

# Severity and Leniency Effect in Assessing Teaching Practicum: Analysis using Multi Facet Rasch Measurement (MFRM)

Siti Wan Aisha W. A, Ahmad Rizal M, Nurfirdawati M. H, Hashima H, Nur Afifah M. S

**Abstract:** Practical teaching assessment is a performance assessment that can be exposed to threat regarding the validity and reliability as the assessment is made subjectively. If it is not measured correctly, it can give effect to a negative score. The purpose of this study is to analyse the severity and leniency effects between raters through a variable map in Multi Facet Rasch Measurement (MFRM) analysis. Instrument of study is an e-PLM assessment rubric that is assessed based on five assessment criteria which are planning, implementation, personality, portfolio and reflection. This study is done quantitatively. Data were collected from 170 of teachers candidate that undergoing teaching practical in Malaysia peninsula and 194 raters that gave the rate towards all the teachers candidate. Raters consists of the lecturers from UTHM and teacher's counselor from Vocational College who are sampled by using purposive sampling technique. Data was analyzed using analysis software Multi Facet Rasch Measurement (MFRM) Version 3.82.2. In the study, it was reached to the results that the rater's severity holds an unexpected effect towards group level and individual level in assessment based on the logit value, infit MNSQ, outfit MNSQ, Model S. E and t-value. Based on results analysis, three rater are found to be more severe and inconsistent in assessing students. While, eight raters are found to be more lenient in making the assessment. It can be concluded that these raters was inconsistent in rating student performance. From the findings, the result shows that there are severity/leniency effects at the individual level while severity/leniency effects does not exist at the group level. However, to summarize this study gives a huge impact on validity and reliability towards performance assessment. Besides that, this study also can assist rater in the assessment field, assessment practice as well as lead towards the valid and quality assessment.

**Keyword:** Teaching practicum, performance assessment, rater severity/leniency, Multi Facet Rasch Measurement (MFRM)

## I. INTRODUCTION

The quality of teachers candidate is important to be developed and expanded in line with the goal of National Philosophy of Education (NPE). A teacher candidate can teach in professional, comprehensive and proactive as well as can apply the theory that been introduced by joining the education

**Revised Manuscript Received on January 15, 2020**

Correspondence Author

**Siti Wan Aisha Wan A'laudden.** Faculty of Technical & Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.

**Ahmad Rizal Madar.** Faculty of Technical & Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.

**Nurfirdawati Muhamad Hanafi,** Faculty of Technical & Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.

**Hashima Hamid.** Faculty of Technical & Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.

**Nur Afifah Mohd Sofian,** Faculty of Technical & Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.

course [1]. During the professional practice, guiding and training is the major focus and the observation towards the teachers candidate is also been exercised. Training and preparing are important in facing the teaching session as it works as the additional guidance towards the teacher [16]. In order to fulfill the goal of NPE in producing a teachers candidate that full of quality and skill, every teachers candidate that undergo the practical teaching will have to face the assessment during the training. The training is held by the teacher's counselor and the supervising lecturer [21]. The assessment of the practical teaching is the bench mark as to access whether the teachers candidate is ready to involve in reality as a teacher [3, 15].

Special assessment is needed in order to trace the capability of the teachers candidate performance in implementing the teaching process in order to know how far the capability has been achieved. The issue of validity and reliability are been emphasized during the assessment [23]. The good assessment can produce a good quality of graduate in TVET course. The performance assessment in the teaching practicum can be exposed to threat regarding the validity and reliability as the assessment is made subjectively. The subjective assessment is difficult because of several factors such as the assessment instrument, the raters, the situation when the assessment been made or the individual that been assessed itself. This matter can be bias and impartial towards the raters. The systematic approach is required during the assessment to provide psychometric quality that is the basis of high validity and reliability [23]. There are many aspects that the raters should consider during the assessment process. Among the aspects are the performance of the teachers candidate, the item difficulty and domain, the raters severity/leniency and score that use the rating scale. All these aspects consist of the relevancy that affect the accuracy during the performance measurement [24].

In the practical teaching programme, the control and assessment process must be transparent and fair and can only be achieved when the criteria and standard that been used is clear to assess the teachers candidate along with the description of work at the end of the practical. The assessment is created so that there will be no bias or impartial issue among the raters and ratees. Not all matters must be assessed, the assessment focused on the achievement in fulfilling the learning outcomes [17]. A rater must make the criteria reference, thorough but fair, choose the suitable assessment, make the trustworthy evaluation and adequate as well as authentic.

