TOTAL SCORE | COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 208.2 | 50.0 | .00 | .25 | .99 | .01 | .99 | .09 |
| SEM | 5.2 | .0 | .29 | .01 | .05 | .22 | .07 | .23 |
| P.SD | 23.3 | .0 | 1.28 | .03 | .22 | 1.00 | .31 | 1.03 |
| S.SD | 23.8 | .0 | 1.31 | .03 | .22 | 1.03 | .32 | 1.06 |
| MAX. | 234.0 | 50.0 | 2.89 | .32 | 1.46 | 2.10 | 1.49 | 2.14 |
| MIN. | 149.0 | 50.0 | -1.69 | .20 | .64 | -1.64 | .57 | -1.50 |
| REAL RMSE | .26 | TRUE SD | 1.25 | SEPARATION 4.73 | Item | RELIABILITY | .96 | |
| MODEL RMSE | .26 | TRUE SD | 1.25 | SEPARATION 4.89 | Item | RELIABILITY | .96 | |
| S.E. OF Item MEAN = .29 | | | | | | | | |

Item RAW SCORE-TO-MEASURE CORRELATION = -.99
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 2. Reliability and Item Separation Values towards the I-KEM Instrument

Based on Figure 2, the study found the item reliability index of 0.96 and the item separation index of 4.73. In this light, an index of 2.0 and above indicates a good separation index (Fox and Jones, 2005). Therefore, the I-KEM instrument has a generally good consistency with a value close to 1.0 (Bond & Fox, 2015). It also shows that the instrument could be used for the actual studies.

2) Polarity Item

The term "polarity item," also known as the "Point-Measure Correlation Coefficient," refers to the point of measurement of the correlation coefficient between a person's aptitude and item difficulty (PTME Corr). This analysis seeks to evaluate whether a predetermined construct can accomplish its objectives. A positive PTME Corr. value (+) indicates that each item can measure the construct that is being assessed. On the other hand, a negative PTME Corr. value represents a negative index. Such a value shows the need for research to determine whether the items need improvement or removal (Bond & Fox, 2007).

Figure 3 shows the PTME Corr. values for the I-KEM instrument are all positive, with no negative values and no values below 0.20. This describes measuring items from a positive direction, parallel in one direction, to measure constructs (Linacre, 2002). A high PTME Corr. value indicates items can differentiate individual abilities (Bond and Fox, 2007). The PTME Corr. should exceed the value of 0 (> 0) and be positive (the item measures one direction in the same direction as the construct being measured). An item polarity of 0 and negative contradicts the measured variable or construct (Linacre, 2007).

| PTMEASUR-AL CORR. | EXP. | EXACT MATCH OBS% | EXP% | Item |
|---|---|---|---|---|
| A | .54 | .62 | 67.3 | 65.6 | B5 |
| B | .64 | .76 | 46.9 | 52.1 | D5 |
| C | .55 | .58 | 77.6 | 68.4 | C3 |
| D | .64 | .72 | 51.0 | 55.8 | D3 |
| E | .66 | .77 | 44.9 | 49.3 | D2 |
| F | .44 | .50 | 73.5 | 76.5 | C2 |
| G | .55 | .56 | 73.5 | 70.3 | B4 |
| H | .69 | .67 | 55.1 | 61.0 | E5 |
| I | .62 | .63 | 75.5 | 64.1 | E1 |
| J | .69 | .73 | 49.0 | 55.2 | D1 |
| K | .70 | .69 | 63.3 | 58.8 | D4 |
| j | .68 | .66 | 61.2 | 61.7 | C4 |
| i | .55 | .53 | 73.5 | 73.3 | C1 |
| h | .73 | .66 | 59.2 | 62.0 | E2 |
| g | .62 | .56 | 73.5 | 70.3 | B3 |
| f | .72 | .67 | 67.3 | 61.3 | C5 |
| e | .67 | .57 | 69.4 | 68.8 | B1 |
| d | .60 | .53 | 73.5 | 73.3 | B2 |
| c | .79 | .70 | 55.1 | 56.9 | E3 |
| b | .75 | .67 | 75.5 | 61.1 | E4 |
| a | .62 | .53 | 77.6 | 73.3 | B6 |
| | | | 64.9 | 63.8 | |
| | | | 10.6 | 7.4 | |

