Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 204 (2015) 212 – 217

Procedia
Social and Behavioral Sciences

4th World Congress on Technical and Vocational Education and Training (WoCTVET), 5th–6th November 2014, Malaysia

Validity and Reliability Learning Transfer Item Using Rasch Measurement Model

Ruhizan Mohd. Yasin[a], Faizal Amin Nur Yunus[a,b*], Ridzwan Che Rus[c], Azmi Ahmad[d], and Mohd Bekri Rahim[b]

[a]*Faculty of Education, The National University of Malaysia, Selangor, Malaysia*
[B]*Faculty Of Technical And Vocational Education, Tun Hussein Onn University, Johor, Malaysia*
[C]*Faculty Of Technical And Vocational Education, Sultan Idris Education University, Perak, Malaysia*
[D]*Centre For Instructor And Advanced Skill Training (Ciast) Shah Alam, Malaysia*

Abstract

This study was conducted to produce an empirical evidence of validity and reliability of the item using a survey questionnaire of learning transfer. Rasch model approach with aided by Winsteps software Version 3.69.1.11 was used to examine the functional items from the reliability and separation of item and respondent, polarity and items fit measuring constructs and standardized residual correlation value. The questionnaire was distributed to 40 trainees from various disciplines at the Centre for Instructor and Advanced Skill Training (CIAST) Shah Alam. The final analysis found that a total of 16 items were eliminated and revealed 81 items that are suitable to measure the five constructs of learning transfer.

Peer-review under responsibility of Faculty of Technical and Vocational Education, University of Tun Hussein Onn Malaysia.

Keywords: Learning transfer; validity; reliability; Rasch Measurement Model

1. Introduction

Rasch measurement model has proven that learning transfer questionnaire has a level of validity and reliability then be used to develop a model of learning transfer. This is because the use of Rasch measurement model is a solution to the issue of validity as Rasch measurement model provides useful statistics and offers a tremendous opportunity to probe the validity (Bond & Fox, 2007). In addition, the application of Rasch measurement model in a

Peer-review under responsibility of Faculty of Technical and Vocational Education, University of Tun Hussein Onn Malaysia.
doi:10.1016/j.sbspro.2015.08.143

study will be able to facilitate and produce a more efficient, reliable and valid measurement while increasing convenience to user (Abdul Aziz et al. 2007). A study to identify the validity and reliability of the instrument is very important for maintaining the accuracy of the questionnaire (Ariffin et al. 2010). This is necessary to determine the questionnaire to measure what is to be measured consistently and accurately. According to Howard and Henry (1988) consistency means that when the same item is tested several time on the same subject at different time interval, the score result given are approximately the same. In conclusion, the reliability is likely to provide a consistent validity.

The purpose of this study is to identify the items of learning transfer constructs for skills training based on NOSS system. This study was performed to produce empirical evidence of the validity and reliability of learning transfer questionnaire using Rasch measurement model. This is because through the Rasch measurement model, it can test the consistency of interpretation of constructs, the reliability of the items and the respondent and the accuracy of the test. Besides that, review the appropriateness of the scale of measurement was also emphasized in the framework of Rasch measurement model. However, for this study, the objective is to; i) test the reliability and item separation index and the respondent; ii) detecting polarity items that measure the constructs; iii) to test the item fit of the instrument items and, iv) determine the item depends on the correlation of the standardized residuals for the items on the transfer of learning.

2. Methodology

This study employs a survey by distributing questionnaires developed as a result of qualitative findings by adapting some transfer models such as Baldwin & Ford Model [5],the conceptual framework of Holton et al. (2000) and learning space factors by Illeris (2004) and Oblinger et al. (2006). The questionnaire consists of 97 items five-point likert scale that measures five major constructs of trainee characteristics, training design, work environment, virtual learning spaces, and learning transfer. Questionnaires were distributed to 40 trainees Malaysian Skills Certificate program of the Centre for Instructor and Advanced Skill Training (CIAST) Shah Alam. The number of respondents in the study was adequate because according to Cooper and Schindler (2011), the number of respondents who suitable to the pilot study is between 25 to 100 people. Whereas Johanson and Brooks (2010) suggest the minimum number of 30 people for a pilot study in which the aim is to study early or development scale. Data were analyzed with the aid of software Winsteps Version 3.69.1.11.