# Severity and Leniency Effect in Assessing Teaching Practicum: Analysis using Multi Facet Rasch Measurement (MFRM)

Raters play important role in making the assessment towards the teacher candidates. If it is not been measured correctly, the assessment can give a huge impact towards the negative score. For instance, the severe raters can possibly rate worse score than it actually is, while the lenient raters may rate better. This case based on the severity/leniency of the group level and individual level of raters [22]. Severity/leniency is one of the rater effects besides the halo effect, central tendency, bias, inconsistency and restriction of rater [19]. The raters severity/leniency is one or more raters making the assessment that depends on the individual performance whether it is higher or lower from what it supposed to be [6]. The raters effect influence the validity and reliability towards the performance score of the teacher candidates [10]. The unfairness that the raters made in giving an assessment can lead to a serious impact towards the credibility of the performance of assessing system although it has a good system [20].

Raters must be proficient in making the assessment so that each teacher candidates can prepare themselves in the spiritual and intellectual aspects in order to carry the responsibility as the future teacher [18]. However, there are some of the teacher candidates that been evaluated by the raters in different courses from the teacher candidates itself. This matter must be taken into consideration since the raters are those who are responsible to ensure that the goal of teaching practicum is achieved [9]. The role of the raters is to give the assessment and score towards the teacher candidates by taking several domain into consideration such as the performance of the teacher candidates, personality and professionalism of the teacher candidates, the knowledge and experience of the raters, the severity/leniency of the raters in rating the score and the suitability of the rubric item. The assessment is been carried out based on the criteria that must be followed, however there are still inconsistency in the assessment process.

The raters severity/leniency can be investigated using the Multi Facet Rasch Measurement (MFRM) model [4]. The rubric is one of the bench mark that can be used to make an assessment with regards to the performance of the teacher candidates. In this study, the analytic rubric assessment is used to perform the assessment as it is more specific with regards to the assessment that require observation to evaluate the performance [15]. The Multi Facet Rasch model has an important role to observe the raters effect in the performance determination process. It provide raters severity/leniency in addition to inter-individual ability differences in the Rasch model [11]. The raters severity/leniency, raters consistency, the scale functioning and the interaction between the variables can be observe using Multi Facet Rasch model [5]. The rater severity/leniency, rater consistency, interaction of variables and scale functioning can be observe by using Multi Facet Rasch model [24]. There are advantages by using Multi Facet Rasch model which is that it allows the raters effect (halo effect, bias, randomness effect, central tendency) to be investigated where more than one raters is used [5].

## II. OBJECTIVES OF THE STUDY

The objective of this study is to identify the effect of the severity and leniency in assessing the practicum teaching towards the teachers candidate of Bachelor in Vocational Education, Faculty of Technical and Vocational Education, UTHM. The assessment is been observed based on the aspect

of rater severity and student performance. This is to help in protecting the validity and reliability of the assessment result that is received by the teachers candidate. The rubric assessment is suitable to use as to evaluate the performance of the teachers candidate during the process of practical teaching. This study can aid the faculty and university in making the right decision towards the real performance of the teachers candidate. In line with the issue that been discussed above, the main objectives of this study are:-

1. To identify the existence of severity at the group-level of rater in teaching practicum assessment.
2. To identify the existence of severity at the individual-level of rater in teaching practicum assessment.

## III. METHODOLOGY

### A. Design of Study

The pattern of this study that been carried out is the empirical study that used the quantitative approach. The study used an previous empirical data set of *e-PLM* analytical assessment rubric in order to receive the information of the raters assessment. The assessment rubric consist of five major aspects which are the planning, implementation, personality, portfolio and reflection.

### B. Group of Study

170 of the teachers candidate and 194 of the raters are the sample of this study by using purposive sampling technique. The study is been done towards the raters that assessing the vocational teachers candidate from Faculty of Technical and Vocational Education, UTHM that undergoing the teaching practicum for 12 weeks at vocational college in Malaysia Peninsula for the semester of 2016/2017.

