The Social Sciences 13 (4): 787-791, 2018
ISSN: 1818-5800

Measuring Validity and Reliability of Collaborative Learning Rubric using Rasch Model

[1]Sharifah Nadiyah Razali, [1]Faaizah Shahbodin, [1]Norasiken Bakar and [2]Hanipah Hussin
[1]Innovative and Sustainable Education (ISTE), C-ACT,
Faculty of Information and Communication Technology,
[2]Centre of Language and Human Development, Universiti Teknikal Malaysia,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

Abstract: Collaborative learning has proven in promoting soft skills development and has been widely implemented in teaching and learning. However, this study addressed the lack of soft skills issue among Malaysia polytechnic graduates causing the graduate to face unemployment. It shows that collaboration does not happen naturally in a group. In previous studies, Online Project Based Collaborative Learning (OPBCL) was developed based on proposed model in order to enhanced student soft skills. A research testing instrument called Collaborative Learning Rubric (CLR) developed in order to evaluate the effectiveness of OPBCL. The main quality indicators for any of testing instrument in research are the validity and reliability. Therefore, this study aims to determine the validity and reliability of CLR. A number of 32 (N = 32) diploma hotel catering students from Politeknik Ibrahim Sultan, Johor (PIS) participated in this study. Data obtained was analysed using WINSTEP Version 3.68 Software. The finding showed that CLR had high reliability with two categories of difficulties items. So, it can be concluded that CLR is reliable and strongly accepted. All items will remain after Rasch analysis. It hoped that this study will give emphasis to other researchers about the importance of analysing items to ensure the quality of an instrument developed.

Key words: Soft skills, collaborative learning rubric, instrument validity, instrument reliability, Rasch Model, PIS

INTRODUCTION

Having excellent academic skills still does not guarantee a graduate person gets a job due to the fierce competition in the career market today. Now, the academic achievement is not the primary criteria for getting a job, but most employers are looking for good soft skills as the selection criteria for choosing the employee. Universities around the world will be increasingly needed to produce highly skilled graduates that meet the needs of employers (Andrews and Higson, 2008). Many published reports have outlined the advantages of collaborative learning suggesting that it improves academic performance, promotes soft skills development (i.e., communications, collaboration, problem-solving and critical thinking skills) and increases satisfaction in the learning experience (Kabilan *et al*., 2011; Lee and Lim, 2012; Shukor *et al*., 2012; Zhu, 2012). But it is difficult for educators to monitor and evaluate student's participation in group project (Zhang, 2012). Therefore, certain strategies must be applied to monitor the learning process. In a previous study by Nadiyah and Faaizah (2015) Online Project Based Collaborative Learning (OPBCL) prototype was developed based on the model proposed by Razali *et al*. (2015) in order to enhance students soft skills. Therefore, research testing instrument must be developed in order to evaluate the effectiveness of OPBCL. Research testing instrument plays an important role in collecting data to answer the research questions that have been set. A research testing instrument called Collaborative Learning Rubric (CLR) developed in order to evaluate the effectiveness of OPBCL. The main quality indicators for any of testing instrument in research are the reliability and validity (Kimberlin and Winterstein, 2008). It is important to get the validity and reliability of an instrument before the actual study carried out to ensure the smooth process of the actual study. Therefore, this study aims to determine the validity and reliability of CLR.

Validity is the extent to which a research testing instrument measures what it is supposed to measure.

Corresponding Author: Sharifah Nadiyah Razali, Innovative and Sustainable Education (ISTE), C-ACT,
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Melaka, Malaysia

Table 1: Cronbach alpha interpretation score

Cronbach alpha score	Value interpreted
0.8-1.0	High reliability
0.7-0.8	Good reliability
0.6-0.7	Fair reliability
<0.6	Poor reliability

Zikmund and Babin (2009)

Therefore, good conclusions can be made from the sample of study (Creswell, 2005). According to Gay and Airasian (2003), review and validation of the expert are needed to ensure that the instrument can meet the objectives that have set. Reliability is the extent to which a research testing instrument can be expected to obtain consistent results when repeated. In this study, expert validation was conducted before performed a pilot study.

