International Journal of Software Engineering & Research Methodology – IJSERM
Volume 1: Issue 1
Publication Date: 09 January 2014

Understanding User Responses towards a Measurement Model using the Rasch Analysis

[1]Prasanna Ramakrisnan, [2]Azizah Jaafar, [3]Noor Faezah Binti Mohd Yatim

Abstract—**In this paper, we propose the Rasch analysis for studying an existing measurement model towards a design. The user satisfaction measurement model (e-Learner Satisfaction (ELS) and Website User Satisfaction (WUS)) was used to show how to understand users' response patterns. A twenty four item questionnaire was adapted based on the user satisfaction measurement model and administered to 86 students. The user satisfaction responses towards the design of the On-line Discussion Site (ODS) in Universiti Teknologi MARA (UiTM) e-learning portal were investigated. Step by step guidance is provided on how data was analyzed in understanding users responses towards the ODS design using the Rasch analysis. The findings from Rasch analysis will reveal the likelihood of user satisfaction with the ODS design. Thus the Rasch analysis was able to provide the most and least agreed items in a measurement model. Thereby system designers can easily identify items that need to be further improved in a design based on user's actual responses.**

***Keywords*— rasch analysis; user satisfaction; on-line discussion site (ODS);**

I. Introduction

The Rasch analysis is applied to measure latent traits (e.g., ability or attitude) in various disciplines. Users' responses are used to understand their latent traits in a measurement scale. Location of items and users of the measurement scale is estimated by the Rasch analysis from the proportion of responses of each user to each item. The probability of success depends on the differences between the ability of the person and the difficulty of the item.

In the Rasch analysis, a user who is more developed has a greater likelihood of endorsing all the items; and easier tasks are more likely to be endorsed by all users [1]. The item difficulty and person ability are expressed in logits through transformation of the raw score (ordinal scale) percentage into success-to-failure ratio or odds. This odds value is then converted to its natural logs (interval scale). The scale resulting from the Rasch analysis of the ordinal response has the properties of an interval scale. This scale is linear, and the numbers tell how much more of the attribute of interest is present.

The basic assumption of the Rasch model is each user is categorized by his or her ability; and each item by a difficulty; user and item can be presented by numbers along one line and lastly the probability of observing any particular scored responses can be computed from the differences between the numbers [1]. Thus the model can be used to link the person to the items that have relative ordering of latent variables.

The purpose of this study is to show how to understand the user's response towards a design. It is done through the Rasch analysis by examining the validity and item hierarchy of a measurement model. Data from previous work on user satisfaction was used for this study purpose. Data analysis includes scrutinizing item fit and establishing the item hierarchy for ordering items from greater to endorse (bottom) to lesser to endorse (top). The findings from this study will provide evidence to support the validity of the instrument and understanding of meaningful activities in a measurement model towards a design.

II. Measurement Model

The relationship between the observable variables and the underlying construct is represented in the measurement model [2]. Development and evaluation of measures in a measurement model uses commonly used procedures called factor analysis [3]. There are two phases involved [4]: (1) exploratory phase involve exploration of possible underlying factor structure of a measurement model and (2) confirmatory phase is used to verify the factor structure of a theory driven model or hypothesized measurement model.

The purpose of this study is to further investigate the validity, reliability; and create understanding on user responses towards the existing user satisfaction model. The items in this study were adapted from measures in the WUS [5] model and the ELS model [6].

The WUS proposes four dimensions; layout, information, connection and language customization [5]. The development of this measurement is based on the IS success theory, hypermedia design theory, a qualitative exploratory pilot study, and a quantitative on-line critical incident technique. This is a general model that can be used to evaluate user satisfaction in any web-based application.

