

Application of Rasch Analysis in Measuring Teacher Collegial Supervisory Instrument's Reliability and Validity

Azmir Hashim^{1,*}, Lokman Mohd Tahir¹, Mohammed Borhandden Musah², Noor Azean Atan¹, Siti Nisrin Mohd Anis¹

¹School of Education, Universiti Teknologi Malaysia, 81300 Scudai, Johore, Malaysia

²Bahrain Teachers College, University of Bahrain, Sakheer, Bahrain

Received May 22, 2020; Revised September 12, 2020; Accepted November 7, 2020

Cite This Paper in the following Citation Styles

(a): [1] Azmir Hashim, Lokman Mohd Tahir, Mohammed Borhandden Musah, Noor Azean Atan, Siti Nisrin Mohd Anis, "Application of Rasch Analysis in Measuring Teacher Collegial Supervisory Instrument's Reliability and Validity," *Universal Journal of Educational Research*, Vol. 8, No. 11C, pp. 6 - 14, 2020. DOI: 10.13189/ujer.2020.082302.

(b): Azmir Hashim, Lokman Mohd Tahir, Mohammed Borhandden Musah, Noor Azean Atan, Siti Nisrin Mohd Anis (2020). *Application of Rasch Analysis in Measuring Teacher Collegial Supervisory Instrument's Reliability and Validity*. *Universal Journal of Educational Research*, 8(11C), 6 - 14. DOI: 10.13189/ujer.2020.082302.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The concept of collegial supervision (CS) is defined as collaborative work beyond their professional sphere of relationships often offered by educators through feedback and sharing platform. However, there is still lack of studies and instruments that evaluated the CS practice within Malaysian context. In measuring the suitability of CS practice in secondary schools in Malaysia, a questionnaire with 26 self-developed items that represent five sub-dimensions/constructs namely, collegial relations, teacher's province, teacher growth, teacher collaboration, and reflective inquiry was developed from series of interviews with secondary teachers. The major aim of this paper is to validate and examined the psychometric elements through the application of Rasch analysis in measuring items'-person reliability, principal component analysis, items - person distribution, fit and dimensionality analyses. The analysis was performed based on feedback of 357 teachers from secondary teachers. Findings revealed on high values on person-items reliability, and the items' difficulty are significantly aligned or matched with teachers' ability. Also, principal component analysis revealed an acceptable value of raw variance explained, and that most teachers agreed with most of the items through structure measurement on the items' validity. Thus, it is concluding on the internal consistencies of the items within Malaysian CS which later contributed to the CS items for Malaysian secondary schools.

Keywords Collegial Supervision, Secondary Schools, Rasch Measurement Model

1. Introduction

Empirical evidence has defined collegial supervision (CS) through the lenses of instructional practice as collaborative efforts made by teachers in refining their instructional practices across knowledge sharing platforms and feedback for teachers' professional growth [1, 2, 3]. In defining the concept of CS, Singh and Manser (2002) [4] believed that it is a learning process towards the practice of shared responsibility and values among school community which includes principals and teachers. In another definition, CS is sometimes referred to as peer supervision [5, 6] for its emphasis on colleagues' assessments and feedbacks as 'informal' supervisors, with broader mechanisms to improve teachers' performance in the instructional practice which is highly concerned with teaching and learning. In the context of this study, CS is referred to as a consistent process of facilitation where colleagues (i.e. principles, administrators, and teachers) work together and offer one another feedback on their performances. According to Glatthorn (1984) [5], the approach is directed towards cooperative professional

growth. Furthermore, CS is deemed as a collective process beyond professional sphere of relationships [6], towards common vision, aiming for the school cultural based improvement and focusing on the teachers' growth and development, interpersonal relationship and collaborative approaches.

As mentioned by Khun-Ineree (2020) [32], the pertinent reason why CS is needed in secondary are based on the lack of knowledge on supervisee practices in helping teachers improve their teaching and learning activities. Thus, it will effect students' performances in their academic achievements. In addition, the trusted person in supervising teachers which is referring to the school's administrators are packed with meetings and bust schedules [32]. In addition, it is cautioned that not all teachers will accept comments and advices provided by their colleagues although the purpose of the CS is to improve other teachers' professional development [32, 33] due to differences in the professional and personal relationship. In addition, Aktas (2018) [34] also stressed although CS is provided to teachers especially to novice teachers, they need to be flexible and selective in choosing their instructional approaches. This is because through the mentoring and collegial assessment approaches, the novice teachers will be shadowed their mentor or their senior teachers who gave assessments towards the improvement in teaching and learning during the collegial supervisory approach.

