Evaluating Construct Validity and Reliability of Intention to Transfer Training Conduct Instrument Using Rasch Model Analysis

Shaliza Shafie*, Faizah Abd Majid, Teoh Sian Hoon and Siti Maftuhah Damio

Faculty of Education, Universiti Teknologi MARA, Puncak Alam Branch, 42300 Puncak Alam, Shah Alam, Malaysia

ABSTRACT

The impact of the Industry Revolution 4.0 (IR4.0) in the workplace requires organisations to ensure clerical employees can effectively transfer their newly acquired knowledge and skills learned in training back into the workplace. Hence, an instrument is required to identify factors influencing the intention to transfer training conduct amongst clerical employees. Thus, this paper presents the evaluation of construct validity and reliability of the new instrument to confirm its objectivity and clarity in measuring the constructs under study as intended. This four-point Likert-type scale instrument consists of 72 self-assessment items that represent 12 constructs. The Rasch Model was then employed to analyse the construct validity and reliability by evaluating the suitability of items in the respective constructs on the instrument. The item and person reliability and strata indices, point-measure correlation, and outfit mean square values were examined. The analysis found that three constructs in the item and person reliability index and eight constructs in the item and person reliability strata index were low but adequate and met the Rasch Model measurement acceptable level. Meanwhile, point-measure correlation values for all constructs fulfilled the criteria. Finally, the outfit mean square values established that 65 items in the constructs were found to be fit, whereas seven items were misfits which require improvement. Subsequently, the seven misfit items were improved as the item and person reliability values could be increased, thus the items were retained. Thereafter, the instrument was ready to be used for data collection in the actual study.

misfit items were improved as the item and person reliability values could be increased, thus the items were retained. Thereafter, the instrument was ready to be used for data collection in the actual study.

Keywords: Clerical employees, construct validity, intention to transfer training, Rasch Model, research instrument, transfer of training

INTRODUCTION

Transfer of training is generally described as the use of knowledge and skills acquired in employee training into the workplace (Baldwin et al., 2017; Burke et al., 2013; Noe & Kodwani, 2018; Saha, 2021). There is a notion that transfer of training can essentially affect training effectiveness (Al Rakhyoot, 2017; Mohanty et al., 2017; Rahman, 2020; Saha, 2021), which indicates that the lessons learned are effectively applied into the workplace. Since organisations' performance depends heavily on employees' performance, employees are expected to utilise the knowledge and skills learned to improve the job quality and productivity level (Abdullah et al., 2019; Noe & Kodwani, 2018; Nusrat & Sultana, 2019; Yaqub et al., 2021).

To meet such expectations, organisations must ensure knowledge transfer can really occur after training sessions (Anjum et al., 2021). It is due to the fact that training is generally considered a significant human resource practice for organisational growth (Kodwani & Kodwani, 2021). In view of that, it is important for organisations to recognise factors that can influence the intention to transfer training amongst their employees. By recognising factors influencing employees' training transfer conduct, organisations can take appropriate actions to ensure the lessons learned are applied more effectively into the workplace.

Moreover, in the present Industry Revolution 4.0 (IR4.0) working environment, employees at all levels require training to enhance their overall job performance.

According to Kumar (2018), there are many levels of employee groups in organisations where the first level is the individual employees including the clerical employee group. Usually, clerical employees are considered as the backbones in organisations but sometimes this employee group is overlooked in the workplace (Muschara, 2018; Robert, 2017). Like other levels of employee groups, clerical employees must be consistently upskilled and reskilled to be more competent in the workplace. Additionally, the World Economic Forum (WEF, 2018) reported that the role of clerical employees would be critically affected by the impact of the IR4.0. For that reason, organisations need to revisit their employee training initiatives. In order to manage the inevitable situation, clerical employees must be upskilled and reskilled through quality employee training programmes. Hence, it is crucial to ensure that clerical employees apply the lessons learned during training back into the workplace effectively, which in turn can produce quality job performance.

In determining a suitable instrument to collect data for this study, the decision was based on several background elements of the samples, such as job positions, locality, and workplace culture. The samples were clerical employees who were mainly Malaysians, working in a Malaysian Government-Linked Company (GLC). Furthermore, the newly developed instrument must be able to capture responses directly related to intention to transfer training conduct. For that reason, the new instrument was developed instead of adopting and adapting previous instruments which were

consistent with Hsu and Sandford (2012) who suggested that researchers must employ appropriate instruments to collect data for a specific study. Thus, it was determined that the need for a new instrument for this study was appropriate.

Therefore, an instrument was developed to collect the feedback from clerical employees after attending the training sessions as an effort to identify factors influencing their training transfer conduct. Given that it was a newly developed instrument, it must be evaluated to confirm its objectivity and clarity, as well as its validity and reliability to measure what it was supposed to measure. To determine the instrument is valid and reliable in measuring the constructs as intended, face validity, content validity, and construct validity are commonly evaluated (Hamed, 2016; Heale & Twycross, 2015). Five experts from the academic and training industry background assessed the face validity and content validity to evaluate whether the items were placed according to the appropriate constructs.