The ELS identified the learner interface, learning community, content and personalization dimension as the measure [6]. The instrument for this measure was developed using samples only from Taiwan. Thus a confirmatory analysis and cross cultural validation using samples collected from outside Taiwan is required for generalization of this instrument. The construct drawn for user satisfaction model includes items about the learner interface, learning community, information and personalization. The measures for the learner interface were derived from the ELS model, including ease of use, user friendliness, operational stability and ease of finding. Layout, guidance, structure as well as hyperlink connotation are included from the WUS model. The measures for learning community were taken directly from the

ELS model. Information was measured using up to date, exact fit, sufficient as well as usefulness from the ELS model and item reliability, ease of understanding plus clear presentation from the WUS model. Lastly, personalization used capability of learning content needed, choosing what needed to be learned, controlling learning progress, recording learning performance from the ELS model together with language customization from the WUS model.

III. Method

A. Item Validity

Closed card sorting techniques were used to test the reliability and validity of the measurement items in this study since the measures came from different measurement models. The participants were provided with four predetermined categories; learners interface, learning community, information and personalization. They were then assigned the questions in the index cards to the given four categories. This helped in identifying the degree to which the participants agreed to the items belonging to the given category. The participants consisted of one IT professional, an academic scholar and a research student. A 93 percent correct hit ratio was achieved in this round, which indicates a sufficient item-construct reliability [7] and as such a second-round of card sorting was not conducted.

B. Instrument

An instrument was developed using 24 measures adopted from the ELS and WUS measurement model. The response categories used for this instrument are (1) strongly disagree, (2) disagree, (3) agree, (4) strongly agree. The instrument was distributed to full time undergraduate students at University Teknologi MARA (UiTM) who use the i-learn portal, the e-learning portal for their academic discussions. A total of 86 responses were collected from the survey distributed manually in four different classrooms and all the 86 responses were used as data set for analysis. The WINSTEPS software version 3.68.2 was used to analyze the data obtained from the responses.

IV. Rasch Analysis

A. Reliability

Reliability is the degree to which a test consistently measures whatever it measures. It depends on the construction of the instrument and distribution of the sample [8]. The Rasch analysis provides the item and person reliability separately. The person reliability depends on the sample ability range and number of items in the instrument. While item reliability is influenced by the item difficulty range and sample size.

B. Fit Statistics

Fits statistics was used to evaluate the quality of items used in the measurement model. Fit statistics are reported as mean squares and provide information about the extent of randomness in user's responses to specific items.

There are two types of fit statistics; infit and outfit. The infit statistics (weighted) report patterns of responses to items targeted on the person while outfit statistics (un-weighted) give the response pattern to items with difficulty far from a person. The value of infit and outfit statistics are reported as MnSq (mean of the squared residuals) and standardized Z values (Zstd). Generally, the criteria used to measure infit / outfit mean square range from 0.6 to 1.4 [9] and Z-standard values ranging from -2 to 2 (Bond & Fox 2007).

C. Item-Person Correlation

The point-measure correlation in the Rasch analysis provides information on the relationship between the observations on an item and the corresponding person measure. Difficult item are more likely to be endorsed by a person with higher ability while easier items are more likely to be endorsed by all persons. To be considered to be functioning in this manner at least a moderate correlation (point measure correlation ≥ 0.30) was expected [10].

D. Unidimensionality

Unidimensional measurement is when all of the non-random variance found in the data can be accounted for by a single dimension of difficulty and ability [11]. The unidimensionality of the instrument is determined by examining the first contrast from the items principle components analysis (PCA) of the standardized residual from the Rasch analysis. The unexplained variance of the first construct has to be below 3.0 [12] to indicate instrument unidimensionality. The secondary dimension only exists if it has at least three items in the unexplained variance of the first contrast.

E. Person-Item Distribution

The Rasch model can be used to establish a hierarchy of person and item together in a single Person-Item Distribution Map (PIDM). It displays the distribution of respondents on the left and the distribution of item agreement towards the measurement model on the right. Additionally, the PIDM also shows the most relevant items for measuring success of the measurement model and are at the bottom of the map.

The item is located by the number of persons getting a specific item correct or endorsing a specific item. While the person is located by number of items they are able to answer correctly or endorse. In applying the Rasch model, item locations are often scaled first. This part of the process of scaling is often referred to as item calibration. The mean for item is always zero because the Rasch model sets the mean of the item as a starting point (0 logits) for the calibration.

