Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/15292
DOI URL: http://dx.doi.org/10.21474/IJAR01/15292

RESEARCH ARTICLE

VALIDITY AND RELIABILITY OF THE ARABIC PRODUCTIVE VOCABULARY KNOWLEDGE INSTRUMENT AT THE HIGHER EDUCATION INSTITUTION LEVEL (I-PVKA IPT) USING THE RASCH MEASUREMENT MODEL

Nur Afiqah Athirah Mohd Rushdi and Lily Hanefarezan Asbulah
Research Centre for Arabic Language and Islamic Civilization, Faculty of Islamic Studies, UniversitiKebangsaan Malaysia, 43600Bangi, Malaysia.

Manuscript Info

Manuscript History
Received: 27 June 2022
Final Accepted: 30 July 2022
Published: August 2022

Key words:-
Arabic Language, Productive Vocabulary Knowledge, Instrument Assessment, Rasch Measurement Model

Abstract

Knowledge of productive vocabulary is something critical in learning Arabic in Malaysia. A precise measurement tool is needed in order to know the level of students' knowledge of productive Arabic vocabulary.Therefore, this research aims to validate the Arabic Productive Vocabulary Knowledge instrument at the IPT level (i-PVKA IPT). Through a quantitative approach, this test was distributed to 40 final year students of Selangor International Islamic University College (KUIS) and Sultan Abdul Halim Mu'adzam Shah International Islamic University (UniSHAMS) who are studying the Arabic language program.Next, the data were analyzed usingthe Rasch Measurement Model with the help of Winsteps version 5.2. The findings recorded a high item reliability index, which is at 0.86 and item separation index at 2.46.The noise level found is at 8.3% which is far from the ceiling value of 15%.Meanwhile, only seven items were found to be out of the mean square fit range, and those items were retained and refined.In conclusion, the i-PVKA IPT instrument was successfully proven to have a high level of validity and reliability by the Rasch Measurement Model and should be used to measure the level of students' knowledge of Arabic productive vocabulary in higher learning institutes.

Introduction:-

Productive Vocabulary Knowledge

Lexical knowledge or vocabulary is seen as something critical in learning Arabic in Malaysia. Vocabulary is an important language center because it plays a big role in the mastery of four types of language skills in foreign language learning, namely reading, writing, listening, and speaking (Nation, 2001). Khatijah Md Khatib dan Hakim Zainal (2018) also admitted that aspects of reading, writing, listening, and speaking cannot be mastered properly without solid knowledge and mastery of vocabulary. Therefore, the importance of vocabulary is often a topic of discussion among linguistic scholars.

Productive vocabulary refers to words that students understand and can use that cover writing and speaking skills. Knowledge of productive vocabulary is central to communicative competence and foreign language acquisition (Schmitt, 2000). The ability to communicate successfully will be hindered if the individual has a limited vocabulary. Likewise, in the aspect of writing, the writing of an individual does not deliver the meaning if the vocabulary used is

Corresponding Author:- Lily Hanefarezan Asbulah
Address:- Research Centre for Arabic Language and Islamic Civilization, Faculty of Islamic Studies, UniversitiKebangsaan Malaysia, 43600Bangi, Malaysia.

inappropriate and out of place. The development of productive vocabulary knowledge needs to be given attention because both speaking and writing skills are indicators of success for the mastery of a language (Zunita Mohamad Maskor, Harun Baharudin & Maimun Aqsha Lubis, 2016).

Every aspect of vocabulary knowledge can be used as a guide in the construction of vocabulary measurement instruments as shown in detail by Nation (2001) in Table 1. Nation (2001) has determined the construction of vocabulary knowledge in a broader perspective by combining three main components namely form, meaning, and use.

Table 1:- Aspects in Vocabulary Knowledge.

Form	Spoken	**R**What does the word sound like? **P**How is the word pronounced?
	Written	**R** What does the word look like? **P** How is the word written and spelled?
	Word parts	**R** What parts are recognizable in this word? **P** What words parts are needed to express meaning?
Meaning	Form and meaning	**R** What meaning does this word form signal? **P** What word form can be used to express this meaning?
	Concepts and references	**R** What is included in the concept? **P** What items can the concept refer to?
	Associations	**R** What others words does this word make us think of? **P** What other words could we use instead of this one?
Use	Grammatical functions	**R**In what patterns does the word occur? **P** In what patterns must we use this word?
	Collocations	**R** What words or types of word occur with this one? **P** What words or types of words must we use with this one?
	Constrains on use	**R** Where, when and how often would we meet this word? **P** Where, when and how often can we use this word?

