DEVELOPMENT AND VALIDATION OF AN INSTRUMENT TO MEASURE LEADERSHIP COMPETENCY AMONG SECONDARY-SCHOOL STUDENTS: A PILOT STUDY

MEI-TENG LING
VINCENT PANG
UNIVERSITY MALAYSIA SABAH

The purpose of this study is to pilot test the newly developed Malaysian Secondary School Students' Leadership Scale (M3SLS) and to determine its psychometric properties. The Rasch Rating Scale analysis was implemented and the correlation coefficient of measurement-point (PTMEA correlation) value of every item was found positive. Eight items with PTMEA correlations below .20, however, needed revision. Since secondary dimension was noticeable, the three domains therefore were analyzed separately in the next study. The item separation and item reliability were 7.49 and .98 respectively which implied that the person sample was sufficient to confirm the item difficulty hierarchy. A Person separation index value of 2.86 indicated there were three levels of respondent ability identified in this pilot study. Person separation of more than two and person reliability of .89 with a relevant person sample implied the instrument was sensitive enough to distinguish between high and low performers. Except for eight items, the infit mean square values for all items were in the range of .50 to 1.50 which was within the acceptable range. As a conclusion, there were items that needed to be revised and modified based on the analysis of item polarity, item fit, and principal component analysis of residual (PCAR). The reliability and separation indices of item and person were within the acceptable range. Based on the outcomes of the pilot test, the instrument is recommended for distribution to a larger population to ensure stability of the scale.

Key words: Psychometric; Rasch model; Rating scale; Secondary school; Student leadership.

Correspondence concerning this article should be addressed to Mei-Teng Ling, Faculty of Psychology and Education, University Malaysia Sabah, 88400 Kota Kinabalu, Malaysia. Email: lingmeiteng@gmail.com

The importance of instilling leadership skills in students has always been a main subject of discussion in Malaysia. Guiding students to fully explore and utilize their leadership potential through co-curricular activities, peer discussion, and leadership training is a major concern in the Malaysian education system, where every student is helped to reach his or her full potential through the creation of formal and informal opportunities for them to work in teams and take on leadership roles. Therefore, leadership development programs in school play an important role in developing leadership traits among students (Bagheri, 2011).

Currently, a common form of measurement used to ascertain student success in school is through academic success. Academic success is assessed through examinations. The Primary School Achievement Test (PSAT/UPSR)[1] is a key academic success indicator in primary school and the Malaysian Certificate of Education (MCE/SPM) assesses secondary-school students' achievement. In 2013, the Ministry of Education Malaysia (MOE/KPM) issued a circular requiring teachers to assess student success in co-curricular activities in school. Every student is rated

by the teacher according to their attendance, involvement, and achievement in co-curricular activities. This assessment is then converted into 10 marks and taken into account as part of the entry requirements to attend public universities. However, the assessment is based on attendance, involvement, and achievement only and does not take into account aspects of student leadership. Furthermore, a student who is successful in co-curricular activity is not necessarily a competent leader. Schools usually only award appreciation certificates to school prefects, librarians, and class monitors for their service but leadership competencies are neither assessed nor recorded.

Outside of Malaysia, several instruments are utilized in schools to assess leadership competencies among students. Examples of scales that have been developed include the Penn State Leadership Competency Inventory (Yoon, Song, Donahue, & Woodley, 2010), Adaptive Leadership Competency Profile (Sherron, 2000), Prospector Instrument (Spreitzer, Kizilos, & Nason, 1997), A 360-Degree Peer Evaluation of Leadership Competencies Scales (Rogers, 2001), and Multifactor Leadership Questionnaire (MLQ; Bass & Avolio, 1994). However, because most of these scales were conceptualized from a Western perspective and based on the model developed by the U.S. Office of Personnel Management (OPM), some aspects of leadership qualities required in the Malaysian context may be different. In addition, using a common or standard questionnaire across cultures may cause cultural bias. Therefore, a conscious effort to translate the instrument into the target language and culture is critical because a good questionnaire developed in a single culture may not necessarily "travel well" across cultures due to differences in meaning and reading (Tuleja, Beamer, Shum, & Chan, 2011).