### C. Data Analysis

The quantitative data will be collected by using the empirical data sets of existing *e-PLM* assessment rubric and been analyzed by using the Multi Facet Rasch Measurement (MFRM) Software Version 3.82.2 (FACETS). Multi Facet Rasch Measurement is used to make the analysis of infit MNSQ and outfit MNSQ, chi square test, the separation index, the separation ratio, rater reliability and logit value. MFRM approach allows to look at group-level and individual-level effects of the various elements within a facet included individual raters, rates or traits in the analysis. The raters makes the assessment that is subjective during the observation and towards the student performance as well as the potential value that they have within them [5]. This study involved three facets which are (i) the rater effect, (ii) student performance and (iii) item difficulty.

## IV. RESULTS AND FINDING

The findings of this study analysis are given below. Measurement models Multi Facet Rasch Measurement (MFRM) used for detecting the severity and leniency effect of the raters [20]. This study using a hybrid 2 model shown in Table 1 to analyze their rating data will obtain the group-level and individual-level statistical indicators described below.

Table 1 shows the analysis of the most severe raters and most lenient raters among 194 raters.

**A. Group-Level Statistical Indicators**

Software Multi Facet Rasch Measurement (MFRM) Version 3.82.2 (FACETS) was used to analyze the data. The output from a MFRM analysis using a hybrid 2 model includes four group-level statistical indicators of a group level severity and leniency effect. The indicators include a fixed chi-square, rater separation ratio, rater separation reliability and rater separation index.

**1. A fixed chi-square test**

From Table 1, the chi-square value of 4277.8 with 198 degree of freedom is statistically significant  $0.00 < 0.005$ , where  $p < 0.05$ , signifying that the raters did not all exercise the same level of severity when evaluating ratees. A significant rater fixed chi-square simply means that the severity measures for 194 raters included in this study are significantly different. However, it is important to emphasize that the rater fixed chi-square test is very sensitive to sample size.

**2. The rater separation ratio**

As seen from Table 1, the rater separation ratio is 3.42 in the whole sample of raters. This means that the differences between rater severity are over three times greater than the error with which these severity are measured.

**3. The rater separation reliability and rater separation Index**

Acceptable value of the rater separation reliability and the rater separation index if the value of respondent reliability  $\geq 0.8$  and separation index  $\geq 2.0$  [12]. The most desirable result is to have a reliability of rater separation close to zero which indicates a very high level of rater severity. The rater separation reliability and rater separation index was shown in Table 2.

The rater separation reliability is shown in Facets output (Table 1). The high degree of rater separation reliability of  $0.92 \geq 0.8$ , implies that raters are differentiated in terms of the levels of severity they exercise [13]. While, the rater separation index does not appear as part of Facets output. The rater separation index is calculated through the formula  $(4G + 1) / 3$  where, G is the rater separation ratio, which is shown in Facets output. The rater separation index in this study computed  $[4(3.42) + 1] / 3 = 4.89$  shown that the rater separation index of 4.89 suggests that there are about five statistically distinct strata of rater severity in the sample of raters. The separation reliability is 0.92 where the reliability

**Table 2: The Rater Reliability of The Separation Index - Facets output**

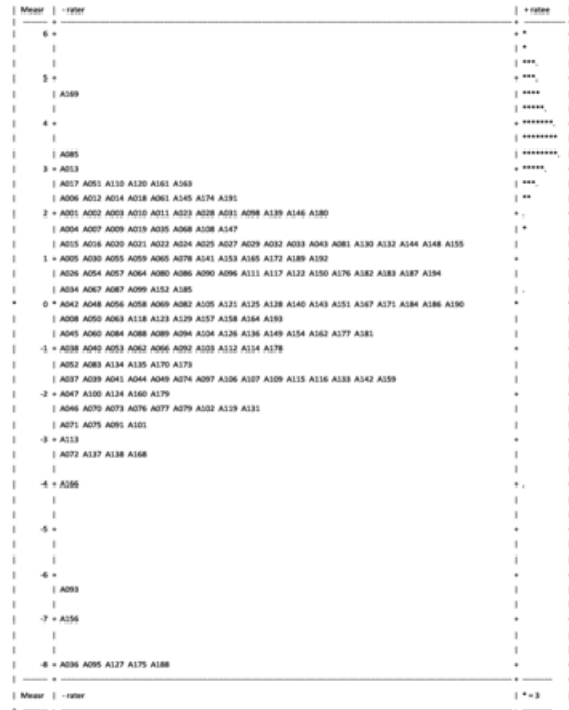
Rater	
Reliability	Separation Index
0.92	4.89

is close to zero, it can be there are no significant differences of raters severity/leniency at group level [7]. The raters severity/leniency does not exist at the group level.