Three experts performed two types of validation which are face and content validation; coordinator at UTeM, a lecturer who teach soft skills subject at Politeknik Ibrahim Sultan and Nutrition subject Lecturer at Politeknik Merlimau Melaka. Face validation aims to check in terms of the language used and the presentation of the overall layout of instruments. For content validation, Aiken (2003) stated that it intended to examine the extent to which the ability of a measuring instrument to measure what should be measured.

Then, a pilot study was conducted to test the reliability of the instrument used. Therefore, Rasch Model approach was implemented to check the reliability of the instruments used. Applications of Rasch Model can produce an instrument that is reliable and valid (Abdul, 2010). In determining the validity and reliability of the instrument, the item functionality checks performed on the reliability and separation item respondents, polarities item, the compatibility (fit) item, the value of standardized residual correlation in determining learning item and the distribution of item difficulty levels and abilities respondents. Cronbach alpha was used to measure the reliability of each testing instrument used in this study. Cronbach alpha value was interpreted based on Table 1.

MATERIALS AND METHODS

The study conducted in the form of a descriptive survey study. According to Cohen and Manion in Ghaffar (1999), the survey is to take the data at a certain time, often using questionnaires. Therefore, researchers choose to distribute a set of questionnaire to each respondent to obtain feedback easily.

CLR was developed to measure learner's core soft skills. The development of the instrument is based on the steps used in Jamil study which are: identify constructs and elements based on document analysis, expert validation on the constructs and elements that have been previously identified, item development expert validation of the developed instrument and pilot test run. A number of 32 (N = 32) diploma hotel catering students from Politeknik Ibrahim Sultan, Johor (PIS) participated in this purposive study. Data obtained was analysed using WINSTEP Version 3.68 Software.

Table 2: Questionnaire content and number of items in section B

Core soft skills	Elements
Critical thinking and problem solving	CTPS1
	CTPS 2
	CTPS3
Collaboration	CL1
	CL2
Communication	CM1
	CM2
	CM3

Table 3: Level of skills score

Score	Level of skills
1	Incompetent
2	Developing
3	Competent
4	Exemplary

Instruments: In creating a rubric for Project Based Learning, Ministry of Higher Education has outlined the steps in the project-based learning handbook which are confined to listing the criteria used in assessing and determining learning outcomes, determine performance levels, write a description for each performance level and use, evaluate and revise the rubric. The instrument was developed based on the core soft skills constructs of critical thinking and problem solving skill, collaboration skill and communication skill that have been determining before. The rubric was adapted and modified based on a developed rubric by polytechnic and was refined using the core soft skills set by MOHE (Table 2). Student soft skills will be evaluated using 4 points of Likert scale which are 1 = Incompetent, 2 = Developing, 3 = Competent, 4 = Exemplary (Table 3).

RESULTS AND DISCUSSION

The development of Collaborative Learning Rubric (CLR) used in this study was developed using 4 Likert scales. Results from the analysis of CLR showed that the value of Cronbach alpha (α) is 0.88 as shown in Table 4. According to Zikmund and Babin (2009), the Cronbach alpha value of the CLR instruments are very good and effective with a high level of consistency and can be used for the real study. Besides that, person reliability value is 0.85, indicating a high reliability with a 2.40 separation index (Table 4) that showed two categories of difficulties items. Bond and Fox (2007) described the reliability of more than 0.8 as very good and strongly acceptable. Meanwhile, Linacre (2011) stated that the separation of

more than two is good value. The value indicates that the item has high reliability and two categories of difficulties items are detected.