Besides that, the Rasch Model was employed to determine the validity and reliability of items developed on the instrument whereby items must show the appropriate measure of the respective constructs. The construct evaluation was crucial to ensure the scale developed was acceptable before it could be employed in the actual study. This is consistent with the fact that researchers should ensure the research instrument can collect the required data and is reliable (Hsu & Sandford, 2012; Mohajan, 2017). This paper only focuses on the construct validity and reliability evaluation in the validation stage of the instrument development process to assess the suitability of items in the respective constructs.

The Rationale of the Study

Most organisations employ training as a mean to improve the job performance of their employees (Kodwani & Kodwani, 2021; Noe & Kodwani, 2018; Saha, 2021; Yaqub et al., 2021). In fact, training investment continues despite the effect of economic recession, such as the present recession caused by the Covid-19 pandemic (Nicola et al., 2020; Prohorovs, 2020). More so, in 2020 the WEF reported that 50% of employees would need to be reskilled by 2025 (WEF, 2020). Then again, training can only produce desired results if the expected outcomes can be achieved (Anjum et al., 2021; Baldwin et al., 2017; Yaqub et al., 2021). On the other hand, the low transfer of training by employees has created an issue as employers are dissatisfied with the training investment on human capital (Brion, 2020; Nusrat & Sultana, 2019). Apparently, only about 10% resulted in the transfer of training by employees after returning to the workplace (Brion, 2020; Iqbal & AlSheikh, 2018; Saha, 2021).

As such, the need to identify factors that could influence the intention to transfer training conduct amongst employees is necessary. This is because any actual action should begin with intention as affirmed by past and current studies (Ajzen, 2019; Code, 2020; Kim & Park, 2019). In fact,

studies on intention to transfer training are available in past and current transfer literature (Al Rakhyoot, 2017; Kim & Park, 2019). However, studies on intention to transfer training conducts by clerical employees in Malaysia are scarce. Despite many studies have been conducted on the transfer of training in a Malaysian working environment (Ab Rahman et al., 2019; Abdullah et al., 2019; Kenayathulla et al., 2019), a specific study on intention to transfer training conducts amongst clerical employees in a Malaysian GLC has yet to be performed. This presents a research gap that the study is attempting to undertake.

Thus, the study aims to determine factors that influence the intention to transfer training conduct amongst clerical employees in a Malaysian GLC. Furthermore, clerical employees need to be developed through training as they play an important role in the workplace (Lal & Singh, 2015; Zarreen, 2018), particularly in supporting organisations to mitigate the effects of the IR4.0. This effort could determine factors influencing clerical employees to apply the lessons learned in the training sessions back into the workplace. Determining factors influencing their intention to transfer training enables improvement in their application of the lessons learned after training. The improvement in clerical employees' job performance can in turn improve organisations' performance. Subsequently, a self-developed instrument was used to identify factors influencing intention to transfer training conduct amongst clerical employees.

Importance of Instrument Validation

The assessment of validity and reliability of an instrument is essential for both quantitative and qualitative research as a measure to ensure the instrument employed is valid and reliable for use in research (Hamed, 2016; Zamanzadeh et al., 2015). The reliability and validity of research outcomes essentially depend on the quality of the instrument used to collect and measure the data. Commonly, the items on the instrument must go through an evaluation process to verify the items' clarity, readability, consistency, and relevance to the specific constructs. Among the common types of assessment in determining the instrument validity and reliability are face validity, content validity, and construct validity (Hamed, 2016; Heale & Twycross, 2015).

According to Coates (2018), there are various validation procedures that can be used to substantiate the feasibility of the instrument employed in research. Among the validation, procedures include face validity that is about the language clarity, feasibility, readability, style, and format used (Hamed, 2016). While content validity is to evaluate if the items are sufficient to measure a specific construct on the instrument (Heale & Twycross, 2015). Whereas construct validity is to determine how well an instrument is supposed to measure what it is intended for (Hamed, 2016; Heale & Twycross, 2015). For that purpose, the Rasch Model was utilised to analyse data collected from the pilot study. In brief, the evaluation of validity and

reliability performed on this new instrument was to ascertain that the items were relevant and adequate to confirm their credibility.

METHODS

This study employed a quantitative approach using a self-developed online survey questionnaire to collect data from clerical employees in a Malaysian GLC. According to Toepoel (2017), quantitative web surveys are appropriate to collect responses from large groups of people and to generalise the results. In view of that, the newly developed instrument needs to be validated to determine its validity and reliability prior to actual study usage. To determine the sample size for the pilot study, a minimum of 30 participants as per a rule of thumb in education and behavioural sciences (Suter, 2014; Whitehead et al., 2016) was adhered to.