Although the practice of CS in school context has begun as early as 1984 by the work of Glatthorn, due to no specific measure of collegiality (Sabharwal, 2011) [7], and the fact that most of the studies conducted are non-quantitative approaches [6], the complexity of collegial practice itself [8, 9, 10], led to the 'paucity' of studies on collegiality until there have been new developments on the collegiality measurement scale [11, 12]. It is apparent that the literature obtained on the CS practice in the context of Malaysia heavily emphasises direct supervision in its clinical mode [13, 14] and pays little attention or less indication to CS practice. Succinctly, clear standard framework, model and items that related to CS also seems to be unavailable within context of secondary schools in Malaysia. This is deemed as a claim that there is limited empirical evidence about the framework; model and instrumentation of CS meant for secondary schools still received little attention among local researchers. In other words, the standard framework and their measuring instrumentation of CS are arguably unknown in the context of secondary schools. This study therefore validates the psychometric - findings of CS scale -through the application of Rasch analysis in measuring items'-person reliability, principal component analysis, items - person distribution, fit and dimensionality analyses.

2. Objectives of the Study

This study was designed to addressed the following

research objectives:

- i To obtain the person and items reliability on CS items.
- ii To access the psychometrics of CS scaling based on principal component analysis, items -person distribution, fit and dimensionality analyses.

3. Conceptual Framework

Theoretical framework of this study is an adaption process from the Zepeda's framework (2007, p. 28) [15] which focuses on formative and cyclical approaches of the instructional supervision. In her framework, Zepeda defined the CS approach as a professional development meant for teachers' development which based in instructional supervision which consisted of formative supervision and evaluations. In this study context, the standard of CS was chosen to replace the professional development aspects due to its similar nature; the CS itself is a professional development's type of supervision [16, 17, 18]. In the formative supervision (observation), it is concerned with the on-going individual's professional development with what is carried out in the CS dimensions. The pre-observation conference, classroom observation and post-observation as formative supervision are elements highlighted by Zepeda (2007) [15] in the CS dimensions [2, 31].

4. Methodology

4.1. Sampling

A total of 357 teachers were selected to provide their feedback based on the listed items. Secondary school teachers were selected using the multistage cluster sampling technique known as multiple probability technique [19] used due to the difficulty in determining the entire population. This technique is appropriate for large populations that are geographically spread and naturally in the population [19] in order to ease the group's identification, locate lists [20] as well as reducing bias and representativeness issues.

4.2. Instrumentation

In this study, the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) questionnaire was designed to assess the effective practice of CS in Malaysian secondary schools. The questionnaire consisted of 28 items which comprised two items on demographics, and 26 items that represent the six sub-dimensions of CS: namely, collegial relations (CR) (5 items), teacher's province (PR) (5 items), teacher growth (TG) (5 items), teacher collaboration (TC) (6 items), and reflective inquiry

(RI) (5 items). In the demographics, two items were constructed: teachers' gender and their years of experiences within the teaching profession.

The items were constructed and derived from transcripts of series of interviews with teachers related to the practice of CS. The senior teachers were purposely selected and asked to provide their comments to all items. Items were initially constructed in the Malay language. However, later, it was decided to provide an English translation given the suggestions from the English language teachers. The translation process from Malay language to English was conducted by a senior English language teacher with the assistance of a Malay language teacher. Later, the items were checked by senior teachers to assess the content validity of all 26 items. In terms of the scaling, SFCSMSS uses a five-point Likert scale: 1: strongly disagree, 2: disagree, 3: Neutral, 4: agree, and 5: strongly agree. The five-point scale was decided use based on the following justifications: (a) it is a common rating scale among social science researchers; (b) the scale provides equal opportunity for all respondents while providing their answers [21].

5. Results

5.1. Teachers' Demographics

The total number of teachers participated in this study was 357 which demonstrated a response rate of 59.01%, which exceed the suggested return rate (49%) as recommended by (Baruch & Holtom, 2008) [22]. Table 1 below illustrates the distribution of teachers' demographics according to their gender and years of experiences. In terms of teachers' experience, teachers were clustered into three major groups: the first group are teachers who had experience between 0 to 10 years of experience, followed by teachers who had 11 to 20 years and the final group of teachers who had experience between 21 to 30 years. Based on their experiences, majority of 149 teachers (41.7 %) that participated were between 11 to 20 years of experience followed by 111 teachers (31.1%) with 21 to 30 years' experience, and only 97 teachers (27.2%) who had experiences between 0 to 10 years. Based on teachers'

gender, 98 teachers (27.5 %) that participated in this study were male teachers and 259 teachers (72.5 %) were female teachers, which is an indication of high numbers of female teachers in local secondary schools in Malaysia compared to their counterpart. Table 1 below indicates the data consisted of teachers' years of experience and gender.

Table 1. Teachers' Demographics

Variables	N	Per centage
<i>Gender</i>		
Male	98	27.5
Female	259	72.5
<i>Years of Experiences</i>		
0 to 10 years	97	27.2
11 to 20 years	149	41.7
21 to 30 years	111	31.1
Total	357	100

5.2. Items and Person's Reliability

The Rasch Person-Item Reliability tests were performed because of its capability in determining the internal reliability of items as well as respondents. As shown in Figure 1, the Rasch person reliability is 0.94, which is considered an acceptable value [23]. A person separation value indicates the value of 4.04 which indicates that the instruments are sensitive enough to distinguish between teachers with many years of experiences and teachers with less experience. Thus, there is no additional items that are needed [24]. Based on the analysis, teachers were categorised into four major classifications; teachers who always received supervision, teachers who received medium amount of supervision, teachers who received least amount of supervision and teachers who never received any type of supervision. In Figure 2 below, the item separation is higher than 3.0, which implies that the person sample is large enough to confirm the item difficulty hierarchy of the instrument [24]. In sum, both reliability values indicate a sufficient sample in determining the item difficulty index of each item [23, 25]. Items were classified into five classifications as too difficult, least difficult, answerable, easy and too easy to answer.