Analysis of the polarities item is intended to test the extent at which the developed construct has achieved its goals and the relationship between the developed items and the respondents. Based on the analysis, the PT-measure corr. showed no negative value items. Therefore, no items should be dropped or defined (Table 5). According to Bond and Fox (2007), to determine whether the item measures the constructs, the value is shown on the PT-measure corr. must be in the positive (+). If the value obtained is negative (-), it means that the developed item does not measure the construct and it should be dropped or refined because it is difficult or not leading to the questions (out of focus). The findings indicate that the items that produced can measure any item to be measured and its move parallel with other items that measure the construct.

Analyses of item fit refer to the value documented in the infit and outfit Mean Square (MNSQ). Observations on the value of the index are required to determine whether the item developed is appropriate (item fit) to measure a latent variable or construct. Based on the Bond and Fox (2007), to determine the suitability item built, the infit and outfit MNSQ should be in the range between 0.6-1.4. According to Gilani (2011), outfit MNSQ should be given more emphasis than infit MNSQ in determining congruity items that measure constructs. If the result indicated value over 1.4 logit, it means that the item is confusing, meanwhile if the result indicated value below 0.6 logit implies that the item is easily expected by the respondents (Linacre, 2007). In addition, the value of infit and outfit ZSTD should be within -2 to +2 (Bond and Fox, 2007). But if the value of infit and outfit MNSQ is accepted then the ZSTD index can be ignored (Linacre, 2007). The findings indicated that all items were in the range of between 0.6-1.4. Therefore, no items should be dropped or defined (Table 6).

Standardised residual correlation measurement value is to determine whether there are items that overlap. High residual correlations for the two items showed that the item is not independent, either because the item has the same characteristics or both the dimensions incorporate several others characteristics. According to Linacre (2011) if the correlation value is above 0.7, it is considered high correlation value therefore only one item is to be maintained while the other items are to be dropped or defined. The analysis indicated that all the items were below 0.7 which means that there was no overlap with the detected item or items are detected not to have similar characteristics to each other (Table 7). Therefore, no items should be dropped or defined.

Person Item Distribution Map (PIDM) provided an illustration of the items or respondents map that indicates whether the instrument produced in accordance with the respondent's ability to agree. According to Rashid *et al*. (2008), PIDM is concerned on the person's ability on the latent trait is responded to item difficulty. Person distribution which shown on the right side of

Table 4: Reliability

Cronbach alpha (α)	Reliability	Separation
0.88	0.85	2.40

Table 5: Item polarity

Items	Entry number	Point measure corr
CM1	6	0.65
CTPS1	1	0.69
CTPS2	2	0.69
CL1	4	0.69
CL2	5	0.72
CM3	8	0.79
CM2	7	0.80
CTPS3	2	0.81

Table 6: Item fit

	Infit		Outfit		
Items	MNSQ	ZSTD	MNSQ	ZSTD	Entry number
CL2	0.99	0.1	1.35	0.8	5
CM3	1.14	0.6	1.17	0.6	8
CL1	1.14	0.6	1.11	0.4	4
CM1	1.02	0.2	1.09	0.4	6
CTPS2	0.91	-0.2	1.04	0.2	2
CM2	0.99	0.0	0.99	0.1	7
CTPS3	0.92	-0.2	0.74	-0.6	3
CTPS1	0.80	-0.7	0.67	-0.8	1

Table 7: Standardized residual correlations

Entry number	MNSQ qutfit	Results	Correlation	Entry number	MNSQ qutfit	Results
7	0.99	Retained	0.41	8	1.17	Retained
2	1.04	Retained	-0.43	7	0.99	Retained
3	0.74	Retained	-0.43	7	0.99	Retained
1	0.67	Retained	-0.36	7	0.99	Retained
3	0.74	Retained	-0.35	6	1.09	Retained
1	0.67	Retained	-0.35	4	1.11	Retained
1	0.67	Retained	-0.33	8	1.17	Retained
5	1.35	Retained	-0.33	8	1.17	Retained
4	1.11	Retained	-0.32	8	1.17	Retained
3	0.74	Retained	-0.29	5	1.35	Retained