In relation to the pilot study, the samples were first selected according to a set of criteria fitting the aim of the study; minimum one-year employment and attended at least one soft skills training. Then a simple random technique was utilised to generate the sampling frame. From the sampling frame produced, 100 samples were invited to participate in the pilot survey of which 31 (31%) participants responded. As for the data collection process involved, an online survey approach was employed to collect the feedback. A link to the survey form was emailed to the selected samples for their responses. Each completed survey form was automatically reverted to the responses folder for analysis. Thereafter, the

data collected were assessed in determining the items developed pertaining to factors influencing intention to transfer training conduct could essentially measure as intended.

Scale Development

Generally, in developing the instrument, the items were formed in line with the operational definitions of the constructs and a suitable number of items were considered. Also, a common rule in formulating items was followed as the questions should be clear to enable participants to understand and answer appropriately (Brinkman, 2009; Tsang et al., 2017). As such, the generated items must be evaluated for clarity, readability, consistency, redundancy, and relevance to the construct under study. An evaluation of the instrument is to indicate that the items are relevant and comprehensive enough to confirm their credibility (Hamed, 2016; Sangoseni, Hellman & Hill, 2013; Zamanzadeh et al., 2015). On that note, the validation process of the instrument was required as it was crucial to ascertain the scale developed was acceptable before it could be tested. Prior to the construct validity evaluation, face validity and content validity of the instrument were performed.

In doing so, a panel of five experts which consisted of two academicians and three training practitioners were involved in the scale development process. The experts were selected based on their strong academic background, extensive knowledge, and familiarity with the concepts, as well as

years of working experience in the respective fields. In the process, the experts commented that a few items on the instrument needed to be rephrased in order to refine the items according to the respective constructs. Consequently, the items were rephrased and maintained on the instrument for testing in the pilot study. In short, all items were acceptable and no items were removed by the experts after the review.

Briefly, the feedback from the experts was analysed to evaluate the face validity and content validity of the instrument. For face validity, Fleiss' Kappa (Fleiss' K) was used due to its suitability for multiple raters and flexibility (Falotico & Quatto, 2015). Fleiss' K revealed the result of 0.658 that suggested the instrument as good and significant at $p = .000 < .005$. Meanwhile, for content validity, a Content Validity Index (CVI) was used where the result indicated that the instrument had an excellent CVI proportion agreement of 0.975. Thus, the items were considered relevant to the constructs being studied and ready to be tested in determining the construct validity.

Survey Instrument

In recent years, the Likert-type scale is the most popular scale format employed by researchers (Joye et al., 2017). Likewise, this study employed the Likert-type scale format to obtain feedback for analysis from the respondents of the survey. The questions prepared on the survey instrument are in the first-person form point of view where respondents are requested to select their agreement to the items using the four-point Likert-type scale categories, namely 1 = strongly disagree to 4 = strongly agree. This newly developed instrument consists of 72 self-assessment items that represent 12 constructs. The instrument comprises of ten independent variables: Personal Intentions – PI (6 items), Personal Expectations – PE (6 items), Training Awareness – TA (4 items), Training Contents – TC (6 items), Training Delivery – TD (6 items), Post-Training Interventions – PTI (6 items), Trainer's Quality – TQ (6 items), HOD's Support – HS (6 items), Workplace Support – HS (6 items), and Workplace Climate – WC (5 items). Whilst the two dependent variables are Self-Efficacy – SEf (7 items) and Self-Motivation – SM (8 items).

For the instrument to be considered valid, it must also undergo construct validity evaluation. The instrument in effect depends on the strength of the instrument to accurately assess the construct being measured, that is the measurement precision of the identified variables by the instrument (Azwani et al., 2016). Therefore, in evaluating the construct validity of the instrument, the Rasch Model with the application of Winsteps 3.72.3 (Linacre, 2011) was utilised to analyse the data collected from the respondents. All the required Rasch analyses performed on the new instrument were measured against the acceptable quality benchmarks. Nevertheless, the benchmarks only provide some indication on the quality of data but by no mean the absolute criterion (Bond & Fox, 2015; Boone, 2016; Boone et al., 2014).

Employing Rasch Model Analysis

There are several approaches to construct an instrument where the Rasch Model measurement is one of the approaches used to construct a scale in a research instrument (Bond & Fox, 2015; Iramaneerat et al., 2011). The Rasch Model has been widely used in various aspects of research to analyse questionnaires and construct validity and reliability (Fitkov-Norris & Yeghiazarian, 2015). Besides that, the model enables an instrument to exclude a middle category for the Likert-type survey responses (Bradley et al., 2015) which seemed to be suitable for this study as the instrument employed was constructed without a middle category.

