THE CONSTRUCT VALIDATION OF LEADERSHIP COMPETENCY SCALE (LCS): INTEGRATION OF RASCH AND PLS-SEM

MEI-TENG LING
VINCENT PANG
UNIVERSITY MALAYSIA SABAH

RAMAYAH THURASAMY
UNIVERSITY SAINS MALAYSIA

This paper aims to verify the construct validity of the Leadership Competency Scale by using Rasch analysis and PLS-SEM. The Malaysian Secondary School Students' Leadership Inventory (M3SLI) was developed based on Tubbs and Schulz's (2006) Leadership Competency Model. The Leadership Competency Scale (LCS) is one part of M3SLI. LCS was administered on a sample of 500 students aged 13-18 at six public schools in Sabah, Malaysia. Items were quantitatively investigated using WINSTEPS, based on the Rasch Measurement Model, to ascertain the fitness of items. The Smart-PLS 3.0 was used to assess the PLS-SEM results of the formative measurement models. Six items were omitted from the data set to produce valid measures. Furthermore, all sub-constructs are unidimensional and the category functions are acceptable based on the Rasch analysis. PLS-SEM analysis confirms the significant relationship between the formative indicators. Consequently, the construct validity of Leadership Competency Scale is established. Compliance of Rasch analysis into PLS-SEM takes into account the contribution of errors by every single indicator and provides an interval scale to run PLS-SEM, thereby reinforcing the construct validity of the instrument.

Key words: Rating scale; Rasch analysis; PLS-SEM; Leadership competency; Secondary school.

Correspondence concerning this article should be addressed to Vincent Pang, Faculty of Psychology and Education, University Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia. Email: pvincent@ums.edu.my

According to the Malaysia Educational Blueprint (MEB) 2013-2025, the education system will create formal and informal opportunities for the student to take on leadership roles and to work in teams (Ministry of Education Malaysia, MOE/KPM,[1] 2012). In the MEB, there are four dimensions of leadership, specifically resilience, entrepreneurship, strong communication skills, and emotional intelligence. An entrepreneur takes the initiative to generate and improve his/her own solutions, makes the effort to transform ambition into reality and is willing to invest his/her own resources. A resilient person is able to withstand obstacles and to develop a practical idea. Emotionally stable people can understand and work effectively with others. Furthermore, they are able to influence others positively. People who have strengths in communication skills have the ability to express their opinions and goals clearly in verbal or non-verbal forms. Leadership is often discussed in terms of competencies (Tubbs & Schulz, 2006) which are associated with certain behaviors, values, personal traits, skills, and knowledge (Jokinen, 2005).

The development of the leadership instrument into a measure for children and young people is strongly encouraged (Shaunessy & Karnes, 2004). Furthermore, Ogurlu and Emir (2013) disclose that the screening and identification instrument in leadership for elementary and secondary students is limited. Posner (2012) also states that valid instruments to measure leadership development designed specifically for college students do not exist. The Malaysian Ministry of Education (MOE) acknowledges that leadership skills are important (it is part of students' attributes), but most schools conduct the development of leadership skills in a slightly "laissez-faire" way, without properly identifying and nurturing the leadership potential of the students. A valid and reliable scale to measure secondary school students' leadership competencies is still absent. Furthermore, student leadership is not specifically evaluated in schools. In 2012, a psychometric assessment (PPsi), which contains an aptitude test, personality trait inventory, and interest toward career inventory was launched in every secondary school in Malaysia.

The function of the psychometric assessment was to collect data that contributed to a student's profile. It was believed that complete and detailed information pertaining to student profiling could help educators to identify and enhance the potential of the students. Despite its advantages, the leadership components were not inserted in the PPsi. During the 2006/2007 intake to public universities, a circular was issued by MOE to access the success of students in cocurricular activities in school.

Every student was assessed by the teachers according to his/her attendance, involvement, and achievement in cocurricular activities. The assessment score was one of the requirements for admission to a public university. However, the marks were calculated based on attendance, involvement, and achievement only. The leadership qualities of a student were not assessed within the scale. Furthermore, there was no empirical evidence to prove that a student who was successful in cocurricular activities was necessarily a competent leader. Later in the year 2013, the Assessment of Physical Activity, Sports and Cocurriculum (PAJSK) was introduced. There was only one section that addressed leadership qualities (3% weight). It has been argued that the scores from the assessment were not fully utilized in planning the leadership programs and that they represented no more than just mere records kept by the schools. In foreign countries, there is a number of tools used to measure leadership competencies. State governments and companies always use leadership instruments for staff recruitment (Yoon, Song, Donahue, & Woodley, 2010). Some of the existing instruments are discussed in the following subsections.

Multifactor Leadership Questionnaire (MLQ)

Avolio and Bass (2004) developed and validated MLQ in 2004. The questionnaire is used to assess the range of transformational, transactional, and non-leadership styles (Rowold, 2005). MLQ is the primary measurement tool used in Multifactor Leadership Theory (Tejeda, Scandura, & Pillai, 2001). Bass (1985) suggests that leadership consists of three main constructs: transactional, transformational, and laissez-faire. In addition, he claims that transactional and transformational leadership are multidimensional concepts and one of concepts involves leader behaviors.

