Jurnal Pendidikan Indonesia
Volume 11, Number 4, 2022 pp. 626-637
P-ISSN: 2303-288X E-ISSN : 2541-7207
Open Access: https://dx.doi.org/10.23887/jpi-undiksha.v11i4.43797

Psychometric Analysis of an Instrument Evaluating Students' Acceptance of Online Platform to Support Online English Learning

Fathia Amalia Sulthonah[1], Herri Mulyono[2*], Wan Fatimah Wan Ahmad[3]

[1,2] Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
[3] Univesiti Teknologi PETRONAS, Malaysia

ARTICLE INFO

Article history:
Received January 19, 2022
Revised January 21, 2022
Accepted September 30, 2022
Available online December 25, 202

Kata Kunci:
Psikometrik, Analisis Rasch, Model Penerimaan Teknologi

Keywords:
Psychometric, Rasch-Analysis, Technology Acceptance Model

ABSTRAK

Penerimaan teknologi mengambil peran penting dalam konteks adopsi teknologi. Sebelum mendorong penggunaan teknologi secara maksimal, penting bagi pengguna untuk terlebih dahulu mengakui penggunaannya dan mereka juga harus menerimanya. Studi saat ini diusulkan bertujuan untuk menguji validitas dan reliabilitas instrumen versi Indonesia yang menyelidiki konstruk TAM seperti persepsi kegunaan, persepsi kemudahan penggunaan, dan niat perilaku. Kuesioner diadopsi untuk mengelaborasi penerimaan teknologi mahasiswa sarjana terhadap WhatsApp, Google Classroom, dan Microsoft Teams sebagai platform online untuk mendukung pembelajaran bahasa Inggris online berdasarkan Model Penerimaan Teknologi (TAM). Data dikumpulkan dari 370 mahasiswa S1 dari berbagai universitas di Indonesia dan studi ini menerapkan teknik analisis Rasch untuk mengatasi asumsi Rasch seperti dimensi item, keandalan orang dan item, pemetaan orang dan item, skala peringkat, dan pengukuran fungsi item diferensial. Temuan penelitian menunjukkan bahwa kuesioner yang diadopsi dan diterjemahkan ke dalam bahasa Indonesia dianggap cukup memadai dalam konteks karakteristik psikometrik dan dianggap memenuhi syarat untuk mengukur penerimaan teknologi dari platform online yang digunakan untuk pembelajaran online bahasa Inggris.

ABSTRACT

Acceptance of technology takes as a crucial role in context of the technology adoption. Before encouraging the maximal use of technology, it is important that users have to first acknowledge its use and that they have to also accept it. The current study was proposed in aims to examine the validity and the reliability of the Indonesian version of instrument investigating TAM constructs such as perceived usefulness, perceived ease of use, and behavioural intention. The questionnaire was adopted to elaborate the undergraduate students' technology acceptance of WhatsApp, Google Classroom, and Microsoft Teams as online platform to support online English learning based on the Technology Acceptance Model (TAM). Data were collected from 370 undergraduate students from different universities in Indonesia and the study applied Rasch analysis technique to address the Rasch assumptions such as items dimensionality, person and item reliability, person and item mapping, rating scale, and differential items functioning measure. The findings of the study suggested that the adopted and translated to Indonesian questionnaire was found to be sufficient in context of psychometric characteristics and was considered eligible to measure the technology acceptance of online platforms used for English online learning.

1. INTRODUCTION

Acceptance of technology takes as a crucial role in context of the technology adoption. Before encouraging the maximal use of technology, it is important that users have to first acknowledge its use and they have to also accept it. This is because technology acceptance is seen as the determiner towards users' decision on whether they would use such technology adoption or not. For instance, the model of Technology Acceptance Model (TAM) implemented the core variables that relate with acceptance, such as perceived usefulness and perceived ease of use of users' technology in determining behavioural intention when adopting the system (Cakir & Solak, 2014; Teo et al., 2019). Perceived usefulness refers to users' belief that

*Corresponding author
E-mail addresses: hmulyono@uhamka.ac.id (Herri Mulyono)

a technology would improve their productivity, while perceived ease of use means that users see a technology as something that can be utilized with less effort. Hence, if users perceive high levels of usefulness and ease of use towards an application, they will most likely have a great score of behavioural intention in implementing the current technology for further learning. For instance, the result of a study observing the levels of Behavioural Intention from Japanese EFL learners showed that the levels of BI encourage positive attitude. They were also found willing to continue using E-learning even though they have had no previous experience (Coskun & Marlowe, 2015; Craig Gamble, 2017). In line with this, Behavioural Intention is also categorised as one of the strong determiners towards users' technology acceptance, either directly or indirectly, and might influence their future use.

Studies observing students' acceptance of a technology in English language education has been captured. For instance, a study focused on the implementation of Google Classroom in unidirectional setting English classroom was conducted by previous study, the total 305 participants have participated and filled the questionnaire (Al-Maroof & Salloum, 2021). The findings showed that variables of technology acceptance such as perceived ease of use and perceived usefulness were seen as the crucial determiners of students' decision makers in the university, where they can choose whether or not continue implementing Google classroom for their educational learning. Thus, another study exploring more about such extended variables was also captured. A total 202 students of undergraduate levels are examined to observe their acceptance towards WhatsApp using the same technology acceptance model (Mulyono et al., 2021). The findings showed the high level of perceived usefulness, availability of learning support, motivation, and their connectedness with their peers, leading students to most likely will continue use WhatsApp as the tool in their English learning classroom. Another study that took the similar topic was also found. More study was also attempted to evaluate the use of Microsoft Teams as an online platform during Covid-19 distance learning (Pal & Vanijja, 2020). According to them, usability is a crucial aspect to determine whether a technology is effective and useful for students. However, the study stated different result where the utilization of such platform was not correlated with the usability aspect.