Researches in the past, such as those by Alias, Yussof, Mustapha, and Ibrahim (2010), Amrin (2007), Don (2009), Fareeza (2010), and Ibrahim and Amin (2014), were more interested in assessing the leadership competency of principals, deputy principals, coaches, and teachers rather than students in secondary school. Yet, secondary-school students also possess leadership potential (Fertman & Linden, 1999; Hine, 2011) and this should be an area of concern for educators so that they could assist students to develop their leadership skills and build character at an early stage of their development.

In order to develop and plan a better training module, it is necessary to ascertain the specific areas of concern. Good student leadership development programs ensure that students not only gain optimum benefits from the training provided, but also avoid unnecessary wastage of time and resources (Zakaria et al., 2008). The purpose of this study is to develop a valid and reliable instrument to measure student leadership competency and then determine a student leadership competency profile. The information that this provides can then help in planning effective interventions and improvements on students' leadership competencies before they enter higher level education or the job market.

In considering an appropriate methodology to ascertain the psychometric properties of an instrument, the Rasch model is considered less preferable than the Classical Test Theory (CTT) in Malaysia. Researchers such as Yau (2007) and Yoon et al. (2010) depended on exploratory factor analysis (EFA) to assess the dimensionality, preliminary validity, and reliability aspects. Although many testing and measurement textbooks present CTT as the only way to determine the quality of an assessment, the Item Response Theory (IRT) does offer a sound alternative to the classical approach (Idowu, 2011). Rasch analysis is a method to obtain measures which are objective, fundamental, and linear. The basic Rasch model is used to separate the ability of respondents and the quality of the test. In this approach, a rating scale is used because the scale is poly-

chromous. According to Sherron (2000), the Rasch Rating Scale model transforms ordinal rating measures to logit scale (interval scale). Therefore, Rasch analysis can produce student position corresponding to the position of the item on the same scale. This pilot study will test the newly developed Malaysian Secondary School Students' Leadership Scale (M3SLS) and determine its psychometric properties by employing an IRT approach.

METHOD

Participants and Procedure

The instrument was administered to 240 students from five schools in the West Coast Division of Sabah, Malaysia. For well-designed pilot studies, 30 participants would be sufficient (Linacre, 1994, as cited in Jackson, Draugalis, Slack, Zachry, & Agostino, 2002). The samples were all from government secondary schools.

In order to secure official ethical clearance for the study, a formal application was made to the Educational Planning and Research Division of MOE. Various documents pertaining to ethical clearance were included with the application such as: application form to conduct research in schools, research instruments, and school list. Each participant took part voluntarily and was presented with an information document on the first page of the questionnaire assuring them of the absolute confidentiality of their data. The researchers took precautions to ensure the safety and confidentiality of the participants.

Teachers in the school administered the questionnaire to the students, who were briefed on the specific instructions as written on the main page of the questionnaire. From these 240 respondents, 21 were excluded from the data analysis because of incomplete responses. The final figure consisted of 136 female students (62.1%) and 83 male students (37.9%), with the majority of participants aged 17-18 years (63%), followed by 13-14 year olds (18.7%), and then 15-16 year olds (18.3%).

The Malaysian Secondary School Leadership Scale Construction

The Malaysian Secondary School Student's Leadership Scale (M3SLS) is a self-administered questionnaire. The scale comprises 73 items with five rating scales. The items on the leadership competency measurement are categorized into three content domains: core personality, values, and leadership skills. In the pilot study, the scale was checked by a counselling expert, two school management experts, and a university lecturer whose expertise was in Management in Education. Their suggestions were taken into account and the scale revised accordingly. The procedures for scale development consisted of three stages: scale development (Stage 1), scale refinement and purification (Stage 2), and scale validation (Stage 3).