Note. R= receptive, P= productive

Types of Productive Vocabulary Tests

Past scholars have used several types of tests to test the level of productive vocabulary knowledge. Koizumi (2003) stated the three main productive vocabulary tests. First, the Lexical Frequency Profile (LFP) test developed by Nation as a way to assess whether a certain text is suitable for use with students at a certain level of proficiency (Meara, 2005). LFP test is a test of writing essays over 200 words, then the percentage of words used by students at different levels of vocabulary frequency will be calculated using computer software, and the ratio obtained is interpreted as "independent productive ability" (Laufer & Nation, 1999).

Second, the Vocabulary-size Test of Controlled Productive Ability (VTCPA) or also known as the Controlled Productive Ability Test (CPAT). This test provides an underlined blank space for the respondent to fill in with the target word (Laufer & Nation, 1999). However, Mochizuki dan Aizawa (2001) thought that this test might not be suitable for beginner and lower intermediate level students.

Third, the Lex30 test (Meara & Fitzpatrick, 2000), which measures context-limited productive vocabulary. In this test, the respondent is given 30 seconds to write as many words as possible that are related to the context of the given stimulus word.

Thus, past scholars have proven that there are various forms of tests that can be implemented to make it easier for researchers to test the level of knowledge of the productive vocabulary of a language being studied. This shows that productive vocabulary is not only limited to tests that require communication alone, but can also be tested in various ways in the form of writing.

Rasch Measurement Models in Vocabulary Instrument Measurement

Studies on the assessment and measurement aspects of vocabulary instruments using the Rasch Measurement Model have been carried out widely by researchers both in Malaysia and abroad. This is because this model is often used by

researchers who conduct quantitative studies based on Item Response Theory (TRI) whereby this model meets the basic requirements of measurement (Juliana Othman, 2017).

For example, Beglar (2009) conducted a study to demonstrate preliminary validation of a 140-item Vocabulary Size Test designed to measure written receptive knowledge of the first 14,000 words of the English language. This study was conducted on 19 native English speakers and 178 native Japanese speakers. After that it was analyzed using the Rasch Measurement Model.

There is also another study conducted on female university students in Japan by Matsuo (2021) which aimed to validate the Phrasal Vocabulary Size Test (PVST) which was built to measure second language (L2) students' receptive knowledge of formulaic language using Rasch Analysis. The results showed that most of the items in the PVST show a good fit for the Rasch model. Both item and individual reliability index estimates for the PVST test were good.

In the context of the Arabic language, there are several studies on Arabic vocabulary that use the Rasch Measurement Model. Among them is the study of Kaseh Abu Bakar et al. (2021) which measured the validity of the Quranic Vocabulary Test (QVT). Their study focused on vocabulary items found in the Quran that have been selected based on the 500 highest frequency vocabulary in the Quranic Arabic Corpus through systematic sampling. Conclusion from the statistical analysis confirmed that the QVT instrument is believed to be able to measure the knowledge of the meaning of Quran verbs with a reliability value exceeding 0.80 for all the test sets studied.

The study of Nadzirah Norudin et al. (2021) was conducted with the aim of measuring the level of item difficulty as well as the level of repetition of various verb vocabulary items using the Rasch Measurement Model. Another objective of their study was to measure the level of students' mastery of the Qur'anic verb vocabulary. The results of the functional examination using the Winstep software successfully proved that the level of difficulty of the item as well as the student's mastery were affected by the frequency of repetition of an item.

Lily Hanefarezan Asbulah et al. (2018) has also conducted a study using the Rasch Measurement Model to validate the Arabic language collocation knowledge instrument. Collocation is part of vocabulary because it is a combination of two or more words that bring a new meaning that has a different meaning from the original single word. The research was conducted on higher learning institutions' students. The results of that analysis scientifically proved that the Arabic language collocation knowledge test instrument has been confirmed to have a high level of reliability which can be used to test students at higher learning institution's level.

In conclusion, the Rasch Measurement Model has been widely used as a measurement tool to validate instruments in quantitative studies. The use of this measurement model is not something new in linguistic studies, especially in testing the vocabulary of a language.

Arabic Vocabulary Mastery Issues

Research from previous studies showed that the level of Arabic language proficiency of students at higher learning institutions is at a worrying level due to the lack of vocabulary. This is proven in the study of Abdul Razif (2015) who found that the average of words mastered by 227 university students was 1231 words and only 9.1% of students mastered words at a high level. The same goes for research of Nurain Syafina Husain and Norhayuza Mohamad (2020), which stated that Arabic language students at university level still lack sufficient vocabulary to undergo the study process at higher learning institution.