In order to understand the potential users' acceptance in using such technologies, various models have been proposed. Technology Acceptance Model (TAM). Four constructs which includes perceived usefulness, perceived ease of use, attitude toward usage, and behavioural intention that can be used as the guidance in identifying user's technology acceptance (Azman et al., 2020; Chung et al., 2015). In terms of technology's behavioural intention, users' perceived usefulness has a direct impact while the perceived ease of use is influenced through students' attitude which will also influence an individual's BI indirectly. Another framework is Technological Pedagogical Content Knowledge (TPACK). In this model, three subscales focusing on knowledge are proposed, namely technological knowledge (TK), pedagogical knowledge (PK), and content knowledge to examine teachers' ability to integrate technology into curriculum (Mishra & Koehler, 2006; Nazari et al., 2019; Tseng, 2018). Furthermore, there are also other extended variables such as anxiety, subjective norm, satisfaction, continuance intention, self-efficacy and facilitating conditions that are correlated to the technology use, with satisfaction played a major effect on users' continuance intention of a technology, and followed by perceived usefulness, attitude, and subjective norm (Lee, 2010; Venkatesh et al., 2003; Venkatesh, 2000).

Many scholars have also developed instruments to explore the levels of technology acceptance. For example, an instrument exploring the factors related to students' behavioural intention to apply mobile English vocabulary learning resources (Bhattacherjee, 2001; Chung et al., 2015; Lee, 2010). The questionnaire was adapted from TAM and analysed through regression which showed that both TAM subscales (PU and PEOU) have correlation in students' behavioural intention along with other variables such as self-efficacy and compatibility. Previous studies using survey have also been attempted to validate the TAM on students' use of social media and LMS as platforms for their EFL teaching and learning, such as Zoom (Alfadda & Mahdi, 2021), Facebook (Rauniar et al., 2014), Google Classroom (Francom et al., 2021), and addressed their correlation to students' learning outputs. However, instruments to expand the overall perspectives of online platforms using the original constructs remain few. Except the study that originated the questionnaire evaluating LINE application (Cakir & Solak, 2014; Chen Hsieh et al., 2017). It incorporates perceived ease of usefulness, perceived ease of use, attitude about us, and behavioural intention of TAM, particularly in examining the four constructs of students' overall perspectives toward a mobile-messaging application named LINE (Cheung & Vogel, 2013; Hu et al., 2003).

This research aims to examine the validity and the reliability of the Indonesian version of TAM adopted questionnaire focusing on perceived usefulness, perceived ease of use, and behavioural intention (Alfadda & Mahdi, 2021; Chen Hsieh et al., 2017). Since most of the previous studies mentioned were all analysed using SPSS, this research is necessary as it will apply another modelling to measure the validity and reliability of the instrument used. In this case, this study attempts to validate the instrument using Rasch analysis which has not been done previously (Chen Hsieh et al., 2017; Francom et al., 2021). The

Rasch technique is a psychometric model to monitor the items' quality in an instrument by measuring its difficulty level and person's ability level (Rauniar et al., 2014; Yu, 2020). In addition, the analysis itself is also proven to provide objective results through invariant measurement in diverse settings. Therefore, the current study is important to be conducted as it will contribute to the literatures the empirical illustration of the modern psychometric model, i.e. Rasch technique, to observe the psychometric characteristics of the Indonesian version of TAM adopted questionnaire (Chen Hsieh et al., 2017; Muntazhimah et al., 2020).

2. METHOD

The current study adopted the design of validation study where the existing responses from data collection were going to undergo a psychometric assessment in order to observe whether the questionnaire applied was able to present consistent (reliability) and true results (validity) or not. The quantitative survey method was aimed to gather data related to TAM variables such as perceived usefulness, perceived ease of use, attitude of use, and behavioral intention (Hackbarth et al., 2003). The participants of this study were 370 Indonesian undergraduate students from different universities, and had experiences in using WhatsApp, Google Classroom, and Microsoft Teams as platforms for their English learning. The sample of the present study was chosen through random sampling and they came from first year up to the fourth-year students. Before they were asked to fill the items, they were asked to answer what application that they often use among Microsoft Teams, WhatsApp, and Google Classroom as well as how long they have been using it and their frequency of their usage.

TAM adopted instrument was employed to explore students' overall perception toward the use of online learning platforms for English learning and teaching activities. The following questionnaire was originally made to examine LINE application using TAM with the two extended variables such as system characteristics and material characteristics (Chen Hsieh et al., 2017; Mulyono et al., 2020). In the present study, the researcher will only focus on the four original variables of TAM which are Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude to Use (AU), and Behavioural Intention (BI). The questionnaire consisted of 51 items, of which 6 items were demographic questions and the remaining 45 items were designed using a 5-point Likert scale, with "strongly agree" (5), "agree" (4), "neutral" (3), "disagree" (2), and "strongly disagree" (1) items. Each of the three applications had 15 items.

