Assessing Arabic Language Exposure Instrument (I-ALEx) in Higher Education Using the Rasch Measurement Model

Nur Afiqah Athirah Mohd Rushdi[1], Lily Hanefarezan Asbulah[*2], Firuz-Akhtar Lubis[3], Ashinida Aladdin[4], Mus'ab Sahrim[5]
[1,2,3]Research Centre for Arabic Language and Islamic Civilization, Faculty of Islamic Studies, Universiti Kebangsaan Malaysia, Malaysia
[4]The Centre for Research in Language and Linguistics, Faculty of Science Social and Humanities, Universiti Kebangsaan Malaysia, Malaysia
[5]Intelligent Cyber-Physical Systems (ICPS), Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, Malaysia
p116037@siswa.ukm.edu.my, lilyhane@ukm.edu.my, firuz@ukm.edu.my, ashi@ukm.edu.my, musab@usim.edu.my

Abstract
Exposure is crucial for language acquisition, yet Arabic language learners in Malaysia face limited exposure. To accurately measure the extent of this exposure among higher education students, the researcher has developed the Arabic Language Exposure instrument (I-ALEx). This study applied the Rasch Model to validate the I-ALEx, which was designed to assess students' exposure to the Arabic language. Data were collected from 374 final-year Arabic language students from eight universities in Malaysia using a cross-sectional survey design. The I-ALEx includes 20 items divided into two sub-constructs: target language and culture, on a four-point Likert scale (strongly disagree to agree strongly). The quality of the items was evaluated using Winsteps version 5.2. The analysis indicated that all items on the I-ALEx satisfied the Rasch model requirements, thereby supporting the reliability and validity of the instrument. This validation supports the I-ALEx as an effective tool for evaluating Arabic language exposure among higher education students, thus enhancing the assessment practices in language education.
Keywords: Exposure; Arabic Language; Instrument Assessment; Higher Education; Rasch Measurement Model

INTRODUCTION

Exposure to a language pertains to the interaction learners have with the target language they seek to master (Al Zoubi, 2018). This exposure plays a vital role in language acquisition, significantly influencing learners' ability to comprehend, speak, and use the language proficiently. One of the key determinants of successful language teaching and learning is the amount and quality of students' exposure to both the target language and its associated culture (Kozhevnikova, 2014). Increased exposure not only enhances linguistic skills but also facilitates an understanding of the cultural contexts necessary for effective communication. However, for Arabic language learners in Malaysia, a significant challenge is the limited exposure to the Arabic language. Baharudin (2017) states that the majority of undergraduate students studying Arabic as a foreign language in Malaysia face limited exposure to the Arabic language.

Exposure to a language encompasses the duration and quality of contact an individual has with the language, whether through verbal or written means, and can occur

via formal or informal communication channels (DeKeyser, 2007). This contact may be active, involving speaking and writing, or passive, involving listening and reading. Exposure takes place whenever individuals engage with the target language, such as reading magazines, books and newspapers; interacting with colleagues, native speakers, and educators; browsing the internet; and accessing information from various media sources. Yet, in numerous regions, language and cultural exposure levels are restricted. Learners often have minimal interaction with speakers from other countries, making their language teachers and available authentic materials, such as newspapers, magazines, literature, films, and internet resources, the primary sources of exposure to the culture and target language (Kozhevnikova, 2014).

The cultural dimensions of a language include the knowledge, clothing, and traditions of individuals within the community where the language is utilized. Cultural factors significantly impact both language usage and comprehension, affecting how individuals express themselves and interpret others. It is essential for students learning a foreign language to understand the target culture, as it adds context and depth to their linguistic knowledge (Zhan, 2016). For Arabic learners, cultural immersion helps them understand idiomatic expressions, social norms, and context-specific meanings, thereby enhancing their overall learning experience.

Given the critical role of exposure in language acquisition, it is essential to evaluate students' interaction with the target language using reliable and valid measurement tools. Accurate assessments ensure that the multifaceted aspects of exposure, both linguistic and cultural are systematically captured and analyzed. While existing studies have underscored the importance of exposure (Kozhevnikova, 2014; Zhan, 2016), there is a notable gap in the availability of instruments specifically tailored to the Arabic language context. Furthermore, psychometric methods, such as the Rasch measurement model, provide a rigorous framework for evaluating the validity and reliability of such instruments., making them essential for advancing research in this area.