3. Findings

Through Rasch measurement model approach, the researchers perform examination the item functional in terms of (i) item reliability and separation of the respondents; (ii) detecting polarity items that measure the constructs based on the PTMEA CORR; (iii) items fit measuring constructs and (iv) determine the item depends on the correlation of the standardized residuals

3.1. Reliability and Separation Items and Respondent

Based on Rasch measurement model approach, the acceptable reliability Cronbach's Alpha (α) is between 0.71-0.99 where it is at the best level (71% - 99%) [6]. The findings of the pilot study found that the reliability obtained based on the Cronbach Alpha (α) is 0.96. So this value shows instruments used are in very good condition and effectively with a high level of consistency thus can be used in the actual research.

Analysis also performed on the instrument as a whole, namely the reliability and the separation of the item and the respondent. Table 1.1 shows the reliability and separation items where the reliability of the items was 0.74, while the separation of items is 1.68 when rounded off is equal to 2.0. Based on the reliability of the items, the value of 0.74 indicates are in good condition and acceptable (Bond & Ford, 2007). While the separation of the item is 1.68 if rounded off is equal to 2.0 and this value can still be used because it shows that the entire item is divided into 2

levels of measurement. According to Linacre (2003), the separation index is better when the value is more than the value of 2.0.

While the reliability of the respondents is 0.94 and the separation of the respondents is 4:13.This shows that the respondents are very high reliability and very good. This is because Bond and Fox (2007) describes the reliability of more than 0.8 is good and stronger acceptable. While the separation of the respondents showed a good separation of the item difficulty level appropriate to the Linacre (2005), which describes the separation of more than 2.0 is a good value.

Table 1.1: Reliability and Separation Item and Respondent for the Entire Construct Instruments: Pilot Study

	Item	Respondent
Separation	1.68	4.13
Reliability	0.74	0.94

3.2. Polarity Item By PTMEA CORR Value

Examination of the Point Measure Correlation (CORR PTMEA) to detect polarity items intended to test the extent to which the construction of constructs to achieve its goal. If the value contained in the PTMEA CORR is the positive (+), it shows the item measure the constructs to be measured (Bond & Ford, 2007). Other hand if value is negative (-), the item is not developed to measure the constructs to be measured. Thus it needs to be improved or dropped because the item is not lead to the question (not focus) or difficult to answer by the respondent. Based on Table 1.2, there are four items that have a negative value in the PTMEA CORR of PK64, PK65, PK66, and PK67. The rest PTMEA CORR. are positive which indicates that the items measuring the constructs to be measured (Bond & Ford 2007). Whereas the negative PTMEA CORR. indicates item should repaired or removed. Thus four items were dropped from the entire 89-item questionnaire.

The rest of PTMEA CORR. is positive despite the lowest positive value of the item PK68 (0.01) and PK63 (0.02). This value should be considered as probable items tend to be difficult be answered by the respondent (Azman Hasan 2011). Thus purification items should be done. However, based on these findings show that positive items moving in one direction with construct and able to measure constructs and does not conflict with the constructs being measured. If the PTMEA CORR. is high, then the item is able to distinguish between respondents capability.