The present study applied Rasch analysis technique through WINSTEP 4.5.1 for transforming the raw data into logit scores, and this is known as "the natural logarithmic scale of the odds". Rasch analysis has been widely used to check the validity of the instrument and to maintain its quality (Mulyono et al., 2020; Yu, 2020). In addition, Rasch analysis also provides an appropriate level of accuracy from the data calculation because the conversion has enabled the technique to measure units at the same interval size to ensure that the length between any two measures is meaningful (Colledani et al., 2020; Huang et al., 2020). Prior to the analysis, the data were retrieved from a questionnaire that was distributed using Google Form. It was downloaded in the form of an Excel file, and all of the answers were converted to raw numbers. To keep the data confidential, the author removed information related to the participants such as the name of the university. After the conversion was completed, the logit data underwent two rounds of analysis. The first analysis was the filtering process where the authors removed data that were misfit. Hence, once these misfits have been observed, the researcher removed them and they were re-analysed using the remaining 210 samples. The next process was focusing on applying the Rasch modelling assumptions such as Rasch model fit summary, Item dimensionality, Rating scale, Item bias, and Wright maps (Chan & Subramaniam, 2020; Huang et al., 2020; Mulyono et al., 2020).

3. RESULT AND DISCUSSION

Result

Item and Person Separation Reliability

Table 1. Summary of *Rasch* Measurement Model

Parameter (with quality criteria)	Global scale (45 items)	Duration (2 items)	Perceived Usefulness (3 items)	Perceived ease of Use (4 items)	Attitude to Use (4 items)	Behavioural Intention (2 items)
Model fit: summary of items						
Item mean in logit (criteria 0.0 logits)	0.00, SD=0.78		0.00, SD =.27	0.00, SD = 0.28	0.00, SD = 0.25	0.00, SD = 0.16
Item reliability	0.99		0.88	0.91	0.89	0.73

Parameter (with quality criteria)	Global scale (45 items)	Duration (2 items)	Perceived Usefulness (3 items)	Perceived ease of Use (4 items)	Attitude to Use (4 items)	Behavioural Intention (2 items)
Item separation reliability	0.99		0.88	0.91	0.89	0.73
Item model fit MNSQ range extremes	Infit: 0.62-1.60 Outfit: 0.62-1.64		Infit: 0.73-1.24 Outfit: 0.72-1.30	Infit: 0.68-1.35 Outfit: 0.68-1.35	Infit: 0.77-1.30 Outfit: 0.77-1.30	Infit: 0.74-1.32 Outfit: .74-1.23
Item separation index	9.13 (9)		2.76 (2)	3.21 (3)	2.91 (2)	1.63
Separate item strata == [(4 x separation index)	36.8 = 36 level		11.37 ≈ 11 level	13.17 ≈13 level	11.97 ≈ 11 level	6.85 ≈ 6 level
Model fit: summary of persons						
Person mean in logits	0.50 SD=1.10		0.89 SD=0.95	0.97 SD=0.73	0.73 SD=0.87	0.65 SD=1.11
Person reliability	0.93		0.77	0.75	0.82	0.76
Person separation reliability	0.93		0.77	0.75	0.82	0.76
Person separation index	3.57 (3)		1.83	1.75	2.12	1.80
Separate Person strata = [(4 x separation index) + 1]/3	14.61 = 14 level		7.65 ≈ 7 level	7.33 ≈ 7 level	8.81 ≈ 8 level	7.53 ≈ 7 level
Rating Scale Analysis						
Responses per category (criteria: >10)	NO	NA	NA	NA	NA	NA
Adjacent threshold distance (criteria 1.4-5 logits)	(i) scale 0-1 was (0)-(-2.48) = 2.48 logits; (ii) scale 1-2 was (2.48) – (-1.12) = 3.6 logits; (iii) scale 2-3 was (-1.12) – (0.11) = 1.23 logits; (iv) scale 3-4 was (-0.11) – (1.07) = 1.18 logits					

Parameter (with quality criteria)	Global scale (45 items)	Duration (2 items)	Perceived Usefulness (3 items)	Perceived ease of Use (4 items)	Attitude to Use (4 items)	Behavioural Intention (2 items)
	(v) scale 4-5 was (1.07) – (2.75) = 1.68 logits					
Outfit MNSQ (Criteria: <2 logits)	YES	NA	NA	NA	NA	NA
Probability curve graph (criteria: decent curve on each response category and each item is higher than 0.5 logits)	NO	NA	NA	NA	NA	NA
Average measure (criteria: jumped significantly over rating scale)	YES	NA	NA	NA	NA	NA

As can be seen from the Table 1, the reliability of item separation was considered excellent for global item separation (α = 0.99), Perceived Usefulness (PU) (α = 0.88), Perceived Ease of Use (PEOU) (α = 0.91), and Attitude Toward Use (ATU) (α = 0.89), whereas Behavioural Intention (BI) was considered as fair (α = 0.73). These values shall be executed along with the dimensionality test to help demonstrate the potential of item and person and place on the latent traits'. The consistency of global scale was categorized as excellent (α = 0.93), yet unfortunately, for the construct PU, PEOU, and BI were examined "not good" (PU with α = 0.77, PEOU with α = 0.75, and BI with α = 0.76). Only ATU with (α = 0.86) was considered as "good" because it scored more than the appropriate value (α = 0.81). Besides item separation index, while the three of four subscales are found below the criteria such as ATU (2.76 logit), ATU (2.91 logit), BI (1.63 logit), the global and one existing subscale (PEOU) were found greater than 3 logits (item separation index > 3) as it scored (global scale = 9.13 logit and PEOU 3.21 logit). The results also provided the person separation index with only global items scored above 3 logits (3,57), while the other constructs such as PU, PEOU, ATU and BI did not match with the criteria because all of them were only producing less than three logits (person separation index > 3).