Even so, the teaching and learning of Arabic in Malaysia focus more on aspects of grammar and sentence structure (Zunita Mohamad Maskor et al., 2016) than vocabulary. This is because in general, society view that vocabulary in a foreign language is basically just words and conclude that learning new vocabulary is just a mnemonic and recall exercise (Bakker, 2020). However, mastering the grammar part alone will not be useful if one does not master the vocabulary of a language (Mc Carthy, 1988).

In addition, past studies have also proven that the field of Arabic vocabulary knowledge has received less attention from researchers. The research by Hasnurol Hashim, Kaseh Abu Bakar and Maheram Ahmad (2020) found that measuring or testing vocabulary mastery is one of the least conducted research, which is only 8%, especially for the aspect of knowledge of meaning and use of vocabulary. Therefore, an accurate measurement tool is needed to

measure the level of students' Arabic vocabulary knowledge. This is to ensure that the instruments used are able to measure and test what is to be measured (Azizi Yahya et al., 2017). Therefore, the purpose of this study is to develop an instrument to measure the level of productive Arabic vocabulary knowledge of students at higher learning institution level (i-PVKA IPT) and further determine the validity and reliability of the instrument.

Methodology:-

Sample Study and Data Collection

This study is a quantitative study conducted on 40 final year students who are currently pursuing bachelor's degree program in Arabic at two private universities, namely Selangor International Islamic University College (KUIS) and Sultan Abdul Halim Mu'adzam Shah International Islamic University (UniSHAMS). Permission was sought from the Dean of Kulliyyah of Arabic Language, UniSHAMS and the Dean of the Faculty of Islamic Civilisation Studies, KUIS to conduct this study. Since the teaching and learning process during the semester when this research took place were conducted completely online in their respective universities, the data collection process was also carried out online. After getting permission, the questionnaire was distributed in the form of Google Form through the student representative and also the head of the department to the respondents. The data findings were collected in a Spreadsheet on Google Form and transferred into an Excel file, then analyzed using the Rasch measurement model with the help of Winsteps version 5.2 software.

Study Instrument

The i-PVKA IPT instrument was built based on the Controlled Productive Ability Test (CPAT) format by Laufer and Nation (1999) as shown in Figure 1, which is a form of question that requires the respondent to fill in the blanks with the correct target word. Prefix or suffix letters are provided as answers to prevent respondents from filling in other semantically appropriate words.

Figure 1:- CPAT Question Format (Laufer & Nation, 1999).

I'm glad we had this opp_____ to talk. [Answer: opportunity]

A total of 20 word items in this test instrument are adapted from the instrument of Zunita Mohamad Maskor, Harun Baharudin and Maimun Aqsha Lubis (2018) and Harun Baharudin (2014) who conducted research on students at the secondary school level. This study takes into account that the words that have been learned by secondary school students should be words that are commonly heard and used by students at the higher institution level.The items consist of verbs and nouns to represent four word levels, namely 1000, 2000, 3000 and 4000 words. Therefore, 20 questions have been adapted from the instrument and modified to follow the CPAT question format with the order of items according to the level of word difficulty in increasing order as indicated in Table 2.

Table 2:- Order of Items According to Difficulty Level.

Difficulty Level	Item
1000	C1 – C5
2000	C6 – C10
3000	C11 – C15
4000	C16 – C20

Validity Of Research Instruments

In a process of instrument validity, content validity and face validity are important in ensuring that the instrument used in the study is accurate and able to reflect the findings with the construct tested (Mohammad Rahim Kamaluddin& Wan Shahrazad Wan Sulaiman, 2019).

For this study, a panel of five experts has been appointed to carry out the content validity process of the i-PVKA IPT instrument to ensure that the selection of vocabulary items in this test is appropriate to the level of difficulty of the tested vocabulary. The selection of a panel of experts to confirm the content of this questionnaire item is based on their areas of expertise in Arabic language education, morphological linguistics, and Arabic linguistics. The experts are lecturers from Kulliyyah of Arabic Language (UniSHAMS), Faculty of Languages and Linguistics (UM), Faculty of Islamic Studies (UKM), Faculty of Revelation and Humanities (UIAM), and Faculty of Education

(UIAM). The expert panel focused on language style, test format, and vocabulary selection tested in this i-PVKA IPT instrument.