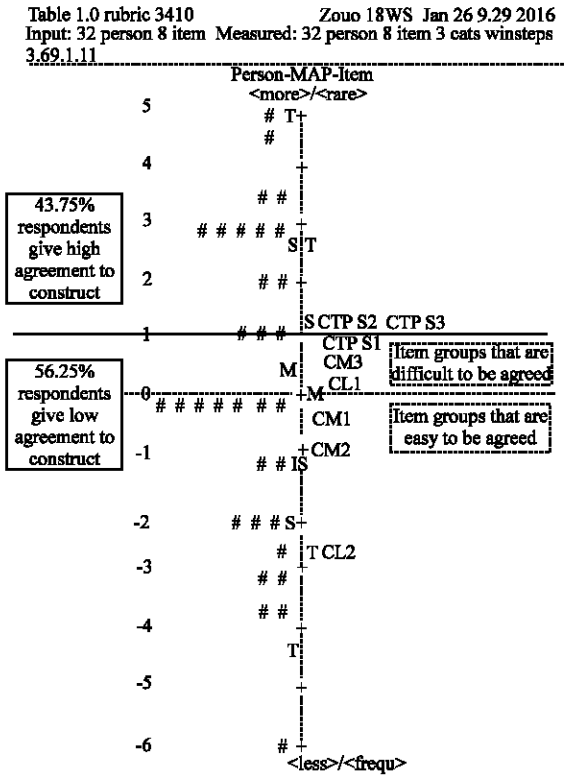

Fig. 1: Distribution of difficulty level

Fig. 1 indicates that 43.75% of the respondents reflect a high agreement to construct and 56.25% respondents give low agreement to the construct. The distributions of the person are indicative the fact that there is difficulty in agreeing upon the item. Meanwhile, item distribution shown on the left side indicates the level of item difficulty. Meanitem acts as the threshold and is set to be zero on the logit scale. The item at a higher location than Meanitem showed a higher level of difficulty compared to he item at a lower location. Item CL2 is the easiest items meanwhile, item CTPS2 and CTPS3 are the most difficult items in this study. The distributions of the items indicated that the level of difficulty of the item is nearly balanced.

CONCLUSION

The findings showed that the, items in the CLR instruments had a high reliability with two categories of difficulties items. So, it can be concluded that CLR instrument is reliable and strongly accepted. The findings on items polarity also indicate that the items that produced in the instrument are able to measure the items that specifically need to be measured and that item are parallel with other items that measure the construct. The analysis on item fit indicated that all items were in the range of between 0.6-1.4. Therefore, no items should be dropped or defined. Moreover, analysis of standard residual correlations showed that there are no items that have similar characteristics. Finally, the distributions of the items indicated that the level of difficulty of the items is nearly balanced and that there are difficult and easy items to be agreed. Therefore, all the items in the CLR instruments will remain as they are after Rasch analysis.

In determining the quality of an instrument developed, the best method used by most researchers is analysing items (Hanafi *et al*., 2014). In this study, Rasch measurement model was used to analyse each item in Collaborative Learning Rubric (CLR). Application of Rasch Model in the instrument can determine the construct validity of items and gave a clear definition of constructs that can be measured are consistent with theoretical expectations. It hoped that this study will give emphasis to other researchers about the importance of analysing items to ensure the quality of an instrument developed.

ACKNOWLEDGEMENTS

Sharifah Nadiyah Razali would like to acknowledge the financial support of Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education of Malaysia for her PhD study.

REFERENCES

Abdul, A.A., 2010. Rasch Model Fundamentals: Scale Construct and Measurement Structure. Sdn Bhd, Kuala Lumpur, Malaysia,.

