Verify the Five Likert Scale for Improvement of Specialist Coaches plus Program Implementation Evaluation Instruments Based on Approach Model Measurement Rasch

Salefah Mat Salleh, Norasmah Othman
Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia

Abstract

This study is to verify the five - point likert scale used in the SISC+ Program Implementation Assessment Instrument (3PSISC+). Instrument 3PSISC+ was established to evaluate the implementation of the School Improvement Specialist Coaches Plus program (SISC+) which was authoritatively propelled in 2014. It includes 100 items using the five-point likert scale and is spreads to 168 teachers engaged with the SISC+ program in the state of Selangor. Assessment scales testing in this study can decide the best scale use based on the Rasch Model measurement approach. The five-point Likert scale in this study has satisfied six set criteria, namely 1) every category has more than ten observations, 2) the arc shape of apiece category has its own peak and is not hidden, 3) the average scale size value growths normally, 4) the value of the MNSQ outfit is less than 2 logit, 5) the threshold value rises lengthways with the rating scale category, 6) the block category is more than the value of one and less than the value 5. Thus, the likert scale used in this study is preserved for being comprehensible and to differentiate by the respondent.

Keywords: Verifying Scale, Rasch Measurement Model, Likert Scale, Threshold Value, Scale Category

Introduction

In order to accomplish the achievement of established realm rank, edification has always been one of the foremost causal reasons in a nation, principally in Malaysia (Jamil et al., 2014). Through an inclusive and real education system, high quality humanoid principal near accomplishing the idea of the nation can be created. The educational policy vicissitudes certainly to encounter the national wants of creating excellence inhabitants (Albury & Aye, 2016). Consolidation professional development is a serious aspect in increase the excellence and effect of teachers in the classroom (Roslee, 2011). However, according to Cansdale (2010), when teachers can enhance their expertise in subject matter, teaching strategies and other key management in teaching, the development of teaching professionalism can be

strengthened. In this regard, an effective and sustainable professional development program is expected to be a major contributor to the improvement of school management, boosting the quality of teaching and learning as well as improving the effectiveness in sharing of expertise among teachers in schools. Therefore, the School Improvement Specialist Coaches Plus program (SISC+) is expected to provide an optimum impact on the information gained in terms of subject matter content, strategies and new teaching skills, working with colleagues and analyzing and planning teacher professional learning.

Literature Review

In order to meet current educational needs, teachers need effective training in services to help improve teaching quality (Piper & Zuilkowski, 2015), always on the right track in producing efficient teaching processes (Hilmi & Jamil, 2017) and can improve work excellence and performance (Azmi, 2016).The development of teaching professionals is one of the major contributors to reforming the state education system (Rashid, Rahman, & Rahman, 2016). Through the School Development Program, the government introduced a new post, a full-time teacher mentor called School Improvement Specialist Coaches Plus (SISC+) (Ministry of Education Malaysia (MOE) 2015). The SISC+ Officers selected to provide teaching guidance consist of experienced teachers and experts in their respective subjects, known as Excellent Teachers (MOE 2015). The involvement of excellent teachers in teacher guidance programs can improve the effectiveness of classroom practice which ultimately leads to increased student achievement (Russo, 2007). In addition, the selection of SISC+ among teachers, in particular the Excellent Teacher is important because coaching counselors can improve their self-confidence (Peter & Richard, 2017), carry out more effective coaching tasks (Knight 2007) and work together to explore teachers' (Richards & Farrell, 2005). Therefore, the task of SISC+ is to provide instructional guidance by making the teacher a primary client and not involve having direct contact with the pupils in the classroom. This is because the teacher is a very important change agent and is considered a variable that needs to be developed in order to improve the education system (Villegas-Reimers, 2003).