After the expert evaluation process was completed, the content validity of the instrument was measured using the Content Validity Ratio (CVR) and the Content Validity Index (CVI), a quantitative measurement procedure of content validity introduced by (Lawshe, 1975). CVR aims to validate items through empirical measurement and focuses on the importance of an item in the instrument and then helps the researcher to determine whether the item on the instrument should be dropped or retained (Mohd Effendi @ Ewan Mohd Matore et al., 2017). The accepted range of CVR values is -1 to +1, and a value close to +1 proves the expert panel's agreement on the importance of the item in content validity. Lawshe (1975) suggested that a CVR value that exceeds the value of zero (CVR>0) proves that the majority of the expert panel involved has agreed to rate the item as very important and has met content validity. Overall, all the i-PVKA IPT items have achieved a good agreement coefficient that is above the zero value. The value shows that all the items are representative of the construct domain and have been confirmed to be important.

The CVI value will be calculated as a whole after identifying the CVR value for each item. Lynn (1986) proposed a three to ten expert panel to carry out the validation process. According to him, the accepted CVI value is >0.80 taking into account the average level of agreement between expert panels. Overall, only one item, which is item C14, (استورد) did not reach a value of 0.80. So, after looking at the researcher's needs and getting expert approval, the item was retained and refined. While the CVI value for the other 19 items reached a suitable agreement coefficient which exceeded the value of 0.80. Next, the i-PVKA IPT instrument has been recognized as capable of testing the content aspects to be tested by the expert panel. Nevertheless, the panel of experts suggested adding fa'il (perpetrator) in some sentence expressions in order to make it a complete sentence that is easy for students to understand.

Thus, the researcher has improved the i-PVKA IPT instrument according to the recommendations and suggestions of experts. Next, the instrument was distributed to the students and the data were analyzed to obtain the reliability of the instrument's items.

Research Findings And Discussion:-

Data findings were analyzed using the Rasch measurement model approach with the help of Winsteps version 5.2 software. The functional examination of items was performed based on several aspects, namely (a) reliability index and item-respondent separation, (b) detecting item polarity based on PTMEA CORR values, (c) detecting item suitability, (d) determining dependent items based on standardized residual correlation values, and (e) measure dimensional uniformity (unidimensionality).

Reliability Index and Item-Respondent Separation

The acceptable Cronbach's alpha value score based on the Rasch Measurement Model is 0.71-0.99 as explained in Table 3 by Bond and Fox (2015).

Table 3:- Cronbach's Alpha Score Interpretation.

Cronbach's Alpha Score	Reliability Interpretation
0.8-1.0	Very good and effective with a high level of consistency
0.7-0.8	Good and is acceptable
0.6-0.7	Acceptable
<0.6	The item needs refinement
<0.5	The item needs to be discharged

After the inspection and functional study was carried out, the reliability value of the test instrument is seen by referring to the reliability index as well as the separation of items and respondents. Item and respondent reliability can verify the extent to which the item conforms to the Rasch measurement model as well as the item and respondent segregation index. The isolation index shows the number of strata of item difficulty and individual ability identified in the studied group. The findings of the study show a very good Cronbach's alpha value for the i-PVKA IPT instrument which is 0.84 as shown in Table 4.

Table 4:- Cronbach's Alpha Value for the i-PVKA IPT Instrument.

```
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .84
```

Next, the item reliability index for the i-PVKA IPT instrument is 0.86 as shown in Table 5, which is considered to be very good and effective with a high level of consistency because it exceeds the value of 0.80 (Bond and Fox ,2015). The item separation index for i-PVKA IPT obtained is 2.46. This value can be interpreted as high because it has exceeded the value of 2.00 as outlined by Linacre (2002).

Meanwhile, Table 6 shows the reliability index of respondents for i-PVKA IPT which is 0.75 with a separation value of 1.74. According to Bond and Fox (2015), a value of 0.71-0.80 is categorized as good and acceptable. As for the respondent's isolation value, the respondent's isolation value must exceed the value of 1.00 to guarantee that students are measured across the spread as stated by Krishnan and Idris (2014). Overall, the reliability and isolation value of the i-PVKA IPT instrument is acceptable and this instrument is suitable for measuring the level of Arabic productive vocabulary knowledge of students at the higher learning institution level.