SUMMARY OF 357 MEASURED (EXTREME AND NON-EXTREME) PERSON									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	100.6	26.0	1.62	.42					
S. D.	12.2	.0	1.83	.15					
MAX.	130.0	26.0	8.78	1.84					
MIN.	130.0	26.0	-3.36	.27	.04	-6.4	.04	-6.3	
REAL RMSE	.50	TRUE SD	1.76	SEPARATION	3.50	PERSON RELIABILITY	.92		
MODEL RMSE	.44	TRUE SD	1.77	SEPARATION	4.00	PERSON RELIABILITY	.94		
S. E. OF PERSON MEAN = .10									

PERSON RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .95

Figure 1. Person's reliability

SUMMARY OF 26 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	1381.1	357.0	.00	.11	.97	-.9	1.00	-.8
S.D.	79.8	.0	.66	.01	.46	3.1	.60	3.4
MAX.	1482.0	357.0	1.84	.12	3.03	9.9	3.68	9.9
MIN.	1025.0	357.0	-1.04	.08	.65	-4.1	.59	-4.4
REAL RMSE	.11	TRUE SD	.65	SEPARATION	5.88	ITEM	RELIABILITY	.97
MODEL RMSE	.11	TRUE SD	.65	SEPARATION	6.13	ITEM	RELIABILITY	.97
S.E. OF ITEM MEAN = .13								

UMEAN=.0000 USCALE=1.0000
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.65

Figure 2. Items' reliability

5.3. The Person-Item Distribution Map

Using *Winsteps* application, the Rasch analysis was performed based on 357 feedbacks from secondary teachers. Given the Person-Item Distribution Map (PIDM), which illustrated in Figure 3, the person-item's distribution map has indicated teachers' abilities to response to the items' difficulty. Using the map which is produced by Rasch Measurement Model, 357 teachers were placed on the right side of the distribution map while all 26 items were plotted in the left side of the distribution map based on *logit* scale distribution. A "*logit*" scale was used to express item difficulty on a linear scale that extends from negative infinity to positive infinity [26]. Using the distribution map, $Mean_{item}$ was plotted and served as a threshold which indicates as zero value on the *logit* scale distribution. In this plot, items that are placed higher than the $Mean_{item}$ formally indicate that items were difficult items compared to items which plotted in lower than the $Mean_{item}$. On the left side, teachers' abilities are matched with item difficulty. If the right side of the map data showed higher than the left side, most of items were considered as difficult for teachers and vice-versa [27]. Based on item-person map, items that are labelled as CR5 and PR3 are considered as difficult items while four items, CR2, PR1, TC6 and TG3 are known as easy items. However, based on overall items and person distribution, the items' difficulties are matched with the teachers' abilities in answering all 26 items given in the questionnaire.

From Figure 3, data showed that Person mean value, $Mean_{person}$ was indicated at 1.62 threshold value while $Mean_{item}$ was indicated at 0.0 values. The highest teacher managed to score 8.78 logit and the lowest scored -3.32 logit. As for the item distribution, the most difficult item perceived by teachers is CR5 with 1.84 logit and the easiest item is noted at -1.04 logit. Based on the findings, a total of 11 items (PR3, RI2, RI4, CR3, RI3, RI5, TC2, TC1, TG2, and TG1) were found above the $Mean_{item}$ which indicated the secondary teachers' abilities in understanding and answering the given items in the questionnaires. Overall, it is assumed that the 26 items within the questionnaire were considered not that difficult

due to the teachers' abilities are above the items' difficulty. Based on the Rasch analysis, a total of 265 teachers (74.5 %) were above the $Mean_{item}$ and 37 teachers (10.5 %) were below the $Mean_{item}$. In conclusion, the items are aligned with teachers' abilities since most of 26 items were seen matching with teachers' abilities in responding and answering all items in the questionnaire. Thus, it is assumed that teachers could understand it well and answer all the questions correctly.

Based on the items-person distribution map which obtained from teachers' feedback towards the five constructs of CS concept, only 11 items were above the $Mean_{item}$ and another 15 items were located below the $Mean_{item}$. All findings which are related to items' plots and locations are presented in Table 2 below. From Table 2, most of items found difficult are related to reflective inquiry which has four items that scatted above the $Mean_{item}$ which are RI2, RI4, RI3 and RI5. As for collegial relations (CR) (CR3, CR5), teacher growth (TG) (TG1, TG2) and teacher collaboration (TC) (TC1, TC2) constructs, each construct has 2 items which fell above the $Mean_{item}$ and teacher's province construct has only one item which is PR3. There are 15 items that plotted below the $Mean_{item}$ which indicated that most items in the questionnaire were easy items compared to 11 items that situated above the $Mean_{item}$.

