Validation of Perceived Community Safety Instrument Using Polytomous Rasch Measurement Model

Shahren Ahmad Zaidi Adruce, Donald Stephen, Dayang Siti Aisah Abang Suhaili, Alexandra Nastassia John, Abg Mohd Heikal Abg Othman and Nur Adila Latif
The Institute of Borneo Studies, Universiti Malaysia Sarawak
azshahren@unimas.my

***Abstract*—This study involves the designing, development and testing of an instrument for capturing perception of community safety in rural communities in Sarawak, Malaysia. Data were collected in Kampung Pulo Salak, Kampung Sebayor, Kampung Tanjong Bako and Kampung Pinggan Jaya. A total of 172 households were interviewed and their responses were recorded accordingly for each item listed in the instrument. The study details the validation of community safety instrument using a Rasch analysis technique. The instrument was adapted, translated from various sources and customized to the need of local communities. The many items in the original instruments were shortlisted and subsequently tested to assess the quality of the measurement used. Aspects investigated include personal and community safety, crime and social disorder, police effectiveness and engagement as well as the sense of community. Based on results from Rasch analysis, thirty items from the initial fifty-one shortlisted items were retained. The final version of the instruments implicates good reliability and validity. There was no differential item functioning detected and the measure was proven to be unidimensional. This instrument is expected to benefit local authorities especially for the formulation of necessary interventions to ameliorate issues pertaining to community safety.**

***Index Terms*—Community Safety; Community Safety Instrument; Rash Analysis; Sense of Community; Crime and Social Disorder.**

I. INTRODUCTION

Community safety involves a wide range of issues. The administration of community safety instrument provides discernment about issues that make people feel safe or unsafe within their community. The community safety instrument provides a platform for the community to voice their concerns and facilitate the implementation of interventions by local authorities [1]. This community safety instrument also incorporates the "sense of community" section, which can be a standalone construct use to measure the extent to which the residents perceived their "sense of belonging" to the community [2, 3]. If a neighbourhood is perceived to be safe by its residents, it is crucial to note that the residents have close knitted relationship and have good sense of community. The issue of community safety has been widely discussed through research and media. However, there is no notable effort to capture the perception of community safety in Sarawak. This is particularly beneficial for local authorities seeking to implement policies pertaining to crime prevention [1]. The instrument thus serve as an imperative tool to ameliorate community safety issues. Using this instrument, Local authorities can request for the administration of this community safety instrument in their area of responsibility to identify critical safety issues in the community.

II. DEVELOPMENT AND CONCEPTUALIZATION

The United States Department of Justice has identified five key components in public safety and law enforcement. The five components are community involvement, procedural justice, performance as well as contact and satisfaction [4]. Community involvement, contact, and satisfaction are components in the sense of community. With the incorporation of the sense of community components, perceived community safety in this study can be categorized into four major components: overall safety perception, crime and social disorder, effectiveness and engagement of law enforcement and sense of community. Initial items pool were adapted from various sources as follows: Personal safety perception [1, 4-6], Crime and Social disorder [1, 4, 7], Police effectiveness, engagement, and procedural justice [4, 8], and Sense of community [2, 3].

A. Personal and community safety

Collective perceived personal safety reflects real safety scenario in a community. Qualitative research regards personal safety as a barrier to local walking in a neighbourhood. Perception of personal safety in a neighbourhood is influenced by several factors, which are social environments, individual factors, physical environments, natural surveillance, and time of day [6].

B. Crime and social disorder

Crime and social disorder components contain a list of social problems that were anticipated to affect the perception on safety in rural communities. Crime and social disorder have been consistently linked to community social order and quality social life. Disorders indicate that the neighbourhood is unsafe, leading to community withdrawal and the increase in sense of insecurity [7].

C. Police effectiveness and engagement

Trust and confident from the community is important for police to perform their duty effectively. Racial and social inequality were cited as factors for the increase of complaints pertaining to police corruption, bias, and abuse of power [8].

D. Sense of community

In community psychology, the understanding about psychological Sense of Community is crucial. The sense of Community represents the interdependent relationship between the individual and the community that they are a part of. It concerns about the feeling of attachment and belonging to the community. High sense of community indicates each member matters to the others. The Sense of Community Index (SCI) is one of the most commonly used measures in Psychological Sense of Community (PSOC) [9].