Moreover, the ability of respondents to answer the questions and the difficulty of each item are usually considered as meeting the construct validity and reliability criteria (Bradley et al., 2015; Norasmah et al., 2014). Additionally, Rasch analysis can identify items' positions in the model, whereby items closer to the hypothetical line indicates a contribution to the construct (Baghaei, 2008; Bond & Fox, 2015). Furthermore, the items can be utilised as an empirical test for construct validity since the items only measure one latent trait which indicates that items on the instrument fit the model (Boone, 2016; Sick, 2011).

However, if the differences are reasonably acceptable, the data can be considered as fitting the model because there are no perfect results (Bond & Fox, 2015; Runnels, 2012). In brief, the item reliability and strata indices indicate the extent to which the items conform with the

Rasch Model measurement. Typically, in evaluating the instrument reliability, item and person reliability and strata indices are analysed. While for construct validity, item polarity is examined by point-measure correlation (PTMEA Corr.), and items fit in the constructs are examined by Outfit MNSQ values (Bond & Fox, 2015). In general, the Rasch quality criteria used for determining the validity and reliability of the instrument are shown in Table 1.

RESULTS

After running the collected data using the Rasch Model analysis, the results for the item and person reliability and strata indices, the PTMEA correlation, and Outfit MNSQ values were generated. Thereafter, the results were analysed to determine whether the items on the instrument sufficiently fulfilled the construct validity and reliability requirement. In the event the items do not meet the requirement, appropriate actions need to be taken such as item modification, replacement, or deletion (Bond & Fox, 2015; Boone, 2016).

Reliability and Separation Index of Constructs

The Rasch Model analysis assessed 72 items on the new instrument against the acceptable quality criteria. To evaluate the construct validity of this new instrument, the item reliability and separation (strata) index and person reliability and separation (strata) index were analysed to confirm the instrument validity and reliability. Based on the criteria, the item and person reliability

Table 1
Summary of Rasch Model quality criteria

Criterion	Values	Denotations	References
Person and Item Reliability	< 0.67	Poor	Boone (2016); Linacre and Fisher (2012); Linacre (2005, cited in Norasmah et al., 2014)
	0.67- 0.80	Fair	
	0.81- 0.90	Good	
	0.91- 0.94	Very Good	
	> 0.94	Excellent	
Person and Item Strata	<0.5	Less productive for measurement, but not degrading.	Bond and Fox (2015); Linacre (2005, cited in Norasmah et al., 2014)
	0.5 - 1.5	Productive for measurement.	
	1.5 - 2.0	Unproductive for construction of measurement, but not degrading.	
	>2.0	Distorts or degrades the measurement system.	
Point Measure Correlation (PTMEA Corr.)	> 0.3	Acceptable	Bond and Fox (2015); Bond and Fox (2007, cited in Norasmah et al., 2014); Boone (2016); Boone et al. (2014)
Outfit Mean Square (MNSQ)	0.5 < x <1.5	Acceptable	
z-Standardised Value (ZSTD)	-2. 0< x <2.0	Acceptable	

values < 0.67 are considered poor (Boone, 2016), whilst for the item and person strata values > 2.0 indicate the instrument is good (Bond & Fox, 2015). Table 2 below displays the results of the item and person reliability and separation indices generated by Rasch analysis.

Referring to the table, the results of the item reliability values for constructs PE, PI, TA, TD, PTI, TQ, HS, WC, SEf, and SM were in the range of fair to very good but two constructs TC (0.62) and WS (0.59) scored poor item reliability. Whilst the person reliability values for PE, TA, TD, PTI, TQ, HS, WC, WS, SEf, and SM were in the range of fair to good but one construct PI showed poor person reliability of 0.62. Normally, low item reliability is caused by insufficient person sample size to confirm item difficulty hierarchy, whereas low person reliability is caused by the small number of items (Boone & Noltemeyer, 2017; Linacre & Fisher, 2012). Although the values for TC and WS (item reliability) and PI (person reliability) did not conform to high-reliability values (> 0.67), the values were adequate and met the acceptable level as the reliability values can be increased when the misfit items are improved or removed (Boone & Noltemeyer, 2017; Pallant, 2011).

The results in Table 2 also show the item and person strata values. In general, the value of the item strata refers to the level of item difficulties, while the person strata is used to classify people's responses to the questions on the instrument. Norasmah

Table 2
Reliability and strata result for constructs

Construct	Total Item (72)	Item		Person	
		Reliability	Strata	Reliability	Strata
Personal Expectations (PE)	6	0.73	1.66	0.77	1.78
Personal Intentions (PI)	6	0.92	3.43	0.62	1.34
Training Awareness (TA)	4	0.79	1.92	0.85	2.26
Training Contents (TC)	6	0.62	1.37	0.75	1.58
Training Delivery (TD)	6	0.80	1.99	0.70	1.54
Post-Training Interventions (PTI)	6	0.68	1.45	0.80	2.08
Trainer's Quality (TQ)	6	0.83	2.24	0.71	1.59
HOD's Support (HS)	6	0.72	1.60	0.81	2.05
Workplace Support (WS)	6	0.59	1.22	0.74	1.68
Workplace Climate (WC)	5	0.88	2.72	0.77	1.75
Self-Efficacy (SEf)	7	0.71	1.58	0.67	1.45
Self-Motivation (SM)	8	0.89	2.91	0.86	2.50

et al. (2014) cited that Linacre (2005) recommended value > 2.0 as good. Ideally, the higher value of the strata indices of the items indicates the instrument is better as the items are separated by levels of difficulty.

