

AIMC 2017
Asia International Multidisciplinary Conference

**RELIABILITY AND VALIDITY OF A MALAY TRANSLATION OF
THE LESSON STUDY SCALE**

Bitty Ansawi (a)*, Vincent Pang (b)
*Corresponding author

(a) Tuaran Education Office, Sabah, Malaysia, atienee2@yahoo.com, +60198118212

(b) Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia, pvincent@ums.edu.my,
+60198618600

Abstract

Lesson study is a collaborative strategy used by School Improvement Specialist Coaches Plus (SISC+) in low performing schools to improve professional learning community practices. Empirical studies with quantitative approach are needed as past studies mainly used a qualitative approach. This study provides a reliability and validity check on a Malay translation of a Lesson Study Scale which was adapted from a previous study and translated to Malay to cater for the cultural context of the Malaysia education system. Content validity and translation of the scale was followed by a pilot study involving 100 samples of randomly selected English teachers in low performing schools in West Coast Sabah. IBM SPSS 23.0 was employed to carry out the confirmatory factor analysis. Sampling adequacy with KMO test yielded an acceptable 0.742 value and Bartlett's test with a zero significance indicated the absence of identity matrix. The revised scale comprises of a 12-item, 3-dimensions, 5-point Likert scale with accepted reliability and validity. The cumulative Eigenvalues for three dimensions yielded 55.97% and due to low factor loading (<0.40), deletion of three items were done. This study concluded with a reliable and valid revised Lesson Study Scale which can be used in the Malaysian context.

© 2018 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Professional Learning Community; Lesson Study; SISC+; Confirmatory Factor Analysis; Low Performing Schools.

1. Introduction

As a basic activity of science, measurement allows the researchers to obtain knowledge about objects, events, people, and processes (Morgado, Meireles, Neves, Amaral, & Ferreira, 2017). In any given research field, the measurement scale development is a crucial step as it contributes to a valid and reliable finding in an empirical study (Crook, Shook, Madden, & Morris, 2009; Slavec & Drnovsek, 2012). In fact, many scholars argue that the foundation of a scientific research lies in the effectiveness of its measurement (DeVellis, 2003; Netemeyer, Bearden, & Sharma, 2003; Slavec & Drnovsek, 2012) and an important part of latent variables assessment (Reynolds, 2010). The scale development is a complex process with systematic procedures to follow which are based on both theoretical and methodological rigor (Clark & Watson, 1995; DeVellis, 2003; Nunnally, 1978; Pasquali, 2010). It is necessary to abide and follow the steps in scale development in order to produce a reliable and valid measurement scale. The quantitative nature of human behaviour and the use of a measurement instrument to understand human behaviour are an integral component of the social science research. In order to understand reality, an empirical-analytical approach or a positive paradigm to understand reality (Smallbone & Quinton, 2004). The reliability and validity of the measurement instruments used in social science studies is highly important.

1.1. Reliability and Validity

Reliability has to do with the ability of a measurement to be repeated again and again. It suggests that if the measurement was carried out by a different researcher in a different event, under a different condition, using supposedly alternative instruments, then it would be measuring the same thing. This, according to Nunnally (1978) represents the stability of the measurement over various conditions, yet attaining the same result. Bollen (1989) refers to this as the consistency of measurement. In general, the range of reliability coefficients is between zero to one, with a greater number representing better reliability and in contrast, a smaller number means weaker reliability. Drost (2011) explained that there are mainly three concerns when conducting a reliability test: equivalence, the stability over time and the internal consistency. Figure 01 presents the methods to show how these concerns are seriously given attention to.

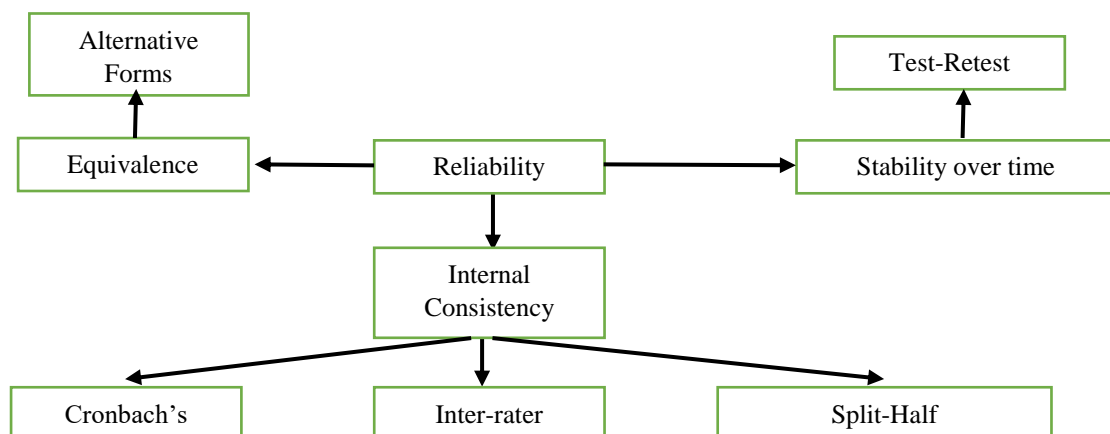


Figure 01. Types of Reliability Source: Drost (2011)