MLQ consists of nine items with a 5-point Likert-type scale namely, *frequently*, *fairly often*, *sometimes*, *once in a while*, and *not at all.* It was adopted to measure leadership behaviors and leadership outcomes of 52 project managers with an average age of 39 years in Thailand

(Kedsuda & Ogunlana, 2008). However, in this study, MLQ measurement was not suitable as not all the target respondents had been leaders before. Further, the researcher believes that leaders are not only influenced by natural factors, but can also be nurtured to become competent leaders. Another important consideration is that the authors of MLQ request that researchers do not adjust the measures (Tejeda et al., 2001). However, if the researcher adopts the MLQ for application to secondary school students, then modifications will become necessary. Moreover, the terms used in the questionnaire are too difficult for secondary school students and therefore bias might be introduced by inadvertently changing the perceived meaning of the terms and questions (Tuleja, Beamer, Shum, & Chan, 2011).

Student Leadership Practice Inventory (S-LPI)

The Student Leadership Practice Inventory (S-LPI) was first established by Kouzes and Posner (2006). Posner investigated the psychometric properties of this instrument in 2012. He claimed that a valid instrument designed specifically to measure leadership development for college students did not exist. The S-LPI determines explicit behaviors and actions for students when they are at their "personal best as leaders." S-LPI looks into five leadership practices, namely inspire a shared vision, enable others to act, model the way, encourage the heart, and challenge the process (Kouzes & Posner, 2006). S-LPI consists of 30 behavioral statements with a 5-point scale, ranging from 1 (*rarely or seldom*) to 5 (*very frequently*). There are six statements comprising each of the five leadership practice measures.

Although the psychometric properties of S-LPI were published as mentioned by Kouzes and Posner (2006), double-barreled questions are still visible in some items. For example, Item 2 "I look ahead and communicate about what I believe will affect us in the future" and Item 3 "I look for ways to develop and challenge my skills and abilities" are double-barreled question. "Look ahead" and "communicate" in Item 2, and "to develop," "challenge," "skills," and "abilities" in Item 3, should be expressed in separate items to avoid confusion. The scale with the response of *rarely or seldom* is also double-barreled, which may confuse the students in how to respond, rarely or seldom. The terms used in the S-LPI such as *conviction* (Item 27), *upbeat* (Item 22), *consensus* (Item 21), *fosters* (Item 4), and *dignity* (Item 14) are difficult for local students to interpret. Besides, a Malay version of S-LPI is not available and the authors of S-LPI do not allow researchers to edit or alter the questions in the S-LPI because the language of S-LPI is constructed based on decades of research, and is specifically worded (Kouzes & Posner, 2015).

Malaysian Secondary School Students' Leadership Inventory (M3SLI)

To fill in the gaps identified in the instruments reviewed above, the M3SLI was developed based on Tubbs and Schulz's (2006) Leadership Competency Model, which was itself derived from interviews and discussions with over 50,000 leaders in Asia, Europe, North America, and South America, over the previous 35 years. Tubbs's and Schulz's Leadership Competency Model consists of three concentric circles. The innermost circle contains individual core personality, the second represents the individual values, and the outer circle covers the individual's leadership competencies.

Tubbs and Schulz (2006) suggest that core personality is acquired at a young age and is not likely to change through leadership development efforts while values are more flexible than personality, but more resistant to change than behaviors. They suggest, moreover, that leadership competencies are the most likely to be modified through leadership development programs. Personality characteristics, values, and leadership competencies among secondary school students are still underdeveloped in Malaysia and elsewhere. Therefore, the researchers adopted the Model and developed the M3SLI questionnaire, which consists of 68 items (15 personality, 18 values, and 35 leadership competencies). In this research, only the leadership competencies part (also named as Leadership Competency Scale, LCS) is discussed. LCS contains 35 self-developed perception items with a 4-point rating scale (1 = *strongly agree*, 2 = *agree*, 3 = *disagree*, and 4 = *strongly agree*). LCS comprises seven domains of competencies, namely understanding big picture (UBP), attitude (ATT), creativity and innovation (CI), communication (COM), driving force (DF), leading change (LC), and teamwork (TMW).

Measurement Theory

Measurement experts can choose a Classical Test Theory (CTT) or an Item Response Theory (IRT), or even a combination of the theories (Hambleton & Jones, 1993) to establish the validity and reliability of an instrument. Researchers such as Bechger, Maris, Verstralen, and Béguin (2003) and Ojerinde (2013) recommend that CTT and IRT should be applied together in the practical determination of the psychometric properties of an instrument to confirm a shared basis of item analysis in which the limitation of one is complemented for by the other. A good theory or model can help in understanding the role that measurement errors play in estimating examinee ability. In line with that, a researcher is able to investigate how the contribution of error might be reduced (Hambleton & Jones, 1993). Many testing and measurement textbooks present CTT as the only way to investigate the quality of an instrument, however, IRT offers a sound alternative to the classical approach (Idowu, 2011). In this study, the validity and reliability analyses have been undertaken using both IRT and CTT.

Measurement theory specifies how the latent variables (constructs) are measured (Hair, Hult, Ringle & Sartedt, 2014). In Partial Least Square Structural Equation Modeling (PLS-SEM), reflective measurement or formative measurement are used to measure latent variables. Reflective measures have an error term associated with each indicator, whereas formative measures are assumed to be error free. Sometimes, latent variables were measured by single item rather than multiitem measures (Hair et al., 2014). The relationships between constructs and indicator variables are considered weights for formative constructs, but are loadings for reflective constructs. In PLS-SEM, latent variables are aggregates of observed indicator variables which always involve some degree of measurement error.