Figure 3. Polarity of the I-KEM Instrument Items

3) Unidimensionality

Unidimensionality is important to detect the second dimension inherent in the measurement of a construct. According to Aziz et al. (2015), Rasch analysis uses Residual Principal Component Analysis (PCA) with the Residual Variant Standard technique to ensure the consistency of instrument dimensions. The optimal value of variance, according to Linacre (2002), is (> 60%). Furthermore, Runnels (2012) asserted that ideal PCA values range from 20% to 40%. Aziz et al. (2015) mentioned that the maximum unexplained variance of first contrast should be 15%. Figure 4 shows the unidimensionality value of the I-KEM instrument with the PCA (raw variance explained by measures) value of 60.5%, compared to the Rasch Model expectation of 61.5%. This indicates a very good value of variance by exceeding the minimum levels of 20% and (> 60%) (Linacre, 2002). The level of interference of the measured item (noise) or unexplained variance in the first contrast is 10%, a controlled value from a maximum value of 15% (Aziz et al., 2015).

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

| | | Eigenvalue | Observed | | Expected |
|---|---|---|---|---|---|
| Total raw variance in observations | = | 53.1971 | 100.0% | | 100.0% |
| Raw variance explained by measures | = | 32.1971 | 60.5% | | 61.5% |
| Raw variance explained by persons | = | 15.5578 | 29.2% | | 29.7% |
| Raw Variance explained by items | = | 16.6393 | 31.3% | | 31.8% |
| Raw unexplained variance (total) | = | 21.0000 | 39.5% | 100.0% | 38.5% |
| Unexplned variance in 1st contrast | = | 5.3377 | 10.0% | 25.4% | |
| Unexplned variance in 2nd contrast | = | 2.7678 | 5.2% | 13.2% | |
| Unexplned variance in 3rd contrast | = | 1.8582 | 3.5% | 8.8% | |
| Unexplned variance in 4th contrast | = | 1.7474 | 3.3% | 8.3% | |
| Unexplned variance in 5th contrast | = | 1.2928 | 2.4% | 6.2% | |

Figure 4. Unidimensionality of the I-KEM Instrument

4) The Standardized Residual Correlation

Overlapping and redundant items can be identified through the standardized residual correlation value. This is intended to avoid any confusion or misplaced objectives in

instrument development. The standardized residual correlation value shall record a value not exceeding 0.70. If two items exceed that value level, only one item is used, and the other needs to be dropped or refined so there is no further correlation between the items. This elaborates that the two items have a high correlation value because they have similar characteristics to each other, and other dimensions are shared (McNamara, 1996; Asbulah et al., 2018). Figure 5 shows the standardized residual correlation values for the I-KEM instrument.