Equivalence suggests the presence of an substitute form of measurement whereby it estimates behaviour that is collected at a different time (Bollen, 1989). The stability over time is determined by the use of test-retest whereby its reliability pertains to the temporal stability of the test as the test is carried out in different sessions. This means that the test is administered to a group of respondents and again, on a later date, repeated to the same group. Then, the correlation between the test and retest scores is determine. A low correlation indicates a considerable measurement error (Drost, 2011). The measurement of the internal consistency assesses the consistency within the measurement test and decide the effectiveness of a group of items to measure a distint behaviour. For internal consistency, Cronbach's Alpha, inter-rater and split-half are the measures normally used. Cronbach's Alpha determines the internal consistency, which according to Nunnally (1978), requires a value of 0.70 and above to indicate acceptable reliability. As for the inter-rater reliability, a number of raters or judge are required to measure a particular behaviour and the combined judgment or ratings assess the internal reliability of the measure (Rosenthal & Rosnow, 1991). In the split-half method, the assumption is that, the measurement consists of a number of item which can be divided into two new measures. The correlation between these two new measures would then determine the reliability of the whole test (Nunnally, 1978). The composite reliability of the ratings is calculated using the Spearman-Brown formula (Drost, 2011). Validity is another aspect of the measurement scales that needs to be greatly considered. An instrument may be reliable despite it being valid but on the other hand, a valid instrument requires that the instrument is reliable. According to Drost (2011), validity ensures that the scale truly measure what it intends to measure. There are four types of validity (Cook & Campbell, 1979). The statistical conclusion validity is about the relationship being assessed whereby it relates to the inferences whether it is reasonable with a presumption that covariance exists at a certain level of alpha and also, the variances that were obtained. The statistical conclusion validity can be compromised by many factors such as violation of assumptions, poor statistical power, treatment reliability, measures reliability and random irrelevancies in the experimental setting as well as arbitrary respondents' heterogeneity. In addition, internal validity of the research regards its own validity in the sense that it considers that there is a representative sample of respondents in contrast to a bias sample. Drost (2011) cautioned that there are many factors threatening internal validity such as history, testing, maturation, selection, instrumentation, diffusion of treatment as well as compensatory equalization rivalry and demoralization.

Another aspect of validity is the construct validity that pertains to how well does the transformation or translation of a concepts, ideas or behaviours into functioning and operating terms. In other words, construct validity is about the operationalization of the construct (Trochim, 2006). Construct validity consists of criterion-related validity which can be divided into predictive validity, concurrent validity, discriminant validity and convergent validity while another group is the translation validity consisting of face validity and content validity (Drost, 2011). Translation validity is about the exact or true meaning of the construct determined by face and content validity. Generally, face validity is the subjective judgment on the operationalizing of a construct whereas content validity is a qualitative approach to validity that ensures clarity of the domain and determines that the measures fully represent the domain (Bollen, 1989). The second group; criterion-related validity relates to whether there is a relationship of a test measure with one or more external criteria or referents, as indicated by the inter-correlation between these measures. Predictive validity and concurrent validity are two types of validity relating to criteria. Concurrent validity

of a scale can predict a present or current event while predictive validity can measure the outcomes of future events. Convergent and discriminant validity as a means of determining construct validity was initially proposed by Campbell and Fiske (1959). Convergent is about testing the same “thing” across different measures or manipulations while discriminant validity is about divergence of the distance and being apart of measures and the manipulation of associated but conceptually different “things” (Cook & Campbell, 1979). Lastly, external validity is the extent to which a relationship between the constructs is able to be generalized across persons, settings and times. Generalizing is useful for a well-explained target population which is not the same as generalizing across populations.

1.2. Scale Development

Slavec and Drnovesek (2012) proposed a scale development procedure with ten steps shown in Figure 02 based on a search of the literature on scale development (Nunnally & Bernstein, 1994; Bagozzi & Edwards, 1998; Netemeyer et al., 2003; DeVellis, 2003).