The Analysis of Dimensionality of the Items

Table 2. Dimensionality

Dimensionality						
Raw variance in data explained by measure (criteria: 20% fair, 40% very good, >60% excellent)	47.7%	NA	41.2%	34.2%	40.0%	50.4%
PCA eigenvalue for first contrast (criteria: >2.0 indicates presence of another dimension, < 2 supports unidimensional scale)	7.4	NA	2.67	3.19	3.31	2.16
Unexplained variance in 1st – 5th contrast of PCA of residuals (criteria: good, 5-10%; very good, 3-5%; excellent, <3%)	8.7%-2.1%	NA	-	17.5%-4.7%	16.6%-3.9%	17.9%-3.7%

Base on Table 2 show items were categorized as dimensional if they are able to measure more than one construct or concept. In Rasch analysis, the assessment of the items' dimensionality was run by monitoring the Principal Component Analysis (PCA) of residuals for the global constructs and each

construct in the instrument (i.e Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude to Use (AU), and Behavioural Intention (BI)). PCA assessment is conducted to identify certain association patterns found among the scales as well as to differentiate the factors that affected the maximum variance in the data.

As shown in Table 2 which illustrated the dimensionality of the items, the outcome of PCA for global scale and all the constructs exceeded the criteria of 20% threshold and were categorised as excellent. The findings revealed the global scale raw variance was 47.7%, and for PU was 41.2%, PEOU 32.2 45.0%, ATU 40.0%, and BI 50.4%. In addition, the PCA eigenvalue for first contrast revealed that all of the items in global scale, PU, PEOU, ATU and BI were indicated as dimensional (Global Subscale = 6.66, PU = 2.67, PEOU = 3.19, ATU = 3.31, and BI = 2.16), as the PCA eigenvalue were all above 2. On the other hand, the residual unexplained variants of PCA in 1st-5th contrast for the global scale and the four subscales were categorised as bad (above 5-10%). These results demonstrated that the TAM adopted questionnaire fits the Rasch model, as it supports the empirical evidence of a dimensionality analysis for the whole and each construct of the items.

Assessment of the Rating Scale

Table 3. Summary of Category Structure of Rating Scales

Rating Scale Step Number	Observed Person	Average Calibration	Outfit MNSQ	Threshold	Threshold Distance
1	824 (9)	-.1.28	1.11	NONE	(-2.48)
2	952 (10)	-0.35	1.04	-0.95	-1.12
3	2358 (27)	0.35	0.93	-0.91	-0.11
4	3051 (32)	0.81	1.16	0.42	1.07
5	2085 (22)	1.34	0.97	1.44	(2.75)

Base on Table 3 show TAM adopted instrument, the questionnaire applied five rating units of Likert scale to measure the participants' perception, with (1) indicates "strongly disagree" to (5) that indicates "strongly agree". Therefore, in the present study, the Rating scale analysis was adopted to examine participants' ability in determining the rating scales used in the questionnaire. As demonstrated in Table 3, the values of an outfit MNSQ for all subscale presented ideal, with each rating unit was less than 2.0, and the adjacent thresholds distance of the rating scale fitted to the criteria of ideal threshold of 1.4–5 logits (i.e. 1.68 logits). Nevertheless, each response rater did not fit the criteria because it did not show a distinct curve that reached the peak < 0.5 logits. Additionally, for the responses per category were seen to also have exceeded the minimum 10 logit.

The result of the Rating scale analysis also showed that the respondents did not feel difficult to distinguish level in each category selected when they were filling out the questionnaire because the average calibration for all subscale was observed to increase monotonically from lowest point (-2.48) to highest (2.75). This finding indicated that the categories supported assessment of latent variables and may suggest accuracy of participants' judgment.

Item and Person Maps

To describe the overall items and person levels, item and person maps (also known as Wright Maps) were conducted to identify the distribution of subjects' perception and the level of difficulty they developed when filling in the questionnaire items. In Figure 1, there are two sides. The boxes on the left side of the map present the distribution of participants' ability level from very high level to very low level. Meanwhile, the right boxes illustrate the difficulties range with the highest level and the lowest level having two meanings. The top box can be interpreted as the most difficult item or that the respondents were very easy to disagree on the rating scale, whereas the bottom part interprets the easy items or items which respondents think were very easy to agree on the Likert scale on the questionnaire. Afterwards, to measure the two different levels between the highest and lowest levels, it can be seen from one standard deviation of the mean of the item.

Based on Figure 1 (N = 210), it was examined that many Indonesian undergraduate students possessed high levels of the Perceived Usefulness, Perceived Ease of Use, Attitude of Use, and Behavioural Intention, and low level of item difficulties. Particularly, most of the answers remained moderate to very low level and with positive affection, where participants showed less trouble in 'choosing agree' and 'strongly agree' to the questionnaire items. However, in terms of Duration subscale, the higher the items placed in the Figure 1, the longer time and frequency of using the three focused applications by participants. In this subscale, the item W1 which asked "how long have you been using WhatsApp" was categorised as a low level item and easily agreed by the participants.

Looking at the details, items G4 and W4 from "PU" scale were found to be low and were categorised as easily agreed among participants, while item W3 from the same construct was considered as a fair item to agree. Most items from all applications in subscale "PEOU" showed to be easy and it had one easiest item to answer by students (i.e W7). It was one standard deviation lower than others in mean item difficulty. However, one item (M7) was still presented as a higher moderate level of item difficulty, but was still categorised as low level as most participants were still able to answer it. All items from "ATU" (i.e M11, M12, G11, G12, G13, and W12) and "BI" (i.e W14 and W15) sub constructs were selected as easy to agree by participants as they were below the mean item difficulty level, while the rest from "ATU" (i.e M10, M13, G18, W10, W11, and W13) and "BI" (M14, M15, G14, and G15) were categorised as moderate. The result stated that the questionnaire items were most likely found to be less difficult in measuring participants' perceived usefulness, perceived ease of use, attitude, and behavioural intention in using Microsoft Teams, Google Classroom and WhatsApp.