Based on the Standard Operating Procedure (SOP), 25 to 30 teachers for the states in Peninsular Malaysia need to be selected by SISC+ for guidance. While in the Federal Territory of Labuan, Sabah and Sarawak there are 20 to 25 teachers per SISC+. The frequency of guidance that SISC+ needs to do is at least three times per teacher for the whole year. Nevertheless, the number of teachers guided by SISC+ is not subject to such provision for districts or states with fewer than 25 schools (MOE 2015). The amount set by this SOP is important to be followed so that every SISC+ officer can identify teachers who need teaching guidance. The SISC+ program is not intended to significantly improve the performance of students' examination results, instead focusing on teaching practices in the classroom so as to meet the needs and demands of the education transformation of the 21st century.

The Inspectorate and Quality Assurance, Ministry of Education Malaysia reported that during the period 2011 to 2013, the quality of teaching of teachers has yet to reach a level of excellence. A total of 30,564 total studies were observed, only 3.22% achieved excellent levels and 14.2% were in good standing. On the contrary, 82.59% only reached the level of expectation, satisfactory, weak, and very weak. Similarly, the study conducted by the Higher Education Leadership Academy, Ministry of Higher Education in 2011 on the quality of teacher teaching found that from observation of 125 teaching in 41 schools across Malaysia, only 12% of the lessons were delivered to achieve excellent levels, 38% yet only achieving satisfactory level, but 50% is at satisfactory level. The situation implies that the learning process is not

entirely focused on the involvement of the students, but rather the teacher-centered in delivering subject content. In this case, teachers are more committed to ensuring that students understand the basic content of the subject rather than the effort towards applying high level thinking skills. Therefore, the SISC+ program officially implemented in 2014 is expected to address the weaknesses inherent in teacher teaching. In this regard, teachers need to prepare themselves for the change (Norazilawati et al., 2014) so that their goals and implementation are well understood and well implemented. Sailors and Price (2015) stated that teaching guidance can have a positive impact on teacher pedagogy and subsequently to student achievement.

Problem Statement

SISC+ officers have been assigned to guide teachers to improve teaching quality in the classroom. The role of SISC+ is considered very important as quality teachers can have a significant impact on the success achieved by students (Darling & Youngs, 2003). However, SISC+'s expertise as a teaching guide is questioned by teachers and school administrators. The question arises when many SISC+ officers are not comprised of Distinguished Teachers subjects. SISC+ Implementation Report (MOE, 2014), 233 people have been appointed as SISC+ in Kedah and Sabah, only 21.9% of them are Excellent Teachers. So, certainly the authority of SISC+ staff as a teaching guide is disputed by teachers or school administrators.

The problem of applying high level thinking skills among students has long been raised by researchers. Therefore, the issue of high level thinking application in teacher teaching is one of the main goals of SISC+ task. Studies conducted by Rajendran (2010) show that more than one-third of college students are not convinced to use high-level thinking skills even for themselves. So the question arises, how can the teacher apply it in the teaching of his own teacher does not have that skill? Based on the Malaysian Certificate of Education examination report for the year 2017, pupils are still unable to master high-level thinking skills in answering the paper Mathematics papers 2 and Additional Mathematics paper 1 (MOE, 2017). Dennis Preston (2011), argued that the objective could not be achieved if instructional guide had no clear goals and lack of planning in the field of work.

In addition, the misconceptions among teachers, SISC+ officers are considered as role models for teachers' teaching (MOE, 2014) and the attitude of teachers who do not want to change and are more comfortable with the practice of teacher-centered methods (Radhiah, Rasid & Azhar, 2016) . This situation explains the negative attitude of teachers who need to be faced by the instructors (Knight, 2011; Richard, 2003).

Therefore, the SISC+ program needs to be assessed to track existing weaknesses so that improvements can be made so that the goal of improving the teaching quality of teachers can be achieved. In order to conduct a SISC+ assessment, the scale used should indicate the level of consent of the respondents (Bond & Fox, 2015) and have different and clearer meanings in the same response (Arce Ferrer, 2006; Schaeffer & Presser, 2003). The use of appropriate scale scores can make it easier for respondents to express perceptions of each item in the questionnaire (Rohrmann, 2003). The five-point likert scale was used in this study based on the perception of teachers involved with the SISC+ program, 1 = Never / Very Low, 2 = Very Rarely / Low, 3 = Occasionally / Medium, 4 = Most Time / Height, 5 = Always / Very High (Ahmad Esa 2006; Uma Sekaran 2000).