Table 5:- Reliability Value and Item Separation of the i-PVKA IPT Instrument

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	9.6	40.0	.68	.74				
S.D.	8.6	.0	2.37	.48				
MAX.	27.0	40.0	4.56	1.83				
MIN.	.0	40.0	-3.23	.41	.62	-1.6	.14	-1.1
REAL RMSE	.89	TRUE SD	2.19	SEPARATION	2.46	ITEM	RELIABILITY	.86
MODEL RMSE	.88	TRUE SD	2.20	SEPARATION	2.49	ITEM	RELIABILITY	.86
S.E. OF ITEM MEAN = .54								

Table 6:- Value of Reliability and Separation of Respondents against the i-PVKA IPT Instrument

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	4.8	20.0	-1.91	.89				
S.D.	3.7	.0	2.02	.39				
MAX.	13.0	20.0	1.93	1.86				
MIN.	.0	20.0	-5.35	.64	.46	-1.8	.17	-1.3
REAL RMSE	1.01	TRUE SD	1.75	SEPARATION	1.74	PERSON	RELIABILITY	.75
MODEL RMSE	.97	TRUE SD	1.77	SEPARATION	1.82	PERSON	RELIABILITY	.77
S.E. OF PERSON MEAN = .32								

Item Polarity

Item polarity analysis (Point Measure Correlation- PTMEA CORR) is an early detection or basic analysis of construct validity which shows that items for an instrument move towards the evaluation of a single sub-construct (Bond & Fox, 2015). According to Linacre (2005), a measurement with a positive index means that the item measures the construct it wants to measure and it works in parallel to measure the construct. However, the item needs to be re-examined if it has a negative index, either needs to be retained and refined, or dropped. Besides, Nunnally and Bernstein (1994) stated that a PTMEA CORR value of less than 0.30 indicates slack items.

Table 7 shows the item polarity values for the i-PVKA IPT instrument. The minimum PTMEA CORR value for the i-PVKA IPT instrument item recorded is (.30) while the maximum PTMEA CORR value is (.72). The results of the analysis show that no item is below the value of 0.30 and no item shows a negative value. Thus, it was found that the items for the i-PVKA IPT instrument move in a set direction and are believed to be able to measure the level of productive Arabic vocabulary knowledge of students at the higher learning institution level.

Table 7:- Item Polarity Value.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
12	1	40	3.28	1.05	.77	.0	.14	-.6	.30	.22	97.1	97.1	C12
9	3	40	1.93	.67	1.13	.4	.60	.1	.34	.36	91.4	91.3	C9
6	9	40	.13	.47	1.45	1.6	1.33	.7	.39	.54	71.4	82.3	C6
19	3	40	1.93	.67	.90	-.1	.33	-.2	.42	.36	91.4	91.3	C19
4	12	40	-.48	.44	1.45	1.9	1.45	1.0	.42	.58	65.7	78.3	C4
20	3	40	1.93	.67	.80	-.4	.30	-.2	.44	.36	91.4	91.3	C20
14	4	40	1.52	.61	.88	-.2	.46	-.1	.46	.41	91.4	89.0	C14
11	3	40	1.93	.67	.69	-.7	.21	-.4	.48	.36	91.4	91.3	C11
2	21	40	-2.06	.42	1.34	1.6	1.63	1.4	.51	.65	71.4	76.2	C2
5	7	40	.61	.51	.83	-.5	.95	.2	.54	.50	88.6	85.1	C5
8	18	40	-1.54	.41	1.11	.6	1.31	.9	.58	.63	71.4	75.0	C8
15	9	40	.13	.47	.91	-.3	.61	-.5	.59	.54	82.9	82.3	C15
3	23	40	-2.42	.43	1.04	.3	.83	-.2	.65	.65	71.4	77.2	C3
1	18	40	-1.54	.41	.93	-.3	.79	-.5	.66	.63	82.9	75.0	C1
18	9	40	.13	.47	.62	-1.6	.40	-1.1	.67	.54	88.6	82.3	C18
13	22	40	-2.24	.42	.91	-.4	.81	-.3	.68	.65	80.0	76.8	C13
7	27	40	-3.23	.47	.84	-.6	.54	-.5	.72	.66	85.7	82.3	C7
MEAN	9.6	40.0	.68	.74	.98	.1	.75	.0			83.2	83.8	
S.D.	8.6	.0	2.37	.48	.24	.9	.44	.6			9.3	6.7	

Item Fit

Analysis using the Rasch measurement model estimates the degree of fit of items that measure a latent variable. The item fit check starts by making sure that the Mean-Square (MNSQ) value of the item is in the range of 0.5 to 1.5. An item with a value less than 0.5 indicates domain overlap with other items while an item is not homogeneous with other items in a measurement scale if it has a value higher than 1.5.