The location of an item on PIDM corresponds with the person's location at which there is a 0.5 probability of a correct response or endorsement of the item. Hence the probability of a person agreeing to the items below them will increase when moving down the PIDM.

V. Finding

Table 1 shows the summary statistics for the analysis of the sample of 86 on the 24 items. The person mean is 2.01 logits which is higher than the item mean 0 logit. Item mean is usually set as 0 point for calibration of item scale.

The summary fit statistics for items and persons shows satisfactory fit to the model. The item and person reliability are 0.77 and 0.91 respectively. This indicates that the items used to measure are reliable and results are reproducible.

TABLE I. SUMMARY STATISTICS OF ITEM AND PERSON

Statistics		Measures (logits)
Mean	Item	0.00
	Person	2.01
Reliability	Item	0.77
	Person	0.91
Mean Infit MnSq	Item	1.00
	Person	0.95
Mean Outfit MnSq	Item	0.95
	Person	0.95
Mean Infit Zstd	Item	0.00
	Person	-0.30
Mean Outfit Zstd	Item	-0.20
	Person	-0.30

Note: MnSq = Mean Square, Zstd = Z-Standard

TABLE II. ITEM ORDER AND FIT STATISTICS IN THE RASCH ANALYSIS OF THE 24 ITEMS (N=86)

Item Number	Item Measures	Item Difficulty (logits)	Point Measure Correlation	Infit Statistics		Outfit Statistics	
				MnSq	*Zstd*	*MnSq*	*Zstd*
23	Operational Stability	1.01	0.49	0.95	-0.2	0.84	-0.8
12	Layout	0.95	0.62	0.88	-0.8	0.81	-1.0
5	Ease of Use	0.48	0.68	1.00	0.0	0.96	-0.1
19	Records Learning Performance	0.37	0.57	0.98	-0.1	0.83	-0.9
20	Sufficient Information	0.37	0.65	0.75	-1.7	0.70	-1.7
16	Hyperlink Connotation	0.18	0.51	0.84	-1.0	0.78	-1.1
21	Guidance	0.18	0.59	1.01	0.1	0.90	-0.4
4	Control Learning Progress	0.13	0.59	1.08	0.5	0.98	0.0
13	Learn Required Content	0.12	0.60	0.76	-1.6	0.64	-2.0
14	Exact Required Information	0.12	0.48	1.01	0.1	0.97	-0.1
17	Discuss with Student	0.12	0.67	1.35	2.0	1.35	1.7
24	Rely on Information	0.12	0.50	0.82	-1.1	0.75	-1.3
2	Structured	0.06	0.69	0.85	-0.9	0.78	-1.1
8	Access Content	0.06	0.57	1.36	2.1	1.35	1.7
10	User Friendliness	0.06	0.67	1.21	1.3	1.26	1.3
22	Up-to-date Information	0.06	0.58	0.98	-0.1	1.0	0.1
9	Easy to Understand	-0.06	0.80	1.06	0.4	0.93	-0.3
15	Choice of Learning	-0.06	0.48	0.99	0.0	1.01	0.1
6	Information Clearly Presented	-0.13	0.66	1.23	1.4	1.15	0.8
18	Ease of Finding	-0.43	0.62	0.84	-0.9	0.75	-1.3
7	Choice of Language	-0.45	0.38	1.17	1.0	1.17	0.9
11	Discuss with Lecturer	-0.71	0.56	0.85	-0.9	0.82	-0.8
3	Useful Information	-1.25	0.47	1.01	0.1	1.09	0.5
1	Share Learning	-1.32	0.52	0.95	-0.2	1.06	0.4

The fit statistics analysis stated a good fit to most of the items (see Table II). Although there is one item (item number 8) with a standardized infit statistics of more than 2.0, the item remained after looking at a combination of statistical findings. Therefore, none of the items were dropped in the analysis, so all of them have to be considered to measure success of a measurement model.

The point-measure correlation values for all the items in the analysis were within the stated range. Thus all the responses towards the item correlated with the ability of the persons.

The unexplained variance in the first construct of the PCA was reported as 2.8. This value indicated that this instrument is unidimensional. From the above findings, the proposed instrument is appropriate, reliable and valid for measuring user satisfaction towards the ODS design.