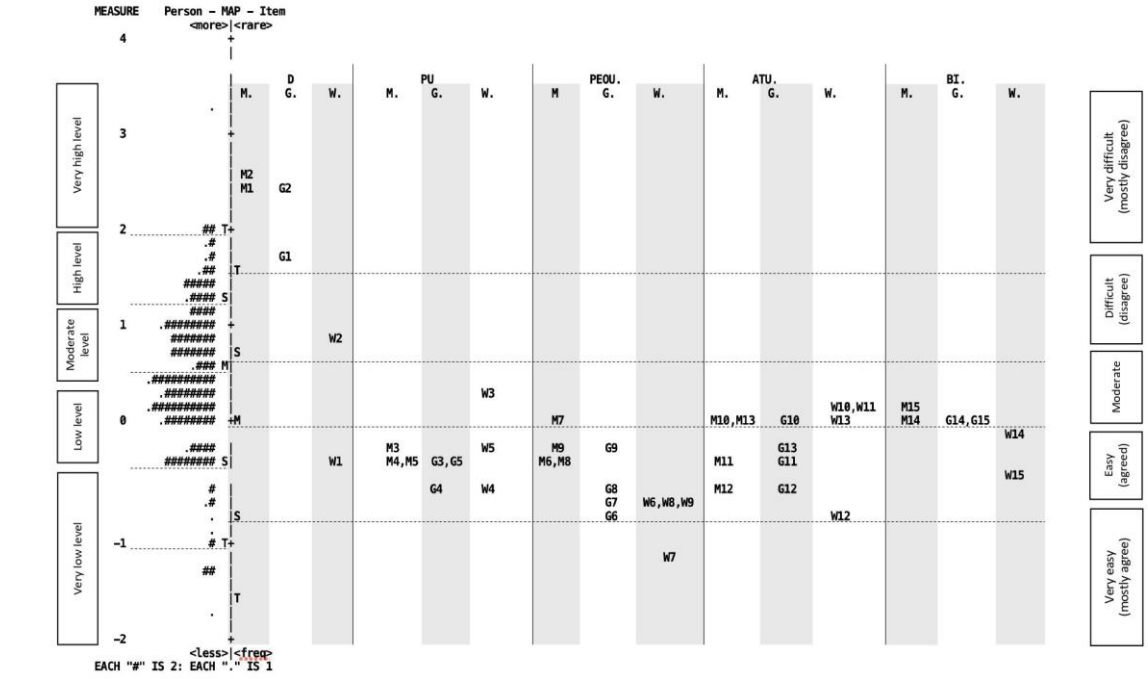

Figure 1. Person Item Mapping

DIF Measure

Base on Table 4, Differential Item Functioning (DIF), or best known as item bias, is one of Rasch assumptions that evaluates items that cause potential result differences when they are given to group characteristics such as gender or educational background. An item is categorized as bias when the score of its DIF contrast is above 0.5 logits and its Rasch-Welch probability value is below 0.05 logits. As presented in the Table 4, there was no differences in terms of gender group as the DIF contrast were all found to score below 0.5 logits, however, some were found in context of "Age", "Year of University", and "Mostly Used Application" groups.

In Age demographic group, item M3 was considered easier for students aged above 22 years old than those under 20 years old (DIF Measure 3 = −0.44 logit, DIF Measure 1 == −0.02 logit, $p < 0.05$). Item W8 (DIF Measure 2 = −0.61 logit, DIF Measure 3 = 0.55 logit, $p < 0.05$), item W11 (DIF Measure 2 = −0.12 logit, DIF Measure 3 = 0.80 logit, $p < 0.05$), and item W13 (DIF Measure 2 = −0.12 logit, DIF Measure 3 = 0.63 logit, $p < 0.05$) were also found to be less difficult for students around 20-22 years but difficult for those who are higher than 22 years old. With regards to the university year, there were three items found to be biased. Participants from 1st year (semester 1 or 2) found item M1 difficult to answer than students from 2nd year (semester 3 or 4) (DIF Measure 1 = 3.05 logit, DIF Measure 2 = −2.22 logit, $p < 0.05$). Item G2 was observed to benefit students from above 4th year than students from 4th year (semester 7 or 8) (DIF Measure 5 = 2.08 logit, DIF Measure 4 = 2.60 logit, $p < 0.05$). Lastly, item W8 was also considered as bias because it was found easier to answer for students from 3rd year, but difficult for students above 4th year (DIF Measure 3 = −1.05 logit, DIF Measure 5 = 0.13 logit, $p < 0.05$).

Table 4. DIF Measure

Item	Age	DIF Measure	DIF Contrast	t	Probability
M3	1 3	-0.02 0.44	0.58	1.76	0.0213
W8	3 2	0.55 -0.61	1.17	3.73	0.0072
W11	3 2	0.80 -0.12	0.67	2.17	0.0334
W13	3 2	0.63 -0.12	0.75	2.43	0.0276
Item	**Year**	**DIF Measure**	**DIF Contrast**	**t**	**Probability**
M1	1 2	3.05 -2.22	0.83	2.87	0.0310
G2	4 5	2.60 2.08	0.52	1.48	0.0357
W8	5 3	0.13 -1.05	1.18	4.16	0.0132
Item	**Mostly Used App**	**DIF Measure**	**DIF Contrast**	**t**	**Probability**
M4	1 3	-0.23 -1.05	0.81	2.90	0.0064
M5	1 3	-0.15 -0.92	0.76	2.80	0.0076
M6	1 3	-0.22 -1.50	1.27	4.10	0.0066
M9	1 3	-0.04 -0.79	0.75	2.83	0.0246
M12	1 3	-0.33 -1.26	0.93	3.16	0.0162
M13	1 3	0.15 -0.57	0.72	2.82	0.0133
G10	1 2	0.09 -0.52	0.61	3.40	0.0169
G11	1 2	-0.10 -0.62	0.53	2.89	0.0064
G13	1 2	-0.07 -0.71	0.64	3.46	0.0007
M2	2 3	2.69 1.52	1.18	4.22	0.0155
M6	2 3	-0.22 -1.50	1.28	3.96	0.0111
M12	2 3	-0.22 -1.26	1.04	3.40	0.0406
W6	2 1	-0.16 -0.86	0.70	3.85	0.0068
G9	3 2	-0.33 -0.60	0.93	3.68	0.0480
G10	3 2	0.50 0.34	1.01	4.06	0.0062
G13	3 1	0.62 -0.07	0.70	3.04	0.0094