Table 2. Items' Plots

Constructs	Above the $Mean_{item}$	Below the $Mean_{item}$
Reflective inquiry (RI)	RI2, RI4, RI3, RI5	RI1
Collegial relations (CR)	CR3, CR5	CR1, CR2, CR4
Teacher collaboration (TC)	TC1, TC2	TC4, TC6, TC3, TC5
Teacher growth (TG)	TG1, TG2	TG3, TG4, TG5
Teacher's province (PR)	PR3	PR1, PR2, PR4, PR5
Total	items	15 items

5.4. Item's Fit and Dimensionality Analysis

Items' fit and misfit analyses were also conducted in this

section. In Figure 4, the findings show that one item from the questionnaire which did not fit with the measurement of RM. Item no 5 has min square (MNSQ) values outside the in-fit range of $0.4 < x < 1.5$ [24] which at 3.68 value. However, the remaining 25 items within this CS questionnaire were located within the acceptable range between 0.4 with 1.5.

Furthermore, items' dimensionality also was inspected and analysed using the principal component analysis

performed through *Winstep*. The expected values are obtained using Rasch measurement which require the measurement to explain at least 40% of raw variance, and that the unexplained variance in the first contrast should not be more than 15% [28, 29]. In Figure 5, the data disclosed a raw variance of 49.7% explained by measures. This value is low compared to the value of the model (51.7%). The 6.7 % of unexplained variance was accepted as it is less than the maximum value 15%.

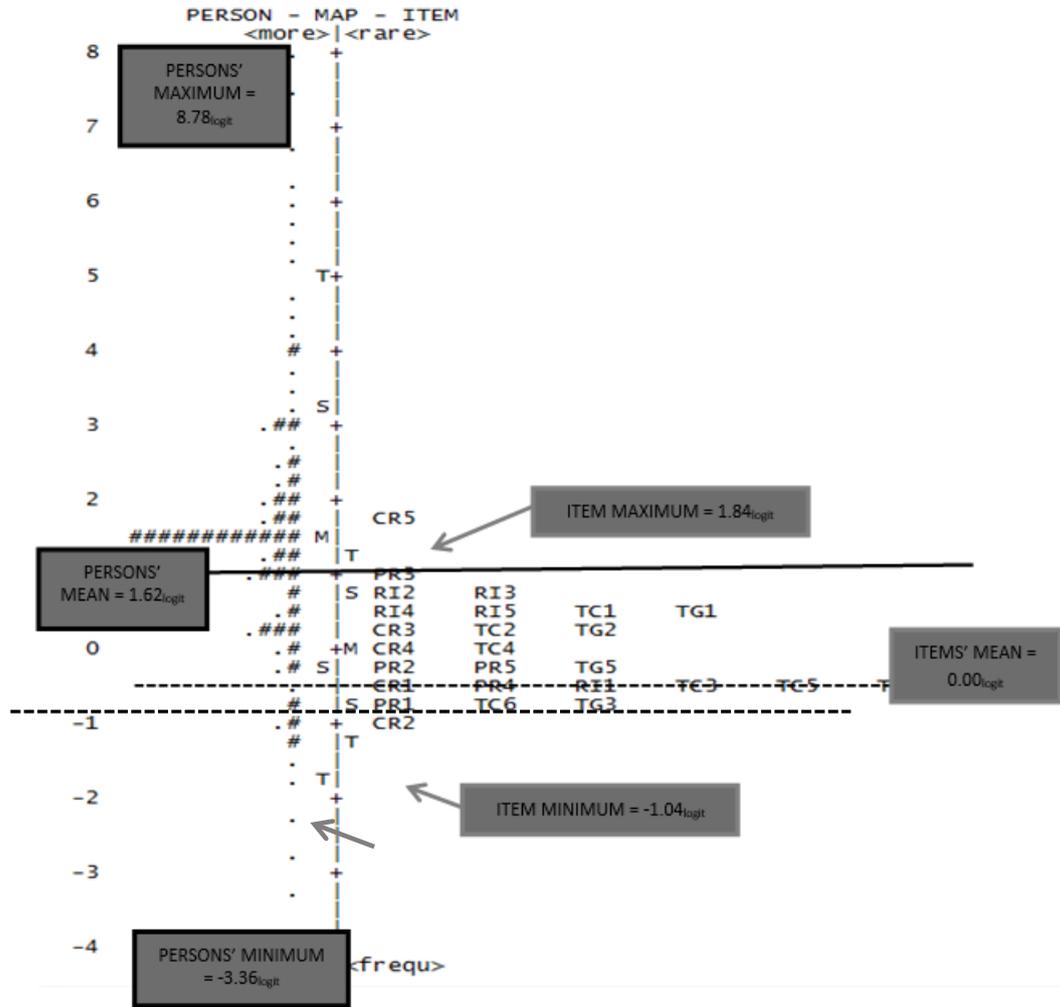


Figure 3. Person-Item Distribution Map.