The PIDM was explored for a better understanding of user responses towards the measurement model. The PIDM linking item difficulties to the person's endorsement ability of the sample across the four response scale (strongly agree, agree, disagree and strongly disagree) is presented in Figure I. More satisfactory items are located at the bottom of PIDM while least satisfactory items are displayed at the top of PIDM.

It was found that from the standpoint of the average group of students (person mean = 2.01 logit), all the items were plotted "much more easy to endorse". The most difficult items to endorse (see Table II) were "Operational Stability" (1.01 logits), "Layout" (0.95 logits) and "Ease of Use" (0.48 logits). While the three least difficult items in ascending order of difficulty included "Share Learning" (-1.32 logits), "Useful

Information (-1.25 logits) and "Discuss with Lecturer" (-0.71 logits).

The items reflecting least satisfaction of activities are aspects of activities from learner interface. The findings offer evidence to suggest that students are least satisfied with some of the elements in the user interface of the current ODS. Thus the designers have to further improve the user interface of the ODS to increase user satisfaction.

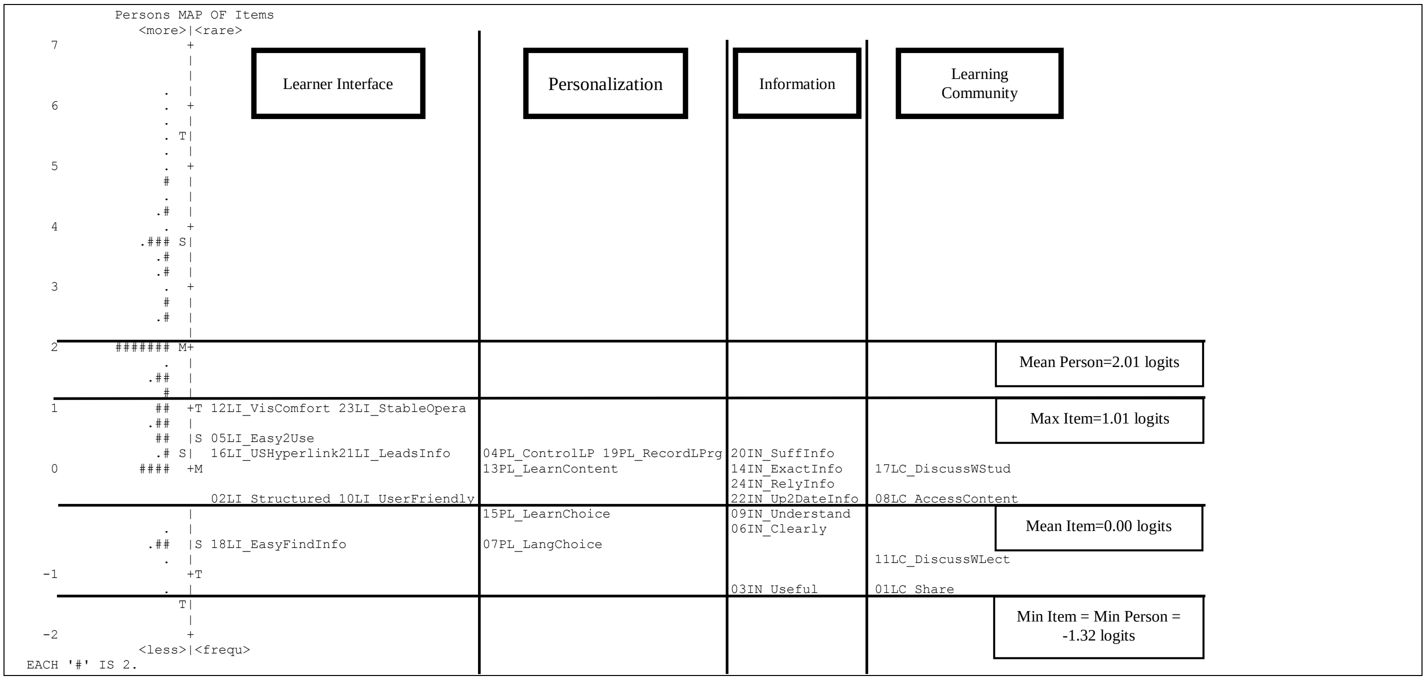

FIGURE I. THE PERSON-ITEM DISTRIBUTION MAP (PIDM)