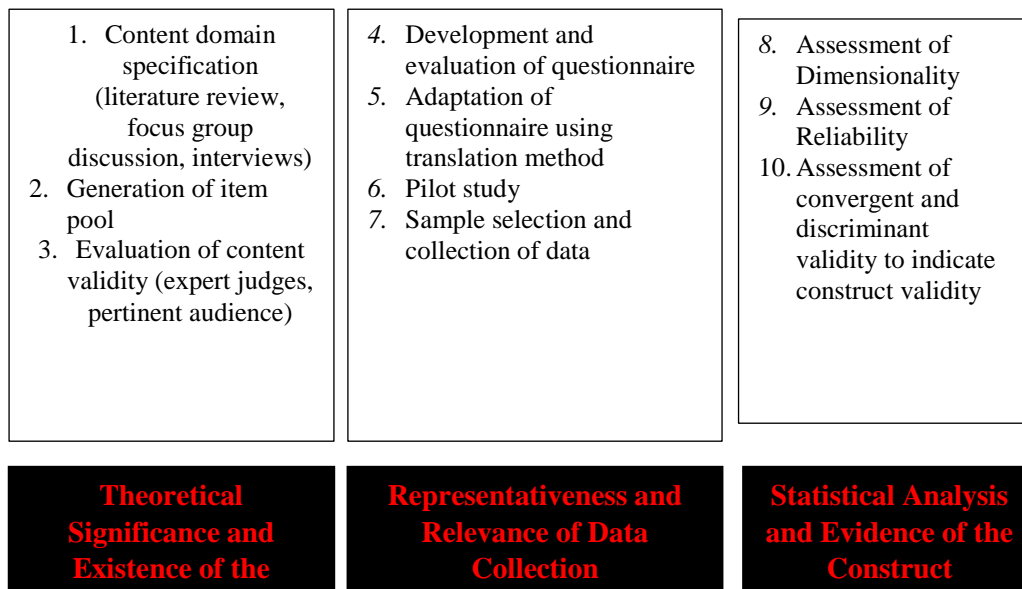


Figure 02. The 10-Step in a 3-Phase Scale Development Process Source: Slavec and Drnovesek (2012)

The scale development process comprises of three phases with the ten steps spread among these phases. In the first phase identifies the theoretical significance as well as the presence of the construct whereas in phase two, it is to determine data collection in terms of representativeness and relevance, and the last phase is to perform statistical analysis and present statistical results to prove the construct. The first step is about content domain specification whereby a new measure may be created and developed which begins by presenting the domain of the new construct based on an in-depth search and review of the literature (Netemeyer et al., 2003). This is then followed by the second step whereby the item pool is generated to create a new scale. After that, there is the content validity assessment which pertains to the sampling adequacy of the construct’s domain (Nunnally & Bernstein, 1994). This ends the first phase of

scale development. The second phase of scale development is about the representativeness and relevance of data collection. The questionnaire undergoes a development process which may include translation and back-translation but these are optional depending on what the research setting is. However, the fifth step is compulsory when conducting cross-cultural studies as there are multiple languages used, thus, translation and back-translation are needed. It is also recommended to run a pilot study to test the proposed questionnaire. Dillman, Smyth, Christian, and Dillman (2009) explained that the pilot study can help to identify potential problems and also to get a pre-result of the newly created measure reliability and the inter-correlation of items in the measure. Sampling is necessary in the second phase to ensure that the result has good quality. The third phase deals with doing analysis statistically and presenting the result to prove the constructs. Dimensionality is about the homogeneity of items (Netemeyer et al., 2003) while reliability is about the ability of the measurement procedure to obtain the same results on repeated trials. Construct validity determines that the construct is providing measurement on what it is purported to measure (Slavec & Drnovsek, 2012). Therefore, the implementation of these steps ensures that the scale developed for use in the study not only reliable but also valid.

In this study, the first stage to develop the scale was skipped as the lesson study scale was taken from a previous study (Mostofo, 2013). However, content validity is necessary to guarantee that the content of the questionnaire to be used in the Malaysian context truly measure what it is supposed to measure. This study also did not complete the third stage as Step 10 will be conducted in the actual study data. This study is limited to report on the result of translation and back translation, dimensionality assessment and reliability assessment.

1.3. Lesson Study Scale

Lesson study originated from the word, *jogyōkenkyū*(授業研究), which is a word in the Japanese language whereby ‘jogyo’ is ‘lesson’ while the word ‘kenkyu’ is ‘study’ or ‘research’ (Mohammad Reza et al., 2010). It has been used in the Japanese education system since the 1900s. An extensive description of lesson study is provided in the book, “The Teaching Gap” which was written by Stigler and Hiebert in 1999 (Hock & Sam, 2010). Zanaton et al. (2014) stated that lesson study enhances not only the teacher’s knowledge about the content of the lesson but also contributes towards improving the teaching methods. Therefore, it enhances both content knowledge and pedagogical knowledge by observing and reflecting on the teaching practices (Dotger et al., 2009; Lewis, 2008; Post & Varoz, 2008). Ewe, Chap, and Munirah (2010) added that lesson study enhances the teachers’ creativity and thinking skills especially in conducting in-depth analysis of the study, promoting innovation and leading to teaching and learning reformation as to seek for answers to an identified issue and also to deepen the understanding skills as well as the teachers and students’ ability. The steps in a lesson are shown in Figure 3.

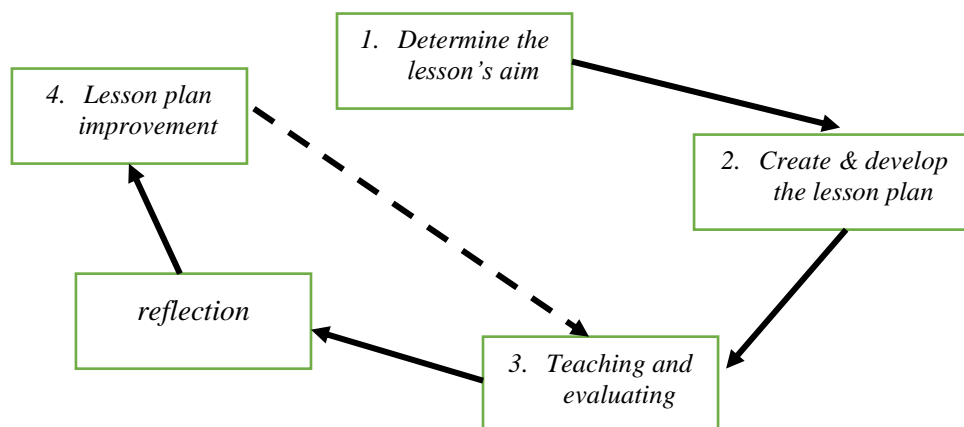


Figure 03. The Cycle of Lesson Study
Source: Zanaton et al. (2014)