Tubbs's leadership competency model was used as a conceptual model for the assessment tool. This model was derived from interviews and discussions in organizations with over fifty-thousand leaders in North America, South America, Europe, and Asia over the past 35 years (Tubbs & Schulz, 2006). In the model, Tubbs and Schulz contend that core personality is fixed at a young age and is unlikely to be changed through leadership development efforts. This is be-

cause value is more flexible than personality, but more resistant to change compared to behavior. Leadership behaviors are the most likely to be changed through leadership development efforts. However, as the target of this study is secondary-school students (young people), personality characteristics, values, and leadership skills would therefore still be underdeveloped. This research explored the measurement of each element in the leadership competency model.

The researchers constructed the items in every domain based on the operational definitions and the blueprint made. Section A in the instrument was used to obtain the demographic data of the respondents. This section consisted of two parts, which were gender and age. Section B was composed of 20 items to measure the level of agreement of possessing some aspects of personality (emotional stability, openness to experience, conscientiousness, extraversion, and agreeableness), and Section C outlined 18 items to measure the importance of 18 values (peace, wealth, happiness, success, friendship, independence, freedom, justice, joy, self-direction, obedience, recognition, family, power, truth, protection, influence, and status) in performing leadership tasks. Value systems are directly related to the individual's world view, which is primarily a person's conscious beliefs about how things are or should be (Hofmann, 2009). The more important the value is to a person, the more the value will influence that person's behavior as a leader. Section D meanwhile measured the level of agreement of respondents on their ability to perform the skills in leadership such as understanding the big picture, attitude, driving force, creativity and innovation, teamwork and followership, communication, and leading change as suggested by Tubbs and Schulz (2006).

After the instrument was developed, it was tested and refined to ensure validity and reliability. Twenty items were prepared and titled by the construct names (emotional stability, openness to experience, conscientiousness, extraversion, and agreeableness). The experts were invited to write comments, concerns, suggestions, and/or questions on the ruled side of the index cards. The items were then sorted and the envelopes returned to the researchers. The researchers took into account Haladyna's (1994) recommendation that 70% of the items need to meet the standardized test-development criteria. The process was repeated for the next dimensions, values and leadership skills. After receiving all envelopes from the experts, the researchers analyzed the comments and suggestions and compared original placements with those suggested by the experts. The items were then rearranged into the constructs. Initially, the questionnaire was written in English, but the Malay language was later added to help students understand the items better. The items were then back-translated to English by language experts to ensure it was suitable for secondary-school students.

Items were quantitatively analyzed using Winsteps to assess their suitability. In item polarity, information on correlation coefficient of measurement-point (PTMEA correlation) is needed. If the responses to a rating scale are incorrectly coded, the PTMEA correlation value will not positively correlate with the latent trait. Therefore, the researcher would need to verify that the data is correctly coded before going further (Linacre & Wright, 2012). Principal component analysis of residual (PCAR) was used to identify whether the items in the construct were unidimensional. Item reliability and separation indices were used to check the extent the empirical scale of items in domain of M3SLS was consistent with the instrument developers' expectation. Person reliability and separation indices meanwhile were used to determine whether items in each construct were able to replicate the layout of the respondents. Finally, item infit mean-square (MNSQ) was used to determine how well the respective items fit the Rasch Rating Scale model.

RESULTS AND DISCUSSION

Scale Refinement

From the envelopes that were returned, the experts commented that two items were considered not relevant as they were in a negative form. The researchers changed the items into positive forms and maintained these in the M3SLS. An Educational Management expert commented that some items in the instrument were "double barreled" in nature and suggested that the researchers split these into two or use only one item for each statement. Most of the items were placed in the correct constructs with the frequency in the M3SLS meeting the standardized test development criteria of 70%. This means that two of the three experts agreed on the item placement (Haladyna, 1994). None of the items was recommended for deletion; therefore the number of items remained the same at 73.

