The credibility analysis of the microteaching assessment instrument based on the responses of biology education students using the Rasch model

Check for updates

Bowo Sugiharto [1, *], **Murni Ramli** [2]

Biology Education Department, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Indonesia
[1] bowo@fkip.uns.ac.id *, [2] mramlim04@fkip.uns.ac.id
* Corresponding author

ARTICLE INFO	ABSTRACT
	The credibility analysis of the microteaching assessment instrument based on the responses of biology education students using the Rasch model. This study aims to reveal the credibility of the microteaching (MT) assessment instrument by analyzing the responses of biology education students using the Rasch model. The biology education students as participants played roles as assessors and practitioners of MT. The instrument used was the Instrument of Assessing the Learning Implementation (IPPP) in Book 4 of the 2013 In-Service Teacher Training Program (PLPG) Guidelines issued by the Human Resources Development Agency (BPSDM) of the Ministry of Education and Culture. The data obtained were analyzed by multirater analysis of Many-Facet Rasch Measurement (MFRM) version 3.83.2. Based on the results and discussion, this study concluded that the MT assessment instrument in biology education using IPPP was quite credible, which was revealed from the following three points: 1) The responses of students practicing MT are based on the results of peer ratings on a logit scale ranging from -0.82 to 0.80, which means that students practicing were quite diverse and can be distinguished by the instruments; 2) Student responses as assessor to the MT assessment instrument in logit scale units ranged -1.99 to 0.94 which means that as an assessor, student were able to assess diversely. The student as assessor was also able to understand the different rating scales used; and 3) The results of the calibration of the instruments showed that 24 instrument items can provide accurate information about the performance of the practitioner with logit values ranged from a scale of -0.72 to 0.66.

Introduction

Microteaching (MT) or commonly referred to as micro-teaching is an obligatory course for the student in Faculty of Teaching and Education (FKIP) especially the Department of Biology Education at Sebelas Maret University. Generally, MT has

been used in Education since 1960s. MT course teaches techniques in teaching for the teacher candidates to apply teaching skills resulted from careful lesson planning (Allen, 1967). MT provides an important part in preparing a teacher's candidate because of its potential to emphasize the relationship between theory and learning practice (Saban & Çoklar, 2013). MT is the first teaching practice for biology teacher candidates using co-student or peers to become their student to gain feedback (Sezen-Barrie, Tran, McDonald, & Kelly, 2014). MT provides a provision to biology teacher's candidates by practicing their teaching skills, so that, they can gain feedback from their peer or supervisor lecturer. According to curriculum structure, MT course in Department of Biology Education is given in Semester 6, before students were participating in Internship Program III in partner schools.

The curriculum in the Department of Biology Education places MT courses in semester 6 by considering that the students have gained many biology lesson material or pedagogic knowledge in the previous semesters. Courses provide pedagogic knowledge are Education Science, Student Development, Learning Technology, Teaching Basic Skill, Educational Profession, Biology Teaching Strategy, Teaching Evaluation, Study of Curriculum, and Biology Lesson Planning. Meanwhile, courses provide Biology knowledge are Diversity and Classification of Invertebrates, Plants Anatomy, Diversity and Classification of Vertebrates, Diversity and Classification of Vertebrates, Animal Anatomy, Animal Physiology, Plant Anatomy, Plant Physiology, Animal Ecology, Plant Ecology, Biochemistry, Genetics, Biotechnology, Histology, Animal Embryology, and Human Physiology Anatomy.

Even though the students have been provided with pedagogic knowledge or biology lesson material, some lecturers suggest that teaching practice in MT still needs to apply teaching skills in one separate stage (only teaches one concept using one or two teaching skills). On the other hand, some lecturers suggest that MT does not need to apply basic teaching skills in the separated stage, but it should provide the whole teaching skill. The term "micro" in the latter notion refers to the aspect of a few student amount, lesson material amount, limited time allocation,

not on the teaching skills practiced in separated. The difference between both opinions surely will impact the way the practitioner gives their assessment.

Until now, the follower of the first notion assesses MT students usually separated from biology basic teaching skills. Assessment is conducted by the lecturer by involving another student to give input on several skills that are used by practitioners and their performance that are practicing. In this study, MT practice uses the second notion, which is the whole teaching practice. MT implementation is conducted by practicing the whole teaching skills and using peer friends as students as well as assessors based on their perception. MT assessment from peer functions as a form of feedback and followed by a discussion to focus on improving teaching strategy mastery by student's practitioners (Faculty Development and Instructional Design Center, 1993). The discussion which focuses on enhancing mastering teaching strategy is useful to increase the competency of the biology teacher's candidate using some indicators such as biology teacher progress in teaching behavior, planning, learning process, class management, communication, and evaluation (Kilic, 2010).