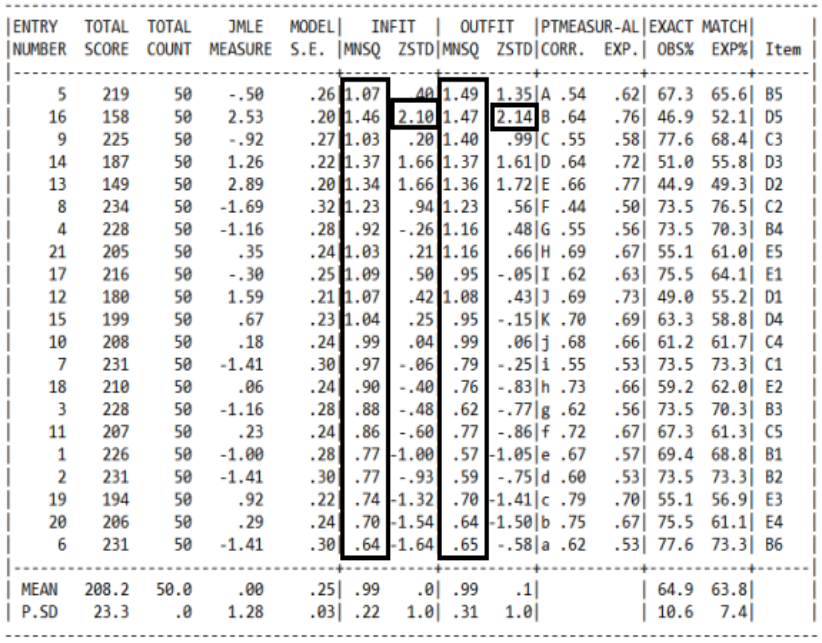
| CORRELATION | ENTRY NUMBER It | ENTRY NUMBER It |
|---|---|---|
| .72 | 2 B2 | 3 B3 |
| .68 | 7 C1 | 8 C2 |
| .64 | 1 B1 | 3 B3 |
| .61 | 1 B1 | 2 B2 |
| .59 | 2 B2 | 6 B6 |
| .56 | 1 B1 | 6 B6 |
| .53 | 13 D2 | 16 D5 |
| .52 | 3 B3 | 6 B6 |
| .52 | 3 B3 | 4 B4 |
| .49 | 4 B4 | 5 B5 |
| .47 | 20 E4 | 21 E5 |
| .46 | 2 B2 | 4 B4 |
| .44 | 12 D1 | 14 D3 |
| -.55 | 3 B3 | 12 D1 |
| -.51 | 3 B3 | 13 D2 |
| -.47 | 3 B3 | 16 D5 |
| -.47 | 2 B2 | 12 D1 |
| -.46 | 3 B3 | 14 D3 |
| -.46 | 1 B1 | 13 D2 |
| -.44 | 1 B1 | 12 D1 |

Figure 5. The Standardized Residual Correlation Values

As shown in Figure 5, the I-KEM instrument contains two items with high standardized residual correlation values (exceeding 0.70), namely items B2 and B3. Both items were filtered by looking at MNQS values approaching a value of 1.00 and ZSTD values approaching a value of 0.00 (Huei et al., 2020). After going through the filtering process, items B2 and B3 were retained according to the researcher and supervisor's discussion as well as expert recommendations.

5) Item Fit

The MNSQ infit value refers to the match corresponding to the response pattern and the measured item. Based on the Rasch Model analysis, item fit values are used to measure a latent variant, and a study by Boone et al. (2014) described the infit and Mean Square outfit (MNSQ) should range between (0.5–1.5). In this regard, ranges higher than the MNSQ value typically have high ZSTD values and fall outside the acceptable range of -2.0 ZSTD +2.0. Meanwhile, the MNSQ outfit values exceeding 1.5 indicates that the responses given by the sample are too random for low-ability samples, simple questions not answered by the sample or negligence of answers for the high-ability sample. The MNSQ infit and outfit values should be between 0.70–1.33 to indicate the suitability of items measuring latent variables or constructs (Bond & Fox, 2015). In this light, the respondents' responses are highly predictable as the MNSQ outfit value is less than 0.5. However, the MNSQ outfit value must be considered before the MNSQ infit value when measuring a construct (Sumintono, 2017). Figure 6 shows the item fit for the I-KEM instrument.



| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | JMLE MEASURE | MODEL S.E. | INFIT MNSQ | INFIT ZSTD | OUTFIT MNSQ | OUTFIT ZSTD | PTMEASUR-AL CORR. | PTMEASUR-AL EXP. | EXACT MATCH OBS% | EXACT MATCH EXP% | Item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 219 | 50 | -.50 | .26 | 1.07 | .40 | 1.49 | 1.35 | A .54 | .62 | 67.3 | 65.6 | B5 |
| 16 | 158 | 50 | 2.53 | .20 | 1.46 | 2.10 | 1.47 | 2.14 | B .64 | .76 | 46.9 | 52.1 | D5 |
| 9 | 225 | 50 | -.92 | .27 | 1.03 | .20 | 1.40 | .99 | C .55 | .58 | 77.6 | 68.4 | C3 |
| 14 | 187 | 50 | 1.26 | .22 | 1.37 | 1.66 | 1.37 | 1.61 | D .64 | .72 | 51.0 | 55.8 | D3 |
| 13 | 149 | 50 | 2.89 | .20 | 1.34 | 1.66 | 1.36 | 1.72 | E .66 | .77 | 44.9 | 49.3 | D2 |
| 8 | 234 | 50 | -1.69 | .32 | 1.23 | .94 | 1.23 | .56 | F .44 | .50 | 73.5 | 76.5 | C2 |
| 4 | 228 | 50 | -1.16 | .28 | .92 | -.26 | 1.16 | .48 | G .55 | .56 | 73.5 | 70.3 | B4 |
| 21 | 205 | 50 | .35 | .24 | 1.03 | .21 | 1.16 | .66 | H .69 | .67 | 55.1 | 61.0 | E5 |
| 17 | 216 | 50 | -.30 | .25 | 1.09 | .50 | .95 | -.05 | I .62 | .63 | 75.5 | 64.1 | E1 |
| 12 | 180 | 50 | 1.59 | .21 | 1.07 | .42 | 1.08 | .43 | J .69 | .73 | 49.0 | 55.2 | D1 |
| 15 | 199 | 50 | .67 | .23 | 1.04 | .25 | .95 | -.15 | K .70 | .69 | 63.3 | 58.8 | D4 |
| 10 | 208 | 50 | .18 | .24 | .99 | .04 | .99 | .06 | j .68 | .66 | 61.2 | 61.7 | C4 |
| 7 | 231 | 50 | -1.41 | .30 | .97 | -.06 | .79 | -.25 | i .55 | .53 | 73.5 | 73.3 | C1 |
| 18 | 210 | 50 | .06 | .24 | .90 | -.40 | .76 | -.83 | h .73 | .66 | 59.2 | 62.0 | E2 |
| 3 | 228 | 50 | -1.16 | .28 | .88 | -.48 | .62 | -.77 | g .62 | .56 | 73.5 | 70.3 | B3 |
| 11 | 207 | 50 | .23 | .24 | .86 | -.60 | .77 | -.86 | f .72 | .67 | 67.3 | 61.3 | C5 |
| 1 | 226 | 50 | -1.00 | .28 | .77 | -1.00 | .57 | -1.05 | e .67 | .57 | 69.4 | 68.8 | B1 |
| 2 | 231 | 50 | -1.41 | .30 | .77 | -.93 | .59 | -.75 | d .60 | .53 | 73.5 | 73.3 | B2 |
| 19 | 194 | 50 | .92 | .22 | .74 | -1.32 | .70 | -1.41 | c .79 | .70 | 55.1 | 56.9 | E3 |
| 20 | 206 | 50 | .29 | .24 | .70 | -1.54 | .64 | -1.50 | b .75 | .67 | 75.5 | 61.1 | E4 |
| 6 | 231 | 50 | -1.41 | .30 | .64 | -1.64 | .65 | -.58 | a .62 | .53 | 77.6 | 73.3 | B6 |
| MEAN | 208.2 | 50.0 | .00 | .25 | .99 | .0 | .99 | .1 | | | 64.9 | 63.8 | |
| P.SD | 23.3 | .0 | 1.28 | .03 | .22 | 1.0 | .31 | 1.0 | | | 10.6 | 7.4 | |

Figure 6. Item Fit for the I-KEM Instrument

Figure 6 illustrates that all items are within the acceptable range of infit and outfit MNSQ values of 0.5-1.5. This explains that these items are in a sufficient range for measurement (Linacre, 2002; 2006). The range of MNSQ is explained in Table 4 below

Table 4
*Description of Infit and Outfit MNSQ*

| MNSQ Values | Measurement Implications |
|---|---|
| >2.0 | Distorting or weakening the measurement system. Probably due to only one or two observations. |
| 1.5 – 2.0 | Less effective for measurement construction but not weakening. |
| 0.5 – 1.5 | Effective enough for measurement. |
| < 0.5 | Less effective for measurement but not debilitating. Likely to result in confusing reliability and separation coefficients. |