The overall results in the table exhibit that the item strata values for the 12 constructs were between the values of 1.22 to 3.43 and the person strata values ranged from 1.34 to 2.50. Statistically, the items on the instrument can be separated into three strata or levels of difficulty and two groups of people. In short, the results indicated that the items and person strata respectively displayed 8 constructs scored strata values < 2.0. Usually, low item separation implies that the person sample size is not sufficient to confirm the item difficulty hierarchy, whereas low person strata is due to the small number of items on the instrument (Boone & Noltemeyer, 2017; Md Yunus et al., 2017). Likewise, the item and person separation

indices can increase if the reliability of items is enhanced by improving or deleting misfit items (Boone & Noltemeyer, 2017; Pallant, 2011).

Item Polarity of Constructs

In the Rasch Model measurement, the validity of items in a polarity item report exhibits if all items move in one direction in a construct. The main output is referred to as a correlation coefficient of measurement point which is known as the point-measure correlation coefficient. PTMEA correlation near zero or negative can mean the items are not consistent with the construct (Bond & Fox, 2015), indicating the items in the construct are not aligned with other items to measure that particular construct (Md Yunus et al., 2017). If the PTMEA correlation value is high, it implies that the items are able to distinguish between respondents' ability to answer the questions, whereas the PTMEA

correlation value < 0.30 denotes that the items do not fulfill the criteria (Bond & Fox, 2015).

Table 3 displays a summary of the PTMEA correlation for 72 items on the instrument. The results showed PTMEA correlation index on the instrument displays > 0.30 with no item nearly zero or negative index with the minimum PTMEA correlation index of 0.54 for item SEf7 and maximum index 0.96 for item WS6. Thus, the results suggested that the items could contribute to the validity measurement of the instrument.

Item Fit in Measuring Constructs

Further analysis was conducted to validate the appropriateness of items in the constructs, that is to determine how well the data fit the model. The analysis technique employed in the study to evaluate item fit was outfit means square (Outfit MNSQ) as used by Boone et al. (2014). This was consistent with Linacre (2012), who highlighted that item fit analysis only needs to report on the outfit unless irrelevant outliers contaminated

the data severely. Additionally, Rasch analysis calculates ZSTD (z-standardised) which measures the probability of MNSQ occurring by chance (Boone et al., 2014). However, in evaluating the item fit, it is recommended to examine MNSQ first then followed by ZSTD as the ZSTD value is based on MNSQ (Boone et al., 2014; Linacre & Fisher, 2012).

If the Outfit MNSQ value is higher than 1.5, it indicates that the item is not consistent with other items in the same construct. Modification or elimination of the items needs to be made to the items that fall out of the quality criteria acceptable range (Boone et al., 2014). This is because problematic items can affect the reliability and validity of the instrument. Nevertheless, Bond and Fox (2015) advised that items should not simply be dropped as such action could cause good items to be lost. Based on the quality criteria, seven items that did not fit the Rasch Model measurement were identified as shown in Table 4. Since the misfit items fulfilled the requirement of the

Table 3
PTMEA correlation results for constructs

Construct/ No. of Items	PE	PI	TA	TC	TD	PTI	TQ	HS	WS	WC	SEf	SM
1	0.92	0.79	0.93	0.89	0.68	0.85	0.79	0.86	0.90	0.95	0.63	0.55
2	0.93	0.69	0.85	0.91	0.67	0.76	0.89	0.83	0.90	0.95	0.79	0.76
3	0.91	0.72	0.95	0.93	0.95	0.82	0.87	0.90	0.93	0.90	0.85	0.90
4	0.83	0.83	0.81	0.93	0.83	0.86	0.79	0.90	0.90	0.87	0.67	0.58
5	0.83	0.75		0.91	0.88	0.86	0.73	0.85	0.85	0.86	0.79	0.75
6	0.81	0.80		0.91	0.68	0.89	0.70	0.81	0.96		0.75	0.89
7											0.54	0.88
8												0.90

Table 4
Misfit items

No	Item	Outfit MNSQ	Outfit ZSTD	Construct
1	SM1	3.22	5.1	Self-Motivation (SM)
2	SM2	2.44	3.7	Self-Motivation (SM)
3	SEf3	2.25	2.3	Self-Efficacy (SEf)
4	PI2	3.54	4.9	Personal Intentions (PI)
5	PTI3	2.07	3.1	Post-Training Interventions (PTI)
6	HS5	1.90	2.4	HOD's Support (HS)
7	HS6	2.47	3.8	HOD's Support (HS)

relevant constructs, improving the items can increase the reliability and strata values of the items in the constructs, thus the items were examined to modify the problems and retained.