The error in the latent variable score therefore induces a bias on the model estimates, where the parameters for the measurement models typically are overestimated (Hair et al., 2014). This statistical summation of Likert-type data is based on the prior assumption that all items are equally difficult to agree with and the space between response options are of equal distance. However, with polytomous data, not all items contribute the same quantity of the latent trait. In the Rasch measurement, fit indices provide information for quality control and measure improvement (Bond & Fox, 2015). The Rasch measurement model is able to convert raw responses into equal-interval Rasch measures in log-odd unit or logit (Bond & Fox, 2015). In other words, this can help PLS-SEM to improve their latent variable scores obtained by the mean of the responses in the underlying indicators.

Prior to the use of PLS-SEM analysis, all the items were diagnosed to ensure that all items fit the statistical, unidimensional, and appropriate category functions. This provides quality control to the measures. Besides, the person measures and standard errors computed by the Rasch model are in equal interval scale. The interval scale is useful to replace the latent variable scores that are claimed to be overestimated by Hair et al. (2014).

METHOD

LCS was administered on a sample of 500 students aged 13-18 at six public schools in Sabah, Malaysia. Items were quantitatively analyzed using WINSTEPS (Linacre, 2003b) based on the Rasch Rating Scale Model (Andrich, 1978) to assess the suitability of items. The Smart-PLS 3 (Hair, Hult, Ringle, & Sartedt, 2017) was used to assess the PLS-SEM results of the formative measurement models. The minimum sample size recommended by Soper (2015) is 100 for the model structure.

RESULTS AND DISCUSSION

Analyzing Using Rasch

Construct Diagnosis

There were five items in each of the seven subconstructs (Figure 1). The mean of item difficulty (0.00 logit) is slightly less than the mean of individual agreement in LC (0.63 logit), COM (0.27 logit), ATT (0.55 logit), CI (0.25 logit), and TMW (0.42 logit). This indicates that the items in COM and CI constructs were slightly easy, while the items in LC, ATT, and TMW were moderately easy to agree with by the respondents involved in this study.

On surveys, a researcher may expect 70% "agreement" due to the normal psychological process of "compliance" (Linacre, 2012b). Conversely, mean of person in UBP (–0.15 logit) and DF (–0.27 logit) are marginally lower than the mean of item difficulty. This indicates that it is slightly harder for the respondents to agree with the items in the two constructs (Table 1).

Item Misfit Diagnosis

Each of the subconstructs was analyzed individually in WINSTEPS. Table 2 shows the fit statistics and PTMEA correlation values for each item in each subconstruct in leadership competencies. Infit MnSq of the items is within 0.6 to 1.4 as recommended by Bond and Fox (2015).