The lesson study is implemented in several teachers' teaching sessions in the classroom. They worked collaboratively to determine the objectives of the lesson and held discussion to create and develop the lesson plan with information about the students' learning needs to guide them (Lewis, 2008; Post & Varoz, 2008). The lesson plan encompasses a detailed information on various aspects of the lesson that will be implemented (Fernandez, 2002). Ideas are derived from the teachers' collaborative efforts in preparing the lesson plan. Then, the lesson plan is implemented in the classroom under another teacher's monitoring and observation in order to assess the teaching (Lewis, 2008; Post & Varoz, 2008). After the lesson has been carried out, reflection is done to identify the good and bad points of the lesson which are then used for the improvement of the lesson plan (Marble, 2007).

Mostofo (2013) developed a lesson study scale in 2011 and revised it in 2013 in a study on improving the proficiency of preservice mathematics teacher in instructional planning and implementation. The questionnaire consisted of 15 Likert scale items divided into three sub-dimensions, each with five items. The three constructs were named: collaborative planning, debriefing session and revising lessons. The 5-point Likert scale has a response of "1" as "strongly disagree" to "5" as "strongly agree". Additionally, the Cronbach's Alpha of the lesson plan yielded an overall 0.93 while for the planning collaboratively, debriefing lessons and revising lessons sub-dimensions, the Cronbach's Alpha was 0.77, 0.96 and 0.96 respectively. The initial items in the lesson study questionnaire are presented in Appendix 1.

2. Problem Statement

Teacher as the main agent delivers the curriculum in the classroom through the teaching and learning process. They play a paramount part to determine the success or failure of the national education system. In the past decades, some international assessment programmes like the Programme for International Student Assessment (PISA) as well as the Trends in International Mathematics and Science Study (TIMSS) were developed and utilized to determine the quality of education in a country. Through this assessment, Japan, Singapore and Finland were regarded as the high performing countries in terms of success in education. Stacey (2010) stated that common factors such as teacher quality and learning about what is happening in the classroom are the determinants for accomplishment of these nations in the international

studies. Lesson study which improves teacher's quality and learning what goes on in the classroom is an effective strategy that could potentially lead to greater performance (Cheah & Lim, 2010). In the past few decades, lesson paradigm is accepted as a new paradigm in the teaching and learning context which gained much popularity among practitioners and researchers. It is a Japanese traditional model to create professional knowledge in the school context (Mohammad Reza, Fukaya, & Lassegard, 2010). It normally involves a group of small teachers who held regular meetings where they engage in a collaborative process of planning, implementing, reflecting and evaluating the lessons in the classroom with the purpose of refining and improving the lesson (Hollingsworth & Oliver, 2005). In fact, lesson study is one of the strategies adopted in the implementation of professional learning community among the teachers and their mentor under the School Improvement Specialist Coaches Plus (SISC+) programme (Zuraidah & Muhammad Faizal, 2013). In Malaysia, Cheah and Lim (2010) reported the first attempt of a lesson study approach carried out at the Universiti Sains Malaysia, Penang. The lesson study project was initiated as a means to determine whether it can serve as an alternative model in the professional development of mathematics teachers in two secondary schools located in the district of Kulim. The study showed that lesson study has the potential to promote pedagogical content knowledge among the teachers using group discussion and peer observation (Chiew, 2009).

The SISC+ programme was initiated in the early 2010 as an initiative under the National Key Regional Areas for Government Transformation Programme 1.0 (NKRA GTP1.0). Master coaches were selected from among excellent teachers to provide mentoring and coaching to the teachers in low performing schools. This programme was among the numerous other programmes implemented in school under the SISC+ programme and the professional learning community development (Zanaton, Siti Nor Aishah, Siti Nordiyana, & Effandi, 2014). Most studies on lesson study application in the classroom however, used a qualitative approach either employing interview or observation. Mostofa (2013) attempted to use both qualitative and quantitative methods. Hence, in his study, a lesson study scale was developed and used to gather information on three dimensions: collaborative planning, debriefing session and lesson revision. However, the English version scale was developed for use in the United States. Cross-cultural studies may contribute to getting incorrect conclusions from empirical data if the respondents are multicultural. By not taking into account the differences in the response pattern, then a systematic measurement error has been done which leads to a biased result (Dolnicar & Grün, 2007). Paulhus (1991) explained that a response bias occurs because there is a systematic tendency to answer the items in the questionnaire based on something besides the specific content of the item, that means, what the item should be measuring (p. 17). It creates a response style which is observed by the consistency of the individual to be biased across time and situations. For instance, the respondent tends to answer consistently the last end of a scale, say for example, the fifth on the scale of five. This is a common response style known as extreme response as referring to the tendency of the individuals to agree with items although there are apparent other possibilities (Dolnicar & Grün, 2007). According to Clarke (2001), technically, extreme response style will increase reliability but the validity will decrease. This is because, the frequency distribution becomes skewed to the end, and this causes the standard deviation to increase, thus the correlation becomes smaller. Hence, all correlation type methods such as factor analysis, simple and multiple regression analysis are affected (Heide & Gronhaug, 2005; Rossi, Gilula, & Allenby, 2001). Therefore, it is considered as