III. METHODOLOGY

In the instrument development process, researchers are mostly concerned about measuring latent constructs of which can only be measured indirectly. Latent construct such as perception is often regarded as an integral part of empirical studies. Traditionally, the classical test theory (CTT), often called "true score model" is used to develop and validate instruments, but it has proven to have several disadvantages. The dependency of items on the sample is one of the major limitations of CTT, resulting in inconsistencies when a different sample is administered [10]. These differences in term of results obtained for the same measure proved to be cumbersome as it decreases the confident in data interpretation.

The community safety instruments utilize Item response theory (IRT), namely Rasch measurement model to overcome this predicament. Rasch measurement model is an example of IRT. The differences in scoring instruments, such as partial credit model and rating scale instruments resulted in different kind of item response models. For rating scales, such Likert, Polytomous Rasch model is suitable. As Rasch measurement model considers the association between person ability, item difficulty, and the probability of a response. The measurement of the latent construct is based on the probability of respondents responding in a certain manner for a particular item [11]. The robustness and probing statistics that Rasch analysis offers provide a solution to the issues of instrument validity [12]. Table 1 lists indicators for instrument quality based on Rasch model.

Table 1
Validity criteria and indicators

Validity Criteria	Indicator
Surface	Item feedback from road users, Item review
Content	Documentary Analysis; item review
Substansive	Item review, Rasch analysis of item behavior.
Structure	Weighted and Unweighted mean squares fit statistics, Item discriminant, Principal Components Analysis of Residuals (PCAR), Internal consistency reliability statistics
Polarity	Item discriminant
External factors	Rasch item functioning
Consequential	Differential item functioning

The instrument was administered in several rural communities in Sarawak. The selection of these rural communities was influenced by accessibility and interest of the stakeholders. It can be difficult and costly to assess some rural areas. Roads to and from certain areas may be impassable depending on environmental factors. At times, people in the community may also be inaccessible due to economic activities (fishing, hunting, and other activities). It was noted that samples greater than 100 would produce stable estimates [13]. Therefore, in this study, overall of 172 households participated in the study. One representative per household was required to respond to the questionnaire. To bridge knowledge and language gap, the data collection was conducted verbally I the language of the respondents whereby respondents were required to respond to Likert-type items verbally.

Initial pool of 51 items were tested. The 51 items on perceived community safety were categorized into 4 aspects (Personal safety = 11 items; Crime and social disorder= 18 items; Police effectiveness, engagement and procedural justice = 4 items; and Sense of Community = 18 items). Each item was explained thoroughly to avoid misunderstanding or confusion. The distribution of participants based on the community is as follow: Pinggan Jaya = 55 households, Kampung Tanjong Bako = 43 households, Kampung Pulo Salak = 40 households, and Kampung Sebayor = 34 households. To avoid response bias, participants were assured that there will be no identifying question and no personal information shared with third party. All information will be kept strictly confidential and participants are free to withdraw from the study or selectively respond to the items. JMetrik V4 software by Meyer [14] was used to run the analysis. Prior to the field test, the instrument was reviewed and translated from the original language.

IV. RESULTS

A. Item analysis and fit statistics

Non-Rasch and Rasch item statistics are computed to show item quality in its measure. Indicators include non-Rasch item difficulty (Diff), Standard Deviation (SD), item discriminant (Discrimin), item difficulty based on joint maximum likelihood estimation (JMLE Diff), standard error (Std Error), weighed mean squares (WMS), standardize weighed mean squares (Std. WMS), unweighted mean squares (UMS), and standardize unweighted mean squares (Std. UMS). A common threshold is applied for items with the same response anchors. Mean squares value in the range of 0.50-1.50 is used and considered productive for measurement. Standardized mean squares of no more than 3 but no less than -2 [15] were used. If all items recorded acceptable fit statistics, they are considered unidimensional [16].

Rasch statistics for the final set of items in the personal and community safety component are as follows: Non-Rasch diff (2.84-3.5), discrimin (0.19-0.58), JMLE diff (-0.59-0.71), Std. Err (0.06-0.12), WMS (0.76-1.20), Std. WMS (-2.01-1.73), UMS (0.70- 1.42), and Std. UMS (-2.38- 2.48). One item was removed from the initial item pool of 11 items. Item a2 was removed (WMS 1.21, UMS 1.57, Std. WMS 2.17 and Std. UMS 2.88). After removal of item a2, WMS and UMS fit statistics show no deviation from productive measurement range (WMS $\geq 0.76 \leq 1.20$, UMS $\geq 0.85 \leq 1.43$, Std WMS ≥ -2-$2 \leq 1.75$, Std UMS $\leq -2.38 \leq 2.49$). However, item a8 shows predictability although it is productive for measurement. On the other hand, item a3 and a5 show signs of unpredictability. All three items were retained because the degree of predictability and unpredictability is mild.