Aiken, S., 2003. Estimation of Item Parameters. In: Rasch Models, Foudations, Recent Developments and Applications, Fisher, G.H. and I.W. Molenaar (Eds.). Springer, New York, USA., pp: 39-52.

Andrews, J. and H. Higson, 2008. Graduate employability, soft skills versus hard business knowledge: A European study. Higher Educ. Eur., 33: 411-422.

Bond, T.G. and C.M. Fox, 2007. Applying the Rasch Model. Fundamental Measurement in the Human Sciences. 2nd Edn., Lawrence Erlbaum Associates Publishers, New Jersey, ISBN: 0805854622.

Creswell, J.W., 2002. Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research. Merrill Prentice Hall, New Jersey, USA., ISBN-13: 9780137905027, Pages: 671.

Gay, D. and Airasian, 2003. The Rating Scale Model. In: Handbook of Modern Item Response Theory, Linden, W.J. and R.K. Hambleton (Eds.). Springer, New York, USA., pp: 67-148.

Ghaffar, M.N.A., 1999. Education Research. Penerbit Utm Press, Skudai, Malaysia,.

Gilani, K.M., 2011. Manual introduction rasch measurement and winstep: Measurement and evaluation in education. Master Thesis, Faculty of Education, National University of Malaysia, Bangi, Malaysia.

Hanafi, N.M., R.A. Ab, M.I. Mukhtar, J. Ahmad and S. Warman, 2014. Validity and reliability of Competency Assessment Implementation (CAI) instrument using rasch model. World Acad. Sci. Eng. Technol. Int. J. Social, Behav. Educ. Econ. Bus. Ind. Eng., 8: 162-167.

Kabilan, M.K., W.F.W. Adlina and M.A. Embi, 2011. Online collaboration of english language teachers for meaningful professional development experiences. English Teaching: Pract. Critique, 10: 94-115.

Kimberlin, C.L. and A.G. Winterstein, 2008. Validity and reliability of measurement instruments used in research. Am. J. Health-Syst. Pharm., 65: 2276-2284.

Lee, H.J. and C. Lim, 2012. Peer evaluation in blended team project-based learning: What do students find important?. Educ. Technol. Society, 15: 214-224.

Linacre, J.M., 2007. A Users Guide to Windteps Rasch Model Computer Programs. Mesa Press, Chicago, Illinois,.

Linacre, J.M., 2011. Winsteps Rasch Measurement Computer Program Users Guide. Beaverton Publisher, Beaverton, Oregon,.

Nadiyah, R.S. and S. Faaizah, 2015. The development of online project based collaborative learning using ADDIE Model. Procedia Social Behav. Sci., 195: 1803-1812.

Rashid, R.A., R. Abdullah, H.A. Ghulman and M.S. Masodi, 2008. Application of Rasch-based ESPEGS model in measuring generic skills of engineering students: A new paradigm. WSEAS Trans. Adv. Eng. Educ., 5: 591-602.

Razali, S.N., F. Shahbodin, H. Hussin and N. Bakar, 2015. Factors Affecting the Effective Online Collaborative Learning Environment. In: Pattern Analysis, Intelligent Security and the Internet of Things, Abraham, A., A.M. Kamilah and C.Y. Huoy (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-17398-6, pp: 215-224.

Shukor, N.A., Z. Tasir and J. Harun, 2012. A theoretical framework for assessing students cognitive engagement through computer-supported collaborative learning. Int. J. Mach. Learn. Comput., 2: 654-657.

Zhang, A., 2012. Peer assessment of soft skills and hard skills. J. Inf. Technol. Educ. Res., 11: 155-168.

Zhu, C., 2012. Student Satisfaction, Performance and Knowledge Construction in Online Collaborative Learning. Educ. Technol. Society, 15: 127-136.

Zikmund, G.W. and J.B. Babin, 2009. Essentials of Marketing Research. 4th Edn., South-Western Cengage Learning, Mason City, Iowa, ISBN:9781439047545, Pages: 456.