In Arabic language research, the Rasch Measurement Model is frequently utilized to evaluate and enhance the validity and reliability of measurement instruments. A notable example is the study by Alsoudi and Abu Shindi (2023), which examined the Critical Thinking Disposition Scale in its Arabic version using the Rasch Measurement Model. The findings revealed robust psychometric properties, including reliability, unidimensionality, and appropriate category functioning, resulting in a refined scale with 25 items. While the scale demonstrated effectiveness in the Omani university context, the study emphasized the need for further validation through convergent and divergent validity measures. Additionally, the researchers recommended extending the scale's application to younger populations, such as school students, to broaden its scope and utility.

The study of Rushdi et al. (2024) applied the Rasch Measurement Model to appraise the psychometric properties of the I-SMA, an instrument devised to gauge students' motivation for learning Arabic within Malaysian universities. The instrument focused on three sub-constructs: intrinsic, extrinsic, and holistic motivation. Analysis using the Winsteps software confirmed that the I-SMA instrument met the Rasch model's criteria, establishing its reliability and validity. These results underscore the instrument's effectiveness as a tool for measuring motivation in Arabic language learning.

Another study of Rushdi and Asbulah (2022) conducted an evaluation of the I-SPPBA instrument, designed to assess students' attitudes toward Arabic language

learning. The instrument encompasses three sub-constructs: affective, cognitive, and conative attitudes. Their analysis demonstrated that the I-SPPBA instrument possesses a high level of reliability, affirming its suitability for measuring students' attitudes in higher education contexts.

The reviewed studies highlight the significant role of the Rasch Measurement Model in validating instruments that measure various constructs related to Arabic language learning, such as motivation and attitudes. Tools like the I-SMA and I-SPPBA have demonstrated reliability and validity, advancing the assessment of these psychological dimensions. However, a critical gap persists in the development of instruments specifically designed to measure language exposure, a fundamental determinant of successful language acquisition. Despite its importance, language exposure remains underexplored in the context of Arabic learning, particularly within higher education settings. To address this limitation, the Arabic Language Exposure instrument (I-ALEx) was conceptualized to provide a comprehensive, reliable, and valid framework for assessing students' exposure to Arabic. This research utilizes the Rasch Measurement Model to evaluate the I-ALEx instrument, aiming to bridge this gap and contribute to the enhancement of assessment practices in Arabic language education.

METHOD

The I-ALEx instrument was developed to evaluate the exposure of higher education students to the Arabic language. It comprises 20 items that reflect two constructs of exposure: (1) Target Language, and (2) Culture. The target language facet assesses the extent of an individual's interaction with the target language, including conversations with colleagues, native speakers, and educators; reading magazines, books and newspapers; browsing the internet; and obtaining information from various media sources (Kozhevnikova, 2014). In contrast, the culture facet examines the amount of time an individual engages with various cultural elements such as language, religion, cuisine, social customs, music, and the arts (Mohammed, 2020).

This instrument utilizes a 4-point scale with the options "strongly disagree," "disagree," "agree," and "strongly agree" to measure levels of agreement, intentionally omitting a neutral choice to avoid respondents opting for an indifferent stance. This approach is informed by Retief et al. (2013), who contends that including a neutral option often introduces issues, recommending a 4-point scale to more clearly define degrees of agreement.

The instrument was subjected to a content validity evaluation by a panel of seven experts, which included senior lecturers, a research fellow, and an excellent teacher from diverse academic institutions specializing in social modeling and measurement, measurement and evaluation, and Arabic linguistics. However, feedback was received from only six panel members, chosen for their expertise in questionnaire construction and Arabic language education. These experts reviewed and provided feedback on the linguistics structure, content and questionnaire's format.

To assess the reliability among experts, the Content Validity Index (CVI) was employed, with a threshold of ≥ 0.83 deemed acceptable for new instruments with six experts, as per Lynn's (1986) recommendations. All items in I-ALEx surpassed this threshold, indicating a satisfactory level of expert consensus. Nonetheless, expert feedback was incorporated, leading to the following item modifications:

1. Refinement of sentence structure in certain items.

2. Clarification of overly general items to provide more detailed information.
3. Review of several items to ensure they align with the instrument's sub-constructs.