Rasch statistics for the final set of items in the crime and social disorder component are as follows: Non-Rasch diff (2.26-3.26), discrimin (0.64-0.84), JMLE diff (-0.74-0.85), Std. Err (0.09-0.12), WMS (0.71-1.26), Std. WMS (-2.02-2.69), UMS (0.74- 1.32), and Std. UMS (-0.35- 2.04). 9 items were retained from the initial pool of 18 items. Results indicate good fitting of WMS and UMS statistics although

item b11 shows some slight predictability. On the other hand, item b9 and b18 shows some tendency to be slightly unpredictable. However, the level of predictability or unpredictability did not warrant for removal.

Item b10 and b17 was removed due to negative polarity with item discriminant value of -0.1027 and -0.0062 respectively. After Rasch analysis, 7 items were further removed due to misfit of Rasch fit statistics. Item removed (in sequence) include item b1 (WMS 1.64, UMS 2.94, Std. WMS 5.52 and Std. UMS 10.41), b4 (WMS 1.47, UMS 2.62, Std. WMS 2.04 and Std. UMS 3.92), b7 (WMS 1. 58, UMS 2.38, Std. WMS 4.07 and Std. UMS 5.47), b6 (WMS 1.91, UMS 2.78, Std. WMS 6.41 and Std. UMS 7.51), b2 (WMS 1.63, UMS 2.36, Std. WMS 4.73 and Std. UMS 6.60), b15 (WMS 0.48, UMS 0.45, Std. WMS -3.74 and Std. UMS -2.73), as well as b5 (WMS 0.48, UMS -.45, Std. WMS -4.18 and Std. UMS -3.12).

Rasch statistics for the final set of items in the police effectiveness and engagement component are as follows: Non-Rasch diff (2.80-3.06), discrimin (0.76-0.86), JMLE diff (-0.47-0.67), Std. Err (0.14-0.16), WMS (0.74-1.12), Std. WMS (-2.12-0.92), UMS (0.71- 1.08), and Std. UMS (-2.28-0.83). None of the items were removed. Item c2 is deemed predictable although WMS and UMS statistics show good fit.

Rasch statistics for the final set of items in the sense of community component are as follow: Non-Rasch diff (0.06-3.63), discrimin (0.34-0.63), JMLE diff (-0.78-0.69), Std. Err (0.10-0.15), WMS (0.80-1.47), Std. WMS (-1.55-2.69), UMS (0.77- 1.46), and Std. UMS (-2.23- 2.34). There was no item that indicate negative polarity. Results indicate productive measurement for WMS and UMS statistics although item d16 shows slight predictability. On the other hand, item d4 and d11 shows some tendency to be unpredictable. Based on other indicators, although there is evidence of predictability and unpredictability, the item is proved to be productive for measurement and not degrading. Therefore, the items were retained and all together a total of 13 items were retained from the initial pool of 18 items.

After performing the Rasch analysis, 6 items were further removed due to the violation of Rasch model based on misfit of Rasch fit statistics. 6 item removed include item d6 (WMS 2.62, UMS 3.14, Std. WMS 11.28 and Std. UMS 12.38), d2 (WMS 1.52, UMS 2.15, Std. WMS 2.93) and Std. UMS 4.92), d8 (WMS 1.32, UMS 1.73, Std. WMS 2.52 and Std. UMS 4.84), d18 (WMS 1.19, UMS 1.76, Std. WMS 1.11 and Std. UMS 3.54), as well as d1 (WMS 1.23, UMS 1.62, Std. WMS 1.40 and Std. UMS 2.85).

B. Reliability Analysis

The comprehensiveness of Rasch model in reliability analysis is apparent, it provides more reliability indicators in addition to the use of Cronbach's alpha test. Reliability value less than 0.6 was not acceptable for any item. The commonly accepted criterion for the separation indices is 3.0 [12]. Table 2 shows reliability indicators for all components. Based on the item and person reliability as shown in the Cronbach's alpha result, it can be concluded that items in all components are reliable (item and person reliability range = 0.68 – 0.96; Cronbach's alpha range = 0.70-0.95).