Item Polarity

Table 1 shows that all the values of PTMEA correlation in M3SLS were positive. There were eight items with PTMEA correlation below .20 (ES4CP16, CS4CP18, ES1CP1, PLE3V2, AFF3V17, EXT4CP19, AGR4CP20, and CPT2V14) with the maximum PTMEA correlation at .54. The positive value of PTMEA correlation proves that the items were finely constructed (Bond & Fox, 2007). A positive value of PTMEA correlation is able to discriminate or differentiate the level of leadership competencies held by the respondents. A high PTMEA correlation meanwhile indicates that an item is able to distinguish between the ability of respondents (Linacre, 2003). Any item with a score of below .20 has to be revised.

TABLE 1
PTMEA correlation for items

Item	PTMEA correlation	Label of items	Item	PTMEA correlation	Label of items	Item	PTMEA correlation	Label of items
16	.06	ES4CP16	35	.29	SEC1V15	47	.41	ATT2LS9
18	.07	CS4CP18	3	.30	CS1CP3	68	.41	ATT5LS30
1	.12	ES1CP1	30	.30	AUT3V10	12	.42	OE3CP12
22	.12	PLE3V2	39	.30	BIG1LS1	36	.42	SEC3V16
37	.12	AFF3V17	2	.31	OE1CP2	40	.42	ATT1LS2
19	.13	EXT4CP19	7	.31	OE2CP7	53	.42	BIG3LS15
20	.13	AGR4CP20	27	.31	AUT2V7	62	.42	DF4LS24
34	.13	CPT2V14	28	.33	SEC2V8	66	.42	TF4LS28
9	.20	EXT2CP9	32	.33	CON2V12	13	.43	CS3CP13
29	.20	PLE1V9	26	.34	AUT1V6	14	.43	EXT3CP14
38	.21	CPT3V18	8	.35	CS2CP8	52	.43	TF2LS14
24	.22	CPT1V4	50	.35	CI2LS12	57	.43	CI3LS19
55	.22	DF3LS17	70	.35	COM5LS32	42	.45	COM1LS4
11	.24	ES3CP11	31	.36	CON1V11	69	.45	DF5LS31

(Table 1 continues)

Table 1 (continued)

Item	PTMEA correlation	Label of items	Item	PTMEA correlation	Label of items	Item	PTMEA correlation	Label of items
17	.25	OE4CP17	64	.36	CI4LS26	73	.45	TF5LS35
6	.26	ES2CP6	71	.36	CI5LS33	41	.46	DF1LS3
21	.26	CON3V1	43	.37	CI1LS5	60	.47	BIG4LS22
23	.26	PLE2V3	44	.37	LC1LS6	45	.48	TF1LS7
54	.26	ATT3LS16	46	.37	BIG2LS8	5	.49	AGR1CP5
25	.27	AFF1V5	56	.37	COM3LS18	63	.50	COM4LS25
48	.27	DF2LS10	49	.40	COM2LS11	58	.51	LC3LS20
4	.28	EXT1CP4	51	.40	LC2LS13	67	.52	BIG5LS29
72	.28	LC5LS34	61	.40	ATT4LS23	65	.53	LC4LS27
33	.29	AFF2V13	10	.41	AGR2CP10	59	.54	TF3LS21

Note. AFF = affiliation; AGR = agreeableness; ATT = attitude; AUT = autonomy; BIG = understand the big picture; CI = leader's voice effective; COM = communication; CON = conformity; CPT = competition; CS = conscientiousness; DF = driving force; ES = emotional stability; EXT = extraversion; LC = leading change; OE = openness to experiences; PLE = pleasure; SEC = security; TF = teamwork and followership.

Principal Component Analysis of Residual (PCAR)

From PCAR analysis, the raw variance explained by items was 29.8% with the unexplained variance in the first contrast at 70.2% as shown in Table 2. The standardized residual variance explained by measures of this data (29.8%) and expectation of model (29.4%) were almost similar. The analysis showed that the Rasch dimension dominated almost three times the secondary dimension, which was notable. A secondary dimension must have the strength of at least three items. If the first contrast has eigenvalue units of less than 3, then the test is probably unidimensional (Linacre, 2003). However, in the pilot study the eigenvalue of the first contrast is 5.3, which indicates that the five items may form another dimension.