inappropriate to adopt and use an existing scale from another culture to be used in another cultural setting. An adaptation of the scale to suit the local culture is required and translating the scale to a language, such as Malay language, which the respondents are more familiar with, ensure that the result from an empirical research is reliable and valid.

3. Research Questions

Based on existing and availability of a lesson study scale previously used in Mostofo (2013), this study is geared to answer the question on the extent to which the lesson study scale can be adapted and used with acceptable reliability and validity for a local study in the Malaysia context.

4. Purpose of the Study

Lesson study is regarded as the unintentional best kept secrets of Japan mainly among the mathematics teacher up till the 1990s (Cheah & Lim, 2010) as it is associated with the success of lessons carried out by Japanese mathematics teachers. Following this success, the teachers and educators in the United States used lesson study as a way to continue the teachers' professional development programme. According to Lewis (2008), this lesson study model has four important features which are: (i) the practicing teachers share long-term goals; (ii) the emphasis is on the importance of the lesson content; (iii) scrutiny is given to the learning and development of the students; and (iv) direct observations of research lessons were carried out. Lim (2006) found in her study that lesson study helped to promote peer collaboration and improved the pre-service mathematics teachers in their knowledge of pedagogical content. Over time, this technique was not confined to mathematics only but have been used in teaching science (Dotger, Moquin, & Hammond, 2009), English (Goh & Fang, 2017) and other subjects. Hence, more studies on lesson plan implementation need to be carried out. The use of a qualitative approach to study lesson study is limited in terms of generalization as compared to using a quantitative approach. Thus, by developing a lesson study measurement scale in Malay version will help to expand more studies within the Malaysian context in terms of coverage and extent of study. Therefore, this study was carried out for the purpose of scale development, particularly a Malay version of the Lesson Study Scale, adapted from Mostofo (2013).

5. Research Methods

This is a cross-sectional study of English teachers in low performing schools in the West Coast areas of Sabah, Malaysia. The scale development activities are summarized in Figure 4.