Separation index describes the number of distinguishable groups from the measurement [17], according to a guideline, item separation should be more than 3. Meanwhile, person separation should be no less than 2 [18]. Generally, the separation index for items is better compared to person index. A lower value for separation index in some of the components as reported in the table warrant further investigations to help understand the situation.

Table 2
Reliability analysis for all components

Indicator	A		B		C		D	
	item	person	item	person	item	person	item	person
Observed Var	0.20	0.58	0.18	1.82	0.19	3.68	0.21	1.14
Observed SD	0.45	0.76	0.42	1.35	0.44	1.92	0.46	1.07
MSE	0.01	0.18	0.01	0.22	0.02	0.72	0.02	0.24
Root MSE	0.09	0.43	0.10	0.47	0.15	0.85	0.12	0.49
Adjusted Variance	0.19	0.39	0.17	1.60	0.17	2.96	0.20	0.90
Adjusted SD	0.44	0.63	0.41	1.26	0.42	1.72	0.45	0.95
Separation Index	4.78	1.46	4.20	2.68	2.80	2.03	3.62	1.94
Number of Strata	6.70	2.28	5.93	3.91	4.07	3.04	5.15	2.91
Rasch Reliability	0.96	0.68	0.95	0.88	0.89	0.80	0.93	0.79
Cronbach's Alpha	0.70		0.95		0.91		0.85	

i. Differential Item Functioning

99 females (57.6%) and 73 males (42.4%) participated in the study. Differential item functioning (DIF) analysis using Mantel-Haenszel Statistics is implemented to examine if different gender responds to items differently. DIF is classified into one of 3 different categories of observed DIF: AA, BB, or CC (negligible, medium, or large). Plus (+) or minus (-) sign is added to show if an item favours focal (+) or reference (-) group [14]. Table 8 to Table 11 shows different item functioning for items on all components. Tables show chi-square statistics, p-value, effect size based on 95% confident interval and DIF class. In this analysis, DIF based on gender is investigated. Female is set to be the focal group.

Table 4 shows DIF analysis for all safety components where item a5 shows slight tendency to show biased towards reference group (male). Item a11 shows a slight tendency of bias towards the focal group. Deletion of items that shows DIF is carefully considered. In this analysis, items were not removed because of item relevance and the extent of DIF shown is not serious. In DIF analysis for crime and social disorder component, only item b14 has the tendency to show biasness towards the focal group. As for the police engagement and effectiveness, the component shows no problem in DIF. Lastly, only one item is shown to have slight DIF (item d11) in the sense of community component. None of the DIF is classified as CC class, indicating it is a good item and the measure is not discriminative. BB category items were therefore retained.

Table 4
DIF analysis of all safety components

Item	χ2	p	E.S (95% C.I)	Class
a1	1.06	0.3	-0.12 (-0.37,0.13)	AA
a3	0.22	0.64	-0.10 (-0.39,0.19)	AA
a4	0.11	0.74	-0.13 (-0.57,0.32)	AA
a5	1.48	0.22	-0.21 -0.60,0.18)	BB-
a6	0.40	0.53	0.10 (-0.08,0.29)	AA
a7	0.86	0.35	0.07 -0.17,0.30)	AA
a8	0.37	0.54	0.05 (-0.18,0.28)	AA
a9	0.57	0.45	0.07 (-0.11,0.26)	AA
a10	0.02	0.89	0.05 (-0.22,0.31)	AA
a11	2.35	0.13	0.24 (-0.11,0.59)	BB+
b3	1.22	0.27	-0.15 (-0.42,0.12)	AA
b8	0.03	0.85	0.05 (-0.18,0.27)	AA
b9	0.10	0.75	-0.13 (-0.49,0.24)	AA
b11	0.38	0.54	0.05 (-0.15,0.26)	AA
b12	0.32	0.57	-0.00 (-0.27,0.27)	AA
b13	1.71	0.19	0.06 (-0.14,0.25)	AA
b14	4.46	0.03	0.28 (0.03,0.53)	BB+
b16	0.29	0.59	0.04 (-0.12,0.21)	AA