**B. Individual-Level Statistical Indicators**

Software Multi Facet Rasch Measurement (MFRM) Version 3.82.2 (FACETS) was used to analyze the data. The effects of rater leniency and severity of individual-level are shown on

the variable map in Facets output (Table 1). Fig 1 shows a variable map for the severity simulation of the distribution of rater severity measures and the distribution of ratees performance measures. Raters are ordered in the variable map in terms of the levels of the severity each exercised. More severe raters appear at the top of column, while more lenient raters appear lower in the column [5]. The rater severity measures for this study shown in the variable map (Fig 1) as below.



**Figure 1 : Variable Map for Severity Simulation - Facets output**

# Severity and Leniency Effect in Assessing Teaching Practicum: Analysis using Multi Facet Rasch Measurement (MFRM)

Total score	Total count	Obsvd average	Fair (M) Average	- Measure	Model S. E	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtMea Correlation	Nu Raters
83	25	3.32	2.88	4.58	0.29	0.49	-2.1	0.55	-1.8	0.10	169 A169
92	25	3.68	3.49	3.30	0.34	1.02	0.1	1.11	0.4	0.10	85 A085
357	100	3.57	3.66	2.85	0.16	0.55	-3.3	0.56	-3.3	0.49	13 A013
124	25	4.96	4.98	-4.06	1.03	1.05	0.3	1.68	0.8	0.25	166 A166
123	25	4.92	5.00	-6.22	0.74	0.88	0.0	0.50	-0.4	0.04	93 A093
107	25	4.28	5.00	-7.07	0.40	0.87	-0.3	0.90	-0.2	0.24	156 A156
125	25	5.00	4.96	-3.44	1.84	Minimum		Minimum		0.00	175 A175
125	25	5.00	4.98	-4.06	1.84	Minimum		Minimum		0.00	127 A127
250	50	5.00	4.99	-4.94	1.83	Minimum		Minimum		0.00	95 A095
125	25	5.00	4.99	-5.18	1.84	Minimum		Minimum		0.00	188 A188
125	25	5.00	4.99	-5.54	1.84	Minimum		Minimum		0.00	36 A036
186.7	43.8	4.42	4.42	-0.12	0.45	0.95	-0.2	0.96	-0.2	0.22	Mean
170.2	43.3	0.39	0.39	1.88	0.28	0.33	1.5	0.39	1.5	0.20	(Count:194) S.D (Sample)
Populn:RMSE (Model) .53 Adj (True) S.D 1.80 Separation 3.41 Strata 4.88 Reliability 0.92											
Sample:RMSE (Model) .53 Adj (True) S.D 1.81 Separation 3.42 Strata 4.89 Reliability 0.92											
Fixed (all same) chi-square: 4277.8 d.f: 193 Significance (probability): .00											

**Table 1: Rater Measurement Report of Most Severe and Lenient Raters from an Analysis Using Hybrid Model - Severity Simulation (Facets output)**

Fig 1 shows graphically depict the manner in which rater severity and leniency are computed by MFRM analysis. Facets also includes as part of its output a table that provides the individual rater severity measure in logit and the standard error of each severity estimate. The individual rater severity measures for all 194 raters. Analysis for whole result demonstrates that the raters severity measure is at good levels, which is medium severity level. However, there are a number of raters which are very severe and very lenient in assessing students performance in teaching practicum.

Based on results analysis, three rater are found to be more severe and inconsistent in assessing students. The raters are A169, A085 and A013. From Fig 1, Rater A169, A085 and A013 stands out as being more severe with severity measure respectively +4.58 logit, +3.30 logit and +2.85 logit. While, eight raters are found to be more lenient in making the assessment. The raters are A166, A093, A156, A175, A127, A095, A188 and A036. The severity measure of the raters which are more lenient from this analysis range from -4.06 logit to -7.07 logit. Among the raters the most severe according to logit values is A169. The rater A156 is the most lenient than other raters. The *t*-values test are performed to determine whether raters severity measures are significantly different [19]. To obtain *t*-values, the severity measure (logit) of each raters is subtracted from the mean logit measure and divided by the standard error. The standard error for each severity measure can be seen in the 'Model S. E' column [19]. The *t*-values for the group of eleven raters which are more severe/lenient are given in Table 3. The individual rater severity measures for the eleven raters are shown in the 'measure' column of Table 1. The larger the measure, the more severe the rater [14]. The mean of the rater severity measures is shown in Table 1. In addition, the severity measure for this eleven raters listed in the table are all within the range of +4.58 to -7.07 logit. The severity measure for Rater A169 (+4.58 logits) is a conspicuous outlier, which is over 16.21 standard errors above the mean severity of the group.