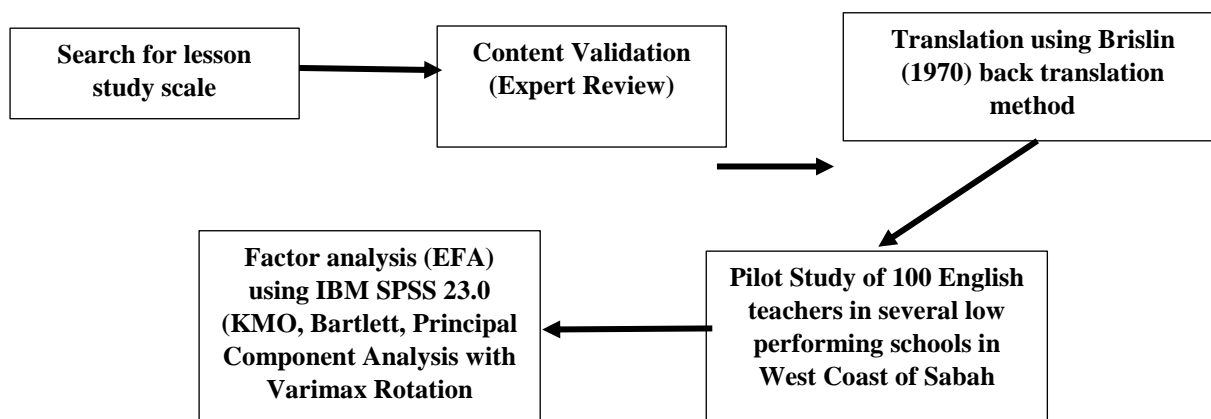


Figure 04. Steps in the Research

An extensive literature search was done to find a suitable lesson study scale that can be adapted for the purpose of research in the Malaysian context for lesson study programme under SISC+. Mostofa's (2013) lesson study scale was selected. The content of the original lesson plan was reviewed by experts and revision of the items was done. Two experts on SISC+ programme from the Education Department assessed the contents of the questionnaire to ensure that the terms and content relate to the lesson study practice. Translation of the original scale from English to Malay and blindly translated again to English, following the Brislin's (1970) back translation method was carried out. The questionnaire was used in a pilot study that involved 100 English teachers in several low performing schools in West Coast of Sabah. Exploratory Factor Analysis (EFA) was conducted using IBM SPSS 23.0 for the assessment of the scale's reliability and validity. EFA is an exploratory method based on data with the purpose of identifying the number of common factors encompassing a set of responses and relationship among individual items and its common factors (Kline, 2011). Generally, the main intention of conducting an EFA is to assess the dimensionality of the items in a scale by determining the factors that can be interpreted and needed to explain the relationships among the responses. In EFA, an observable variable refers to an indicator or an extracted factor that is regarded as the reason for the observed response (Brown, 2006). Hence, EFA is used to ascertain the number of dimension to a set of responses, the subjective meaning of each dimension, and how these items are inter-related with the dimension and how the dimensions are related to one another (Osborne & Costello, 2009). In EFA, Bartlett's sphericity test was used to determine if there is a certain redundancy between the variables ($p < 0.05$ is considered acceptable). Kaiser-Meyer-Olkin test was used to determine sampling adequacy ($KMO > 0.50$ is acceptable). Then, Principal Components Analysis with Varimax Rotation was employed to assess the scale construct validity (Factor loading of less than 0.60 subjecting the item for deletion). The reliability of the scale was determined based on internal consistency shown by Cronbach' Alpha.

6. Findings

6.1. Content Validity and Back Translation of the Lesson Study Scale

Table 01 presents the items in the revised lesson study scale, presented in Malay language after being examined and translated by relevant experts.

Table 01. Revised Lesson Study Questionnaire in Malay

Dimensions	Items
Perancangan Kolaboratif	1. Membina Rancangan Pengajaran Harian secara kolaboratif dapat menambahkan pengetahuan pengajaran saya
	2. Membina Rancangan Pengajaran Harian secara kolaboratif dapat menambahkan pengetahuan kandungan/kurikulum mata pelajaran yang saya ajar
	3. Membina Rancangan Pengajaran Harian secara kolaboratif membantu saya dalam merancang pelajaran dan pembelajaran saya
	4. Membina Rancangan Pengajaran Harian secara kolaboratif meningkatkan keyakinan diri dalam penyampaian pelajaran saya
	5. Membina Rancangan Pengajaran Harian secara kolaboratif meningkatkan keyakinan diri semasa saya mengajar
Sesi Penerangan	6. Aktiviti refleksi penyeliaan selepas sesi pengajaran membantu menganalisis pedagogi saya
	7. Menganalisis pedagogi orang lain semasa sesi perbincangan selepas sesi pengajaran membantu saya menilai aspek pedagogi dengan lebih efektif
	8. Maklumbalas aktiviti refleksi penyeliaan selepas sesi pengajaran daripada rakan sebaya membantu saya membina Rancangan Pengajaran Harian dengan lebih efektif
	9. Aktiviti refleksi penyeliaan selepas sesi pengajaran meningkatkan keyakinan saya dalam membina Rancangan Pengajaran Harian
	10. Maklumbalas yang saya terima semasa aktiviti refleksi selepas sesi pengajaran daripada pegawai SISC+ membantu saya membina Rancangan Pengajaran Harian saya
Penyemakanpelajaran	11. Semakan susulan daripada aktiviti refleksi bersama pegawai SISC+ membantu saya untuk merancang Rancangan Pengajaran Harian dengan efektif
	12. Semakan susulan daripada aktiviti refleksi bersama pegawai SISC+ membantu saya untuk mengajar semula menggunakan RPH yang telah ditambah baik
	13. Semakan susulan daripada aktiviti refleksi bersama rakan sebaya dapat membantu untuk membina Rancangan Pengajaran Harian dengan lebih baik
	14. Setelah membuat refleksi selepas perbincangan tentang pengajaran dan pembelajaran, saya yakin boleh merancang Rancangan Pengajaran Harian dengan lebih berfokus
	15. Saya lebih berkeyakinan mengajar kerana pedagogi saya telah disemak.

6.2. Exploratory Factor Analysis

Table 02 presents the Kaiser-Meyer Olkin (KMO) and Bartlett's Test result from IBM SPSS 23.0. The result shows that KMO has a value of 0.742, which means there is adequacy of sampling. The Bartlett's Test of Sphericity showed a zero-significant value, thus indicating the absence of an inverse matrix.

Table 02. Sampling Adequacy and Test of Sphericity

KMO Measures of Sampling Adequacy		.742
Bartlett's Test of Approximated Chi-Square		589.203
Sphericity	Degree of freedom	105
	Significance	.000

Table 03 presents the Cronbach's Alpha for each sub-dimension and overall construct. The Cronbach's Alpha for Planning Collaboratively is 0.790, for debriefing session is 0.713 and for lesson revision is 0.635. Overall. The Cronbach's Alpha is 0.784. Thus, all Cronbach's Alpha coefficients are more than 0.60, indicating acceptable internal consistency.

Table 03. Internal Consistency of the Lesson Study Scale

Dimension	Number of Items	Cronbach's Alpha
Planning Collaboratively	5	0.790
Debriefing Session	5	0.713
Lesson revision	5	0.635
Overall	15	0.784

Table 4 shows the SPSS output for Total Variance Explained. It showed that there are three dimensions with Total Extraction Sums of Squared Loadings more than one and the cumulative percentage of the rotation sums of squared loadings is 55.966%. The first dimension was able to explain 22.218% while the second dimension explained 17.235% and the third dimension explained 16.513%. The finding is further supported with a scree plot in Figure 5. Therefore, the lesson study scale can validly be divided into three categories.