As for the assessment of mostly used app, there were several items that found to be easier to fill by students who use Microsoft Teams more often than students who use WhatsApp, such as; item M4 (DIF Measure 3 =−1.05logit, DIF Measure 1 = −0.23 logit, $p < 0.05$), item M5 (DIF Measure 3 =−0.92 logit, DIF Measure 1 = −0.15 logit, $p < 0.05$), M6 (DIF Measure 3 =−1.50 logit, DIF Measure 1 = −0.22 logit, $p < 0.05$), M9 (DIF Measure 3 =−0.79 logit, DIF Measure 1 = −0.04 logit, $p < 0.05$), M12 (DIF Measure 3 =−1.26

logit, DIF Measure 1 = −0.33 logit, $p < 0.05$), and M13 (DIF Measure 3 =−0.57 logit, DIF Measure 1 = 0.15 logit, $p < 0.05$). Additionally, items M2 (DIF Measure 3 =1.52 logit, DIF Measure 2 =2.69 logit, $p < 0.05$), M6 (DIF Measure 3 = −0.22 logit, DIF Measure 2 = −1.50 logit, $p < 0.05$) and M12 (DIF Measure 3 = −1.26 logit, DIF Measure 2 = −0.22 logit, $p < 0.05$) were also discovered to be bias and only advantaged students who use Microsoft Teams more often than students who use Google Classroom.

Furthermore, there were also observed items that benefited students who use Google Classroom more frequently than those who use WhatsApp such as items G10 (DIF Measure 1 = 0.09 logit, DIF Measure 2 =−0.52 logit, $p < 0.05$), G11 (DIF Measure 1 = −0.10 logit, DIF Measure 2 =−0.62 logit, $p < 0.05$), and G13 (DIF Measure 1 =−0.07 logit, DIF Measure 2 =−0.71 logit, $p < 0.05$) and more advantaged students who use Google Classroom often than students who use Microsoft Teams such as items G9 (DIF Measure 3 =−0.33 logit, DIF Measure 2 =−0.60 logit, $p < 0.05$), G10 (DIF Measure 3 =0.50 logit, DIF Measure 2 =0.34 logit, $p < 0.05$), and G13 (DIF Measure 3 =0.62 logit, DIF Measure 2 =−0.07 logit, $p < 0.05$). Lastly, students who use WhatsApp more often also answered item W6 easily than students who use Google Classroom who perceived it as difficult (DIF Measure 1 = −0.86 logit, DIF Measure 2 =−0.16 logit, $p < 0.05$).

The measurement of the dimensionality showed that the PCA of residuals for the global constructs and each construct in the instrument (i.e Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude to Use (AU), and Behavioural Intention (BI)) were all exceeded the minimum 20% of variance, while the PCA eigenvalue were all found to be higher than 2.0 logits, indicating the presence of dimensionality. These results contribute to the literature that the adopted and Indonesian version of TAM adopted questionnaire can be used to measure a single latent variable of students' overall perception of using Microsoft Teams, WhatsApp and Google Classroom as online platforms in EFL classrooms. Furthermore, the item and person separation reliability test discovered that all of the TAM constructs and each of it were reported high.

Discussion

Based on the data analysis, the current findings presented the consistency of the whole items as excellent (α = 0.99), and for each TAM variable item were high, with Perceived Usefulness (α = 0.88), Perceived Ease of Use (α = 0.91), and Attitude Toward Use (α = 0.89). This means that the instrument was able to gather consistent answers from the participants. However, the result of Behavioural Intention (BI) consistency value was considered as fairly reliable (α = 0.73). This result is in line with the previous findings which observed the internal consistency of every construct of TAM including perceived ease of use as highly reliable for the Cronbach alpha value were more than 0.70 (Chan & Subramaniam, 2020; Colledani et al., 2020; Ningsih et al., 2021).

The current study also found that the items were categorized as dimensional, for it was able to measure more than one construct. The PCA eigenvalue for Global Scale, PU, PEOU, ATU and BI were indicated as dimensional (Global Subscale = 6.66, PU = 2.67, PEOU = 3.19, ATU = 3.31, and BI = 2.16). Hence, the present result showed that the items were multidimensional because the first contrast lower than 2, and hence were categorised as one-dimensional (Kant, 2018; Yu, 2020). The following constructs which were found to be dimensional also do not corroborate with the study observed the students' acceptance of mobile assisted language learning (MALL) using the same TAM variables (Hoi & Mu, 2021; Mulyono et al., 2021). In their study, the value for each scale, particularly Perceived Usefulness, Perceived Ease of Use, and Behavioural Intention were found to be unidimensional as the first construct were all lower than 2 (PU 1.55, PEOU 1.85, and BI 1.85).