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S. E.	INFIT		OUTFIT		PT-MEASURE		EXACT OBS%	MATCH EXP%	ITEM	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
5	1025	357	1.84	.08	3.03	9.9	3.68	9.9	.14	.74	28.0	54.8	CR5	0
8	1350	357	1.07	.10	.84	-1.9	.83	-1.8	.71	.65	77.1	70.5	PR3	0
24	1378	357	.72	.12	.71	-3.5	.65	-3.7	.75	.63	79.7	75.4	RI3	0
23	1375	357	.63	.11	.78	-2.7	.75	-2.9	.73	.65	78.8	72.0	RI2	0
11	1398	357	.51	.11	.76	-2.8	.70	-3.1	.73	.64	80.2	74.0	TG1	0
26	1380	357	.47	.10	1.19	2.3	1.18	2.0	.59	.66	64.4	67.5	RI5	0
25	1396	357	.43	.12	.72	-3.3	.62	-4.0	.74	.63	81.1	76.5	RI4	0
16	1407	357	.41	.12	.85	-1.6	.78	-1.9	.68	.63	81.6	78.2	TC1	0
17	1422	357	.34	.11	.65	-4.1	.59	-4.3	.76	.64	83.1	74.4	TC2	0
12	1422	357	.30	.11	.80	-2.1	.72	-2.7	.72	.64	80.5	75.9	TG2	0
3	1413	357	.23	.11	.98	-2	1.01	.1	.65	.64	74.0	72.7	CR3	0
19	1307	357	.07	.10	1.11	1.4	1.16	1.8	.61	.66	64.7	65.6	TC4	0
4	1443	357	-.02	.11	.83	-2.0	.79	-2.1	.69	.63	79.1	73.8	CR4	0
10	1352	357	-.16	.10	1.48	4.7	1.68	5.8	.48	.65	70.9	70.5	PR5	0
15	1349	357	-.19	.10	.92	-9	.94	-.6	.69	.66	75.7	68.6	TG5	0
7	1372	357	-.19	.09	1.03	.4	1.15	1.4	.64	.67	75.7	68.0	PR2	0
22	1394	357	-.40	.11	.75	-2.7	.68	-3.0	.72	.64	79.7	75.1	RI1	0
14	1365	357	-.47	.10	.93	-.7	.94	-.6	.68	.66	75.7	69.9	TG4	0
9	1373	357	-.48	.10	.76	-2.8	.70	-3.1	.74	.65	79.1	71.7	PR4	0
20	1388	357	-.56	.11	.73	-3.0	.66	-3.4	.74	.64	79.9	74.8	TC5	0
18	1418	357	-.60	.10	.67	-4.0	.60	-4.4	.76	.65	74.9	70.8	TC3	0
1	1452	357	-.60	.10	1.32	3.2	1.32	2.7	.55	.64	69.8	68.2	CR1	0
21	1409	357	-.72	.11	.79	-2.2	.75	-2.4	.71	.64	80.8	75.0	TC6	0
13	1417	357	-.78	.10	.68	-3.8	.62	-4.1	.76	.65	78.2	71.8	TG3	0
6	1422	357	-.81	.10	.99	-1	1.36	3.1	.64	.65	72.0	70.5	PR1	0
2	1482	357	-1.04	.11	.99	.0	1.03	.3	.61	.62	73.2	71.1	CR2	0
MEAN	1381.1	357.0	.00	.11	.97	-.9	1.00	-.8			74.5	71.4		
S. D.	79.8	.0	.66	.01	.46	3.1	.60	3.4			10.5	4.5		

Figure 4. Items' fit analysis

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

	Empirical	Modeled
Total raw variance in observations	51.7 100.0%	100.0%
Raw variance explained by measures	25.7 49.7%	51.7%
Raw variance explained by persons	17.0 32.8%	34.1%
Raw variance explained by items	8.8 16.9%	17.6%
Raw unexplained variance (total)	26.0 50.3%	48.3%
Unexplned variance in 1st contrast	3.5 6.7%	13.4%
Unexplned variance in 2nd contrast	2.3 4.4%	8.8%
Unexplned variance in 3rd contrast	2.0 3.8%	7.6%
Unexplned variance in 4th contrast	1.7 3.3%	6.6%
Unexplned variance in 5th contrast	1.6 3.0%	6.0%

Figure 5. Items' dimensionality

SUMMARY OF CATEGORY STRUCTURE. Model="R"
FOR GROUPING "0" ITEM NUMBER: 1 CR1

ITEM DIFFICULTY MEASURE OF -.60 ADDED TO MEASURES

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INAVRGE	OUTFIT EXPECT	IN MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE
1	1	4	1	-.80	-2.19	2.58	4.65	NONE	(-3.81)
2	2	8	2	-.65	-1.08	1.46	2.10	-1.72	(-2.41)
3	3	43	12	.24	.15	1.07	1.11	-1.55	(-1.24)
4	4	207	58	1.32	1.32	.97	1.23	-.21	(1.07)
5	5	95	27	3.01	3.16	1.40	1.04	3.48	(3.99)

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

Figure 6. Summary of the rating scale category structure

In addition, the communication validity which represents the structure calibration calculated from the rating scale used by the instrument (e.g. Likert scale) was examined. Rasch analysis helps to determine the validity of the scale used by 'zero setting' and calibrate the rating scale

used. Rasch analysis also verifies the probability of even spreading (i.e. equal interval) between the specified scale [30]. A summary of the rating scale category structure and structural measures at intersections are shown in Figures 6 and Figure 7.