Verifying the Use of Five Points Likert Scale Based on Rasch Model Measurement Approach

Measurement scales commonly used in studies such as Thurstone, Guttman and Likert scale. The Thurstone scale takes the average percentage of standard normal distribution based on the proportion of respondents' priorities towards one question (Lipovetsky, 2007). The use of the Guttman scale involves the relevance of the level of consent, in which case the individual agrees with the particular item, then the individual agrees with all other questions previously. Therefore, the Guttman scale is rarely used by researchers as it is difficult to obtain legitimate questions (Uhlaner, 2002). The most easy-to-use measurement scale is Likert scale and preferred by many researchers. Likert scale is used to measure individual behavior by responding to 5 points of choice on each item, which is strongly agreed, agreed, neutral, disagreeable, and highly disagreeable (Likert, 1932). Likert scale is not only used to measure individual behavior by measuring ideological variables, perspectives, personal training, and other people's training (Kelly & Tincani, 2013) even in agriculture, Likert scale is also often used to measure consumer acceptance of modified food products (Herath et al., 2013) and peasant selection on the characteristics of wheat crops (Nelson, 2013).

Scale rating is one of the most widely used data collection methods in various disciplines to allow for measurable phenomena that can't be viewed and measured directly, ie the latent variables (Hagell, 2014). The rating (partial credit) scale is used to measure scale rating tests. The results shown are Observedd Average and Andrich Threshold which illustrates the accuracy of the choice given to respondents (Bambang, 2014). Most Likert scales tend to be unevenly assuming the distance between points in a set of responses. Rasch analysis enables respondents' feedback using rating scales to determine the distance that actually occurs during the rating. Rasch measurements diagnostics are used to assess the extent to which these five categories make a set of response functions to make the steps understandable (Kim & Kyllonen, 2006).

Scale review analysis is based on six criteria outlined by Linacre (2002). The use of existing scale can be maintained if the criteria are met. First, each category has more than ten observations. This criterion is important because if the value of the frequency is low, the expectation of Rasch's score accuracy can be a problem. Hence, the higher the frequency value of a category, the better the expectation for the accuracy of a score. Second, the arc shape of each category must have a peak and is not shielded. This situation illustrates the uniform distribution of each category.

Third, the average size of scales increases normally according to the category scale. For example, likert five scales must have a higher-than-low value category, ie likert four, three, two and one. Fourth, the value of MNSQ outfit must be less than 2 logit. Outfit size statistics are about variance. Therefore, the score exceeds 2 logit indicates that the degree of interference with variance is not explained.

Fifth, the threshold value must increase in line with the rating scale category. Sixth, the category of the blockade must be more than one and less than the value of 5. If the difference in value is less than one, then the assumed rating needs to be merged. Whereas if the difference value is more than five, the rating should be split separately.

The five-point Likert scale used in this study is based on Bruner et al. (2001) and Weijters et al., (2010), ie all categories of responses are clearly labeled. The five-point scale used in substructure Role, Knowledge, Feedback, Personal Properties, Role Model, Implementation of Assignment and Effectiveness of Guidance are 1 = Never, 2 = Very Rarely, 3 = Occasionally, 4 = Most Time, 5 = Always . Whereas for substructures Knowledge, the scale is 1 = Very Low, 2 = Low, 3 = Medium, 4 = High, 5 = Very High.