Respondents and Procedure

This quantitative study included 374 final-year students pursuing Bachelor's Degrees majoring in Arabic language from eight Malaysian universities, including the International Islamic University of Malaysia (UIAM), Universiti Kebangsaan Malaysia (UKM), Universiti Teknologi Mara (UiTM), Universiti Sultan Zainal Abidin (UniSZA), Universiti Putra Malaysia (UPM), Universiti Malaya (UM), Universiti Pendidikan Sultan Idris (UPSI) and Universiti Sains Islam Malaysia (USIM). Prior approval was obtained from the Dean and the Head of the Arabic Language Program. The study instrument was administered during designated class periods. Respondents were instructed to read instructions before privately completing the questionnaire within an hour. Completed questionnaires underwent scrutiny to ensure compliance with instructions and completeness before analysis using the Rasch Measurement Model via Winsteps version 5.2 software.

Data Analysis

In this research, the Rasch Measurement Model was applied using Winsteps version 5.2 to enhance the evaluation of the instrument's validity, going beyond the traditional dependence on Cronbach's alpha. This analytical approach enables a comprehensive evaluation of item quality and performance, with the goal of strengthening the instrument's validity with accuracy and thoroughness (Azrilah Abdul Aziz et al., 2013). Item analysis comprised several aspects, including: (1) Item polarity detection, (2) Reliability index and item-respondent separation, (3) Item fit detection, (4) Dimensionality uniformity, (5) Local independence, (6) Rating scale diagnostics, and (7) The Wright Map (Boone et al., 2014).

RESULTS AND DISCUSSION

Item Polarity Detection

Item polarity analysis, also referred to as Point Measure Correlation (PTMEA CORR) or item parallelism, assesses whether the direction of each item aligns consistently with the overall construct being measured. When all items have a positive index, it suggests they are coherently aligned with the construct, confirming that the items are functioning as intended. Conversely, negative values suggest that an item may be misfitting or poorly aligned with the construct, prompting a closer examination to determine whether the item should be revised or removed.

The Point Measure Correlation (PTMEA CORR) values for the items in I-ALEx range from 0.54 to 0.73, demonstrating positive alignment with the measured construct and the absence of negative values. This uniformity emphasizes the alignment and consistent direction of the items, demonstrating the importance of PTMEA CORR analysis in confirming item similarity and supporting construct validity.

Reliability Index And Item-Respondent Separation

Rasch analysis uses item and respondent reliability to determine how well the data fits the model. The separation index then reveals the number of distinct difficulty levels the items represent and the range of abilities within the group of respondents. For I-ALEx,

DOI: 10.18860 /ijazarabi.V8i3.30108
ISSN(print): 2620-5912 | ISSN(online):2620-5947

the person separation index is 3.12 with a reliability coefficient of 0.91, while the item separation index is 8.87 with a reliability coefficient of 0.99. These findings suggest that the I-ALEx instrument demonstrates high reliability and effectiveness, with strong consistency in measuring higher education students' exposure to the Arabic language. This is supported by robust reliability indices approaching 1.00 and separation indices exceeding the threshold of 2.00, as recommended by Linacre (2002). Table 1 provides a summary of person metrics, while Table 2 presents a summary of item metrics. These results indicate a high level of trustworthiness and precision in the instrument's assessment capabilities (Bond & Fox, 2015).

Table 1. Summary of Measured Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	47.8	20.0	-.23	.37				
S.D.	11.2	.0	1.50	.18				
MAX.	80.0	20.0	6.39	1.84				
MIN.	20.0	20.0	-6.49	.32	.23	-3.9	.22	-3.9

REAL RMSE	.45	TRUE SD	1.43	SEPARATION	3.18	PERSON RELIABILITY	.91
MODEL RMSE	.41	TRUE SD	1.44	SEPARATION	3.48	PERSON RELIABILITY	.92

S.E. OF PERSON MEAN = .08

Table 2. Summary of Measured Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	893.3	374.0	.00	.08	1.00	-.2	1.01	.0
S.D.	116.6	.0	.74	.00	.17	2.4	.19	2.5
MAX.	1126.0	374.0	1.52	.09	1.34	4.4	1.43	5.2
MIN.	663.0	374.0	-1.45	.08	.72	-4.4	.74	-3.7

REAL RMSE	.08	TRUE SD	.73	SEPARATION	8.87	ITEM RELIABILITY	.99
MODEL RMSE	.08	TRUE SD	.73	SEPARATION	9.16	ITEM RELIABILITY	.99

S.E. OF ITEM MEAN = .17

Item Fit Detection

Assessing the mean square outfit index value (MNSQ) is essential for detecting potential outliers or inconsistencies in construct measurement. MNSQ infit indicates the degree to which response patterns conform to the expected measurement, while MNSQ outfit identifies significant measurement issues, which are generally easier to address. Initially, the MNSQ outfit index should be evaluated to ascertain the suitability of items (Item Fit) for measuring the construct. The optimal range for productive items is typically between 0.5 and 1.5. According to Boone et al. (2014), deviations from this range are frequently signaled by elevated z-Std values, which should ideally stay within the acceptable range of -2.0 to +2.0.