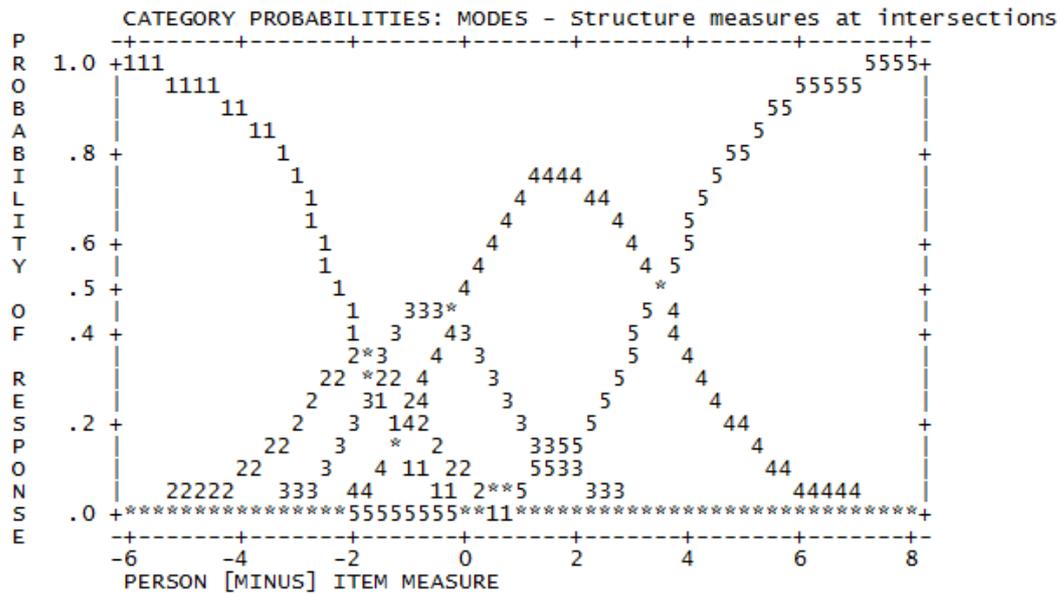


Figure 7. Structure measures at intersections.

Figure 6 demonstrates that the most answered response was rating scale 4, based on observed count of 207 (58%) which indicates that most teachers agreed on most of the 26 item statements within the questionnaire. On the contrary, the rating scale 1 had the lowest number of responses, with an observed count of 4 (1 %) which indicates least strongly disagree of teachers with reference to the 25 item statements.

Based on the findings, the pattern of the observed responses ranged between -2.19 *logit* and improved in one direction to +3.16 *logit*. This showed that the pattern of the teachers' responses is considered as normal due to the increase from negative to positive value. In this reliability analysis, the values of deviation between scale 1 and 2, 2 and 3, and 3 and 4 are 4.0, 4.0 and 4.5, respectively. These results confirmed the validity of the scales, indicating that items are differentiated by the teachers. In this sense, teachers clearly understood the difference between all scales. Also, they knew how to answer the questions by rating their answers through the given scales. This result confirmed that the validity of the structure calibration is rejected as the value of deviation is more than 1.4 and less than 5 ($1.4 < s < 5$) [23].

6. Discussion

In examining on the psychometrics elements of the collegial supervision practice items which are labelled as the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS), a total of 357 secondary teachers were asked to give feedback on the internal consistency of the SFCSMSS items. In the first phase, items were later analysed quantitatively using the Rasch analysis in determining the reliability of the items within the CS questionnaire. Later, secondary teachers' feedback was

analysed in measuring the items' internal consistency using the principal-component analysis, items -person distribution, fit and dimensionality analyses. In answering the objectives of the study, Rasch measurement model analysis was conducted throughout the study in determining the reliability analysis followed by principal-component analysis, items-person distribution, and later the fit and dimensionality analysis was performed and reported.

In determining the reliability analysis, the Rasch analyses indicated that items of the the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) have indicated a suitable and acceptable values which are considered as acceptable, sufficient and have high consistency in measuring secondary teachers' collegial supervision practice. Based on the results, it showed that items from the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) are considered acceptable and measuring the collegial practice among teachers in secondary schools in Malaysia. In fact, using the standard items of Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS), teachers' collegial practice can be examined and investigated. Based on the items' analysis, items analysis was separated into five three major classifications: Items were classified into five classifications as too difficult, least difficult, answerable, easy and too easy to answer items. As for the person separation, the Rasch analysis which was employed to analyse the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) items has revealed that there are four major classification according the teachers' demographics; from senior teachers to novice or less experienced teachers. In addition, through the analysis, Rasch analysis also showed the segregation of teachers who have been supervised by their school administrators: teachers who always received supervision,

teachers who received medium amount of supervision, teachers who received least amount of supervision and teachers who never received any type of supervision. Through the analysis, results imply that the person sample is large enough to confirm the item difficulty hierarchy of the instrument [24]. In sum, both reliability values indicate a sufficient sample in determining the item difficulty index of each item [23, 25].