Differential Item Functioning or also known as items that only benefitting one group among others in the present study were also observed. In reference of age, the usefulness of Microsoft Teams to improve English ability was more significantly perceived by students above 22 years old than students under 20 years old. Students above 22 years old also showed low attitude than students around 22-20 years old in terms of their preference in using WhatsApp and thinking that implementing the application in a classroom is a good day. Unfortunately, this finding contradicts to what have been found by related study which concluded that age has no significant affect in forming participants' attitude of utilizing WhatsApp.

The present findings found that item bias benefitting participants who used certain application were most likely caused by knowledge and confidence they retrieved when using one application more often than the others. For example, students who use Microsoft Teams more frequently perceived more usefulness than students who use WhatsApp more often. This mainly in terms of receiving feedbacks that can be helpful as well as strengthening critical thinking through work peers. This finding supports the finding from related study that Microsoft Teams supports the collaborative knowledge building learning among students. The results of their study demonstrated how students were provided and received feedback and share their thoughts and ideas confidently towards their peers (Hackbarth et al., 2003; Mpungose, 2020). Additionally, students who use Microsoft teams more often perceived high level of ease of use in reference to receiving clear guidance and conversing with instructor and friends as easy and not

stressful. This finding contradicts to the other study observed the same field, revealing that students were struggling in conversing and interacting with their peers and instructors because most of them were neglected by other peers and that they were hesitant to express their thoughts (Buchal & Songsore, 2018; Linh & Ngo, 2021).

For students who use Google classroom more often than those who use WhatsApp frequently showed greater attitude. This specifically apply to context where students like using Google classroom as a tool to learn English, and that they showed positive attitude, and were looking forward to apply it in the future learning. This corresponds to the study evaluating Google classroom, stating that students found the platform as a useful tool because they provides productivity, improve communication among peers and engagement, as well as providing usability (Francom et al., 2021; Iftakhar, 2016). For example, the application allows students to get a reminder about their assignments or tasks through email.

The implications of this study provide an overview related to the application of the Psychometric analysis instrument in evaluating students' acceptance of online platforms to support online English learning which can become reference for educators, especially English teachers in evaluating students' acceptance of online platform. However, during the analysis of this study, several limitations were still encountered. The participants of this study were dominated by female students with only a few numbers of male students participating. Additionally, there were few students who came from an age above 22 years old. Therefore, further investigation adopting the matched setting is highly recommended to accommodate a more balanced distribution in terms of sample groups.

4. CONCLUSION

The adopted and translated Indonesian questionnaire of TAM was found to be sufficient in context of psychometric characteristics. In reference to Wright Maps, the Rasch measurement also examined that the Indonesian undergraduate students possessed a high level of all constructs and they had less difficulty in choosing "agree" or "strongly agree" to the questionnaire items. However, several items were observed to be biased and only offer advantages to students coming from certain groups such as age, year of university, and mostly used application. Moreover, the study also measured the effectiveness of the 5-Likert scale used in the instrument. Based on the finding of the Rating scale analysis, it was found that the outfit MNSQ for each construct as well as the adjacent threshold distance of the rating scale fits to the criteria was appropriate. The average calibration is also found to rise significantly and is signified that the scale categories supported assessment of latent variables and may suggest accuracy of participants' judgment according.

5. REFERENCES

Al-Maroof, R. S., & Salloum, S. A. (2021). *An Integrated Model of Continuous Intention to Use of Google Classroom*. https://doi.org/10.1007/978-3-030-47411-9_18.

Alfadda, H. A., & Mahdi, H. S. (2021). Measuring Students' Use of Zoom Application in Language Course Based on the Technology Acceptance Model (TAM). *Journal of Psycholinguistic Research*, *50*(4). https://doi.org/10.1007/s10936-020-09752-1.

Azman, M., S, K., Puad F, Kamis A, Jerusalem M, & Sari M. (2020). The Usage of E-Learning Platform between MyGuru Malaysia and BeSmart Indonesia towards Perceived Usefulness and Behavioral Intention to Use. *Psychology And Education*, *57*(8), 214–220. https://doi.org/10.21831/jpe.v11i1.55533.

Bhattacherjee, A. (2001). An empirical analysis of the antecedents of electronic commerce service continuance. *Decision Support Systems*, *32*(2), 201–214. https://doi.org/10.1016/S0167-9236(01)00111-7.

Buchal, R., & Songsore, E. (2018). Collaborative Knowledge Building using Microsoft SharePoint. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.v0i0.13043.

Cakir, R., & Solak, E. (2014). The Factors Influencing E-Learning Of Turkish Efl Learners Through TAM. *Turkish Online Journal of Educational Technology*, *13*(3), 79–87. https://eric.ed.gov/?id=EJ1034230.

Chan, M., & Subramaniam, R. (2020). Validation of a Science Concept Inventory by Rasch Analysis. In *Rasch Measurement* (pp. 159–178). Springer Singapore. https://doi.org/10.1007/978-981-15-1800-3_9.

Chen Hsieh, J. S., Wu, W.-C. V., & Marek, M. W. (2017). Using the flipped classroom to enhance EFL learning. *Computer Assisted Language Learning*, *30*(1–2), 1–21. https://doi.org/10.1080/09588221.2015.1111910.

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the

technology acceptance model for e-learning. *Computers & Education*, *63*, 160–175. https://doi.org/10.1016/j.compedu.2012.12.003.

Chung, H.-H., Chen, S.-C., & Kuo, M.-H. (2015). A Study of EFL College Students' Acceptance of Mobile Learning. *Procedia - Social and Behavioral Sciences*, *176*, 333–339. https://doi.org/10.1016/j.sbspro.2015.01.479.

Colledani, D., Anselmi, P., & Robusto, E. (2020). Rasch models in the analysis of repgrid data. *Journal of Constructivist Psychology*, 1–21. https://doi.org/10.1080/10720537.2020.1852461.