For I-ALEx, eleven items, specifically D2.4, D2.5, D2.6, D2.7, D2.8, D2.9, D2.10, D2.11, D2.12, D2.13, and D2.18, exhibit ZSTD values beyond the accepted range, as indicated in Table 3. However, their MNSQ outfit and PTMEA CORR values remain within acceptable limits, leading to their consideration for refinement and retention. Boone et al. (2014) recommend evaluating item consistency using three primary criteria: MNSQ, ZSTD, and PTMEA CORR. Items that meet at least one of these criteria should be considered for retention (Sumintono & Widhiarso, 2015). Abdul Aziz et al. (2014) further assert that an item is considered unsuitable only if it fails to meet the acceptable standards across all three criteria.

DOI: 10.18860 /ijazarabi.V8i3.30108
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

Table 3. Item Misfit

Item	Measure	Outfit		PTMEA CORR
		MNSQ	ZSTD	
D2.8	.47	1.43	**5.2**	.59
D2.7	-.77	1.30	**3.8**	.57
D2.10	-.63	1.21	**2.8**	.54
D2.13	.26	1.19	**2.5**	.65
D2.18	.03	1.16	**2.1**	.61
D2.11	1.52	.81	**-2.4**	.68
D2.5	-.76	.84	**-2.3**	.64
D2.4	.78	.78	**-3.2**	.69
D2.12	1.07	.74	**-3.6**	.73
D2.6	.39	.75	**-3.6**	.72
D2.9	.42	.75	**-3.7**	.69

Note: MNSQ = mean square, ZSTD = standardized z-scores, PTMEA CORR = Point Measure Correlation

Dimensionality Uniformity

Dimensional uniformity is crucial because it allows an instrument to measure a single construct clearly and unambiguously (Azrilah Abdul Aziz et al., 2013). Instruments that exhibit ambiguity or cause confusion among respondents necessitate thorough review and refinement to achieve precise measurement objectives. The Rasch model, utilizing Residual Principal Component Analysis (PCA), is effective in evaluating the unidimensionality of an instrument, ensuring that it maintains a tolerable degree of item noise.

Linacre (2002) recommends that the highest proportion of variance explained by the Rasch model should exceed 60%. However, a minimum of 20% raw variance is generally acceptable to demonstrate construct uniformity. As shown in Table 4, the I-ALEx construct explains 47.6% of the variance, surpassing Linacre's (2012) minimum threshold of 40%. Additionally, the first contrast of the residual PCA shows very little unexplained variance, only 6.7%, demonstrating strong control and staying significantly under the maximum acceptable level of 15%.

Table 4. Dimensionality Map

	-- Empirical --		Modeled
Total raw variance in observations =	38.2 100.0%		100.0%
Raw variance explained by measures =	18.2 47.6%		47.6%
Raw variance explained by persons =	8.7 22.8%		22.8%
Raw Variance explained by items =	9.5 24.8%		24.8%
Raw unexplained variance (total) =	20.0 52.4%	100.0%	52.4%
Unexplned variance in 1st contrast =	2.6 6.7%	12.9%	
Unexplned variance in 2nd contrast =	2.1 5.5%	10.5%	
Unexplned variance in 3rd contrast =	1.6 4.3%	8.1%	
Unexplned variance in 4th contrast =	1.5 3.8%	7.3%	
Unexplned variance in 5th contrast =	1.4 3.6%	6.9%	

Local Independence

Local independence refers to the assumption that the responses to each item are independent of one another, given the respondent's ability level. Violations of local independence can lead to inaccurate measurement outcomes, hence the importance of assessing and ensuring this aspect in instrument development and evaluation. The Standardized Residual Correlation test is utilized to detect possible redundancy or overlap between item pairs in the instrument, thereby assuring its clarity and relevance to the study's aims. A correlation value exceeding 0.7 between two items signifies substantial similarity in characteristics, prompting consideration for retaining only one item for measurement purposes. In the case of I-ALEx, all ten item pairs demonstrate standardized residual correlations that fall within the acceptable thresholds, indicating that the

assumption of local independence is upheld across the instrument, ranging from -0.23 to 0.41. These results indicate the absence of inter-item relationships and confirm that no item pairs confuse respondents.