One of the instruments, that is parallel with the second opinion and can be used to give MT practice assessment, is the Instrument of Assessing the Learning Implementation (IPPP) issued by BPSDM in 2013. Therefore, the study intends to convey an instrument reliability description based on the student's response analysis result on the instrument used. The description response mentioned is in the form of assessment map logarithm odd unit (logit) from assessor aspect (peer friend), practitioner, and instrument including score range scale used. The whole complete description of student response toward instrument is very required to give accuracy certainty and a sense of fairness in MT assessment because every instrument items are already well-calibrated.

MT is a teaching practice activity with many micro/small limitations. What is meant by the limitations include: lesson taught was little, the students were few, the skills practiced were little, and the time was few. Because many aspects are limited, so the student practitioners are demanded

to decide not only what strategy will be applied, but also the time allocated (Pauline, 1993). MT practice is intended to improve the basic skill and knowledge of a teacher's candidate in teaching (Cheng, 2017).

MT learning practice is guided by a lecturer who also gives assessment. Moreover, assessment and input are also carried out by students in the group as a learning process. MT evaluation by involving peer shows weakness. MT evaluation involving peer commonly will only give advantages among the participant because there are factors of subjectivities among participants (Cheng, 2017). Even though MT is not similar to a regular class, but it was quite effective to improve teaching skills (Pauline, 1993). Teaching skill improvement can be created through the MT environment which is by studying and understanding various student's characteristics (Seidel, 2007).

The instrument used in this study was the Instrument of Assessing the Learning Implementation (IPPP). This instrument was regarded as quite comprehensive in assessing teaching skill which related to the pre-learning aspect, skills in involving student, assessing learning process and result, language usage, and closing skill (Badan Pengembangan Sumber Daya Manusia Pendidikan Kebudayaan dan Penjaminan Mutu Pendidikan, 2013). IPPP was explained in detail into 24 items of assessment, each of the items was given a 1 – 5 value scale (attached Instrument).

IPPP has been used nationally in implementing PLPG and also it could be used as an MT assessment instrument outside the activity. Giving the final score for the practitioner was carried out by summed the entire raw score from the existing 24 items. This type of assessment application was also known as the classical scoring theory. The number selected by the assessor in applying classical grading assessment theory actually cannot be summed up as usual mathematical operation does because it was ordinal scale. The way of classical scoring theory has a weakness because it never attained information of interaction among the assessor aspect, the person assessed, and the assessment instrument. Besides that, assessment using classical grading also did not attain instrument calibration.

Some notes from the classical scoring theory which solely use raw score of other instrument are: 1) raw score is not measurement result, but it is merely described frequency from assessor perception, 2) raw score is initial information, while perception amount from every item is ordinal data type which is giving rank symbol, 3) raw score has a weak quantitative meaning which cannot be operated as in mathematics, 4) raw score does not provide clear description toward student's ability in performing the task, 5) raw score and percentage amount answer do not always linear (Sumintono & Widhiarso, 2015).

If assessment carried out in MT only provides information on score/value gained for the practitioner, then the reliability is questioned because perception value symbol from the assessor is ordinal scale (grade/level) which cannot be operated in mathematics. After all, it contains objectivity and subjectivity element mixture. Thus, the raw score could be more significant by analyzing it through the Rasch Model Multifacet Analysis. The Rasch model analysis can be carried out toward three aspects simultaneously, which are 1) practitioner (a student who practices teaching), 2) assessor (other students who observe practitioner), and 3) instrument used in the assessment. The three aspects can generate an equal interval measurement scale because it is stated in the logit unit. Rasch Model Multifacet analysis can measure inter-aspect interaction. Afterward, the model can detect other assessor effects such as range limitation, hallo effect, and internal consistency through fit statistic utilization (Kudiya, Sumintono, Sabana, & Sachari, 2018). Measurement result using the Rasch model will obtain a logit unit (logarithm odd unit). Logit unit will form a logit ruler with the same length (equal interval), so that, it is similar to measurement result as in physics (Sumintono & Widhiarso, 2015).

Rasch model Multifacet analysis is a solution to give a complete description and logical measurement result toward those three MT aspects, which are: assessor, practitioner, and assessment instrument. This analysis can be said as an effort to gain instrument reliability through students' response in Biology Education toward the complete and logical MT assessment instrument.

Instrument credibility is important to be conveyed to give certainty in measurement results on MT practitioner's

competence. Based on the above background, the study is important to be conducted to convey the credibility of the MT assessment instrument in biology education students' response using the Rasch model through the following formulation of problems. 1) How is the actual description (in logit scale/logarithm odd unit) of the practitioner's response toward MT assessment instrument based on peer assessment/peer in MT practice?; b) how is the actual description of student's response as assessor toward MT assessment instrument which is stated in logit scale unit (logarithm odd unit)?; and c) how is the calibration result based on logit scale score of every instrument item which is used in MT assessment?.

Method

The study was conducted in the even semester of academic year 2018/2019; it was undertaken in the Department of Biology Education FKIP UNIS located in D Building FKIP UNS, Jl. Ir. Surami No 36A Surakarta. The research subjects were semester 6 of biology education students who were taking the MT course as much as 9 students (practitioners) using 24 instrument items, so that, it was derived 216 data for analysis.