Each level of scale used in this study has different and clearer meanings in the same reaction (Ferrer, 2006; Schaeffer & Presser, 2003). Therefore, it can facilitate respondents to express perceptions of each question in the questionnaire (Rohrmann, 2003). Studies related to measuring perceptions and views are extremely difficult and require precise definition (Rashidi et al., 2014). Validation of measurement instruments, in particular the use of the likert scale to be used is very important in order to produce a better quality study. Likert scale is widely used in instruments to measure opinions, beliefs and attitudes (DeVellis, 2017). The Rasch measurement model consists of a set of instructions to meet the scientifically measurable requirements to be applied in the study (Bond, 2015).

Objective

This study was conducted to verify the five-point likert scale to meet the six criteria set by Rasch's model approach. The objective of this study is to verify:

1. each category has over ten observations,
2. the curve forms of each category have their respective peaks and are not shielded,
3. the average scale size value increases normally,
4. the value of MNSQ outfit is less than 2 logit,
5. the threshold value increases with the rating scale category,
6. block category is more than one and less than value 5.

Methodology

This study uses a quantitative approach by distributing questionnaire instruments to research respondents. The respondents of the study were 168 Malay Language, English and Mathematics teachers involved with the SISC+ teaching program in Selangor. The results of the study were analyzed using the Winsteps Version 3.69.1.11 software based on Rasch's measurement model approach.

This questionnaire has 100 items consisting of four parts. Part A describes respondents' demographics; Section B on SISC+ Features; Part C is in conjunction with the Duty Implementation Process and part D on the Effectiveness of Guidance.

Research Findings and Discussions

Six appropriate criteria are used to assess the effectiveness of a measurement scale using the Rasch model (Linacre, 2002). If the criteria are met, existing scale usage can be maintained. Table 1 shows each category has more than ten observations. The higher the frequency value for a category, the better the expectation for the accuracy of a score. Therefore, the first criterion of Linacre (2002) has been fulfilled.

Table 1
Categories that have over ten observations

| Category | | Observed | | Obsvd | Sample | MNSQ | | Andrich | Category |
Label	Score	Count	%	avrge	Expect	Infit	Outfit	Threshold	Measure
1	1	580	4	-3.25	-3.63	1.40	1.37	NONE	(-5.34)
2	2	1915	13	-2.28	-2.15	1.01	1.04	-4.20	-2.78
3	3	3477	23	.76	.73	1.05	1.10	-1.33	.06
4	4	5232	35	2.79	2.88	.84	.79	1.46	2.78
5	5	3916	26	4.67	4.58	.94	.95	4.06	(5.21)