Rating Scale Diagnostic

Calibrating the scale stands as a crucial aspect in any validity and measurement system. Consistent and even increases in the observed average indicate uniformity in response patterns. Evaluating the Andrich Threshold value becomes essential to ascertain the accuracy of the polytomy in this study. The transition from negative to positive indicates the validity of the respondent's choice, yet discrepancies less than 1.4 between restrictions call for summarizing scale categories. Conversely, if the disparity surpasses 5, separating scale categories is necessary due to a substantial gap.

As shown in Table 5, the results reveal an absence of statistically significant differences in the Andrich Threshold segment that exceed the threshold of 5 or fall below 1.4. This confirms the suitability of the scales for each construct, eliminating the need for separation or combination. Furthermore, the Observed Average section in Table 6 shows that the I-ALEx construct exhibits a typical response pattern, characterized by a consistent transition from negative to positive values. This validation ensures the suitability of the chosen scales, aiming for a balanced distribution of responses on the scales. Figure 1 illustrates a visual depiction of the response option categories, depicting a pattern where the categories display interconnectedness and overlap, reinforcing the appropriateness of the scales.

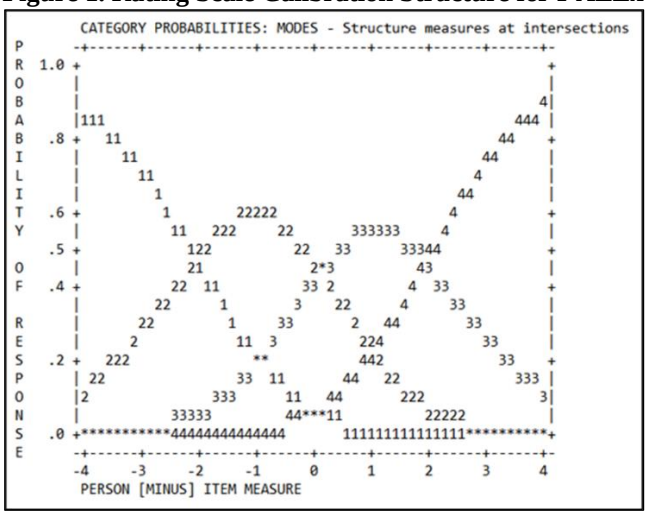
Table 5. Thresholds Width Between Categories

Thresholds between category	Threshold's width (logits)
Category 1 and 2	2.05
Category 2 and 3	2.16
Category 3 and 4	1.83

Table 6. Diagnostic Rating Scale

Category	Threshold	Observed		Observed average	Infit MNSQ	Outfit MNSQ
		Count	%			
Strongly disagree	None	1298	17	-1.97	0.93	0.94
Disagree	-2.05	2910	39	-0.59	0.94	0.93
Agree	0.11	2341	31	0.46	0.96	1.00
Strongly agree	1.94	931	12	1.46	1.17	1.18