TABLE 2
Table of standardized residual variance (in eigenvalue units)

		Empirical		Modeled
Total raw variance in observations	104.0	100.0%		100.0%
Raw variance explained by measures	31.0	29.8%		29.4%
Raw unexplained variance (total)	73.0	70.2%	100.0%	70.6%
Unexplained variance in 1st contrast	5.3	5.1%	7.3%	

Table 3 shows that the five items labelled A, B, C, D, and E measured the third domain, leadership skills. The items labeled a, b, c, d, and e, on the other hand, measured the second domain, values. Five items were enough to split them into a separate instrument (Linacre, 2003). Therefore, it was suggested that the three domains be analyzed separately in the real study.

TABLE 3
Standardized residual loadings for items

Loading	Measure	Infit MNSQ	Outfit MNSQ	Entry number	Item
.56	.08	.85	.85	A	ATT5LS30
.53	1.15	.81	.81	B	DF5LS31
.49	–.01	.83	.82	C	TF5LS35
.49	–.06	.96	.96	D	ATT4LS23
.40	.43	.93	.92	E	LC1LS6
–.48	.85	1.35	1.36	a	AFF3V17
–.47	.34	1.34	1.34	b	CPT3V18
–.42	–.56	1.29	1.28	c	PLE1V9
–.42	.58	1.26	1.25	d	CPT2V14
–.40	–1.33	1.87	1.75	e	PLE2V3

Note. MNSQ = mean square. AFF = affiliation; ATT = attitude; CPT = competition; DF = driving force; LC = leading change; PLE = pleasure; TF = teamwork and followership.

Item Reliability and Separation Indexes

Table 4 shows the summary of item statistics in Winsteps. The item reliability for 73 items in M3SLS was .98. This high item reliability is due perhaps to the wide difficulty range of items and large sample size. Winsteps' item reliability has no traditional equivalent. When the value is high, it indicates the sample size is enough for stable comparisons between items (Linacre & Wright, 2012). The item separation was 7.49. The higher the number, the more confidence the researcher can place in the replicability of item placement across other samples (Bond & Fox, 2007). High item separation (> 3, item reliability > .9) implies the person sample is enough to confirm the item difficulty hierarchy, which is the construct validity of the instrument (Linacre & Wright, 2012). The separation value of item was 7.49 indicating that personality items in this scale can be statistically differentiated to seven levels of difficulty.

TABLE 4
Summary of item statistics

					Infit		Outfit	
	Raw score	Count	Measure	Model error	MNSQ	ZSTD	MNSQ	ZSTD
M	560.4	219.0	.00	.08	1.04	.0	1.02	–.2
SD	106.3	.0	.65	.01	.35	3.4	.32	3.3
Max	781.0	219.0	1.32	.11	2.16	7.8	1.97	6.3
Min	320.0	219.0	–1.58	.07	.47	–7.2	.47	–7.2

Real RMSE .09	Adjusted *SD* 0.65	Separation 7.49	Item reliability .98
Model RMSE .08	Adjusted *SD* 0.65	Separation 8.23	Item reliability .99
SE of item mean = 0.08			

Note. MNSQ = mean square; ZSTD = standardized fit statistics. Umean = .000; Uscale = 1.000. Item raw score-to-measure correlation = –.99. Data points: 15987. Log-likelihood chi-square: 39431.75 with 15736 *df*. *p* = .0000.

Person Reliability and Separation Indexes

The person raw score reliability index was .89, determined through the internal consistency method. Therefore, the person reliability is equivalent to Cronbach's alpha. As stated by Chua (2006), the person reliability is considered satisfactory when the alpha value is within .65 and .95. A higher person reliability might be due to better sample targeting, wider ability range of respondents, and longer instrument (Linacre & Wright, 2012). The person separation index value of 2.86 indicates that there are three levels of respondent ability identified in this pilot study. High person separation (> 2, person reliability > 0.8) with a relevant person sample implies the instrument is sensitive enough to distinguish between high and low performers (Linacre & Wright, 2012; Sherron, 2000). Table 5 shows the summary of person statistics in Winsteps.