**Conclusion**

The pilot study results indicate that validity and reliability tests should be performed as the initial step of instrument development to ensure that the instrument is reliable and will yield precise data. Each item in the instrument should be tested according to standard indices and conditions of the Rasch Measurement Model. Items exceeding the range of fit items should either be refined or removed according to experts' views and consensus. In this regard, the refined instruments showed better reliability. Overall, the validity and reliability tests based on the Rasch Model showed that the I-KEM instrument has good validity and high reliability. This indicates that all 21 items can measure the constructs. Therefore, these findings explain that the I-KEM instrument is suitable for use by university-level students. The findings from the analysis can guide researchers to create highly valid and reliable instruments to ensure that the measurements can meet the study's goals. Accordingly, this study proposes the development of pedagogical applications such as media kits related to Arabic communication

skills that are relevant and appropriate to help students master the Arabic language. In addition, this study also aims to guide Arabic language teachers to develop innovative and student-centered teaching tools to further enhance the communication competence of non-Arabic speakers in Malaysian tertiary institutions.

**Acknowledgment**
Special thanks to the Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia, for permission and support in conducting this research. The highest appreciation is also due to Zakat Dana As-Syakirin PETRONAS and Research Management and Innovation Centre (RMIC), UPSI sponsoring this research. We would also like to thank the respondents at UiTM, UM, and UPM and our enumerators for this study.

**Corresponding Author**
Siti Rahmah Borham
Department of Modern Languages, Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris (UPSI), 35900 Tanjong Malim, Perak, Malaysia.
Email: rahmahborham@gmail.com

**References**

Arshad, M., & Bakar, K. A. (2012). Penggunaan Strategi Pembelajaran Kemahiran Bertutur Bahasa Arab: Kajian di Pusat Asasi UIAM. Paper presented at Persidangan Kebangsaan Pengajaran dan Pembelajaran Bahasa Arab 2012 (PKEBAR'12), UKM, Bangi, Malaysia.

Asbulah, L. H., Lubis, M. A., Aladdin, A., & Sahrim, M. (2018). Kesahan dan Kebolehpercayaan Instrumen Pengetahuan Kolokasi Bahasa Arab IPT (i-KAC IPT) Menggunakan Model Pengukuran Rasch. *ASEAN Comparative Education Reasearch Journal on Islam and Civilization (ACER-J), 2*(1), 97-106.

Ashour, R. Q., & al-Hawamidat, A. F. (2014). *Asalib Tadris al-Lughat al-'Arabiyyat baina al-Nazoriyyat wa al-Tatbiq*, 4th Eds. Amman: Dar al-Masirat li al-Nasyari wa at-Tauzi'i.

Aziz, A. A., Masodi, M. S., & Zaharim, A. (2015). *Asas Model Pengukuran Rasch: Pembentukan skala & struktur pengukuran*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Blaschke, L. M. (2012). Heutagogy and Lifelong Learning: A Review of Heutagogical Practice and Self-Determined Learning. *The International Review of Research in Open and Distance Learning, 1,* 56-71.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences 3rd Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. USA: Springer Dordrecht.

Borham, S. R. (2018). *Kesediaan untuk berkomunikasi bahasa Arab dan variabel psikologi yang mempengaruhinya: pendekatan Model Rasch dan Model Jalur*. Unpublished master dissertation. Universiti Kebangsaan Malaysia.

Che Hat, N., Sha'ari, S. H., & Abdul Hamid, M. F. (2013). Persepsi Pelajar Terhadap Penggunaan Animasi dalam Pembelajaran Bahasa Arab. *Jurnal Teknologi, 63*(1), 25-29.

Darusalam, G., & Hussin, S. (2016). *Metodologi penyelidikan dalam pendidikan*. Kuala Lumpur: Penerbit Universiti Malaya.

Daud, N., & Pisal, N. A. (2014). Permasalahan Pertuturan dalam Bahasa Arab sebagai Bahasa Kedua. *GEMA Online Journal of Language Studies, 14*(1), 117-133.

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194–197.

Fox, C. M., & Jones, J. A. (2005). Uses of Rasch Modeling in Counseling Psychology Research. *Journal of Counselling Psychology, 45*(1), 30-45.