TPM Vol. 24, No. 2, June 2017
269-283

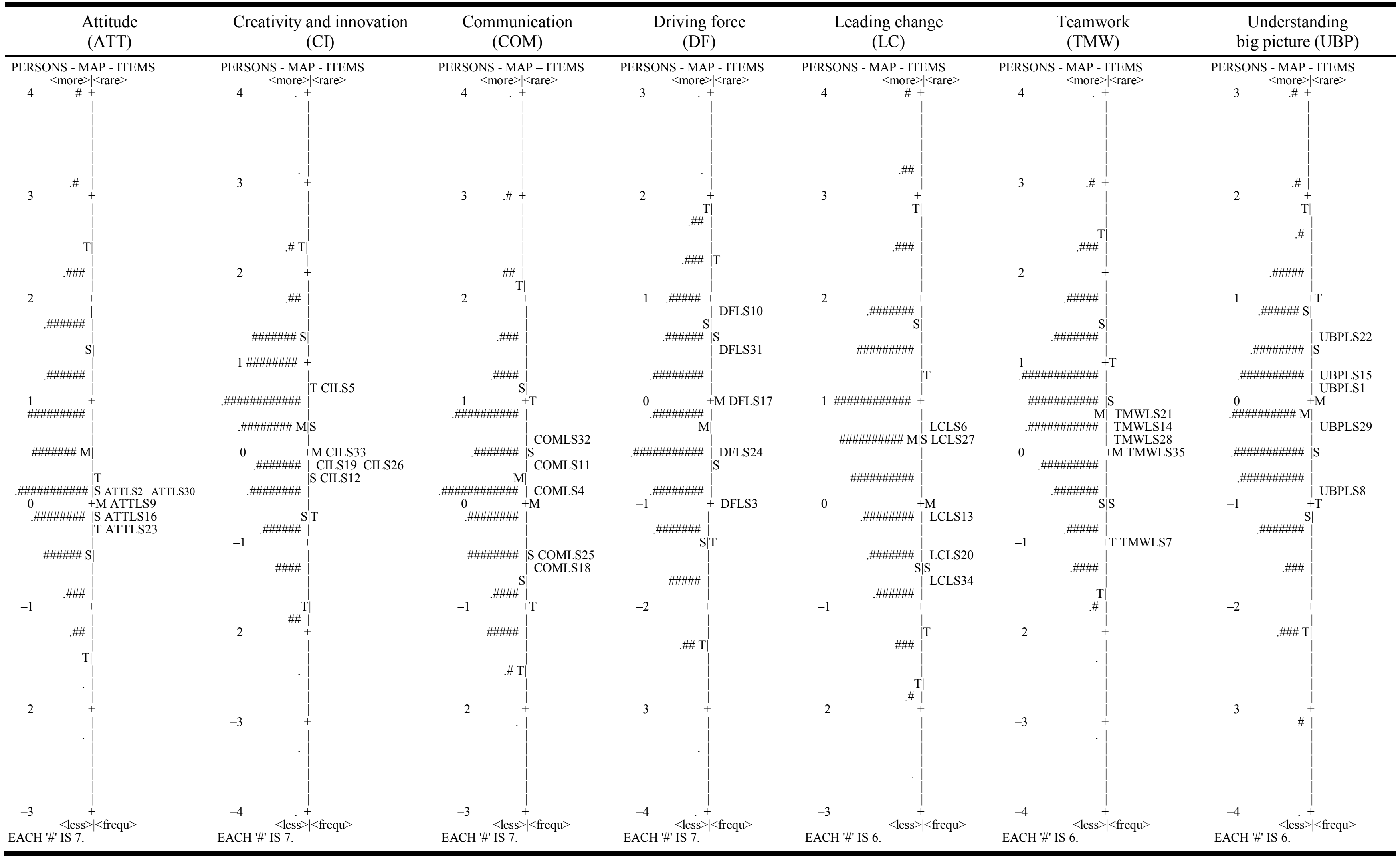

FIGURE 1
Item-person map.

TABLE 1
Effect size of mean of item difficulty and mean of person agreement in seven constructs

Constructs	Mean of person agreement	*SD*	Mean of item difficulty	*SD*	Effect size	Indication
LC	0.63	1.23	0.00	0.60	–0.65	Medium
COM	0.27	1.04	0.00	0.48	–0.33	Small
ATT	0.55	1.63	0.00	0.28	–0.47	Medium
CI	0.25	1.05	0.00	0.33	–0.32	Small
TMW	0.42	1.03	0.00	0.51	–0.51	Medium
UBP	–0.15	1.07	0.00	0.50	0.18	Small
DF	–0.27	1.13	0.00	0.68	0.29	Small

Note. LC = leading change; COM = communication; ATT = attitude; CI = creativity and innovation; TMW = teamwork; UBP = understanding big picture; DF = driving force.

TABLE 2
Item misfit and item polarity

Sub-constructs	Items	Infit		Outfit		PTMEA correlation
		MnSq	Z-Std	MnSq	Z-Std	
Understanding big picture (UBP)	UBPLS1	1.20	3.20*	1.21	3.30*	0.51
	UBPLS8	0.88	–2.20	0.87	–2.30	0.61
	UBPLS15	0.97	–0.50	0.96	–0.70	0.64
	UBPLS22	1.09	1.40	1.05	0.70	0.60
	UBPLS29	0.89	–2.10	0.89	–1.90	0.62
Attitude (ATT)	ATTLS2	0.95	–0.80	0.96	–0.60	0.57
	ATTLS9	1.06	1.00	1.05	0.90	0.59
	ATTLS16	1.34	5.20*	1.33	5.00*	0.48
	ATTLS23	0.83	–2.90	0.84	–2.80	0.61
	ATTLS30	0.80	–3.60	0.81	–3.40	0.67
Driving force (DF)	DFLS10	0.88	–1.90	0.86	–2.20	0.60
	DFLS17	1.31	4.80*	1.30	4.60*	0.56
	DFLS24	0.95	–0.90	0.97	–0.50	0.56
	DFLS3	0.89	–2.00	0.92	–1.30	0.66
	DFLS31	0.96	–0.60	0.96	–0.60	0.63
Communication (COM)	COMLS11	1.00	0.00	0.98	–0.30	0.63
	COMLS18	0.88	–2.10	0.91	–1.50	0.57
	COMLS25	0.82	–3.20	0.84	–2.80	0.57
	COMLS32	1.03	0.60	1.02	0.40	0.61
	COMLS4	1.22	3.70*	1.22	3.50*	0.48

(table 2 continues)

Table 2 (continued)

Sub-constructs	Items	Infit		Outfit		PTMEA correlation
		MnSq	Z-Std	MnSq	Z-Std	
Creativity and innovation (CI)	CILS12	1.01	0.10	1.00	–0.10	0.60
	CILS19	0.93	–1.20	0.94	–1.10	0.59
	CILS26	1.11	1.90	1.10	1.70	0.57
	CILS33	0.97	–0.40	0.97	–0.50	0.59
	CILS5	0.98	–0.40	0.98	–0.30	0.58
Leading change (LC)	LCLS13	1.14	2.40*	1.14	2.30*	0.54
	LCLS20	0.95	–0.90	0.94	–0.90	0.68
	LCLS27	1.01	0.10	1.01	0.20	0.63
	LCLS34	0.87	–2.20	0.84	–2.50	0.69
	LCLS6	1.00	0.10	1.01	0.20	0.59
Teamwork (TMW)	TMWLS14	0.93	–1.30	0.93	–1.20	0.66
	TMWLS21	1.00	0.00	1.01	0.20	0.56
	TMWLS28	0.86	–2.50	0.86	–2.50	0.68
	TMWLS35	1.00	0.10	1.00	0.10	0.56
	TMWLS7	1.19	3.00*	1.24	3.40*	0.45

Note. MnSq = Mean square; Z-Std = standardized fit statistics; * item with Z-std of more than 2.0 are not within the accepted range (Linacre & Wright, 1999).