**Table 3: Rater *t*-value test - Facets output**

Rater	<i>t</i> -value
A169	16.21
A085	10.06
A013	18.56
A166	-3.83
A093	-8.24
A156	-17.38
A175	-1.80
A127	-2.14
A095	-2.62
A188	-2.75
A036	-2.95

Based on the most severe/lenient raters measurement report analysis (Table 1), the raters range of the infit values from minimum value to 1.05, while the values range of outfit from minimum value to 1.68. Acceptable range fixed values are between 0.5 and 1.5 [12]. In this study, it can be declared that the infit value of the raters except for rater A169, A175, A127, A095, A188 and A036 are acceptable. As can be seen the outfit value of the most severe rater A169 are lower than the accepted range limit and the outfit value of the most lenient rater A156 are in acceptable range limit. From the result shows that the rater A169 gave a low score in assessment and rater A156 gave high marks in assessment. The raters severity/leniency does exist at the individual level.

## V. DISCUSSION AND CONCLUSION

The study aimed was to to analyse the severity and leniency effects between the group-level of raters and individual-level of raters in assessing student performance of teaching practicum using Multi Facet Rasch measurement hybrid model. The purpose of the analysis made have resulted the significant differences between the raters severity/leniency. As shown in the result, the rater A169 is the most severe, and rater A156 is the most lenient than other raters. It can be concluded that these raters was inconsistent in rating student performance.

In producing a skilled professional teacher, teaching practicum program opens up the opportunity for teachers candidate to improve their teaching skills, broaden the experiences, gain theory or practice of teaching and learning process as these training programs can create quality teacher [8, 25]. A quality teacher should be competent in terms of skills, knowledge and attitudes.

Based on the results from the variable map, it is found that the trait distribution pattern of the teacher candidates is normalized and has a higher ability than the rater severity. Over 80% of teacher candidates have high proficiency. This finding is in line with the findings of previous study [2]. From the results of the study, it can be assumed that it is possible that the teachers candidate has high ability, very lenient raters, very low scale or the raters cannot distinguish the rating scale well. What is most likely to happen is that teachers candidate have high abilities or that the raters rates high scores to low ability candidates.

The rater severity/leniency was measured based on a logit scale whether it is severe or lenient. On the variables map, the rater at the top is a very severe rater, while the rater at the bottom is a very lenient rater. According to the variables map, all raters was consistent in assessing except eleven raters who was very severe and very lenient in making assessment. In this study, only 3 out of 194 raters were severe. However, the number of severe raters is small but these 3 raters still give the teachers candidate a score. Although only 3 raters are severe but it affects the teachers candidate score. The findings of this study show that each rater has a high degree of firmness but a high degree of consistency. This finding is in line with previous studies suggesting that the effects of severity/leniency may exist because the raters are from various backgrounds [2]. The difference of raters severity/leniency effect can also be seen based on the varying rating scales. Some examiners can be generous in some situations and firm in others [22]. The examiner should have the same assurance that the assessment made is fair and transparent [5].

Based on the findings of the above study, the three most severe rater are the supervisors from UTHM, while the most lenient rater is the teacher's counselor from vocational college. The effect of raters in this study can be clearly identified from the analysis of group level and individual level. Besides, the raters have difference severity/leniency effect even though training has been given in assessment. Preliminary training should be given to the raters to improve the quality of the assessment transparently and fairly. The severity/leniency effect does not exist at the group level but does exist at the individual level. These findings indicate that raters interaction with teachers candidate is significant. The findings of this study can contribute to assessment practices, knowledge, areas of assessment that lead to valid and quality assessments. High severity/leniency effects will exist in subjective assessment if not controlled. Therefore, briefing on assessment is essential to achieve consistency in assessment between raters. The integrity of the rater is also emphasized. If the rater is biased towards the rate such as giving high scores to low ability students or giving low scores to high ability students, it may threaten the validity of the scoring of scores.