Fraenkel, J. R., & Wallen, N. E. (1996). *How to design and evaluate research in education*. 3rd Eds. New York: Mc-Graw Hill, INC.

Fraenkel, J. R., & Wallen, N. E. (2003). *How to design and evaluate research in education*. 5th Eds. New York: Mc-Graw Hill, INC.

Fuad, A., & Al-Yahya, M. (2022). AraConv: Developing an Arabic Task-Oriented Dialogue System Using Multi-Lingual Transformer Model mt5. *Appl. Sci. 14,* 1-16.

Ghani, M. T. A., Hamzah, M., Daud, W. A. A. W., & Romli, T. R. (2022). The Impact Of Mobile Digital Game in Learning Arabic Language at Tertiary Level. *Contemporary Educational Technology. 14*(1), 1-18.

Hai, A. A., Amiruddin, A. Z., Rahman, A. A., & Daud, W. A. A. W. (2020). Penggunaan Aplikasi Web 2.0: GoAnimate.com dalam Aktiviti Pembelajaran Bahasa Arab di Universiti Malaysia Kelantan*. Sains Humanika, 12*(1), 11-17.

Hamzah, M., Ghani, M. T. A., Daud, W. A. A. W., & Ramli, S. (2019). Digital Game-Based Learning as an Innovation to Enhance Student's Achievement for Arabic Language Classroom. *International Journal of Recent Technology and Engineering (IJRTE), 8*(3), 2108-2112.

Haron, S. C. (2011). The role of learning strategies Arabic speaking skills. In S. Hussien. (Eds.), *Current issues and themes in education: A handbook for practitioners*. Selangor: IIUM Press.

Hashimoto, Y. (2002). Motivation and Willingness To Communicate as Predictors of Reported L2 Use: The Japanese ESL Context. *Second Language Studies, 20*(2), 29-70.

Hassan, N. W. M., Tahar, M. M., & Yasin, M. H. M. (2017). Teaching and learning Using Software "Let's Reading" for Malay Language Subjects for Students with Learning Disabilities. *Jurnal Penelitian dan Pengembangan Pendidikan Luar Biasa, 4*(1), 63-63.

Huei, O. K., Rus, R. C., & Kamis, A. (2020). Construct Validity and Reliability in Content Knowledge of Design and Technology Subject: A Rasch Measurement Model Approaches for Pilot Study. *International Journal of Academic Research in Business and Social Science, 10*(3), 497-511.

Ibrahim, M. (2013). Ta'thīr al-Ittijahat wa al-Infiʿalat fī Istikhdam al-Lughat al-ʿArabiyyat fī al-ʿAmaliyyat al-Ittiṣaliyyat: Dirasat Ḥalat Mutaʿallimī al-Lughat al-ʿArabiyyat Biwaṣfiha Lughat Thaniyat fī Jamiʿat al-ʿUlūm al-Islamiyyat al-Malīziyyat. *Majallat al-Dirasat al-Tarbawiyyat wa al-Nafsiyyat*, *7*(3), 330-343.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, *3*(1),85-106.

Linacre, J. M. (*2006*). Rasch Analysis of Rank-Ordered Data. *Journal of Applied Measurement, 7*(1), 129–139.

Linacre, J. M. (2007). *A User's Guide to WINSTEPS Rasch-Model Computer Programs*. Chicago, Illinois: MESA Press.

Mahmuda, S. (2018). Media Pembelajaran Bahasa Arab. *An-Nabighoh, 20*(1), 129-138.

Marpuah, S. (2015). *Penguasaan Komunikasi Bahasa Arab Melalui Kaedah Aktif Komunikatif di Sekolah Menengah Agama*. Unpublished doctoral thesis. Universiti Tun Hussien Onn, Malaysia.

Mazlan, A. (2018). *Pembangunan dan Penilaian Keberkesanan Modul Pendekatan Pengajaran Berasaskan Otak dengan Integrasi I-Think dan Brain Gym untuk Meningkatkan Kefahaman Konseptual dan Motivasi Belajar Fizik Pelajar Matrikulasi*. Unpublished doctoral thesis. Universiti Sains Malaysia.