Item	χ2	p	E.S (95% C.I)	Class
c1	0.49	0.48	-0.07 (-0.27,0.13)	AA
c2	0.25	0.62	-0.05 (0.19,0.08)	AA
c3	1.71	0.19	0.18 (-0.02,0.38)	AA
c4	0.03	0.87	-0.05 (-0.24,0.13)	AA
d3	0.49	0.48	-0.07 (-0.27,0.13)	AA
d4	0.25	0.62	-0.05 (0.19,0.08)	AA
d5	1.71	0.19	0.18 (-0.02,0.38)	AA
d7	0.03	0.87	-0.05 (-0.24,0.13)	AA
d9	0.00	0.98	-0.06 (-0.25,0.13)	AA
d10	0.91	0.34	-0.15 (-0.36,0.05)	AA
d11	2.06	0.15	0.32 (0.01,0.64)	BB+
d12	0.13	0.72	0.07 (-0.14,0.29)	AA
d13	0.48	0.49	-0.15 (-0.34,0.05)	AA
d14	1.18	0.28	-0.09 (-0.25,0.06)	AA
d15	0.01	0.93	0.01 (-0.20,0.22)	AA
d16	1.90	0.17	0.18 (-0.02,0.39)	AA
d17	0.08	0.78	-0.03 (-0.24,0.18)	AA

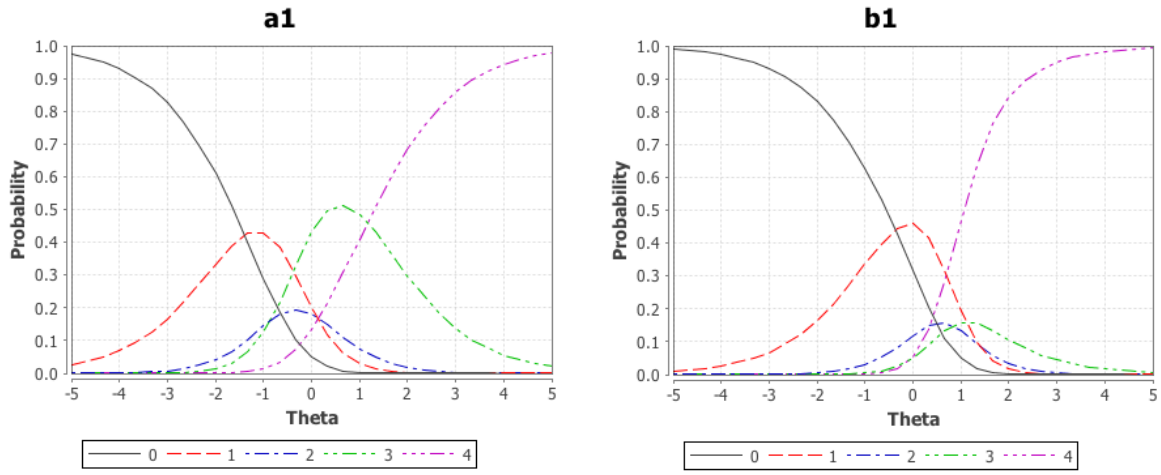

Figure 1: Option characteristic curves for example of well-functioning item (a1) and ill-functioning item (b1)

C. Overall functioning of items.

From Figure 1, there is no item that left that shows serious differential functioning. Item discriminant and fit statistics somehow have removed all problematic items. This is critical to ensure that the measure used is valid.

D. Dimensionality

Rasch fit statistics in the form of unweighted and weighted mean squares serve as indications of a single dimension. Another indicator used is the eigenvalue of first rash residuals. The eigenvalue of first residuals for Personal and community safety is 2.65, crime and social disorder component yield eigenvalue of 1.61. Three components recorded acceptable eigenvalue of less than 3 [18]. The eigenvalue of the sense of community component is rather high at 4.48. This prompted further analysis by examining the polarity of first factor loading of Rasch residuals. The items were separated based on polarity with the separation of negative and positive factor loading between subsets [19, 20]. Negative items are d5(-0.50), d7(-0.55), d9(-0.73), d10(-0.73), d1(-0.60), d12(-0.40). Meanwhile positive items are d3(0.18), d4(0.09), d13(0.74) d14(0.53), d15(0.74), d16(0.74), d17(0.61). Person estimated of from the two groups of items are then calculated. A paired-samples t-test was conducted to compare person estimates of the two groups of items based on loading of the first factor. There was not a significant difference in the scores for IV level 1 (M=2.95, SD=2.46) and IV level 2 (M=2.70, SD=1.81) conditions; t(171)=1.43, p = 0.16. This analysis concluded that there is not enough evidence to disregard the assumption based on unidimensional.

V. Conclusion

Rasch analysis is gaining popularity over classical techniques to assess reliability and validity of the instrument. Based on the results of the polytomous Rasch model, the instrument shows good reliability. There was no issue of Differential Item Functioning in the final version of the instrument. Items in all components are proven to be unidimensional. Overall, 36 items were retained from the initial shortlisted pool of 51 items.