TABLE 5
Summary of person statistics

					Infit		Outfit	
	Raw score	Count	Measure	Model error	MNSQ	ZSTD	MNSQ	ZSTD
M	189.0	73.0	.70	.13	.92	–.6	.91	–.7
SD	23.0	.0	.43	.01	.28	1.8	.27	1.8
Max	265.0	73.0	2.51	.20	1.49	2.7	1.76	3.9
Min	138.0	73.0	–.16	.13	.50	–3.9	.50	–3.8

Real RMSE .14	Adjusted *SD* 0.40	Separation 2.86	Item reliability .89
Model RMSE .14	Adjusted *SD* 0.40	Separation 2.99	Item reliability .90
SE of item mean = 0.03			

Note. MNSQ = mean square; ZSTD = standardized fit statistics. Deleted: 43 persons. Person raw-score-to-measure correlation = 1.00. Cronbach's alpha person raw score reliability = .89.

Item Fit

The infit mean square for all items was between .50 and 1.50, which is in the acceptable range (Linacre & Wright, 2012) except for AGR3CP15 (.47), AFF1V5 (1.51), CON3V1 (1.52), EXT4CP19 (1.58), SEC2V8 (1.8), PLE2V3 (1.87), AFF2V13 (2.07), and CPT1V4 (2.16). Items with mean square value of more than 1.5 are unproductive for construction of measurement. However, the items do not degrade the measurement, but still might have to be revised with modification of sentence structure and terms used. Figure 1 shows the item fit map generated by Quest with all items' fit noted to be within the range.

CONCLUSIONS

In conclusion, there are items that need to be revised and modified based on the analysis of item polarity, item fit, and PCAR. The reliability and separation indexes of item and person are