The second criterion is the arc shape for each category must be vertex and not protected from each other. The findings in Figure 1 show each category have their respective peaks and the peaks for each category are clear and unprotected.

```
         CATEGORY PROBABILITIES: MODES - Structure measures at intersections
          P    -+---------+---------+---------+---------+---------+---------+-
              R  1.0 +                                        +
              O    |                                          |
              B    |                                          |
              A    |11                                      55|
              B  .8 + 1                                    55  +
              I    |  11                                    5  |
              L    |   1      2                             5  |
          I    |     1    222 222     333333     4444     5    |
              T  .6 +    1  2    22  33    3   44   44   5     +
              Y    |     1 22     2 3      33 4     44 5      |
              .5 +       *        23        34        45       +
              O    |      2 1       32        43        54       |
              F  .4 +    2  1      3 2      4 3      5 4      +
                 |    2   1     3  2     4  3     5  4     |
              R    |   2    1    3    2    4    3    5    44   |
              E    |  22       1 33     2 44    3  5     4  |
              S  .2 + 2        1 3       224      355      44  +
              P    |22        3*1        42       533       44|
              O    |         33 11    444 22     55 33        |
              N    |       3333    111*444   222*555   3333      |
          S  .0 +*********************5*************1******************+
              E    -+---------+---------+---------+---------+---------+---------+-
                  -6       -4       -2        0       2       4       6
                        PERSON [MINUS] ITEM MEASURE
```
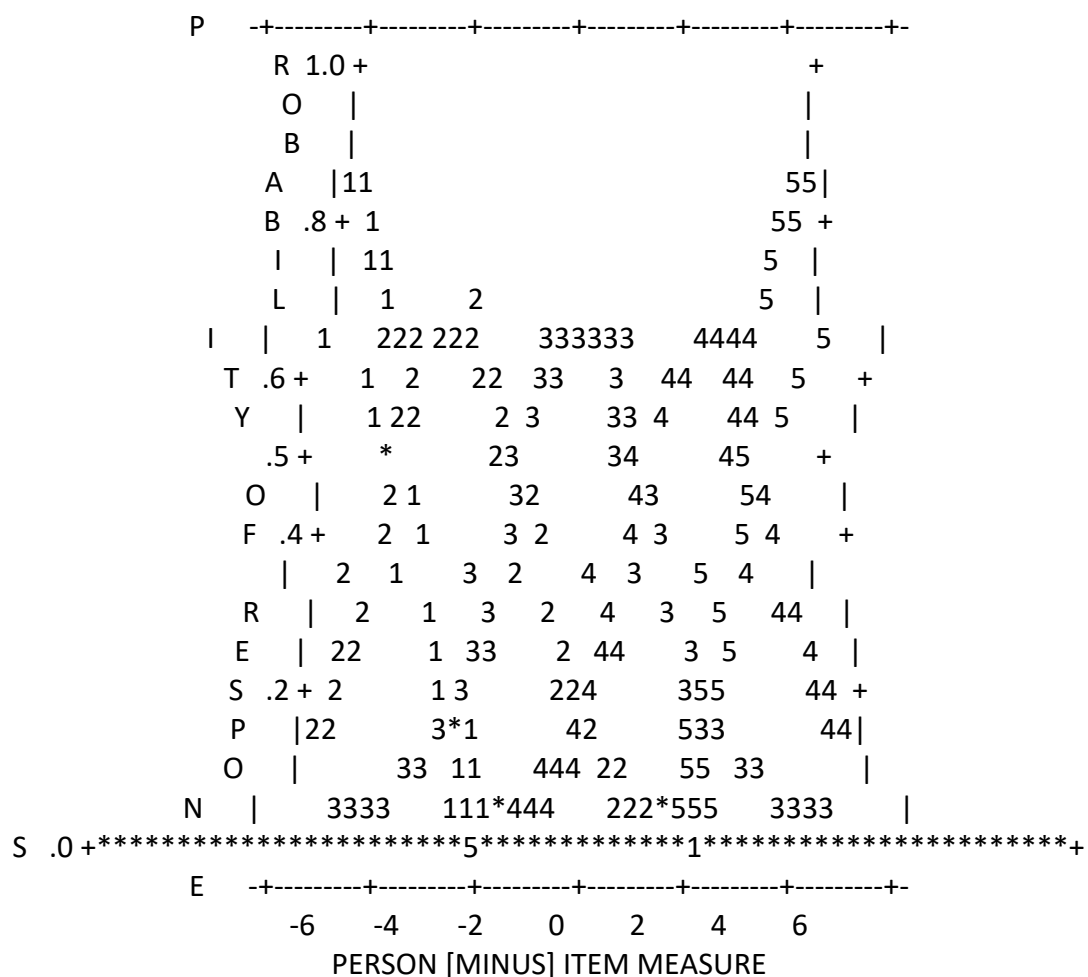
Figure 1 Arms and peak shape

For the third criteria, the average size value for each category should increase in line with the category scale. The findings in this study show that increasing the value of the scale for each scale is increasing uniformly. The proof is that the measurement value for category one is -3.25 logits, category two is -2.28 logits, category three is 0.76 logits, category four is

2.79 logits, and category five is 4.67 logits (Table 2). This increase shows the normal and even pattern of respondent pattern.