The outfit and infit Z-Std of all the items are within the accepted range (Linacre & Wright, 2012), except Item UBPLS1 ("I am alert to changes in global needs"), ATTLS16 ("I accept different ideas from members"), DFLS17 ("I show rather than tell the ways"), COMLS4 ("I handle my emotion well"), LCLS13 ("I encourage individual changes in club/class"), and TMWLS7 ("I praise my classmate/club members when they succeed in something"). Linacre (2012a) states that then Z-Std could be ignored if mean-squares are acceptable when there are more than 300 observations. This is because it would probably be too sensitive, that is, "everything misfits."

Nevertheless, for the items under investigation, crossplot was done to check the fit statistics before and after omitting the items. After omitting UBPLS1 in UBP subconstruct, UBPLS8 and UBPLS29, which initially showed Z-Std lower than –2, now are fit to the range. Similarly, after omitting DFLS17, DFLS10 shows Z-Std within the range. Both COMLS18 and COMLS32 show Z-Std within the range after omitting COMLS4. On the other hand, after omitting ATTLS16 and TMWLS7, one of the items in each subconstruct (ATT and TMW), Z-Std is still higher than 2.0. Only after omitting ATTLS9 and TMWLS35, all infit and outfit Z-Std are within the acceptable range.

PTMEA correlation with positive values shows that the items are carefully developed (Bond & Fox, 2001). Minimum PTMEA correlation is 0.45 (TMWLS7), while maximum PTMEA correlation is 0.69 (LCLS34). This indicates that the items contribute to the measurement as PTMEA correlation with positive values is able to discriminate the level of leadership. Zero or a negative value indicates that the respondents or item responses are in conflict with the variables or constructs (Linacre, 2003a).

Unidimensionality Diagnosis

Principal component analysis of residuals (PCAR) of each construct in LCS is shown in Table 3. The analysis shows that the raw variance explained by measures of each construct is between 33.4%-45.2%, which is near to the raw variance predicted by the model. Variance of raw observations between 40%-50% is a typical value (Linacre, 2003a). Noise is present if the variance does not reach 40%. However, the eigenvalue for unexplained variances in the first contrast is less than three. A secondary dimension must have the strength of at least three items, therefore the eigenvalue of less than three indicates that the test is almost certainly unidimensional (Linacre, 2005). After the five misfit items were omitted from the data set, the variances increase.

TABLE 3
Principal component analysis of residuals (PCAR) of each construct in LCS

Subconstruct	Raw variance explained by measures (%)		Unexplained variance in 1st contrast (eigenvalue)	
	Before	After	Before	After
UBP	40.9	46.2	1.4	1.4
ATT	33.4	48.3	1.6	1.6
DF	44.4	50.4	1.6	1.6
COM	38.3	45.8	1.4	1.4
CI	36.2	–	1.5	–
LC	45.2	50.9	1.5	1.6
TMW	40.3	41.2	1.6	1.4

Note. UBP = understanding big picture; ATT = attitude; DF = driving force; COM = communication; CI = creativity and innovation; LC = leading change; TMW = teamwork.

Separation Diagnosis

Table 4 shows the summary of the item statistics in WINSTEPS. The item reliability for 35 items in Leadership Competencies Assessment is 0.99, which indicates that the sample size is sufficient for constant comparisons between items (Linacre & Wright, 2012). The item separation and reliability is 8.11, and 0.99, respectively. The higher the number, the better it is in the replicability of the item placement across other samples (Bond & Fox 2007). The separation value of items is 8.11, which means that the leadership competencies items in this scale can be statistically differentiated to eight levels of difficulties.

The person raw score reliability is 0.89 (Table 5). The person reliability is considered satisfied as Chua (2006) states that alpha values of within .65 and .95 are considered acceptable. Higher person reliability can be due to good sample targeting, broader ability range of respondents, and a lengthier instrument (Linacre & Wright, 2012). A person separation of 2.64 (Table 5) and person reliability of 0.87 imply that the instrument is sensitive to differentiate between high, medium, and low performers.

TABLE 4
Item separation and reliability

	Raw score	Count	Measure	Model error	Infit		Outfit	
					MnSq	Z-Std	MnSq	Z-Std
Mean	799.4	500.0	.00	.06	1.00	–.1	1.00	–.0
SD	149.9	.0	.47	.00	.12	2.1	.12	2.1
Max	1093.0	500.0	1.16	.06	1.33	5.4	1.31	5.0
Min	442.0	500.0	–.95	.06	.82	–3.5	.83	–3.1
Real RMSE		.06	Adjusted *SD*	.47	Separation	8.11	Item reliability	.99
Model RMSE		.06	Adjusted *SD*	.47	Separation	8.33	Item reliability	.99

Note. RMSE = root mean square error; MnSq = Mean square; Z-Std = standardized fit statistics.

TABLE 5
Person separation and reliability

	Raw score	Count	Measure	Model error	Infit		Outfit	
					MnSq	Z-Std	MnSq	Z-Std
Mean	56.0	35.0	.20	.21	1.00	–.2	1.00	–.2
SD	14.4	.0	.67	.03	.44	2.1	.43	2.1
Max	103.0	35.0	4.07	.71	2.50	4.7	2.44	4.6
Min	22.0	35.0	–1.45	.20	.13	–7.0	.14	–7.0
Real RMSE		.24	Adjusted *SD*	.62	Separation	2.64	Item reliability	.87
Model RMSE		.22	Adjusted *SD*	.63	Separation	2.91	Item reliability	.89
Person raw score-to-measure correlation = .99								
Cronbach's alpha (Kr-20) person raw score reliability = .89								

Note. RMSE = root mean square error; MnSq = Mean square; Z-Std = standardized fit statistics.