Overall, assessment is an important element in translating student achievement. The implication on assessment must ensure to quality assessment. Therefore, the practice of giving

students appropriate scores is very important. Although the raters are of the same background, have degrees in the same field and have several teaching experience, it is not possible to consistently assess students without a specific guide. The findings of the study did not eliminate the effect of raters severity/leniency but were able to maintain the internal consistency of the raters. The raters have different level of severity/leniency and require training on assessment practices in order to control the consistency during the assessment.

## ACKNOWLEDGEMENT

The success and final outcome of this paper required a lot of guidance from many people and I am extremely fortunate to have got this all along the completion. I would like to express my gratitude and appreciation to my supervisor Dr Nurfirdawati Binti Muhamad Hanafi who gave me the possibility to complete the paper for many simulating discussions concerning the design of study, analysis, evaluation of writing, comments and guidance and for comments that greatly improved the manuscript. I would also deeply thank to University Tun Hussein Onn Malaysia who had done the arrangements of publishing this article under the PPG Research Grant with Vot. No. K012.

## REFERENCES

1. Abdul Majid Khan, A. R. (2007). *Teacher As A Researcher*. Kuala Lumpur: PTS Professional Publishing Sdn Bhd.
2. Ab Rahman, A (2017). *Build And Enhancing Rubric Competence Of Electronic Course Practice At Vocational College*. Faculty of Education, National University of Malaysia.
3. Che Nudin, C. M. N. (2007). *Lecturer's Supervision of Teacher Teaching - Teacher Training During Teaching Practice*. University of Technology Malaysia.
4. Eckes, T. (2009). On common ground? How raters perceive scoring criteria in oral proficiency testing. In A. Brown & K. Hill (Eds.), *Tasks and criteria in performance assessment: Proceedings of the 28th Language Testing Research Colloquium* (pp. 43-73). Frankfurt, Germany: Lang.
5. Eckes, T. (2011). Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments, 30-32. *Internationaler Verlag Der Wissenschaften Frankfurt: Peter Lang*.
6. Engelhard, G. (2003). Monitoring raters in performance assessment. In G. Tindal & T.M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 261-287). Mahwah, NJ: Erlbaum.
7. Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge
8. Hashim, S., Jacob, R. & Ahmad, M. Z. (2007). *Pedagogy: Strategy and Techniques Teaching Effectively*. Bentong, Pahang: PTS Publications & Distributors.
9. Ibrahim, M.A., Ramli, M.R., Kamaruddin, .M.I., Megat Zakaria, M.A.Z. (2011). *Comparison of Assessment Between Lecturers and Teachers Towards the Competence of UTM Teacher Candidates*. University of Technology Malaysia.
10. Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.) *Educational measurement* (4th ed., pp. 387-431). Westport, CT: American Council on Education/Praeger.
11. Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.
12. Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.

# Severity and Leniency Effect in Assessing Teaching Practicum: Analysis using Multi Facet Rasch Measurement (MFRM)

13. Linacre, J. M. (2010). Facets Rasch measurement computer program, version 3.66.2. Chicago, IL: Winsteps.com.
14. Linacre, J. M. (2017). A user's guide to FACETS. Rasch-model computer programs. Chicago IL.
15. Madar, A.R., Sun, C.E. & Hamid, H. (2019). Facilitating tolerance test of creative thinking use in Malaysian TVET research: The initial step of inter-rater reliability determination. *Journal of Technical Education and Training*, 11 (1): 100-108.
16. Ministry of Education Malaysia (2013). Malaysian Education Development Plan. Putrajaya: Ministry of Education Malaysia.
17. Moh, E. (2016). Cultivating Integrity in School-Based Assessment (PBS): Educators' Guide to Effective Assessment. ISBN 978-967-14184-6.
18. Muhamad Hanafi, N. (2016). Development and Verification of Evaluation Section Achievement for Assessing Architectural Design Projects In Polytechnic Malaysia. Faculty of Education and Human Development, UPSI.
19. Myford, C.M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4.
20. Myford, C.M., & Wolfe, E.W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5.
21. Othman, N., & Amiruddin, M.H. (2010). Different Perspective Of Learning Styles From VARK Model. *Procedia Social and Behavioral Sciences*, 7 (C), 652-660.
22. Prieto, G., & Nieto, E. (2014). Analysis Of Rater Severity On Written Expression Exam Using Many Faceted Rasch Measurement. *Psicologica*, 35, 385-397. Saal, F. E., Downey, R. G., & Lahey, M. A.
23. Wortham S.C., (2012). *Assessment in Early Childhood Education*. (6th Ed). United States: Pearson.
24. Yilmaz, F. N (2017). Analysis of the Rater Effect on the Scoring of Diagnostic Trees Prepared by Teacher Candidates with the Many Facet Rasch Model. Faculty of Education Nevşehir Hacı Bektaş Veli University, Nevşehir, Turkey.
25. Yunus, J. M., Rubani, S. N. K., Alias, M., Sumarwati, S., Yunus, F. A. N., Hamid, H. & Rizal, F. Vocational pedagogical decisions of Malaysian and Indonesian non-engineering TVET teachers. *Journal of Technical Education and Training*, 2019, 11 (3).