Figure 1. Rating Scale Calibration Structure for I-ALEx

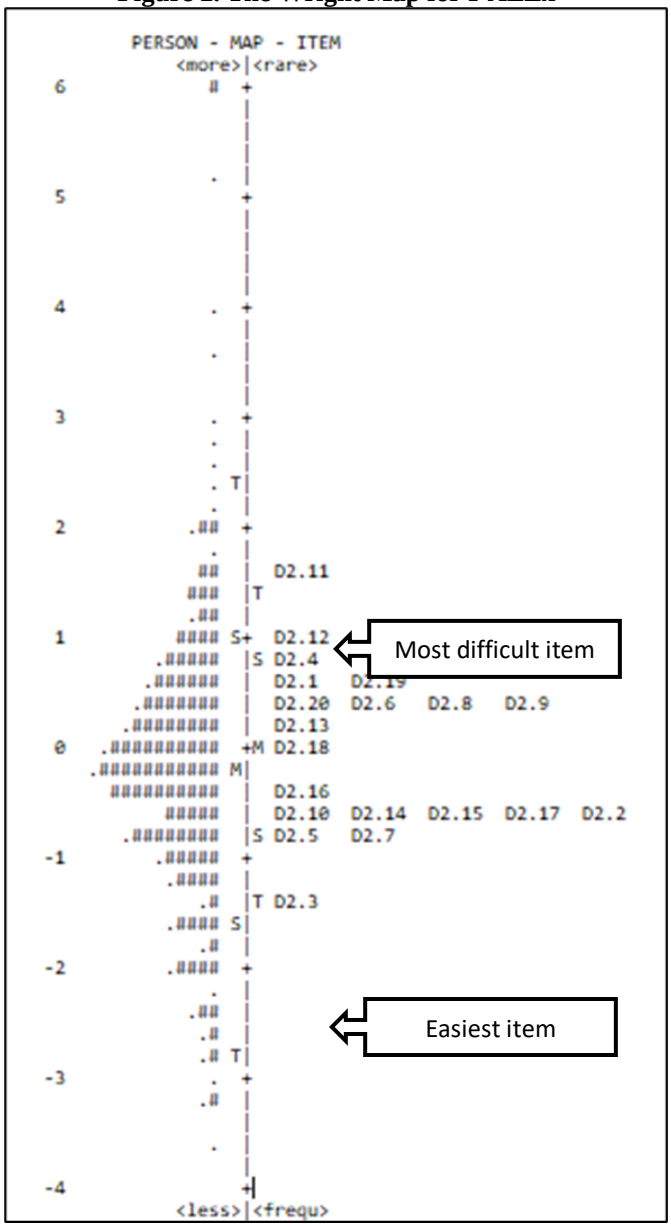


DOI: 10.18860 /ijazarabi.V8i3.30108
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

The Wright Map

Figure 2, known as the Wright Map, displays how both the rubric items and the study's respondents are distributed. The right side of the map organizes the items by their level of difficulty, placing the hardest items at the top and the easiest at the bottom. Conversely, the map's left side depicts the students' abilities in descending order. For I-ALEx, the map demonstrates that item means exceed person means, indicating that the items, on average, are perceived as challenging by the respondents. The item in the I-ALEx that is most easily endorsed is D2.3 (I listen to Arabic songs) and the most difficult item to endorse is D2.11 (I interact in Arabic at home) in the sub-construct target language. This shows that according to students' opinions about their self-evaluation, students prefer listening to Arabic songs rather than interacting in Arabic at home.

Figure 2. The Wright Map for I-ALEx



Source: Winsteps

The Rasch Measurement Model is used in this study to examine the psychometric properties of the I-ALEx instrument to assess its efficacy in evaluating higher education students' exposure to the Arabic language. This method offers a solid framework for assessing the validity and reliability of the instrument, ensuring it precisely measures the targeted construct. Data, processed with Winsteps software version 5.2, are examined against several criteria, including item fit, reliability, separation indices, unidimensionality, local independence and Wright Map to confirm the instrument's accuracy and consistency. This thorough analysis seeks to validate the I-ALEx as a dependable tool for educational assessment.

First, the analysis of the I-ALEx instrument reveals impressive reliability and separation indices, underscoring its robustness as an assessment tool. Specifically, a person separation index of 3.12 and a reliability coefficient of 0.91 indicate that the instrument effectively differentiates between respondents with varying levels of exposure to the Arabic language. This level of separation suggests that the instrument can categorize students into multiple distinct ability groups, providing nuanced insights into their language exposure. Similarly, an item separation index of 8.87 with a reliability coefficient of 0.99 reflects a broad range of item difficulty levels, affirming the instrument's capability to assess diverse aspects of language exposure comprehensively. These high separation indices, well above the threshold recommended by Linacre (2002), indicate that the I-ALEx is not only consistent in its measurements but also precise in distinguishing between different levels of both item difficulty and respondent ability. This dual capability enhances the instrument's utility in educational settings, providing educators with detailed, reliable data to inform their teaching strategies and interventions. Furthermore, the robust reliability coefficients nearing 1.00 further reinforce the precision and consistency of the I-ALEx, making it a trustworthy tool for evaluating higher education students' exposure to the Arabic language.

Next, the item fit analysis of the I-ALEx reveals a nuanced picture of item performance and fit within the Rasch model. While eleven items exhibited ZSTD values outside the accepted range, their acceptable MNSQ outfit and PTMEA CORR values suggest that these items are not entirely misfitting. The ZSTD value, or standardized fit statistic, is sensitive to sample size and can indicate unexpected response patterns, but it should not be the sole criterion for item removal. The MNSQ outfit values remaining within the acceptable range indicate that the majority of the responses align well with model expectations, minimizing noise or unpredictability in the data. PTMEA CORR values, which assess the correlation between item scores and the overall measure, further support the retention of these items as they contribute positively to the construct being measured (Wieser et al., 2022).