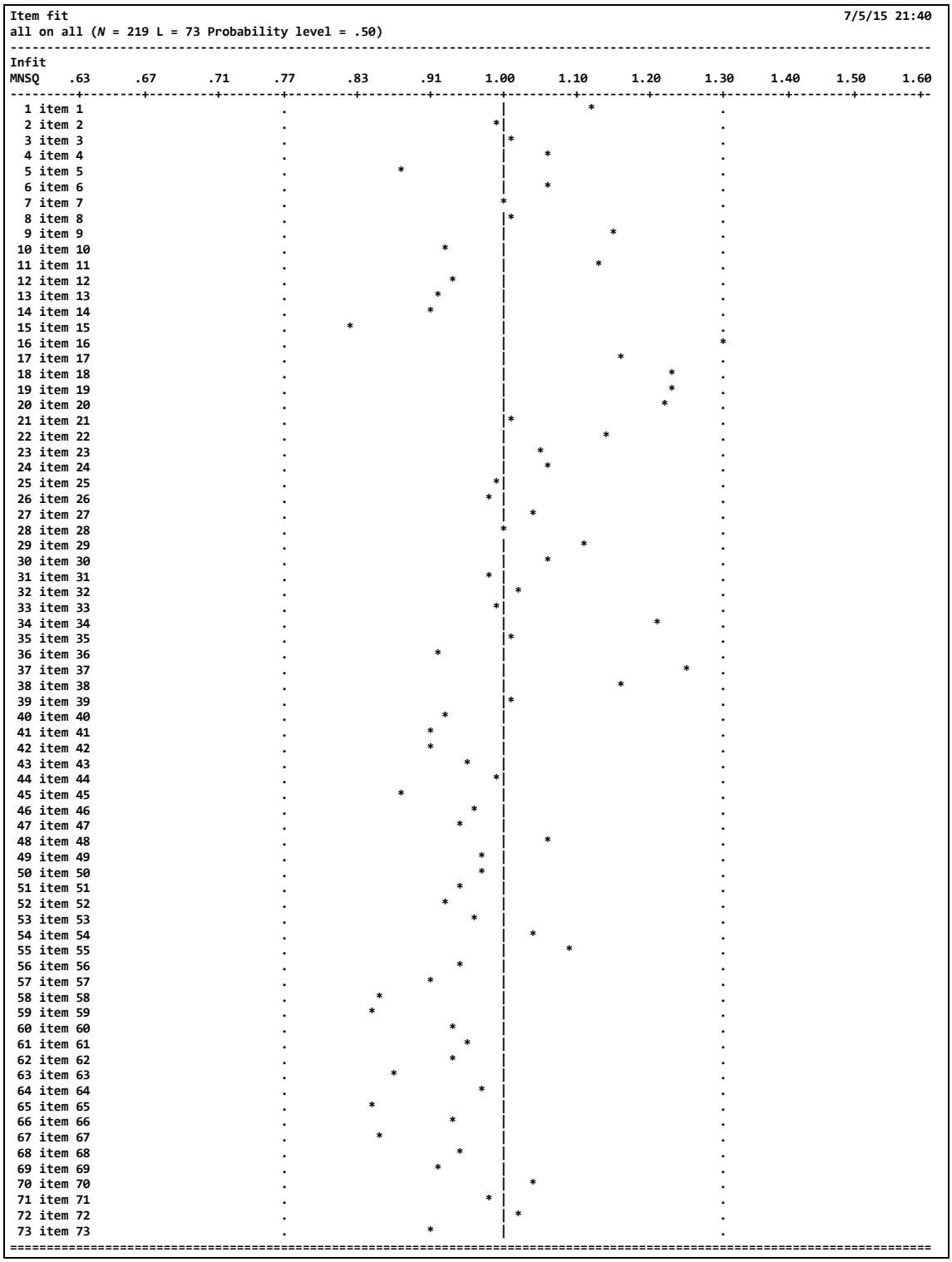

FIGURE 1
Item fit map generated by Quest.

within the acceptable range. Following the pilot test, it is recommended that the instrument is now distributed to a larger population to ensure the stability of the scale. Differential item functioning (DIF) is recommended for inclusion in future studies to ensure the instrument is not biased to any demographic element such as gender and age.

NOTE

1. UPSR, SPM, and KPM are the Malay language equivalent of the English abbreviations PSAT, MCE, and MOE.

ACKNOWLEDGMENTS

We wish to express our gratitude and appreciation to Dr. Stewart Tubbs who gave us the permission to develop an instrument based on Tubbs' model. Furthermore, we would like to acknowledge the generosity, assistance, and support of the staff and students of the School of Psychology and Education, University Malaysia Sabah.

REFERENCES

Alias, B. S., Yussof, A. B., Mustapha, R., & Ibrahim, M. S. (2010). Analisis kompetensi pengetua berdasarkan kualiti peribadi, pengetahuan, kemahiran dan amalan dalam bidang pengurusan sekolah menengah Malaysia [Principal competency analysis based on personal qualities, knowledge, skills and practices in Malaysia secondary school management]. *Jurnal Pendidikan Malaysia*, *35*, 31-41.

Amrin, H. A. (2007). *Penilaian pelajar terhadap kompetensi pemimpin program pendidikan luar di Institutsi Pendidikan Guru Malaysia* [Students' evaluation on the instructors' competencies in the teaching and learning process of the outdoor education program at the Malaysian Teacher Education Institutes] (Unpublished doctoral dissertation). University Putra Malaysia, Serdang, Malaysia.

Bagheri, A. (2011). *Development of entrepreneurial leadership competencies among Malaysian university students* (Unpublished doctoral dissertation). University Putra Malaysia, Serdang, Malaysia.

Bass, B. M., & Avolio, B. J. (1994). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage Publications Inc.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Routledge.

Chua, Y. P. (2006). *Kaedah dan statistik: Asas statistik penyelidikan* [Methods and statistics: Fundamentals statistics research]. Kuala Lumpur, Malaysia: McGraw-Hill.