McCroskey, J. C., & Baer, J. E. (1985, November). *Willingness to Communicate: The Construct and Its Measurement*. Paper presented at the Annual Meeting of the Speech Communication Association, California.

McNamara, T. F. (1996). *Measuring Second Language Performance: Applied Linguistic and Language Study*. Harlow, Essex, UK: Addison Wesley Longman Ltd.

Mohamad, A. H. (2009). Tahap Komunikasi Bahasa Arab dalam Kalangan Pelajar Sarjana Muda Bahasa Arab di IPTA Malaysia. *Journal of Islamic and Arabic Education, 1*(1), 1-14.

Noor, S. S. M., Osman, N., Rouyan, N. M., Hat, N. C., & Saad, K. N. M. (2014). Kemahiran Bertutur Bahasa Arab Luar Kelas dalam Kalangan Penutur Bukan Asli Bahasa Arab. *BITARA International Journal of Civilization Studies and Human Sciences, 4*(2), 60-69.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing & Health, 30*(4), 459–467.

Punch, K. F. (1998). *Introduction to Social Research*. London: SAGE.

Rahman, A. A., & Ahmad. N. (2020). Pembinaan Model Kemahiran Bertutur Bahasa Arab "MGG@TFMODKTBA@INNOV" di Institut Pengajian Tinggi Awam (IPTA) Malaysia. *Jurnal Kesidang, 5,* 77-87.

Ramli, S., Atoh, N., Zakaria, Z., & Ariffin, M. (2017). I-Kit Bahasa Arab Dalam Kalangan Pelajar ISM Bahasa Arab dengan Pendidikan UPSI: Satu Analisis. *Journal of Global Business and Social Entreprenuership (GBSE), 3*(7), 79-89.

Rahman, A. A., Azizan, K. F. K., & Jamali, H. N. (2015). Persepsi Pelajar di IPT Malaysia Terhadap Pengajaran dan Pembelajaran Bahasa Arab. *E-Jurnal Pendidikan, 1*(2), 1-14.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.

Runnels, J. (2012). Using the Rasch Model to Validate a Multiple-Choice English Achievement Test. *International Journal of Language Studies, 6*(4), 141-153

Samin, S. M., Pebrian, J., & Zulkifli, A. (2020, September). *Heutagogy Approach for Arabic Learning in Higher Education in Industrial Revolution 4.0*. Paper presented at the Second International Conference on Social, Economy, Education and Humanity (ICoSEEH 2019), Riau, Indonesia.

Sumintono, B. (2017). Rasch Model Measurement as Tools in Assessment for Learning. *Advances in Social Science, Education and Humanities Research, 173*, 38-42.

Tohlong, F. (2015). *Tahap Kesediaan Berkomunikasi Bahasa Arab Pelajar ILA Universiti Fatoni, Thailand*. Unpublished master dissertation, Universiti Kebangsaan Malaysia.

Tu'aimat, R. A. (2006). *Al-Maharah al-Lughawiyyah: Mustawiyatuha Tadrisuha Su'ubatuha*. Kaherah: Dar al-Fikr al- 'Arabi.

Wright, B. D., & Linacre, J. M. 1992. Combining and Splitting Categories. *Rasch Measurement Transactions*, 6(3), 233-235.

Yahaya, M. F., Halim, A. Z., Sahrir, M. S., & Hamid, M. F. A. (2021). Need Analysis on Developing Arabic Language M-Learning Basic Level During Covid-19. *Journal of Contemporary Issues in Business and Government, 27*(2), 5452-5461.

Yahaya, M. F., Sahrir, M. S., & Nasir, M. S. (2013). Pembangunan Laman Web EZ-Arabic Sebagai Alternatif Pembelajaran Maya Bahasa Arab Bagi Pelajar Sekolah Rendah Malaysia. *Jurnal Teknologi, 6*(1). 11-18.

Zaini, A. R., Ghazali, A. R., Ismail, A. R., Zakaria, N., Hamdan, H., & Azizan, M. R. (2019). Pengajaran Bahasa Arab di Malaysia: Permasalahan dan Cabaran. *Jurnal Pengajian Islam (JPI), 1*(12), 47-57.