Table 2
Average scale size value increases normally

Category		Observed		Obsvd	Sample	MNSQ		Andrich	Category
Label	Score	Count	%	avrge	Expect	Infit	Outfit	Threshold	Measure
1	1	580	4	-3.25	-3.63	1.40	1.37	NONE	(-5.34)
2	2	1915	13	-2.28	-2.15	1.01	1.04	-4.20	-2.78
3	3	3477	23	.76	.73	1.05	1.10	-1.33	.06
4	4	5232	35	2.79	2.88	.84	.79	1.46	2.78
5	5	3916	26	4.67	4.58	.94	.95	4.06	(5.21)

The fourth criteria, the value of MNSQ outfit must be less than two logs. The value of MNSQ outfit for the instrument in this study ranges from 0.95 to 1.37 and shows less than two logs. MNSQ outfit size statistics explain about variance, ie the score above two logs indicates the level of interference for the unexplained variance (Table 3).

Table 3
Value of MNSQ outfit is less than 2 logit

Category		Observed		Obsvd	Sample	MNSQ		Andrich	Category
Label	Score	Count	%	avrge	Expect	Infit	Outfit	Threshold	Measure
1	1	580	4	-3.25	-3.63	1.40	1.37	NONE	(-5.34)
2	2	1915	13	-2.28	-2.15	1.01	1.04	-4.20	-2.78
3	3	3477	23	.76	.73	1.05	1.10	-1.33	.06
4	4	5232	35	2.79	2.88	.84	.79	1.46	2.78
5	5	3916	26	4.67	4.58	.94	.95	4.06	(5.21)

The fifth criterion, the threshold must increase along with the rating scale category. The findings in Table 3 indicate that the threshold thresholds are structured (-4.20, -1.33, 1.46 and 4.06) and can be seen on the size structure (Table 4). This criterion explains the tendency of respondents to select a scale evenly and the findings do not suffer from step disordering problems, ie the low probability of a category chosen by the sample.

Table 4
Threshold value increases with the rating scale category

Category		Observed		Obsvd	Sample	MNSQ		Andrich	Category
Label	Score	Count	%	avrge	Expect	Infit	Outfit	Threshold	Measure
1	1	580	4	-3.25	-3.63	1.40	1.37	NONE	(-5.34)
2	2	1915	13	-2.28	-2.15	1.01	1.04	-4.20	-2.78
3	3	3477	23	.76	.73	1.05	1.10	-1.33	.06
4	4	5232	35	2.79	2.88	.84	.79	1.46	2.78
5	5	3916	26	4.67	4.58	.94	.95	4.06	(5.21)

The sixth criteria, the category of restrictions must exceed one but less than the value of five (Bond & Fox, 2007; Rasch, 1961, 1980). If the difference is less than one, the assumed rating should be incl. And if it exceeds the value of five, then the rating must be separated. Figure 1 shows curves of categories that indicate the blockage for a five-point scale, which has a clear peak of each category and is not shielded. The boundaries between the categories shown below are above the value of one and less than five to be maintained as a scale (Table 5). Borders between one and two categories are broader than other category borders. This finding shows that the probability of one and two categories selected by respondents is high (Mohd Effendi @ Ewan, 2015).

Table 5
Block category is more than one and less than value 5.

Category		Observed		Obsvd	Sample	MNSQ		Andrich	Category
Label	Score	Count	%	avrge	Expect	Infit	Outfit	Threshold	Measure
1	1	580	4	-3.25	-3.63	1.40	1.37	NONE	(-5.34)
2	2	1915	13	-2.28	-2.15	1.01	1.04	-4.20	-2.78
3	3	3477	23	.76	.73	1.05	1.10	-1.33	.06
4	4	5232	35	2.79	2.88	.84	.79	1.46	2.78
5	5	3916	26	4.67	4.58	.94	.95	4.06	(5.21)

$$S1\text{-}2 = 0.00 - (-4.20) = 4.20 \ (> 1.0)$$
$$S2\text{-}3 = -4.20 - (-1.33) = 2.87 \ (> 1.0)$$
$$S3\text{-}4 = -1.33 - (-1.46) = 2.79 \ (> 1.0)$$
$$S4\text{-}5 = 1.46 - (4.06) = 2.60 \ (> 1.0)$$

Restrictions in this study have been able to meet the criteria set by Linacre (2002), which exceeds the value of 1.0 and less than the value of 5.0.