DISCUSSION

The new instrument was developed to gather data in determining factors influencing intention to transfer training conduct amongst clerical employees. Then the items on the instrument must be validated to ensure they could measure as intended. The evaluation could indicate whether the items are relevant to confirm the instrument's credibility which is crucial to produce valid and reliable outcomes (Hamed, 2016; Zamanzadeh et al., 2015). The Rasch Model measurement was employed to verify the validity and reliability of items on the newly developed instrument to ensure the suitability of items in the respective constructs. Essentially, the Rasch Model analysed the item reliability and strata index for the instrument reliability and examined the point-measure correlation (PTMEA Corr.) and Outfit MNSQ values for the construct validity.

From the analysis on the item reliability and strata index for the instrument reliability, all items and persons reliability generally achieved high-reliability values (> 0.67). However, item reliability values for TC (0.62) and WS (0.59) and person reliability for PI (0.62) did not conform to high-reliability values. Whereas for the item and person strata indices, eight constructs showed a low index (< 2.0). The low values in the item reliability and strata index were due to insufficient person sample size to confirm item difficulty and the low values in the person reliability and strata index were due to the small number of items on the instrument. Nevertheless, the values are sufficient in meeting the acceptable level as both the item reliability and item strata index values can be increased once the misfit items are improved (Boone & Noltemeyer, 2017; Pallant, 2011).

Meanwhile, for the construct validity, Rasch analysis generated positive PTMEA Corr. values for all the 72 items in the constructs (> 0.30) with no item nearly zero or negative index. The results implied that all the items moved in the same direction of the measurement scale and correlated

with the other constructs on the instrument (Bond & Fox, 2015). To sum up, the findings implied that the items are able to distinguish between respondents' ability to answer the questions which fulfilled the criteria.

As for the item fit analysis, only Outfit MNSQ was analysed as there were no irrelevant outliers that contaminated the data. The Outfit MNSQ results exhibited that a total of 65 items were found to be fit and seven items (SM1, SM2, SEf3, PI2, PTI3, HS5, and HS6) in the constructs were misfit which, hence, require improvement. Since the misfit items are relevant to the constructs assessed, the seven items are retained on the instrument after the items were checked for weaknesses and modified.

CONCLUSION

The study intends to determine factors influencing intention to transfer training conduct amongst clerical employees. Using this instrument, the data collected in the study could provide a better insight into factors influencing their intention to transfer training conduct into the workplace. As a result, the application of lessons learned in training into the workplace could be improved. This effort enables organisations to make required improvements if needed so that clerical employees could apply the lessons learned more effectively after training. Accordingly, clerical employees could improve their job performance which, to an extent, could support organisations in mitigating and navigating the IR4.0 challenges.

For that purpose, the new instrument was developed for data collection whereby the instrument must undergo the validity and reliability test. The test was required to ascertain the instrument could measure the constructs as intended. Employing the Rasch Model, the analysis was performed to determine the validity and reliability of the items in the constructs. Thereafter, the analysis produced a valid and reliable instrument to evaluate relevant constructs under study. Now, the instrument is ready to be utilised to investigate factors influencing intention to transfer training conduct amongst clerical employees in a Malaysian GLC.

ACKNOWLEDGEMENT

REFERENCES

Ab Rahman, F., Mohamed, A. H., Saidin, K., & Nasir, N. S. A. (2019). Exploring perceptions of employers on communication skills among fresh graduates. *Practitioner Research, 1,* 69-85.

Abdullah, A. R., Muhammad, M. Z., & Md Nasir, N. A. (2019). The role of soft skills on business graduates employability. *Journal of Entrepreneurship and Business, 7*(2), 83-94. https://doi.org/10.17687/JEB.0702.07

Ajzen, I. (2019). *Behavioral interventions based on the theory of planned behavior.* https://people.umass.edu/aizen/pdf/tpb.intervention.pdf

Al Rakhyoot, A. A. M. (2017). *Conceptualising the antecedents of employee intentions to transfer training to the job: An application of the Theory of Planned Behaviour in health education institutions* [Doctoral dissertation, University of Strathclyde]. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.725157

Anjum, R., Khan, H. H., Bano, S., Nazir, S., Gulraiz, H., & Ahmed, W. (2021). A latent factor (PLS-SEM) approach: Assessing the determinants of effective knowledge transfer. *Journal of Asian Finance, Economics and Business, 8*(2), 851-860. https://doi.org/10.13106/jafeb.2021.vol8.no2.0851

Azwani, M., Nor'ain, M. T., & Shah, S. N. (2016). Evaluating the face and content validity of a Teaching and Learning Guiding Principles Instrument (TLGPI): A perspective study of Malaysia teacher educators. *Malaysian Journal of Society and Space, 12*(3), 11-21.