2016, she received her Ph.D degree from Sultan Idris Education University in evaluation education. Her research area concerns is on assessment and evaluation education, studio based learning and architectural education. She is the author for more than 10 publications and books which is a major focus in assessment and evaluation. She has several research publications in well-known international Journals and conferences.



**Hashima Hamid** is a lecturer in the Department of Educational Affairs at the Faculty of Technical & Education, University Tun Hussein Onn Malaysia. She received the B.Deg in Information Technology (Management) from the Universiti Malaya in 2001 and the M.(Ed) and Ph.D in Vocational Education and Technology Management from University Tun Hussein Onn Malaysia, in 2007 and 2015, respectively. Her area of research interest include continuous quality improvement in TVET, management education research in TVET and quality management. She has more than 20 publications, books and proceedings which focuses on quality improvement in TVET, vocational pedagogy and technical and vocational education. She has several years work experience in teaching, research, students affairs and arranging seminars, workshops and events.



**Nur Afifah Mohd Sofian** was born in Perak, Malaysia in 1992. She is a master's degree student by research in Technical and Vocational Education, Faculty of Education from University Tun Hussein Onn Malaysia (2018-2019). In 2017, she previously has completed on her first degree of B.(Ed) Hons in Vocational programme majoring in Building Construction from University Tun Hussein Onn Malaysia. Her current research study is in assessment and evaluation education which focuses on examiner bias in assessing undergraduate projects using analytical rubrics among University Tun Hussein Onn Malaysia students. Her previous research study during her first degree is also in assessment and evaluation area which is more concentrated on practice formative assessment in the environment of vocational lecturers around Johor state.

## AUTHORS PROFILE



**Siti Wan Aisha Wan A'lauddin** is a master's degree student by research in Technical and Vocational Education, at the Faculty of Technical and Vocational Education from University Tun Hussein Onn Malaysia (2018-2019). She previously received her first degree of B.(Ed). Hons in Vocational majoring in Building Construction from University Tun Hussein Onn Malaysia in 2017. In 2013, she received her diploma in civil engineering field from Polytechnic Kota Bharu. Her current research area of study is on the assessment and evaluation education which focuses on the raters effect in the teaching practicum assessment for teachers candidate.



**Ahmad Rizal Madar** is an associate professor in the Department of Engineering Education at the Faculty of Technical & Vocational, University Tun Hussein Onn Malaysia. He hold Ph.D from University of Technology Malaysia in technical and vocational education, in 2009, master's degree in education and bachelor's degree in electrical engineering, in 2000 and 1999, respectively from University Tun Hussein Onn Malaysia. He has several research publications in well-known international journals and conferences for more than 30 publications and books. His research interest include technical and vocational education training (TVET), engineering education, instructional design and CSF for entrepreneurship program. He is the professional technologist from Malaysia Board of Technologist (MBOT), in 2018.



**Nurfirdawati Muhamad Hanafi** is a senior lecturer in the Department of Engineering Education at the Faculty of Technical & Vocational Education, University Tun Hussein Onn Malaysia and appointed as a group leader of Focus Group (FG), Testing, Evaluation, Assessment and Measurement (TEAM). She finished her bachelor's degree in building technology from University of Technology Mara, in 2001 and M.(Ed) in vocational from University of Tun Hussein Onn Malaysia, in 2003. In a year