Therefore, the reliance on multiple criteria for item evaluation underscores the complexity of maintaining a balanced measurement model. Items that display misfit on one statistic but not others may still hold valuable information and maintain the integrity of the scale when refined. This holistic approach to item analysis, supported by Boone et al. (2014) and further explained by Abdul Aziz et al. (2014), ensures that items are not prematurely discarded based on a single indicator of misfit. Consequently, the decision to retain and refine the eleven items is justified, emphasizing a rigorous yet flexible application of the Rasch measurement principles to enhance the validity and reliability of the I-ALEx instrument.

The assessment of unidimensionality for the I-ALEx instrument demonstrates its capability to measure a single construct effectively (Brentani & Golia, 2007). The use of Residual Principal Component Analysis (PCA) within the Rasch model framework supports this evaluation by identifying the proportion of variance explained by the construct and the degree of noise in the data. The results indicate that the instrument aligns well with the intended construct, as evidenced by a high percentage of variance explained and minimal unexplained variance in the residuals. This alignment suggests that the items work together cohesively to measure students' exposure to the Arabic language without introducing significant ambiguity or confusion (Alnahdi et al., 2025). Consequently, the I-ALEx instrument is validated as a reliable and precise tool for this specific educational context, ensuring that it accurately reflects the targeted construct and provides meaningful insights for educators and researchers.

Additionally, the analysis of the rating scale for the I-ALEx instrument underscores its effectiveness in maintaining valid and reliable measurements. A key aspect of this process is the evaluation of the Andrich thresholds, which confirm that the scale categories are functioning as intended, without necessitating any modifications (Andrich, 2011; Bradle & Massof, 2018). The absence of discrepancies in the Andrich thresholds, falling within the acceptable range, ensures that respondents can differentiate between scale categories clearly, supporting the instrument's precision (Andrich, 2004; Wind, 2022). Additionally, the consistent transition of observed averages from negative to positive values indicates a well-calibrated scale that captures a uniform response pattern. This uniformity enhances the instrument's reliability by ensuring that each category is appropriately utilized by respondents, reducing the likelihood of ambiguous or inconsistent responses. The graphical representation further substantiates the effectiveness of the scale, showing distinct yet intersecting categories that validate the suitability of the rating scale employed in the I-ALEx instrument. Consequently, these findings highlight the robustness of the rating scale, affirming its capability to provide accurate and meaningful data for assessing higher education students' exposure to the Arabic language.

Finally, the Wright Map analysis reveals that a greater number of students are located lower on the map, differ from those positioned higher, indicating that many students have relatively low exposure to the Arabic language. The bottom part of the map represents students with lower exposure, while the top part represents those with higher exposure. This distribution suggests that students find the items overall challenging, as evidenced by the item means surpassing the person means. In practical terms, this means that many students experience limited interaction with the Arabic language, particularly in more immersive contexts. For instance, students find it easier to listen to Arabic songs (an easier item) than to interact in Arabic at home (a more difficult item). This clustering at the bottom highlights a need for enhanced educational strategies to increase students' exposure to Arabic, fostering a more comprehensive and immersive language learning experience. Overall, this instrument reliably and validly assesses Arabic language exposure in higher education students.

CONCLUSION

In conclusion, the comprehensive analysis of the I-ALEx instrument using the Rasch Measurement Model demonstrates its robustness and reliability as an assessment tool for measuring higher education students' exposure to the Arabic language. The high

reliability and separation indices confirm the instrument's capacity to distinguish between varying levels of language exposure, while the item fit analysis ensures that the majority of items align well with the model expectations. The assessment of unidimensionality and the assessment of the rating scale further validate the instrument's precision and consistency. These findings collectively endorse the I-ALEx as a reliable and effective tool for educational assessment, providing educators with accurate and meaningful data to inform their teaching strategies and interventions. This study underscores the importance of rigorous validation processes in developing reliable educational measurement instruments.

ACKNOWLEDGEMENT

This study is fully funded by Faculty of Islamic Studies, Universiti Kebangsaan Malaysia.

REFERENCES

Abdul Aziz, A., Jusoh, M. S., Omar, A. R., Amlus, M. H., & Awang Salleh, T. S. (2014). Construct Validity: A Rasch Measurement Model Approaches. *J. Appl. Sci & Agric*, *9*(12), 7–12.