Table 1.2: A Part of Point Measure Correlation Value

Entry Number	Point Measure Corr.	Item	Entry Number	Point Measure Corr.	Item
67	-.27	PK67	68	.01	PK68
65	-.16	PK65	63	.02	PK63
66	-.06	PK66	26	.07	RB26
64	-.01	PK64	53	.19	PK53
61	.27	PK61	27	.27	RB25
38	.29	RB38	58	.29	PK58
52	.33	PK52	18	.34	CP18
27	.36	RB27	55	.40	PK55
32	.43	RB32	70	.43	PK70
75	.43	RP75	50	.45	PK50
31	.45	RB31	74	.46	RP74
73	.49	RP73	69	.35	PK69
57	.19	PK57	59	.41	PK59
49	.20	PK49	56	.43	PK56
62	.25	PK62	37	.45	RB37
6	.27	CP6	51	.47	PK51
28	.28	RB28	11	.33	CP11

3.3. Item Fit Measure Constructs

Items fit is measuring the constructs that can be seen through the infit and outfit Mean Square (MNSQ). According to Bond and Fox (2007), the outfit and infit MNSQ should be in the range of 0.6 to 1.4 to ensure the items are suitable for measuring the constructs. But the outfit index MNSQ noteworthy in advance compared infit MNSQ for determining congruity of items that measure a construct or latent variable (Kashfi Mohd. Jailani 2011). If the infit or outfit MNSQ value more than 1.4 logit, then it gives meaning confusing item. If the MNSQ value is less than 0.6 logit, it shows that the item is too easily anticipated by the respondents (Linacre 2007). Beside that the outfit and infit ZSTD value should also be within -2 to +2 (Bond & Fox, 2007). But if the outfit and infit MNSQ be accepted, the ZSTD index can be ignored (Linacre 2007).

Therefore, if this condition is not met, then the item can be considered to be removed or having purified. Table 1.3 below shows the misfit oder featuring 14 items having the largest MNSQ and 6 items of value resulting from the smallest MNSQ item analysis statistics: misfit oder. Based on Table 1.3, found at least 20 items that are not in the specified range and it should be purified or dropped. Items that exceed the value of 1:40 in column outfit MNSQ is CP6(1.51), RB26(1.86), RB27(1.46), RB28(1.75), RB38(1.72), PK49(2.42), PK58(2.11), PK62(1.51), PK63(2.81), PK64(2.67), PK65(3.57), PK66(2.52), PK67(2.65), and PK68(1.65). Whereas a value less than 0.6 are item CP1(0.50), CP4(0.53), RB34(0.54), RB35(0.50), RB44(0.52), and PK60(0.53). Thus from this diagnosis, there were 15 items dropped while 5 items having purified by looking at the needs of researchers and expertise.

Table 1.3: Item Fit Based on MNSQ Value

Entry Number	Infit		Outfit		Items
	MNSQ	ZSTD	MNSQ	ZSTD	
65	3.28	6.8	3.57	7.2	PK65
63	2.49	4.9	2.81	5.6	PK63
64	2.36	4.7	2.67	5.4	PK64
67	2.47	5.2	2.65	5.5	PK67
66	2.16	4.2	2.52	5.0	PK66
49	1.80	2.8	2.42	4.5	PK49
58	1.88	3.1	2.11	3.8	PK58
26	1.68	2.6	1.86	3.1	RB26
28	1.62	2.3	1.75	2.8	RB28
38	1.66	2.4	1.72	2.6	RB38
68	1.48	1.8	1.65	2.5	PK68
6	1.33	1.5	1.51	2.1	CP6
62	1.24	1.0	1.51	1.9	PK62
27	1.37	1.5	1.46	1.9	RB27
34	.51	-2.4	.54	-2.3	RB34
4	.54	-2.3	.53	-2.4	CP4
60	.50	-2.5	.53	-2.4	PK60
44	.52	-2.4	.53	-2.4	RB44
1	.48	-2.6	.50	-2.6	CP1
35	.47	-2.7	.50	-2.6	RB35

3.4. Measurement Standardized Residual Correlations Value

The measurement of the correlation of the standardized residuals are able to detect local dependence whether items are dependent or not with other items. Local dependence may occur if there is a high positive correlation values. According to Linacre (2010), if the correlation of the two items above 0.7, it indicates that the items are inter-dependent and singular. Thus Linacre (2010) suggest that only one item is selected to be used in the measurement. So to produce a good-quality instrument, an item should be dropped. Retention items refer to the MNSQ value, where a value close to 1.00 will be retained (Linacre 2010).