Table 04. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.348	28.989	28.989	4.348	28.989	28.989	3.333	22.218	22.218
2	2.129	14.197	43.186	2.129	14.197	43.186	2.585	17.235	39.453
3	1.917	12.780	55.966	1.917	12.780	55.966	2.477	16.513	55.966
4	1.102	7.346	63.312						
5	1.032	6.880	70.191						
6	.828	5.523	75.714						
7	.714	4.761	80.475						
8	.662	4.415	84.890						
9	.570	3.798	88.688						
10	.401	2.673	91.362						
11	.378	2.522	93.884						
12	.284	1.893	95.777						
13	.234	1.562	97.339						
14	.220	1.467	98.806						
15	.179	1.194	100.000						

Extraction Method: Principal Component Analysis.

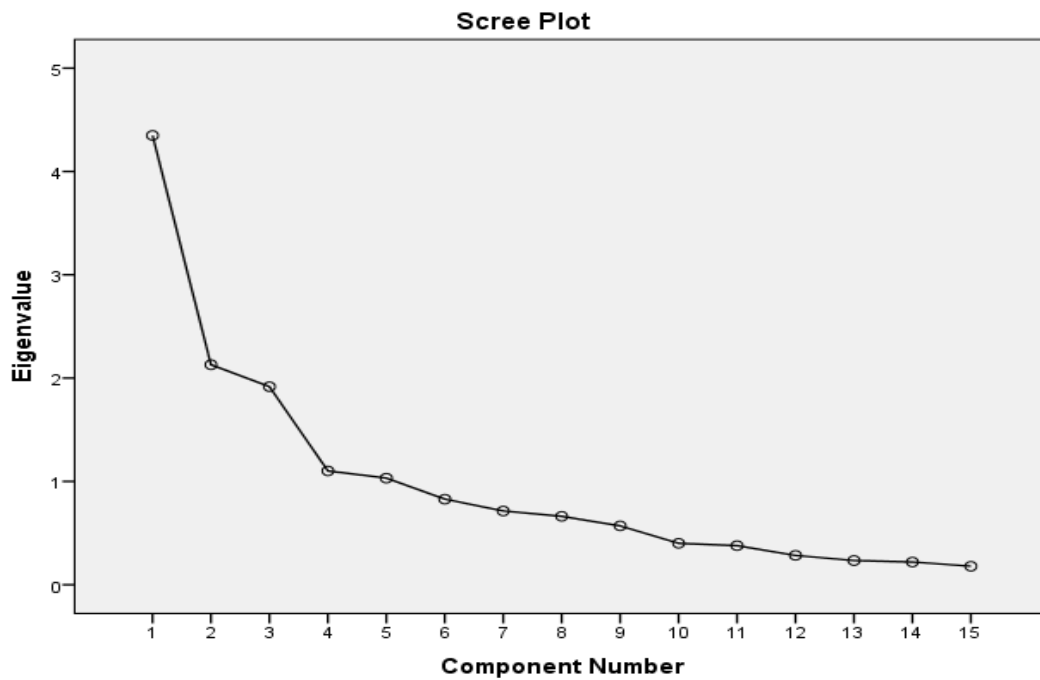


Figure 05. Scree Plot Showing Dimensionality of the Lesson Study Scale

Table 5 presents the factor loadings from the rotated component matrix. The result shows that for the Planning Collaboratively dimension, only four items were retained and one item (LS-5) was deleted due to factor loading less than 0.60. For the second dimension, Debriefing Session, four items were also retained and one item (LS-6) was deleted due to factor loading less than 0.60. For Lesson Revision dimension, four items were retained and one item deleted (LS-15). Thus, the new revised lesson study scale has 12 items with four items each for the three dimensions.

Table 05. Factor Loading in the Rotated Component Matrix

Item	Component		
	1	2	3
LS-1	.859		
LS-2	.887		
LS-3	.663		
LS-4	.806		
LS-5			-.454
LS-6	.412		.460
LS-7			.603
LS-8			.726
LS-9			.712
LS-10			.681
LS-11	.409	.616	
LS-12		.779	
LS-13		.780	
LS-14		.763	
LS-15		-.305	

7. Conclusion

Findings from the pilot study concluded that the scale used for the actual study has high reliability and validity, thus ensuring that the result obtained will provide an accurate representation of the teachers' practice relating to lesson study. This validated scale can be used to collect of data for the final study, and therefore to provide larger coverage. This scale can therefore be used to assess lesson study practices used in SISC+ programme as well as in other related field.