Baghaei, P. (2008). The Rasch Model as a construct validity tool. *Rasch Measurement Transactions American Educational Research Association, 22*(1), 1145-1146.

Baldwin, T., Ford, J. K., & Blume, B. D. (2017). The state of transfer of training research: Moving toward more consumer-centric inquiry. *Human Resource Development Quarterly, 28*(1), 17-28. https://doi.org/10.1002/hrdq.21278.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental measurement in the human sciences* (3rd ed.). Routledge Taylor & Francis Group.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education, 15*(4), 1-7.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences* (e-book). Springer.

Boone, W. J. & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education, 4*(1), 1-13.

Bradley, K. D., Peabody, M. R., Akers, K. S., & Knutson, N. M. (2015). Rating scales in survey research: Using the Rasch Model to illustrate the neutral middle category measurement flaw. *Survey Practice, 8*(1), 1-12.

Brinkman, W.-P. (2009). Design of a questionnaire instrument. In L. Steve (Ed.), *Handbook of mobile technology research methods* (pp. 31-57). Nova Publisher.

Brion, C. (2020). The role of culture in the transfer of training. *International Journal of Training and Development, 24*(4), 384-393.

Burke, L. A., Hutchins, H. M., & Saks, A. M. (2013). Best practices in training transfer. In M. A. Paulidi (Ed.), *Psychology for business success* (Vol. 3, pp. 115-132). Praeger.

Coates, H. (2018). Instrumentation. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (p. 839). SAGE Publications, Inc. http://dx.doi.org/10.4135/9781506326139.n335

Code, J. (2020). Agency for learning: Intention, motivation, self-efficacy and self-regulation. *Frontiers in Education, 5*(19), 1-15. https://doi.org/10.3389/feduc.2020.00019

Falotico, R., & Quatto, P. (2015). Fleiss' Kappa statistic without paradoxes. *Quality Quantitative, 49,* 463-470.

Fitkov-Norris, E. D., & Yeghiazarian, A. (2015). Validation of VARK learning modalities questionnaire using Rasch analysis. *Open Access. Journal of Physics: Conference Series, 588,* 012048. https://doi.org/10.1088/1742-6596/588/1/012048

Hamed, T. (2016). Validity and reliability of the research instrument; How to test the validation

of a questionnaire/survey in a research. *International Journal of Academic Research in Management, 5*(3), 28-36.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative research. *Evidence-Based Nursing, 18*(3), 66-67.

Hsu, C., & Sandford, B. A. (2012). Instrumentation. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 608-610). SAGE Publications, Inc. http://dx.doi.org/10.4135/9781412961288.n189

Iqbal, M. Z., & AlSheikh, M. H. (2018). Factors affecting the transfer of training to the workplace after a faculty development programme: What do trainers think? *Journal of Taibah University Medical Sciences*, *13*(6), 552-556.

Iramaneerat, C. Smith Jr., E. V., & Smith, R. M. (2011). An introduction to Rasch measurement. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 50-70). SAGE Publications, Inc. http://dx.doi.org/10.4135/9781412995627.d6

Joye, D., Wolf, C., Smith, T. W., & Fu, Y.-C. (2017). *Handbook of survey methodology.* SAGE Publications Ltd. http://dx.doi.org/10.4135/9781473957893

Kenayathulla, H. B., Ahmad, N. A., & Idris, A. R. (2019). Gaps between competence and importance of employability skills: Evidence from Malaysia. *Higher Education Evaluation and Development, 13*(2), 97-112. https://doi.org/10.1108/HEED-08-2019-0039

Kim, E.-J., & Park, S. (2019). Support, training readiness and learning motivation in determining intention to transfer. *European Journal of Training and Development, 43*(3/4), 306-321. https://doi.org/10.1108/EJTD-08-2018-0075

Kodwani, A. D., & Kodwani, M. (2021). Exploring determinants of pre-training motivation and training effectiveness: A temporal investigation. *Evidence-based HRM*, ahead-of-print. https://doi.org/10.1108/EBHRM-05-2020-0070

Kumar, S. (2018). Understanding different issues of unit of analysis in a business research. *Journal of General Management Research*, *5*(2), 70-82.

Lal, R. S., & Singh, A. P. (2015). Role of personality factors in clerical employees' satisfaction with special reference to Banaras Hindu University. *South Asian Journal of Marketing & Management Research, 5*(10), 10-19.