Don, Y. (2009). *Kolerasi dan pengaruh kompetensi emosi terhadap kepimpinan sekolah: Perbandingan antara sekolah berkesan dengan sekolah kurang berkesan* [Correlation and effect of emotional competence of school leadership: Comparison between the effective school and less effective school] (Unpublished doctorate dissertation). University Malaya, Kuala Lumpur, Malaysia.

Fareeza, A. R. (2010). *Assistant principals leadership capacity building through the Malaysia education quality standards instrument* (Unpublished master's thesis). University Malaya, Kuala Lumpur, Malaysia.

Fertman, C. I., & van Linden, J. A. (1999). Character education for developing youth leadership. *Education Digest*, *65*, 11-16.

Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Erlbaum.

Hine, G. (2011). *Exploring the development of student leadership potential within a catholic school: A qualitative case study* (Doctoral dissertation). Retrieved from http://researchonline.nd.edu.au/cgi/viewcontent.cgi?article=1062&context=theses

Hofmann, J. G. (2009). *The multidimensional structure and function of human values* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (UMI No. 3389616).

Ibrahim, M. Y., & Amin, A. (2014). Model kepemimpinan pengajaran pengetua dan kompetensi pengajaran guru [Instruction leadership of principals and instructional competency of teachers]. *Jurnal Kurikulum & Pengajaran Asia Pasifik*, *2*, 11-25.

Idowu, O. (2011). Evaluation of Mathematics Achievement Test: A comparison between Classical Test Theory (CTT) and Item Response Theory (IRT). *Journal of Education and Social Research*, *1*, 99-106.

Jackson, T. R., Draugalis, J. R., Slack, M. K., Zachry, W. M., & Agostino, J. D. (2002). Validation of authentic performance assessment: A process suited for Rasch modeling. *American Journal of Pharmaceutical Education*, *66*, 233-243.

Linacre, J. M. (2003). Dimensionality: Contrasts & variances. Retrieved January 20, 2015, from http://www.winsteps.com/winman/principalcomponents.htm

Linacre, J. M. (2011). Winsteps further tutorial 1. Retrieved December 10, 2014, from http://www.winsteps.com/a/winsteps-tutorial-further-1.pdf

Linacre, J. M., & Wright, B. D. (2012). *A user's guide to WINSTEPS ministeps Rasch model computer programs*. Chicago, IL: Mesa Press.

Rogers, A. (2001). *A 360-degree peer evaluation of leadership competencies* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (UMI No. 3022343).

Sherron, C. T. (2000). *Psychometric development of the adaptive leadership competency profile*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (UMI No. 3019211).

Spreitzer, G. M., Kizilos, M. A., & Nason, S. W. (1997). A dimensional analysis of the relationship between psychological, empowerment and effectiveness, satisfaction and strain. *Journal of Management*, *23*, 679-704. doi:10.1177/014920639702300504

Tubbs, S., & Schulz, E. (2006). Exploring a taxonomy of global leadership competencies and meta-competencies. *Journal of American Academy of Business*, *8*, 29-35.

Tuleja, E. A., Beamer, L., Shum, C., & Chan, E. K. Y. (2011). Designing and developing questionnaires for translation tutorial. *IEEE Transactions on Professional Communication*, *54*, 392-405. doi:10.1109/TPC.2011.2172834

Yau, O. H. M. (2007). Developing a scale for stakeholder orientation. *Journal of Marketing*, *41*, 1306-1327. doi:10.1108/03090560710821198

Yoon, H. J., Song, J. I. H., Donahue, W. E., & Woodley, K. K. (2010). Leadership competency inventory: A systematic process of developing and validating a leadership competency scale. *Journal of Leadership Studies*, *4*, 39-50. doi:10.1002/jls

Zakaria, S., Aziz, A. A., Mohamed, A., Arshad, N. H., Ghulman, H. A., & Masodi, M. S. (2008). Assessment of information managers' competency using Rasch measurement. *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology*, *ICCIT 2008*, *1*, 190-196. doi:10.1109/ICCIT.2008.387