Category Function Diagnosis

In the level of agreement scale, the observed average value increases from –.51 to .84 logit when the category value increases from zero to three (Table 6). For the step calibration value, it is in ascending order from –1.31 to 1.11 when the category value increases. Step calibration with ascending order reflects the equal probability of observances (Linacre, 2002).

It can indicate that all categories represent a similar division of the latent variable to a concept that is well defined in the thoughts of the respondents. Figure 2 shows the category probability of response in level of agreement. The probability curve for four categories shows that all categories show clear curves. Besides, the segments of the categories are relatively similar. It can be assumed that the category measure is acceptable.

TABLE 6
Summary of category structure for level of agreement on a 4-point rating scale

Category label	Observed		Observed average	Sample expect	Infit MnSq	Outfit MnSq	Structure calibration	Category Measure
	Count	%						
0	2221	13	–.51	–.54	1.04	1.05	None	(–2.55)
1	6041	35	–.11	–.08	.94	.95	–1.31	–.70
2	5775	33	.40	.37	.90	.88	.19	.77
3	3463	20	.84	.86	1.04	1.05	1.11	(2.44)

Note. MnSq = Mean square; Z-Std = standardized fit statistics.

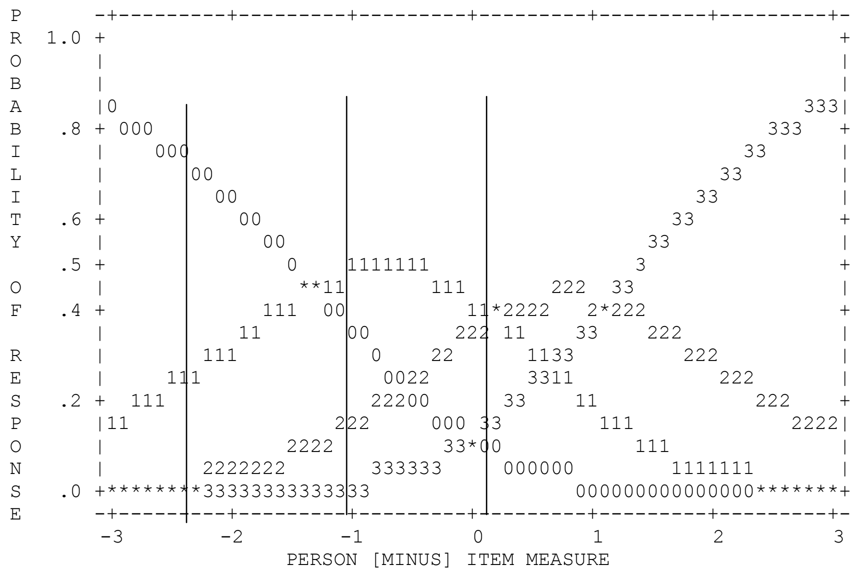

FIGURE 2
Category probability of response in level of importance on a 4-point rating scale.

Analyzing using Rasch and PLS-SEM

The measurement model is an element of a path model that contains the indicators and their relationships with the constructs and is also known as outer model in PLS-SEM (Hair et al., 2014). The structural model defines the relationships between latent variables (constructs). However, hypothesis tests involving the structural relationship among constructs will only be reliable or valid when the measurement model explains how these constructs are measured. Leadership competencies are considered as hierarchical component models (HCM) because they involve testing second-order structures that contain two layers of components (Hair et al., 2014). Leadership competencies can be defined at different levels of abstraction. Specifically, leadership competencies can be represented by seven first-order components that capture separate attributes of leadership competencies (Figure 3). In the context of competencies, these might include UBP, ATT, DF, CI, COM, LC, and TMW.

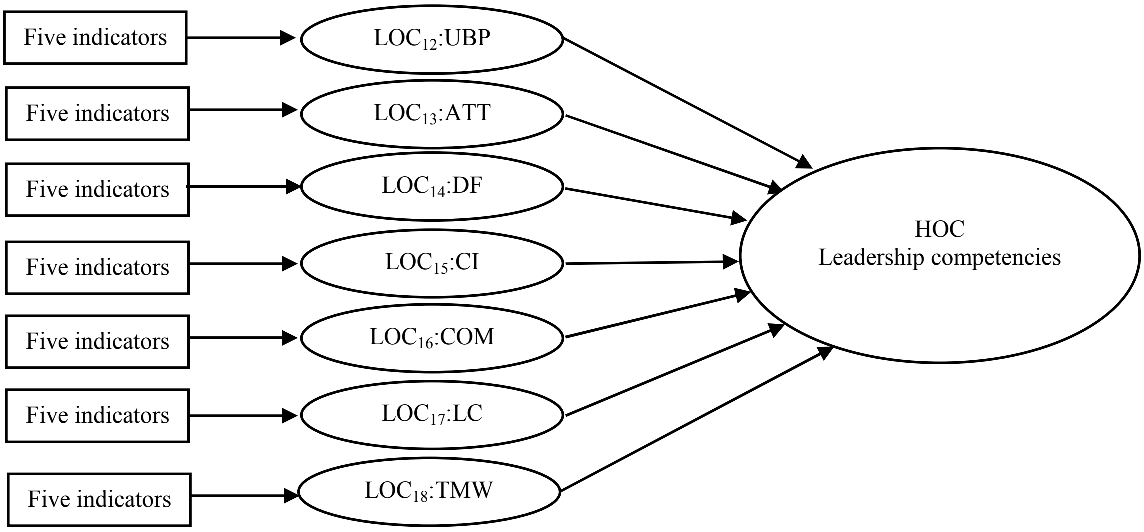

FIGURE 3
Hierarchical component models (HCM) of leadership competencies.
HOC = higher-order construct; LOC = lower-order component.
UBP = understanding big picture; ATT = attitude; DF = driving force; CI = creativity and innovation;
COM = communication; LC = leading change; TMW = teamwork.