Acknowledgement

This study is a section of community studies conducted by the Institute of Borneo Studies, Universiti Malaysia Sarawak. Authors would like to express their heartfelt gratitude to the Nusantara Chair at the Institute for Borneo Studies at UNIMAS for funding the study.

REFERENCES

[1] M. Willis, "Community safety in Australian Indigenous communities: Service providers' perceptions," *Canberra: Australian Institute of Criminology*, 2010. Retrieved from papers2://publication/uuid/BF6199B1-946B-4446-BF10-EB14C40F65B0

[2] D. W. McMillan, "Sense of Community," *Journal of Community Psychology*, vol. 24, no. 4, pp. 315–325, 1996.

[3] D. W. McMillan and D. M. Chavis, "Sense of community: A definition and theory," *Journal of Community Psychology*, vol. 14, no. 1, pp. 6–23, 1986.

[4] U.S. Department of Justice, *Community Survey on Public Safety and Law Enforcement*. Washington, DC 20530, 2014.

[5] S. Christakopoulou, J. Dawson, and A. Gari, "The Community Well-being Questionnaire: Theoretical context and initial assessment of its reliability and validity," *Social Indicators Research*, vol. 56, pp. 321–351, Aug. 2001.

[6] S. Foster and B. Giles-Corti, "The built environment, neighborhood crime, and constrained physical activity: An exploration of inconsistent findings," *Preventive Medicine*, vol. 47, no. 3, pp. 241–251, 2008.

[7] E. Gracia and J. Herrero, "Perceived neighborhood social disorder and attitudes toward reporting domestic violence against women," *Journal of Interpersonal Violence*, vol. 22, no. 6, pp. 737–752, 2007.

[8] D. P. Rosenbaum, D. S. Lawrence, S. M. Hartnett, J. McDevitt, and C. Posick, "Measuring procedural justice and legitimacy at the local level: the police–community interaction survey," *Journal of Experimental Criminology*, vol. 11, no. 3, pp. 335–366, 2015.

[9] P. L. Obst and K. M. White, "Revisiting the sense of community index: A confirmatory factor analysis," *Journal of Community Psychology*, vol. 32, no. 6, pp. 691–705, 2004.

[10] S. E. Embretson and S. P. Reise, *Item response theory*, Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 2000.

[11] M. Curtin, J. Browne, A. Staines, and I. J. Perry, "The Early Development Instrument: an evaluation of its five domains using Rasch analysis," *BMC Pediatrics*, vol. 16, no. 1, p. 10, 2015.

[12] T. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Third Edition*. Taylor & Francis, 2015. Retrieved from https://books.google.com.my/books?id=s17LCQAAQBAJ

[13] D. A. Revicki, W. H. Chen, and H. Gelhorn, "Rasch Model Analysis and Small Sample Size," In D. Neasham & A. Cox (Eds.), *Evidence Matters: Innovative approaches to database study design and analysis*, Blue Bell, Pennsylvania: United BioSource Corporation, pp. 19–22, 2013.

[14] J. P. Meyer, *Applied Measurement with JMetrik*. Taylor & Francis, 2014. Retrieved from https://books.google.com.my/books?id=MJbOAwAAQBAJ

[15] J. M. Linacre, "What do infit and outfit, mean-square and standardized mean," *Rasch Measurement Transactions*, vol. 16, no. 2, 878, 2002.

[16] J. M. Linacre, "Detecting Multidimensionality: Which Residual Data-type Works Best?," *Journal of Outcome Measurement*, vol. 2, no. 3, pp. 266–283, 1998.

[17] M. Wilson, *Constructing measures: An item response modeling approach*, Mahwah, NJ: Erlbaum, 2005.

[18] J. M. Linacre, *A user's guide to Winsteps Ministeps Rasch-model computer programs [version 3.74. 0]*, Chicago IL: Winstep.com, 2012.

[19] J. F. Pallant, R. L. Miller, and A. Tennant, "Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis," *BMC Psychiatry*, vol. 6, p. 28, 2006.

[20] E. V. Smith, "Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals," *Journal of Applied Measurement*, vol. 3, no. 2, pp. 205–231, 2002.

[21] B. D Wright, "Reliability and separation," *Rasch Measurement Transactions*, vol. 9, no. 4, p. 472, 1996. Retrieved from http://www.rasch.org/rmt/rmt94n.htm.