Based on Table 1.4 below, there are 10 pairs of items that have a high correlation like correlation value of 0.85 between item PK64 with PK66, and item PK64 with PK65, on correlation value 0.82 between PK65 with PK67, on correlation value 0.79 between PK65 with PK66, on correlation value 0.75 between PK66 with PK67, on correlation value 0.74 between RP84 with RP85 and RP82 with RP84, on correlation value 0.72 between PK64 with PK67, on correlation value0.70 between PK63 with PK65 and on correlation value 0.69 between RB25 with RB26. This

means that these items have the same measurement meaning or incorporate other dimensions of shared. Therefore, these items need to be addressed and have dropped one of the items for each pair of items involved.

Table 1.4: Standardized Residual correlation Largest item

Correlation	Entry Number	MNSQ Outfit	Result	Entry Number	MNSQ Outfit	Result
.85	PK64	2.67	remove	PK66	2.52	remove
.85	PK64	2.67	remove	PK65	3.57	remove
.82	PK65	3.57	remove	PK67	2.65	remove
.79	PK65	3.57	remove	PK66	2.52	remove
.75	PK66	2.52	remove	PK67	2.65	remove
.74	RP4	0.90	remove	RP85	1.01	retained
.74	RP82	1.05	retained	RP84	0.90	remove
.72	PK64	2.67	remove	PK67	2.65	remove
.70	PK63	2.81	remove	PK65	3.57	remove
.69	RB25	1.26	retained	RB26	1.86	remove

If referred to MNSQ value such items are involved, then the item should be removed only RB26, PK63, PK64, PK65, PK66, PK67, and RP84. Selection of items to be removed should also be aligned with the items removed based on negative PT MEASURE CORR. value as discussed on the previous analysis. Yet MNSQ value closest to 1:00 has been retained which are item RB25, RP82, and RP85.

4. Discussion

After data analysis, following the standard index and the conditions that must be followed to achieve the standards of validity and reliability of the instrument based on the Rasch measurement model does revision of each item. The removal and the purification items done by referring and consider the views and expert evaluation.

Based on the results obtained, there are 16 items that do not meet the requirements analysis and should be removed. Whereas 8 items having purified appropriate to the context and significance of the study. Overall summary related items in the questionnaire are shown in Table 1.5 below.

Table 1.5: Summary of Items Dropped and Retained

Construct	Retained Item	Total Items Retained	Item Dropped	Total Item Dropped
B. Trainee characteristic (CP)	B2, B3, B4, B5, B7, B8, B9, B10, B11,B12, B13, B14, B15, B16, B17, B18, B19,B20.	21	B1, B6	2
C. Training Design (RB)	B21, B22, B23, C24, C25, C29, C30, C31, C32, C33, C34, , C36, C37, C39, C40, C41, C42, C43, C45, C46, C47, C48	19	C26, C27, C28, C35, C38, C44	6
D. Work Environment (PK)	D50, D51, D52, D53, D54, D55, D56,D57, D58, D59, D61,D62, D68, D69, D70, D71, D72,	17	D49, D60, D63, D64, D65, D66, D67	7
E. Learning Spaces (RP)	E73, E74, E75, E76, E77, E78, E79, E80, E81, E82, E83, E85	12	E84.	1
F. Learning Transfer (PP)	F86,F87, F88, F89, F90, F91, F92, F93, F94, F95, F96, F97	12	None	0
TOTAL	**81**			**16**

Based on this research, it can be concluded that the validity and reliability are very important aspect to consider in developing a new instrument for a study. Overall from this analysis found 16 items were dropped are questionable items on validity and reliability.