Linacre, J. M. (2011). *Winsteps* (Version 3.72.3). http://www.winsteps.com.

Linacre, J. M. (2012). A user's guide to Winsteps-Ministep: Rasch Model computer programs. Program manual 3.68.0. http://www.winsteps.com/manuals.htm

Linacre, J. M., & Fisher, W. P., Jr. (2012). Harvey Goldstein's objections to Rasch measurement: A response from Linacre and Fisher. *Rasch Measurement Transactions, 26*(3), 1383-1389.

Md Yunus, J., Alias, M., Mukhtar, M. I., Lee, M. F., Tee, T. K., Rubani, S. N. K., Nur Yunus, F. A., Hamid, H., Sulaiman, J., & Sumarwati, S. (2017). Validity of vocational pedagogy constructs using the Rasch measurement model. *Journal of Technical Education and Training, 9*(2), 35-45.

Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University, 17*(3), 56-82.

Mohanty, P. C., Dash, M., & Dash, M. (2017). Impact of trainee characteristics and organizational climate on training effectiveness with special reference to financial organization in India. *International Journal of Applied Business and Economic Research, 15*(18), 233-241.

Muschara, T. (2018). *Risked-based thinking: Managing the uncertainty of human error in operations.* Routledge, Taylor & Francis Group.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R.

(2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery, 78,* 185-193.

Noe, R. A., & Kodwani, A. D. (2018). *Employee training and development* (7th ed.). McGraw-Hill Education India Private Limited.

Norasmah, O., Suria, M. S., Haliza, H., & Haryaty, A. W. (2014). *Assessing construct validity and reliability of competitiveness scale using Rasch Model approach.* The 2014 WEI International Academic Conference Proceedings, Bali, Indonesia. https://www.westeastinstitute.com/wp-content/uploads/2014/06/Suria-Mohd-Salleh.pdf

Nusrat, M., & Sultana, N. (2019). Soft skills for sustainable employment of business graduates of Bangladesh. *Journal Higher Education, Skills and Work-Based Learning, 9*(3), 264-278.

Pallant, K. (2011). Transfer of training: A review and direction for future research. *Personnel Psychology, 41*(1), 63-105.

Prohorovs, A. (2020, May 7). Getting ready for recession: How SMEs prepared for Covid-19 pandemic and economic recession. *Forbes (Latvian edition),* (11), 46-49. http://dx.doi.org/10.2139/ssrn.3596611

Rahman, A. A. (2020). Tracing the evolution of transfer of training: A review article. *Annals of Social Sciences & Management Studies, 5*(4), 71-79.

Robert, C. E. (2017). *Clerical turnover in the K-12 campus office* [Doctoral dissertation, The University of Texas at San Antonio]. ProQuest Dissertations Publishing.

Runnels, J. (2012). Using the Rasch Model to validate a multiple choice English achievement test. *International Journal of Language Studies, 6*(4), 141-153.

Saha, M. (2021). "I will be aware of the appraisal, which will automatically change my work behaviour": Evaluating training transfer. *Journal of Teaching and Teacher Education, 9*(1), 31-39. https://dx.doi.org/10.12785/jtte/090104

Sangoseni, O., Hellman, M., & Hill, C. (2013). Development and validation of a questionnaire to assess the effect of online learning on behaviors, attitudes, and clinical practices of physical therapists in the United States regarding evidence-based clinical practice. *The Internet Journal of Allied Health Sciences and Practice, 11*(2), 1-13.

Sick, J. (2011, March). Rasch measurement in language education Part 6: Rasch measurement and factor analysis. *SHIKEN: JALT Testing & Evaluation SIG Newsletter, 15*(1), 15-17.

Suter, W. N. (2014). *Introduction to educational research: A critical thinking approach; sampling in research.* SAGE Publications, Inc. https://dx.doi.org/10.4135/9781483384443

Toepoel, V. (2017). *Doing surveys online.* SAGE Publications Ltd. https://dx.doi.org/10.4135/9781473967243

Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia, 11*(1), 80-89.

Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2016). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research, 25*(3), 1057-1073.

World Economic Forum. (2018). *The future of jobs.* http://www.weforum.org/report/the-future-of-jobs-report-2018

World Economic Forum. (2020). *The future of jobs.* http://www.weforum.org/report/the-future-of-jobs-report-2020

Yaqub, Y., Singh, A. K., & Dutta, T. (2021). An empirical study of factors influencing training transfer in the management training intervention. *Journal of Workplace Learning,* Ahead-of-print. https://doi.org/10.1108/JWL-02-2020-0034

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., AlaviMajd, H., Nikanfar, A-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences, 4*(2), 165-178.

Zarreen, Z. (2018). Impact of psychological capital on organizational citizenship behaviour: A study of clerical employees. *Asian Journal of Multidimensional Research, 7*(11), 142-146.