In investigating the second objective which is related to measurement of items psychometrics, these analyses were conducted which comprised the principal-component analysis, items-person distribution, and lastly inspecting the fit and dimensionality of the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) items. From the Rasch's item and person distribution, 357 teachers were placed on the right side of the distribution map while all 26 items were plotted in the left side of the distribution map based on *logit* scale distribution. A "*logit*" scale was used to express item difficulty on a linear scale that extends from negative infinity to positive infinity [26]. Using the item-person distribution map, items and teachers' abilities in answering the items were matched in a distribution map. Through the *logit* scale, results indicated that only two items were classified as difficult items for teachers to provide feedback. Thus, items have higher abilities than teachers' abilities. In addition, four items are considered as easy items. Thus, 20 items are matched with teachers' abilities in answering the 26 items.

Additionally, a total of 11 items from the Rasch analysis were reported matched with secondary teachers' abilities in answering the items within the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) items. Hence, it is assumed that items within the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS) items are considered as items that matched with teachers' abilities. Thus, it is assumed that teachers could understand it well and answer all the questions correctly. Through in-depth analysis on the items' descriptions, most difficult items were mostly from the reflective inquiry construct which has four items. However, another four constructs which are collegial relations, teacher growth, teachers' province and teacher collaboration have items that matched with teachers' abilities and items that below the teachers' abilities which are labelled as easy items. In measuring the items' fit and misfit analysis, only one item that reported did not matched with acceptable measurement of Rasch. Therefore, another 26 items are ranging within the acceptable values and range. Using the the principal component analysis, the variance that accounted are also reported the acceptable raw variance which also indicate the internal consistency of all 26 items on the Standard Framework of CS for Malaysian Secondary Schools' (SFCSMSS)

In this study, we are acknowledged on the limitation of the study. Firstly, the study is limited to the feedback provided by 357 secondary teachers. Therefore, the feedbacks are limited to the 357 secondary teachers which

did not represent the whole secondary teachers in Malaysian schools. In order to generalise the findings, it is suggested that future study to replicate the study with a larger sample size in order to obtain the overall perceptions of secondary teachers related to their practice of collegial supervision whether CS is considered beneficial their instructional tasks and enhance their competencies. The next limitation is related to the items used in the questionnaire which considered very simple and being analysed with descriptive statistics to determine the collegial practice within the secondary schools setting. As for future study, it is suggested to replicate the study with the other context of schools such as technical and vocational schools, religious-based schools, primary and even international schools which also practice the collegial supervision approach.

7. Conclusions

Based on the comprehensive analysis using *Winstep*, this study has established evidence that items measured the secondary school CS practice exhibited acceptable values in measuring the practice of CS across the sampled secondary schools. From these analyses, teachers are classified according to their abilities in answering all items in the questionnaire which reflected their performance in understanding and providing responses to the items within the questionnaire. Rasch analysis reached the conclusion that there is a linkage between items' difficulty with teachers' understanding across the 26 items in the questionnaire. Thus, Rasch analysis is considered as suitable analysis in measuring items difficulty that matched with teachers understanding and probability of teachers in providing responses to the provided scales. As such, Rasch analysis potentially provides researchers with mechanisms in monitoring respondents' categorisation.

Acknowledgements

The authors wish to acknowledge the financial support received from the Research Management Centre, Universiti Teknologi Malaysia under the Fundamental Research Grant of Q.J.130000.2553.21H34.

REFERENCES

- [1] Darling-Hammond, L. (2005). Teaching as a profession: Lessons in teacher preparation and professional development. *Phi Delta Kappan*, 87(3), 237–240.
- [2] Greene, M.L. (1992). Teacher supervision as professional development: Does it work? *Journal of Curriculum and Supervision*, 7(2), 131–148.