References

- Bagozzi, R. P., & Edwards, J. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, 1(1), 45–87.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons, Inc.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185–216.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105.
- Cheah, U. H., & Lim, C. S. (2010). Disseminating and popularising lesson study in Malaysia and Southeast Asia. In *APEID Hiroshima Seminar "Current Status and Issues on Lesson Study in Asia and the Pacific Regions" held* (pp. 1–9). Hiroshima University in Japan.
- Chiew, C. M. (2009). *Implementation of lesson study as an innovative professional development model among mathematics teachers*. Universiti Sains Malaysia.
- Clark, L. A., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Association September*, 7(3), 309–319.
- Clarke, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review*, 18(3), 301–324.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Crook, T. R., Shook, C. L., Madden, T. M., & Morris, M. L. (2009). A review of current construct measurement in entrepreneurship. *International Entrepreneurship and Management Journal*, 6(4), 387–398.
- DeVellis, R. F. (2003). *Scale development: theory and applications*. Sage Publications.
- Dillman, D. A., Smyth, J. D., Christian, L. M., & Dillman, D. A. (2009). *Internet, mail, and mixed-mode surveys: the tailored design method*. Wiley & Sons.
- Dolnicar, S., & Grün, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review*, 24(2), 127–143.
- Dotger, S., Moquin, F. K., & Hammond, K. (2009). Using lesson study to assess student thinking in science. *Educator's Voice*, 5, 22–31.
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123.
- Ewe, G. O., Chap, S. L., & Munirah, G. (2010). Examining the changes in novice and experienced mathematics teachers questioning techniques through the Lesson Study process. *Examining the Changes in Novice Journal of Science and Mathematics Education in Southeast Asia*, 33(1), 86–109.
- Fernandez, C. (2002). Learning from Japanese approaches to professional development: the case of Lesson Study. *Journal of Teacher Education*, 53(5), 393–405.
- Goh, R., & Fang, Y. (2017). Improving English language teaching through lesson study. *International Journal for Lesson and Learning Studies*, 6(2), 135–150. <http://doi.org/10.1108/IJLLS-11-2015-0037>

- Heide, M., & Gronhaug, K. (2005). The impact of response styles in surveys: A simulation study. *Journal of the Market Research Society*, 34(3), 215–223.
- Hock, C. U., & Sam, L. C. (2010). Disseminating and popularising Lesson Study in Malaysia and Southeast Asia. In *APEID Hiroshima Seminar. Current Status and Issues on Lesson Study in Asia and the Pacific Regions* (pp. 1–9).
- Hollingsworth, H., & Oliver, D. (2005). Lesson study: a professional learning model that actually makes a difference. *2005MAV Annual Conference*, 1–8.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Press.
- Lewis, C. (2008). Does lesson study have a future in the United States?, (October), 1–11.
- Lim, C. S. (2006). Promoting peer collaboration among pre-service mathematics teachers through Lesson Study process. In *Proceedings of XII IOSTE Symposium: Science and Technology in the Service of Mankind* (pp. 590–593). Universiti Sains Malaysia, Penang.
- Marble, S. (2007). Inquiring into teaching: Lesson Study in elementary science Mmethods. *Journal of Science Teacher Education*, 18(6), 935–953.
- Mohammad Reza, S. A., Fukaya, K., & Lassegard, J. P. (2010). “Lesson Study” as professional culture in Japanese schools: an historical perspective on elementary classroom Practices. *Japan Review*, (22), 171–200.
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão E Crítica*, 30(1), 1–20.
- Mostofo, J. (2013). *Using lesson study with preservice secondary mathematics teachers: effects on instruction, planning, and efficacy to teach mathematics*. Arizona State University.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: issues and applications*. Sage Publications.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGrawHill Book Company.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (Third). New York: McGraw Hill.
- Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, 12(2), 131–146.
- Pasquali, L. (2010). *Instrumentação psicológica: Fundamentos e práticas*. Brasil, Artmed: Porto Alegre. Retrieved from <https://www.passeidireto.com/arquivo/3701868/pasquali-l-e-cols-instrumentacao-psicologica-fundamentos-e-praticas-porto-alegre>
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Post, G., & Varoz, S. (2008). Lesson-study group with prospective and practicing teachers. *Teaching Children Mathematics*, 4(1), 472–478.
- Reynolds, C. R. (2010). Measurement and Assessment: An editorial view. *Psychological Assessment*, 22(1), 1–4.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (second). New York: McGraw Hill.
- Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association*, 96(453), 20–31.
- Slavec, A., & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review*, 14(1), 39–62.
- Smallbone, T., & Quinton, S. (2004). Increasing business students’ Confidence in Questioning the Validity and Reliability of their Research. *Electronic Journal of Business Research Methods*, 2(2), 153–162.
- Stacey, K. (2010). Mathematical and Scientific Literacy Around the World. *Kaye Stacey Journal of Science and Mathematics Education in Southeast Asia*, 33(1), 1–16.
- Trochim, W. M. (2006). Reliability and Validity. *The Research Methods Knowledge Base*, 1–4.
- Zanaton, I., Siti Nor Aishah, M. N., Siti Nordiyana, M., & Effandi, Z. (2014). Applying the principle of “Lesson study” in teaching science. *Asian Social Science*, 10(4), 108–113.
- Zuraidah, A., & Muhammad Faizal, A. G. (2013). Professional learning community: a guideline to improve education system in Malaysia. *Hope Journal of Research*, 1(4), 1–26.