VI. Conclusion

Applying the Rasch analysis to understand user responses provides potential information on the assessment of constructs in the activities participated by respondents. The findings indicate how a validity and unidimensional analysis was conducted towards a measurement model. The reliability, fit statistics, item-person correlation, unidimensionality and person-item distribution was used as criteria to understand user's responses towards items in a measurement model.

Developer and designers can use the Rasch analysis for studying user's actual responses on the system or the interface design. Any existing instrument can be adapted to understand users' responses for understanding items or constructs that need to be further improved in a design. Therefore we propose the Rasch analysis for instrument validation and exploration of user responses.

Acknowledgment

The authors are grateful to Associate Professor Dr. Mohd Nor Mamat from the i-Learn Centre, Universiti Teknologi MARA (UiTM) for his effort in providing valuable critiques during the data analysis stage.

References

[1] T. Bond and C. Fox, *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum, 2007.

[2] O. Götz, K. Liehr-Gobbers, and M. Krafft, "Evaluation of structural equation models using the partial least squares (PLS) approach," in *Handbook of Partial Least Squares*, Springer, 2010, pp. 691–711.

[3] F. J. Floyd and K. F. Widaman, "Factor analysis in the development and refinement of clinical assessment instruments," *Psychological assessment*, vol. 7, no. 3, pp. 286–299, 1995.

[4] K. D. Mackenzie and R. House, "Paradigm development in the social sciences: A proposed research strategy," *Academy of Management Review*, pp. 7–23, 1978.

[5] S. Muylle, R. Moenaert, and M. Despontin, "The conceptualization and empirical validation of web site user satisfaction," *Information & Management*, vol. 41, no. 5, pp. 543–560, May 2004.

[6] Y.-S. Wang, "Assessment of learner satisfaction with asynchronous electronic learning systems," *Information & Management*, vol. 41, no. 1, pp. 75–86, Oct. 2003.

[7] G. c. Moore and I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Information Systems Research*, vol. 2, no. 3, pp. 192–222, 1991.

[8] W. P. J. Fisher, "Reliability, Separation, Strata Statistics," *Rasch Measurement Transactions*, vol. 6, no. 3, p. 238, 1992.

[9] B. D. Wright, J. M. Linacre, J. E. Gustafson, and P. Martin-Lof, "Reasonable mean-square fit values," *Rasch measurement transactions*, vol. 8, no. 3, p. 370, 1994.

[10] M. L. Finlayson, E. W. Peterson, K. A. Fujimoto, and M. A. Plow, "Rasch Validation of the Falls Prevention Strategies Survey," *Archives of physical medicine and rehabilitation*, vol. 90, no. 12. W.B. Saunders, pp. 2039–2046, 01-Dec-2009.

[11] J. Sick, "Assumptions and requirements of Rasch measurement," *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, vol. 14, no. 2, pp. 23–29, 2010.

[12] A. M. Eakman, "Measurement characteristics of the Engagement in Meaningful Activities Survey in an age-diverse sample," *The American Journal of Occupational Therapy*, vol. 66, no. 2, pp. e20–e29, 2012.

About Author (s):

Prasanna Ramakrisnan
Faculty of Computer and Mathematical Sciences (FSKM), University Technology MARA (UiTM), 40450 Shah Alam, Selangor MALAYSIA
prasanna@fskm.uitm.edu.my

Associate Prof. Dr. Azizah Jaafar
Institute of Visual Informatics, National University of Malaysia, 43600 Bangi, Selangor MALAYSIA
aj@ftsm.ukm.my

Dr. Noor Faezah Binti Mohd Yatim
Faculty of Infomation Science and Technology (FTSM), University Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor MALAYSIA
nfmy@ftsm.ukm.my