Coskun, A., & Marlowe, Z. (2015). Technology in ELT: English Teachers Investigate Animoto and Fotobabble. *International Journal of Higher Education*, *4*(3). https://doi.org/10.5430/ijhe.v4n3p119.

Craig Gamble. (2017). Exploring EFL University Students' Acceptance of E-learning Using TAM. *Kwansei Gakuin University Humanities Review*, *22*, 23–37. https://core.ac.uk/download/pdf/151651395.pdf.

Francom, G. M., Schwan, A., & Nuatomue, J. N. (2021). Comparing Google Classroom and D2L Brightspace Using the Technology Acceptance Model. *TechTrends*, *65*(1), 111–119. https://doi.org/10.1007/s11528-020-00533-0.

Hackbarth, G., Grover, V., & Yi, M. Y. (2003). Computer playfulness and anxiety: positive and negative mediators of the system experience effect on perceived ease of use. *Information & Management*, *40*(3), 221–232. https://doi.org/10.1016/S0378-7206(02)00006-X.

Hoi, V. N., & Mu, G. M. (2021). Perceived teacher support and students' acceptance of mobile-assisted language learning: Evidence from Vietnamese higher education context. *British Journal of Educational Technology*, *52*(2), 879–898. https://doi.org/10.1111/bjet.13044.

Hu, P. J.-H., Clark, T. H. K., & Ma, W. W. (2003). Examining technology acceptance by school teachers: a longitudinal study. *Information & Management*, *41*(2), 227–241. https://doi.org/10.1016/S0378-7206(03)00050-8.

Huang, F., Huang, L., & Oon, P.-T. (2020). Constructs Evaluation of Student Attitudes Toward Science—A Rasch Analysis. In *Rasch Measurement* (pp. 139–157). Springer Singapore. https://doi.org/10.1007/978-981-15-1800-3_8.

Iftakhar, S. (2016). Google Classroom: What Works And How? *Journal of Education and Social Sciences*, *3*(1), 12–18. https://jesoc.com/wp-content/uploads/2016/03/KC3_35.pdf.

Kant, R. (2018). Whatsapp Usage: Attitude And Perceptions Of College Students. *Conflux: Journal Of Education*, *5*(9). https://www.researchgate.net/profile/Ravi-Kant-42/publication/323935784.

Lee, M.-C. (2010). Explaining and predicting users' continuance intention toward e-learning: An extension of the expectation–confirmation model. *Computers & Education*, *54*(2), 506–516. https://doi.org/10.1016/j.compedu.2009.09.002.

Linh, H. G., & Ngo, T. T. C. (2021). Challenges in learning listening comprehension via Microsoft Teams among English majors at Van Lang University. *International Journal of TESOL & Education Home*, *1*(3), 142–175. http://i-jte.org/index.php/journal/article/view/36.

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x.

Mpungose, C. B. (2020). Is Moodle or WhatsApp the preferred e-learning platform at a South African university? First-year students' experiences. *Education and Information Technologies*, *25*(2), 927–941. https://doi.org/10.1007/s10639-019-10005-5.

Mulyono, H., Saskia, R., Arrummaiza, V. S., & Suryoputro, G. (2020). Psychometric assessment of an instrument evaluating the effects of affective variables on students' WTC in face-to-face and digital environment. *Cogent Psychology*, *7*(1). https://doi.org/10.1080/23311908.2020.1823617.

Mulyono, H., Suryoputro, G., & Jamil, S. R. (2021). The application of WhatsApp to support online learning during the COVID-19 pandemic in Indonesia. *Heliyon*, *7*(8). https://doi.org/10.1016/j.heliyon.2021.e07853.

Muntazhimah, Putri, S., & Khusna, H. (2020). Rasch Model untuk Memvalidasi Instrumen Resiliensi Matematis Mahasiswa Calon Guru Matematika. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 65. https://doi.org/10.30998/jkpm.v6i1.8144.

Nazari, N., Nafissi, Z., Estaji, M., & Marandi, S. S. (2019). Evaluating novice and experienced EFL teachers' perceived TPACK for their professional development. *Cogent Education*, *6*(1), 1632010. https://doi.org/10.1080/2331186X.2019.1632010.

Ningsih, S. K., Mulyono, H., Ar Rahmah, R., & Fitriani, N. A. (2021). A Rasch-based validation of EFL teachers' received online social support scale. *Cogent Education*, *8*(1). https://doi.org/10.1080/2331186X.2021.1957529.

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform

during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review*, *119*. https://doi.org/10.1016/j.childyouth.2020.105535.

Rauniar, R., Rawski, G., Yang, J., & Johnson, B. (2014). Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. *Journal of Enterprise Information Management*, *27*(1), 6–30. https://doi.org/10.1108/JEIM-04-2012-0011.

Teo, T., Doleck, T., Bazelais, P., & Lemay, D. J. (2019). Exploring the drivers of technology acceptance: a study of Nepali school students. *Educational Technology Research and Development*, *67*(2), 495–517. https://doi.org/10.1007/s11423-019-09654-7.

Tseng, J.-J. (2018). Exploring TPACK-SLA interface: insights from the computer-enhanced classroom. *Computer Assisted Language Learning*, *31*(4), 390–412. https://doi.org/10.1080/09588221.2017.1412324.

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540.

Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*, *11*(4), 342–365. https://doi.org/10.1287/isre.11.4.342.11872.

Yu, C. H. (2020). Objective measurement: How Rasch modeling can simplify and enhance your assessment. In *Rasch Measurement* (pp. 47–73). Springer Singapore. https://doi.org/10.1007/978-981-15-1800-3_4.