According to Bond, Yan and Heene (2021), deviating MNSQ infit statistics will usually cause more concern than MNSQ outfits that are out of range. MNSQ outfit works to help track responses to item difficulty levels. Therefore, MNSQ outfit needs to be given priority in determining the appropriateness of items that measure a construct. Typically, an item that is out of the accepted MNSQ range will show a z-Std value that is out of the accepted range as well, i.e. -2.0<ZSTD<+2.0.

Based on recommendations by Boone et al. (2014), inappropriate items need to be identified based on three criteria, namely MNSQ, ZSTD and PTMEA CORR. If an item meets at least one of the three criteria, then it should be retained (Sumintono&Widhiarso, 2015). The statement is supported by Abdul Aziz et al. (2014), which mentioned that an item is only considered unsuitable if it does not meet the suitability range for all three criteria. For the i-PVKA IPT instrument, there are seven items that are out of the accepted MNSQ range as shown in Table 8. The items are C2(1.63), C14(0.46), C18(0.13), C19(0.33), C20(0.30), C11(0.21), and C12(0.14). However, the ZSTD and PTMEA CORR values for all these items are still within the acceptable range. Therefore, all these items have been preserved and purified.

Table 8:- List of Item Misfit.

Item	**MEASURE**	**Outfit MNSQ (0.50-1.50)**	**Outfit ZSTD (-2.0-2.0)**	**PTMEA-CORR**
C2	-2.06	**1.63**	1.4	0.65

C14	1.52	**0.46**	-0.1	0.41
C18	0.13	**0.40**	-1.1	0.54
C19	1.93	**0.33**	-0.2	0.36
C20	1.93	**0.30**	-0.2	0.36
C11	1.93	**0.21**	-0.4	0.36
C12	3.28	**0.14**	-0.6	0.22

Standardized Residual Correlation Value

The correlation residual standard determination test was performed to identify if there are two overlapping and confusing items. When there are two items with a high residual correlation value, then it shows that the items are not independent, with the reason that the two items have the same characteristics or the items share some other dimensions. Linacre (2005) put the value 0.7 as the maximum residual correlation value. If there is a pair of items with a value above 0.7, then one of the items should be dropped because it has a high correlation value. Table 9 shows that the ten matching items in the i-PVKA IPT instrument have a standardized residual correlation value that meets the condition of local independence, which is between -0.37 and 0.63. Therefore, all items are believed to be unrelated to each other and no pair of items is seen as confusing by respondents.

Table 9:- The Largest Standardized Correlation Values on Items.

CORRELATION	ENTRY NUMBER	ITE	ENTRY NUMBER	ITE
.63	19	C19	20	C20
.39	9	C9	18	C18
.37	11	C11	19	C19
-.51	11	C11	12	C12
-.49	4	C4	19	C19
-.46	9	C9	15	C15
-.39	2	C2	14	C14
-.38	6	C6	14	C14
-.38	4	C4	11	C11
-.37	1	C1	20	C20

Dimensional Uniformity (Unidimensionality)

The aspect of dimensional uniformity is one of the critical aspects to achieve and strengthen the objective of the instrument to measure something in one direction. An instrument can give misleading results when it has an ambiguity in the measurement process (Azrilah Abdul Aziz et al., 2013). The Residual Principal Component Analysis (PCA) technique has been used in Rasch analysis to detect the ability of an instrument to measure in a uniform dimension with an acceptable noise level.

The best value of the variance outlined by Linacre (2002) is >60%. Table 10 shows that the gross variance explained by PCA for the i-PVKA IPT instrument is 47%, which exceeds the minimum value outlined by Linacre J. M. (2012) which is as much as 40%. While the level of item interference in contrast one recorded by the i-PVKA IPT instrument is as high as 8.3% where the value is far from the ceiling value outlined by Fisher (2007) which is 15% and sufficient.

Table 10:- Dimensional Uniformity of the i-PVKA IPT Instrument.

		-- Empirical --		Modeled
Total raw variance in observations	=	32.1 100.0%		100.0%
Raw variance explained by measures	=	15.1 47.0%		48.0%
Raw variance explained by persons	=	5.5 17.2%		17.5%
Raw Variance explained by items	=	9.6 29.9%		30.5%
Raw unexplained variance (total)	=	17.0 53.0%	100.0%	52.0%
Unexplned variance in 1st contrast	=	2.7 8.3%	15.6%	
Unexplned variance in 2nd contrast	=	2.2 6.9%	12.9%	
Unexplned variance in 3rd contrast	=	1.9 5.8%	11.0%	
Unexplned variance in 4th contrast	=	1.6 5.1%	9.7%	
Unexplned variance in 5th contrast	=	1.4 4.5%	8.5%	

Conclusion:-

Overall, it can be concluded that the process of validity and reliability must be carried out against each new instrument that is developed in order to form a valid and accurate instrument to obtain a precise measurement. This study has successfully concluded the discussion of findings using the Rasch Measurement Model to identify the extent to which the i-PVKA IPT test items are valid through unidimensional diagnosis, item suitability, item polarity, reliability index, and item-respondent classification as well as standardized residual correlation values. Thus, this study is expected to contribute and be used as a guide for teachers or researchers who wish to build valid and accurate assessment and measurement instruments.