Therefore, the five-point likert scale in this study can be maintained, ie 1 = Never, 2 = Very Rarely, 3 = Occasionally, 4 = Most Time, 5 = Always (substructure Role, Knowledge, Feedback, Role Model, Implementation of Task and Effectiveness of Guidance). Whereas for substructures Knowledge, the scale is 1 = Very Low, 2 = Low, 3 = Medium, 4 = High, 5 = Very High. Based on the results obtained, the calibration structure in this study is within a normal range.

Conclusion

In conclusion, testing of scale scales that can determine the best use of scale is one of the important aspects to consider in developing new instruments for a study. The five-point likert calibration in this study has met six criteria set out based on Rasch's model measurement approach. Testing of scale rating for each item in this study shows that the built-in instrument is able to evaluate the implementation of the SISC+ program more effectively.

Based on the findings of the analysis, the use of five-point likert scale has the proper use of quality to evaluate the implementation of the SISC+ program as it is understood and distinguished by the respondents. This instrument can provide a clearer and more accurate descriptive analysis of the implementation of the SISC+ program that you want to assess. Implications from this analysis can help researchers to assess the strengths and weaknesses

that exist throughout the implementation of the SISC+ program. According to Rensburg (2005), an instrument to measure the role of counselors can be an important tool in the effective selection process of counselors. This is important because the objectives and goals are the standard of success and failure of a program implemented (Zawawi, 2008). However, this study is limited in Selangor only. To obtain more accurate and consistent findings, it is proposed that future studies be conducted throughout Malaysia. Further studies are also suggested to cover both parties, namely teachers involved with guidance and SISC+ officers.

Acknowledgement

Corresponding Author

Salefah Mat Salleh, Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia.
Email: salefahms@gmail.com

References

Ahmad, E. (2006). Penerapan Kemahiran Generik Menerusi Kokurikulum Di Politeknik Bagi Memenuhi Keperluan Industri Di Malaysia. PhD's Thesis. Universiti Tun Hussein Onn Malaysia.
Albury, N. J., & Aye, K. K. (2016). Malaysia's national language policy in international theoretical
context. *Journal of Nusantara Studies*, *1*(1), 71-84.
Bambang, S. & Wahyu, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. (Bambang Trim, Ed.). Cimahi, Indonesia: Trim Komunikata Publishing House.
Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates, Publisers. Mahwah, New Jersey. London.
Bruner, G. C., James, K. E., & Hensel, P. J. (2001). *Marketing Scales Handbook. A compilation of Multi Item Measures*. Volume III. Chicago: American Marketing Association.
Darling-Hammond, L. & Youngs, P. (2003). Defining "highly qualified teachers": what does "scientifically-based research" actually tell us? *Educational Researcher.*
Dennis Preston. 2011. Executive Coaching for HR Professionals? Article *HDA*. U.K.
Devellis, R. F. (2017). *Scale development: Theory and Application*. Edisi ke-4. Los Angeles: Sage Publication.
Jamil, H., Razak, N. A., Raju, R., & Mohamed, A. R. (2014). Teacher professional development in Malaysia: Issues and challenges. *Africa-Asia University Dialogue for Educational Development Network*, 85-102.
Kelly, Amy & Tincani, M. (2013). Collaborative Training and Practice among Applied Behavior Analysts who Support Individuals with Autism Spectrum Disorder. *Education and Training in Autism and Developmental Disabilities* 48(1) pp: 120–131
Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oak, CA: Corwin.
Knight, J. (2011). *Unmistakable impact*. Thousand Oaks, CA: Corwin.
Likert, R. (1974). The method of constructing an attitude scale. Dalam*,* Marannell, G.M. (Eds)*, Scale: A Sourcebook for behavioral*. Chicago, IL: Aldine

Publishing Company.