By using the reliability items and respondent test indicate that the set of questionnaires is valid and reliable to measure learning transfer. Thus, there is no mismatch of items and respondent (over 50% fit) found during the process of data analysis. This is because the advantage of using Rasch model measurement is the ability to identify the fitness of items and respondents. According to Bond and Fox (2007) this method can identify the difficulty level of items and the ability of the respondents. Then the problematic questionnaires items can be improved or removed to ensure that it measure the constructs. Thus the result obtained related to the construct reliability and validity of this questionnaire is acceptable to answer the research question.

5. Conclusion

Thus based on the examination of the validity and reliability of these instruments, these instruments indicate the quality of the fit to be used by trainees. The implications of this analysis help researchers in developing learning transfer model for skills training based on NOSS system and learning in the workplace. This is the first step in ensuring transfer of learning can take place effectively thus helping all parties involved in producing highly skilled workers.

References

Abdul Aziz, A., Arshad, A., Zakaria & N, Masodi, S. (2007). Appraisal of Course Learning Outcomes Using Rasch Model: A Case Study in Information Technology Education. *International Journal of Systems Applications, Engineering and Development*,Vol. 1.(4).

Ariffin, S. R., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability multiple intelligent item using rasch measurement model. *Procedia-Social and Behavioral Sciences*, Vol. 9, pp.729-733.

Azman Hasan (2011). Kesahan Dan Kebolehpercayaan Item Penilaian Pembimbing Dalam Pembelajaran Berasaskan Kerja (PBK) Menggunakan Model Pengukuran Rasch. USM, *Psychometrics Centre, MIMOS & Malaysian Examination Syndicate*, MOE.

Azmi Ahmad, Wan Mohd Rashid Wan Ahmad, Zakaria Kasa, Gessler, M., & Spöttll, G. (2010).Effectiveness of learning transfer in the National Dual Training System (NDTS) in Malaysia. Paper presented at the Persidangan Kebangsaan Pendidikan Kejuruteraan dan Keusahawanan 2010, Kota Bahru, Kelantan.

Baldwin, T. T., & Ford, J.K. (1988). Transfer of Training: A Review and Direction for Future Research. *Personnel Psychology*, Vol. 41(1), pp. 63-105.

Bond, Trevor G, & Fox, Christine M. (2007). Applying the Rasch Model: Fundamental Measurement in the Human Sciences.

Cooper, D. R., & Schindler, P. S. (2011). Business research methods (11th ed.). New York: McGraw-Hill/Irwin.

Holton, E. F., Bates, R. A. & Ruona, W. E. A. (2000). Development of a Generalized Learning Transfer System Inventory. *Human Resource Development Quarterly*Vol. 11(4).pp.333-360.

Howard, W. & Henry, I.B. (1988). Test Validity. New Jersey: Lawrence Erlbaum Associates Publishers.

Illeris, K. (2004). *Adult Education and Adult Learning*. Roskilde University Press.

Johanson, G.A., & Brooks, G.P. (2010). Initial scale development: sample size for pilot studies. *Educational And Psychological Measurement*, Vol.70(3), pp. 394-400.

Kashfi Mohd Jailani (2011). Manual Pengenalan Pengukuran Rasch & Winstep. Pengukuran dan Penilaian dalam Pendidikan. Fakulti Pendidikan Universiti Kebangsaan Malaysia

Linacre, J. M. (2005). Test validity, and Rasch measurement: Construct, content, etc. Rasch measurement transactions.

Linacre, J. M. . (2007). A user's guide to WINDTEPS Rasch-model computer programs. Chicago, Illinois: MESA Press.

Linacre, J.M. (2010). User's guide to Winsteps Ministep Rasch-Model Computer Programs.Avialable at: http://www.winsteps.com/winman/ (Accessed 15 March 2011).

Oblinger, D. G., Lomas, C. & Educause. (2006). Student Practices and Their Impact on Learning Spaces.