Based on the Rasch measurement model, the functionality of each item of the i-PVKA IPT instrument can be analyzed in depth to determine its validity and reliability. The inspection results show that all i-PVKA IPT items have been validated and fair to measure the level of productive Arabic vocabulary knowledge. Statistically, the results of this research have successfully proved that the level of validity and reliability of the i-PVKA IPT items is high and appropriate to be used for students at the higher learning institution level.

Acknowledgement:-

This research was supported by *Geran Galakan Penyelidik Muda*(GGPM-2021-016), Faculty of Islamic Studies, Universiti Kebangsaan Malaysia.

References:-

1. 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.
2. Abdul Razif. 2015. PenguasaanKosa Kata dalamKalanganPelajarMelayu di PeringkatKolejUniversiti
3. Azizi Yahya, Peter Voo, Ismail Maakip&Mohd Dahlan A. Malek. 2017. KaedahPenyelidikanDalam Pendidikan. Universiti Pendidikan Sultan Idris.
4. Azrilah Abdul Aziz, MohdSaidfudinMasodi&AzamiZaharim. 2013. Asas Model Pengukuran Rasch: Pembentukan Skala &StrukturPengukuran. PenerbitUniversitiKebangsaan Malaysia.
5. Bakker, B. 2020. Learning arabic vocabulary: The effectiveness of teaching vocabulary and vocabulary learning strategies. Journal of Higher Education Theory and Practice 20(10).
6. Beglar, D. 2009. A Rasch-based validation of the vocabulary size test. Language Testing 27(1): 101–118.
7. Bond, T.G. & Fox, C.M. 2015. Applying the Rasch Model Fundamental Measurement in the Human Sciences. Routledge.
8. Bond, T.G., Yan, Z. & Heene, M. 2021. Applying The Rasch Model Fundamental Measurement in the Human Sciences. Routledge.
9. Boone, W.J., Staver, J.R. & Yale, M.S. 2014. Rasch Analysis in the Human Sciences. Dordrecht: Springer Netherlands.
10. Fisher, J.W.P. 2007. Rating scale instrument quality criteria. Rasch Measurement Transactions 21(1): 1095.