- [3] Wanzare, Z. and da Costa, J.L. (2000). Supervision and staff development: Overview of the literature. *NASSP Bulletin*, 84(618), 47–54.
- [4] Singh, P. and Manser, P.G. (2002). Collegiality in education: a case study. *South African Journal of Education*, 22(1), 56–64.
- [5] Hopfengardner, J.D. and Walker, R. (1984). Collegial support: An alternative to principal-led supervision of instruction. *NASSP Bulletin*, 68(471), 35–40.
- [6] Shah, M. and Abualrob, M.M. (2012). Teacher collegiality and teacher professional commitment in public secondary schools in Islamabad, Pakistan. *Procedia - Social and Behavioral Sciences*, 46, 950–954.
- [7] Sabharwal, M. (2011). Job satisfaction patterns of scientists and engineers by status of birth, *Research Policy*, 10 (6), 853-863.
- [8] Barth, R.S. (2006). Improving relationships within the schoolhouse. *Educational Leadership*, 63(6), 8–13.
- [9] Kelchtermans, G. (2006). Teacher collaboration and collegiality as workplace conditions. A review. *Zeitschrift für Pädagogik*, 52(2), 220–237.
- [10] Kligyte, G. and Barrie, S. (2014). Collegiality: Leading us into fantasy – the paradoxical resilience of collegiality in academic leadership. *Higher Education Research and Development*, 33(1), 157–169.
- [11] Miles, M.P., Shepherd, C.D., Rose, J.M. and Dibben, M. (2015). Collegiality in business schools. *International Journal of Educational Management*, 29(3), 322–333.
- [12] Shah, M. (2011). The dimensionality of teacher collegiality and the development of teacher collegiality scale. *International Journal of Education*, 3(2), 1- 20.
- [13] Rasit, R.A. (2007). Teaching and learning practices at primary schools at central Malacca district. Unpublished master dissertation, Universiti Teknologi Malaysia.
- [14] Taib, M.R., Abdullah, Z., Mat Ail, N.M., Mat Jusoh, N., and Yahya, M.R. (2015). Clinical supervision of teaching Mara Junior Science College (MJSC) in northern zone of Malaysia. *Procedia-Social Behaviour*, (191), 452 – 458.
- [15] Zepeda, S.J. (2007). *Instructional Supervision: Applying Tools and Concepts*. Eye on Education. Routledge.
- [16] Glatthorn, A.A. (1984). *Differentiated Supervision*. Alexandria: ASCD.
- [17] Hammersley-Fletcher, L. and Brundrett, M. (2008). Collaboration, collegiality and leadership from the head: The complexities of shared leadership in primary school settings. *Management in Education*, 22(11), 11–16.
- [18] Sergiovanni, T. & Starratt, R.J. (2002). *Supervision: a redefinition*. Boston: McGraw-Hill.
- [19] Teddlie, C. and Yu, F. (2007). Mixed Methods Sampling: A Typology with Examples. *Journal of Mixed Methods Research*, 1(1), 77–100.
- [20] Creswell, J.W. (2012). *Educational Research: Planning, Conducting, And Evaluating Quantitative and Qualitative Research*. Boston: Pearson.
- [21] Dillman, D. A., Smyth, J. D. and Christian, L. M. (2009) *Internet, mail and mixed-mode surveys: The tailored design method*. Hoboken, N.J: John Wiley & Sons.
- [22] Baruch, Y. and Holtom, B.C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139–1160.
- [23] Bond, T. G., and Fox, C. M. (2015). *Applying the Rasch model: Fundamental Measurement in the Human Sciences (3rd ed.)*. Mahwah, NJ: Erlbaum.
- [24] Linacre, J. M. (2016). Winsteps® (Version 3.92.1) [Computer Software]. Beaverton, Oregon.
- [25] Kathy, E.G. and Catherine, G.F. (2002). Survey development and validation with the Rasch model. A paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC, November 14-17.
- [26] Boone, W.J., Staver, J.R., and Yale, M.S. (2014). *Rasch Analysis in the Human Sciences*. Netherlands: SpringerLink
- [27] Rozeha, A. R., Azami, Z. and Mohd Saidfudin, M. (2007). Application of Rasch Measurement in Evaluation of Learning Outcomes: A Case Study in Electrical Engineering. Paper published at the Regional Conference on Engineering Mathematics, Mechanics, Manufacturing and Architecture (EM3ARC).
- [28] Azrilah, A.A., Azlinah, M., Azami, Z., Sohaimi, Z., Hamzah, A. G., and Mohd Saidfudin, M. (2008). Evaluation of Information Professionals Competency Face Validity Test Using Rasch Model. Paper presented at the 5th WSEAS / IASME International Conference on Engineering Education (EE'08). July 22-24.
- [29] Azrilah A.A., Nuraini, K., Khairul Anwar, Azami, Z. and Siti Hanani, M.Y. (2012). A Rasch Analysis on Total Learning Experience of UKM Engineering Students. *International Educational Studies*, 6(6), 202 – 212.
- [30] Perkins, H. W. (2002). Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *Journal of Studies on Alcohol*, 14, 91–100.
- [31] Dixon-Houston, MI. (2012). Collegiality and instructional Supervision: A case study of relationships in the context of walkthrough observations in six rural South Carolina schools. Doctoral dissertation, University of South Carolina.
- [32] Hareesol Khun-Inkeeree, Mohd-Hibatul-Hakimie Mahmood, Siti-Saadiah Haji-Mohd-Noor, Muhammad Dzahir Kasa, Mohd Faiz Mohd Yaakob, M. S Omar-Fauzee, Fatimah Noor Rashidah Mohd Sofian. (2020). Increasing teachers' self-efficacy through regular teaching and learning supervision. *Universal Journal of Educational Research* 8(7): 3002-3013.
- [33] H. Khun-inkeeree. P. D. Dali. Y. Daud, M.S.B.O. Fauzee, R. Khalid. (2018). Effects of Teaching and Learning Supervision on Teachers Attitudes to Supervision at Secondary School in Kubang Pasu District, Kedah. *International Journal of Instruction*, 12(1), 1335-1350.
- [34] Aktas, B.C. (2018). Assessment of induction to teaching program: Opinions of novice teachers, mentors, school administrators. *Universal Journal of Educational Research* 6(10): 2101-2114.