Al Zoubi, S. (2018). The Impact of Exposure to English Language on Language Acquisition. *Journal of Applied Linguistics and Language Research*, *5*(4), 151–162. http://www.jallr.com/

Alnahdi, A. H., Alsubiheen, A. M., & Aldaihan, M. M. (2025). Rasch Measurement Model Supports the Unidimensionality and Internal Structure of the Arabic Oswestry Disability Index. *Journal of Clinical Medicine*, *14*(4), 1259. https://doi.org/10.3390/JCM14041259

Alsoudi, S., & Abu Shindi, Y. (2023). Rasch Rating Scale Modelling of the Arabic Version of the Critical Thinking Disposition Scale. *Journal of Educational and Psychological Studies [JEPS]*, *17*(4), 359–369. https://doi.org/10.53543/jeps.vol17iss4pp359-369

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care*, *42*(1 Suppl), 7. https://doi.org/10.1097/01.MLR.0000103528.48582.7C

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research*, *11*(5), 571–585. https://doi.org/10.1586/ERP.11.59

Azrilah Abdul Aziz, Mohd Saidfudin Masodi, & Azami Zaharim. (2013). *Asas Model Pengukuran Rasch: Pembentukan Skala & Struktur Pengukuran*. Penerbit Universiti Kebangsaan Malaysia.

Baharudin, H. (2017). Strategi dan teknik terjemahan novel Arab dalam kalangan pelajar universiti. *GEMA Online Journal of Language Studies*, *17*(4), 225–243. https://doi.org/10.17576/gema-2017-1704-15

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences*. Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4

Bradle, C., & Massof, R. W. (2018). Method of successive dichotomizations: An improved method for estimating measures of latent variables from rating scale data. *PLoS ONE*, *13*(10), e0206106. https://doi.org/10.1371/JOURNAL.PONE.0206106

Brentani, E., & Golia, S. (2007). Unidimensionality in the Rasch model: how to detect and interpret. *Statistica*, *67*(3), 253–261. https://doi.org/10.6092/ISSN.1973-2201/3508

DeKeyser, R. M. (2007). Study abroad as foreign language practice. *Practice in a Second Language*, 208–226. https://doi.org/10.1017/CBO9780511667275.012

Kozhevnikova, E. (2014). Exposing Students to Authentic Materials as a Way to Increase Students' Language Proficiency and Cultural Awareness. *Procedia - Social and Behavioral Sciences*, *116*, 4462–4466. https://doi.org/10.1016/j.sbspro.2014.01.967

Linacre, J. (2002). What Do Infit, Outfit, Mean-Square, and Standardization Mean? *Archives of Rasch Measurement*, *16*, 871–882.

Linacre J. M. (2012). *A User's Guide to WINSTEPS: Rasch Model Computer Programs*. MESA Press.

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*(6).

Mohammed AbdAlla AbdAlgane Mohammed. (2020). The Impact of Culture on English Language Learning. *International Journal on Studies in English Language and Literature*, *8*(1), 21–27. https://doi.org/10.20431/2347-3134.0801003

Retief, L., Lutz, M., & M Potgieter. (2013). The usefulness of the rasch model for the refinement of likert scale questionnaires. *African Journal of Research in Mathematics, Science and Technology Education*, *17*(1–2), 126–138. https://doi.org/10.1080/10288457.2013.828407

Rushdi, N. A. A. M., & Asbulah, L. H. (2022). Validity and Reliability of the Student Attitudes towards Arabic Language Learning Instrument (I-SPPBA) using the Rasch Measurement Model. *International Journal of West Asian Studies*, *14*, 98–110.

Rushdi, N. A. A. M., Asbulah, L. H., Bakar, K. A., Lubis, F.-A., & Aladdin, A. (2024). Rasch Analysis of Students' Motivation in Learning Arabic Instrument (I-SMA) for Malaysian Universities. In Q. Zhang (Ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)* (pp. 299–310). Atlantis Highlights in Social Sciences, Education and Humanities. https://doi.org/10.2991/978-94-6463-494-5_19

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata Publishing House.

Wieser, H., Mischo-Kelling, M., Cavada, L., Lochner, L., Fink, V., Naletto, C., & Vittadello, F. (2022). Evaluating Measurement Properties of the Adapted Interprofessional Collaboration Scale through Rasch Analysis. *Healthcare*, *10*(10), 2007. https://doi.org/10.3390/HEALTHCARE10102007

Wind, S. A. (2022). Detecting Rating Scale Malfunctioning With the Partial Credit Model and Generalized Partial Credit Model. *Educational and Psychological Measurement*, *83*(5), 953. https://doi.org/10.1177/00131644221116292

Zhan, C. (2016). The Importance of Culture Factor in Foreign Language Teaching. *Theory and Practice in Language Studies*, *6*(3), 581. https://doi.org/10.17507/tpls.0603.17