11. Harun Baharudin. 2014. Strategi PembelajaranKosa Kata Bahasa Arab PelajarSekolahMenengah Agama di Malaysia
12. Hasnurol Hashim, Kaseh Abu Bakar &Maheram Ahmad. 2020. PenguasaanKosa Kata Bahasa Arab MenerusiPengetahuanMakna Dan Penggunaannya. Malim: JurnalPengajianUmum Asia Tenggara (Sea Journal of General Studies). 21(1): 160–174.
13. Juliana Othman. 2017. Good Language Learners Are Born, Not Made. Journal of The English Teacher 10.
14. Kaseh Abu Bakar, Hakim Zainal, NadzirahNorudin& Mohamad AzwanKamarudin. 2021. Aplikasi Model Rasch BagiKesahan Item Kosa Kata Kerja Quranic Vocabulary Test (QVT). International Journal of West Asian Studies 13: 259–287.
15. Khatijah Md Khatib & Hakim Zainal. 2018. Penguasaan Kata Nama dan Kata Kerja Arab dalamPelaksanaanKurikulumBersepadu Dini. BITARA International Journal of Civilizational Studies and Human Sciences 1(1).
16. Koizumi, R. 2003. A Productive Vocabulary Knowledge Test for Novice Japanese Learners of English: Validity and its Scoring Methods. Japan-Britain Association for English Teaching Journal 7: 23–52.
17. Krishnan, S. & Idris, N. 2014. Investigating Reliability and Validity for the Construct of Inferential Statistics. International Journal of Learning, Teaching and Educational Research 4(1): 51–60.
18. Laufer, B. & Nation, P. 1999. A vocabulary-size test of controlled productive ability. Language Testing 16(1): 33–51.
19. Lawshe, C.H. 1975. A Quantitaive Approach To Content Validity. Personnel Psychology 28(4): 563–575.
20. Lily HanefarezanAsbulah, MaimunAqshaLubis, Ashinida Aladdin &Mus'abSahrim. 2018. Kesahan dan KebolehpercayaanInstrumenPengetahauanKolokasi Bahasa Arab IPT (i-KAC IPT) Menggunakan Model Pengukuran Rasch. ASEAN Comparative Education Research Journal on Islam and Civilization (ACER-J) 2(1): 97–106.
21. Linacre, J. 2002. What Do Infit, Outfit, Mean-Square, and Standardization Mean? Archives of Rasch Measurement 16: 871–882.
22. Linacre J. M. 2012. A User's Guide to WINSTEPS: Rasch Model Computer Programs. Chicago: MESA Press.
23. Linacre, J.M. 2005. A User's Guide to Winsteps/Ministeps Rasch Model Programs. Chicago, IL: MESA Press.
24. Lynn, M.R. 1986. Determination and Quantification of Content Validity. Nursing Research 35(6).
25. Maskor, Z.M., Baharudin, H. &Lubis, M.A. 2018. Measurement Validity and Reliability of the Productive Vocabulary Knowledge Instrument for Arabic Learners in Malaysian Secondary Schools. Article in Journal of Computational and Theoretical Nanoscience 24: 3423–3426.
26. Matsuo, T. 2021. A Rasch-based Validation of the Phrasal Vocabulary Size Test. 大阪女学院大学紀要第 18: 93–108. https://www.researchgate.net/publication/359082203.
27. Mc Carthy, M. 1988. Vocabulary and Language Teaching. New York: Longman.
28. Meara, P. 2005. Lexical frequency profiles: A monte carlo analysis. Applied Linguistics 26(1): 32–47.
29. Meara, P. & Fitzpatrick, T. 2000. Lex30: An Improved Method of Assessing Productive Vocabulary in An L2. System 28: 19–30.
30. Mochizuki, M. & Aizawa, K. 2001. A Validity Study of the Vocabulary Size Test of Controlled Productive Ability. Reitaku University Journal 73: 85102.
31. Mohammad Rahim Kamaluddin& Wan Shahrazad Wan Sulaiman. 2019. Teknik Kesahan Dan Kebolehpercayaan Alat UjianPsikologi. PenerbitUniversitiKebangsaan Malaysia.
32. Mohd Effendi @ Ewan MohdMatore, Hisyamsani Idris, Normawati Abdul Rahman & Ahmad ZamriKhairani. 2017. KesahanKandunganPakarInstrumen IKBAR BagiPengukuran AQ MenggunakanNisbahKesahanKandungan. Proceeding of International Conference on Global Education V (ICGE V), Padang, Indonesia (10 –11 April 2017)"Global Education, Common Wealth and Cultural Adversity,". 979–997.
33. NadzirahNorudin, Kaseh Abu Bakar, Hakim Zainal & Mohamad AzwanKamarudin. 2021. Kesukaran Item Kosa Kata Kerja Al-Quran DalamUjianKosa Kata Al-Quran (QVT) Berdasarkan Model Pengukuran Rasch. Al-Basirah 11(1): 47–54.
34. Nation, I.S.P. 2001. Learning Vocabulary in Another Language. Cambridge University Press.
35. Nunnally, J.C. & Bernstein, I.H. 1994. The Assessment of Reliability. Psychometric Theory 3: 248–292.
36. NurainSyafina Husain &Norhayuza Mohamad. 2020. SaizKosa Kata Bahasa Arab DalamKalanganPelajar Ijazah Sarjana Muda Bahasa Arab Komunikasi Professional, UiTM. International Journal of Modern Languages and Applied Linguistics 4(2): 38.
37. Schmitt. 2000. Vocabulary in Language Teaching. Cambridge: Cambridge University Press.

38. Sumintono, B. &Widhiarso, W. 2015. AplikasiPemodelan Rasch Pada Assessment Pendidikan. Cimahi: Trim Komunikata Publishing House.
39. Zunita Mohamad Maskor, Harun Baharudin&MaimunAqshaLubis. 2016. PengetahuanKosa Kata Produktifdalam Kemahiran Menulis Sorry, yes, the same Project View project Validating Productive Vocabulary Knowledge Inventory for Arabic language among Malaysia secondary students View project. https://www.researchgate.net/publication/311668345.
40. Zunita Mohamad Maskor, Harun Baharudin, MaimunAqshaLubis& Nurul Kamalia Yusuf. 2016. Teaching and Learning Arabic Vocabulary: From a Teacher's Experiences. Creative Education 07(03).