Instead of modeling the attributes of leadership competencies as drivers of the respondent's overall satisfaction on a single construct layer, higher-order modeling involves summarizing the lower-order component (LOCs) into a single multidimensional higher-order construct (HOC). This modeling approach improves theoretical parsimony and lessens the complexity of the model (Hair et al., 2014). In this study, a reflective-formative type of HCM is implemented. All the numbers of indicators per LOC are equal to avoid bias. Formative relationships between the LOCs and the HOC reveal the relative contribution of each LOC in explaining the HOC. Hence, HOC is a general construct that represents all the LOCs, thereby fully mediating their relationships with other target variables.

Traditionally, PLS-SEM is used to construct measures for each variable and then determine the causal links between the variables. In this study, Rasch measurement is used to construct measures while PLS-SEM is used to determine the causal links between the variables. The person measures and standard errors computed by WINSTEPS are in equal interval scale. The person measures and the standard errors were then imputed into SMART-PLS to replace LOCs scores. The formative model (between LOCs and HOC) is evaluated as usual. Therefore, in this study, there is only one type of measurement model to evaluate in PLS-SEM. Formative measurement models (between LOCs and HOC) are evaluated by looking at collinearity among indicators and significance and relevance of outer weights.

Formative Measurement Models for Collinearity Issues

In this study, each of the LOCs (UBP, ATT, DF, CI, COM, LC, and TMW) is represented by a single composite indicator (person measure with standard error instead of factor loading score) to construct formative HOC (leadership competencies). To assess the level of collinearity, tolerance is computed. Tolerance represents the amount of variance of one formative indicator that is not explained by the other indicators in the same block (Hair et al., 2014). From the analysis, all the variance inflation factors (VIF) are greater than 1: UBP (1.537), ATT (1.770), CI (1.861), COM (1.770), DF (1.852), LC (1.583), and TMW (1.804). The level of collinearity as indicated by a tolerance value is within the range 0.20 to 5. Therefore, none of the indicators needs to be removed.

Significance and Relevance of the Formative Indicators

When there is no critical value of collinearity (VIF < 5), the significance of outer weights is analyzed and the formative indicators' absolute and relative contribution are interpreted. From the analysis, all the outer weight are significant (Table 7). Therefore, the interpretation of the outer weight's absolute and relative size is continued. Looking at the significance levels, all formative indicators are significant.

TABLE 7
Outer weight significance testing results

Formative constructs	Formative indicators	Outer weights (outer loadings)	*t*	Significance level	*p*	CI
Leadership competencies	ATT	0.139 (0.717)	15.599	***	.000	[.674, .762]
	CI	0.221 (0.776)	13.686	***	.000	[.752, .814]
	COM	0.189 (0.747)	21.262	***	.000	[.708, .789]
	DF	0.223 (0.731)	20.273	***	.000	[.685, .774]
	LC	0.214 (0.682)	22.025	***	.000	[.631, .722]
	TMW	0.176 (0.738)	17.094	***	.000	[.691, .775]
	UBP	0.210 (0.709)	25.424	***	.000	[.653, .754]

Note. ATT = attitude; CI = creativity and innovation; COM = communication; DF = driving force; LC = leading change; TMW = teamwork; UBP = understanding big picture. Bootstrap confidence intervals for 5% probability of error (α = 0.05).
*** $p < .01$.

CONCLUSION

In relation to this study, construct validity in Rasch analysis is established by item misfit diagnosis, construct diagnosis, unidimensionality, separation, and categorical functions, while construct validity in PLS-SEM is established by collinearity and significance and relevance of formative measurement model. From the result of the Rasch analysis, six items were omitted from the data set to produce valid measures. Otherwise, all subconstructs are unidimensional and the category functions are acceptable. As for the PLS-SEM analysis, it confirms the significant relationship between the formative indicators. Therefore, the construct validity of Leadership Competency Scale is established.

As mentioned before, a good measurement theory or model can help in understanding the role of measurement errors in estimating examinee ability and how the contribution of error can be minimized (Hambleton & Jones, 1993). Compliance of Rasch analysis into PLS-SEM should take into account the contribution of errors by every single indicator and provide interval scale to run PLS-SEM. Bond and Fox (2015) argue that analyses based on the Rasch model are not meant to replace statistical techniques, but instead are necessary prerequisites to those techniques. This is because interval level measures are exactly the sort of data clearly required by many data analysis techniques, including PLS-SEM. Furthermore, the vast majority of these statistical techniques require interval data.

The Rasch measurement, which is a basic sort of scientific quality control, provides exactly the tools required for the construction and control of interval level measures in the human sciences (Bond & Fox, 2007). Besides, Rasch analysis is uniquely positioned to fulfil the requirement of converting ordinal data into interval measures. Measures of leader behaviors are potentially restricted by dependence on self-reporting, rather than visible behaviors derived by using a questionnaire. Survey research also conveys extra limitations such as the halo effects of stereotypes and attributions biases. Some responses can be retrospective. Another possible limitation is that M3SLI technically measures the leader's self-perception of leadership; the perceptions do not totally represent the actions. In-depth interviews with more student leadership experts and outstanding student leaders should be conducted to assess the accuracy of the conceptualization of Secondary School Students' Leadership Competency Model so that it is clearer and more effective.