Likert, R. (1932). Technique for the measurement of attitudes. Archives of Psychology, 140 pp: 1-55

Lincare, J. W. (2002). What do infi t and outfi t, mean-square and standardized mean?, Rasch Measurement Transaction 16(2): 878. http:www.rasch.org/rmt/rmt162fhtm

Liposvetsky, S. (2007). Thurstone Scaling in Order Statistics. *Mathematical and Computer Modelling* 45 pp: 917 – 926

Ministry of Education Malaysia. (2015). Panduan Pengurusan Program Transformasi Daerah Putrajaya: Kementerian Pendidikan Malaysia.

Ministry of Education Malaysia. (2017). Panduan Pengurusan Program Transformasi Daerah (DTP). Putrajaya: Kementerian Pendidikan Malaysia.

Azmi, M. Y. (2016). Penilaian Program Latihan Dalam Perkhidmatan Pentaksiran Berasaskan Sekolah (PBS). PhD's Thesis. Universiti Kebangsaan Malaysia.

Ewan, M. E. M. M. (2015). Pembinaan instrumen kecerdasan menghadapi cabaran (IKBAR) bagi pelajar politeknink menggunakan model Rasch. PhD's Thesis. Universiti Sains Malaysia.

Piper, B., & Zuilkowski, S. S. (2015). Teacher coaching in Kenya: Examining instructional support in public and nonformal schools. *Teaching and Teacher Education*, 47, 173-183.

Rajendran, N. (2010). Teaching thinking skills at Institutions of Higher Learning: Lessons learned. *Pertanika J. Soc. Sci. & Hum.* Vol. 18 (S).

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test.* Copenhagen: Denmarks Paedagogiske Institute.

Rasch, G. (1980). *Probabilistic model for some intelligence and attainment test*. Chicago, IL: University of Chicago Press.

Rashid, R. A., Rahman, M. F. A., & Rahman, S. B. A. (2016). Teachers' engagement in social support process on a networking site. *Journal of Nusantara Studies*, *1*(1), 34-45.

Rashidi, M. N, Ara Begum, R., Mokhtar, M. & Pereira, J. J. (2014). Implementation of Data Analysis Using Rasch Measurement Model for Determining the Weights of Items. *Journal of Advanced Research Design ISSN* (online): 2289-7984 | Vol. 2, No.1.

Rohrmann, B. (2003). Verbal Qualifiers for Rating Scales: Sociolinguistic considerations and psychometric data. *Working paper*.

Radhiah, R. S., Rasid, A. J., & Azhar, M. S. (2016). Pengetahuan dan Kefahaman Skop Pengajaran dan Pembelajaran Bahasa Melayu dalam Kalangan Jurulatih Pakar Pembangunan Sekolah (SISC+). *International Journal of Education and Training* (InjET) 2(2) November: 1- 9.

Russo, A. (2007). School-based coaching: A revolution in professional development or just the latest fad? *Harvard Education Letter*.

Sailors, M., & Price, L. (2015). Support for the improvement of practices through intensive coaching (SIPIC): A model of coaching for improving reading instruction and reading achievement. *Teaching and Teacher Education*, 45, 115e127.

Uhlnaer, Lorraine, M. (2002). The Use of the Guttman Scale in Development of a Family Bussiness Index. *Research Report H2002003*. SCALES: Zoertemeer

Uma, S. (2000). *Research Methods for business: A Skill Building Approach*. Singapore: John Wiley & Sons, Inc.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on

response styles: The number of response categories and response category labels. *International Jurnal of Research in Marketing* 27 (2010) 236–247. doi:10.1016/j.ijresmar.2010.02.004.

Zawawi, I. (2008). *Penilaian Pelaksanaan Bahasa Kurikulum Kemahiran bertutur Bahasa Arab Komunikasi di sekolah Menengah Kebangsaan Agama*. PhD's Thesis. Universiti Kebangsaan Malaysia.