The findings of this research imply that CTT can be integrated with the Rasch measurement model in ensuring the construct validity of an instrument. It is therefore suggested that future psychometric studies take into account the combined use of the two models.

NOTE

1. KPM, PPsi, PAJSK are the Malay language abbreviations which stand for Ministry of Education, psychometric assessment, Assessment of Physical Activity, Sports and Cocurriculum, respectively.

REFERENCES

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, *43*(4), 561-573.

Avolio, B. J., & Bass, B. M. (2004). *Multifactor Leadership Questionnaire. Manual and sampler set* (3rd ed.). Redwood City, CA: Mind Garden.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, NY: Free Press.

Bechger, T. M., Maris, G., Verstralen, H. H. F. M., & Béguin, A. A. (2003). Using Classical Test Theory in combination with Item Response Theory. *Applied Psychological Measurement*, *27*, 319-334. doi:10.1177/0146621603257518

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* New Jersey, NJ: Lawrence Erlbaum Inc. Publishers.

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

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Chua, Y. P. (2006). *Kaedah dan Statistik: Asas Statistik Penyelidikan* [Methods and statistics: Fundamental research statistics]. Kuala Lampur, Malaysia: McGraw-Hill.

Hair, J. F., Hult, G. T. M., Ringle, M. C., & Sartedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: Sage Publications Inc.

Hair, J. F., Hult, G. T. M., Ringle, M. C., & Sartedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.
Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47.
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*(4), 99-106.
Jokinen, T. (2005). Global leadership competencies: A review and discussion. *Journal of European Industrial Training*, *29*, 199-216. doi:10.1108/03090590510591085
Kedsuda L., & Ogunlana, S. O. (2008). Linking personal competencies with transformational leadership style evidence from the construction industry in Thailand. *Journal of Construction in Developing Countries*, *13*(1), 27-50.
Kouzes, J. M., & Posner, B. Z. (2006). *The five practices of exemplary student leadership*. San Francisco, CA: Jossey-Bass.
Kouzes, J., & Posner, B. (2015). *Student Leadership Practices Inventory Permissions FAQs*. Retrieved from http://www.leadershipchallenge.com/Research-section-Use-the-LPI-in-Your-Research.aspx
Linacre, J. M. (2002). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85-106.
Linacre, J. M. (2003a). *Dimensionality: Contrasts & variances*. Retrieved from http://www.winsteps.com/winman/principalcomponents.htm
Linacre, J. M. (2003b). *Winsteps Computer Program* [Computer software]. Chicago, IL: Winsteps.
Linacre, J. M. (2005). *A users's guide to WINSTEPS ministeps Rasch model computer programs*. Retrieved from http://ifile.hkedcity.net/1/001/950/public/Secondary/EI0020070012/winsteps.pdf
Linacre, J. M. (2012a). *Misfit diagnosis: Infit outfit mean-square standardized*. Retrieved from http://www.winsteps.com/winman/diagnosingmisfit.htm
Linacre, J. M. (2012b). *Winsteps Rasch Tutorial 2*. Retrieved from http://www.winsteps.com/a/winsteps-tutorial-2.pdf
Linacre, J. M., & Wright, B. D. (1999). *A user's guide to WINSTEPS ministeps;Rasch model computer programs*. Chicago, IL: MESA Press.
Linacre, J. M., & Wright, B. D. (2012). *A user's guide to WINSTEPS ministeps: Rasch model computer programs*. Chicago, IL: MESA Press.
Ministry of Education Malaysia (MOE). (2012). *Malaysia Education Blueprint 2013-2025*. Kuala Lumpur, Malaysia: Kementerian Pendidikan Malaysia.
Ogurlu, Ü., & Emir, S. (2013). Development of leadership skills scale for students. *International Online Journal of Educational Sciences*, *5*(3), 539-550.
Ojerinde, D. (2013). *Classical Test Theory (CTT) VS Item Response Theory (IRT): An Evaluation of the comparability of item analysis results*. Retrieved from http://ui.edu.ng/sites/default/files/PROF%20OJERINDE'S%20LECTURE%20(Autosaved).pdf
Posner, B. Z. (2012). Effectively measuring student leadership. *Administrative Sciences*, *2*, 221-234. doi:10.3390/admsci2040221
Rowold, J. (2005). *Multifactor Leadership Questionnaire. Psychometric properties of the German translation by Jens Rowold*. Redwood City, CA: Mind Garden.
Shaunessy, E., & Karnes, F. A. (2004). Measuring leadership in children and youth. *Gifted Child Today*, *27*(1), 42-47.
Soper, D. S. (2015). *A-priori sample size calculator for structural equation models* [Computer software]. Retrieved from http://www.danielsoper.com/statcalc
Tejeda, M. J., Scandura, T. A., & Pillai, R. (2001). The MLQ revisited: Psychometric properties and recommendations. *The Leadership Quarterly*, *12*, 31-52. doi:10.1016/S1048-9843(01)00063-7
Tubbs, S. L., & Schulz, E. (2006). Exploring a taxonomy of global leadership competencies and meta-competencies. *Journal of American Academy of Business*, *8*(2), 29-34.
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
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.20176