The study is descriptive research that tries to give the whole description of MT assessment instrument credibility through students' response analysis. Responses which are analyzed include students' responses as assessors toward the instrument, practitioners' response toward the instrument, and calibration of every instrument item in a new analysis way, which is the Rasch Model especially using Multifacet or Multirater.

The research instrument used was the Instrument of Assessing the Learning Implementation (IPPP) (attached). The instrument was considered quite comprehensive in assessing teaching skills which is consisted of 8 aspects, they are; learning, lesson material mastery, utilizing learning/media resource, skill in involving students, assessing learning process and result, language usage, and closing skill (Badan Pengembangan Sumber Daya Manusia Pendidikan Kebudayaan dan Penjaminan Mutu Pendidikan, 2013). IPPP is explained in detail into 24 assessment items, each of the items was given a 1-5 value scale.

The detailed stages and procedures in the research were as follows. 1) Every week, the students participated in the MT course under the condition if there was a student practiced teaching upfront, so the other students took a role as both students and assessor toward practitioner using instrument provided. 2) Every week, there were 3 out of 9 students who practiced teaching with an approximate duration of 30 minutes for every student/practitioner, so that, in one semester every student will undertake 5 times teaching practice; 3) assessment was conducted using an instrument which referred to peer teaching assessment in PLPG Manual Book 2013; 4) the result of the assessment was tabulated according to the rules in Rasch Model Multirater/Multifacet analysis (Linacre, 2018); 5) it was derived logit map of assessor aspects, practitioners, and instrument assessment as well as the interaction among those three aspects; 6) the result of Multifacet was used to measure student's response as assessor, practitioner, and instrument items.

The logit scale map appeared on the logit ruler. It was used to observe every practitioner position, every assessment item, and every assessor. Using the logit ruler, a detailed description of every practitioner can be implied that the bigger logit value means that the practitioner's performance is better and conversely, the smaller logit value means the poorer their performance. Logit ruler also has placed an assessor based on its logit values, the greater logit value means the assessor is more parsimonious and conversely; the smaller its logit value means the assessor is lower (generous). Logit ruler also places every assessment item in its position, the greater its logit value means the easier it will be achieved by the practitioner. Therefore, using a logit ruler in every instrument item is able to give information in detail about the practitioner skills and assessor responses, this is what is called by instrument calibration (Boone & Staver, 2020).

Multifacet used in this study included practitioner, assessor, and assessment instrument aspects (Maryati, Prasetyo, Wilujeng, & Sumintono, 2019). The three aspects can be measured on its consistency, so that, assessment pattern can be explained thoroughly from the three aspects (Andrich & Marais, 2019). Software

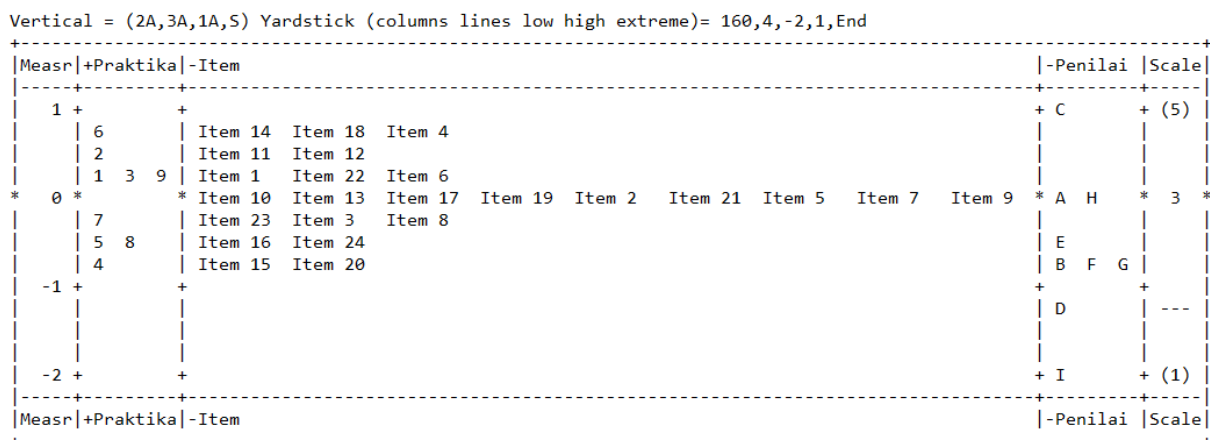used to analyze was Many-Facet Rasch Measurement (MFRM) version 3.83.2.

Multirater analysis with the Rasch model can use some parameters according to the requirement and objective of the study. MT assessment instrument credibility from multitrater analysis was a logistic measurement model based on logarithm probability which was capable on giving information about accuracy, reliability, and validity of the assessors. In this case, the index which will be used as criteria was outfit mean square (MnSq), outfit Z-standard (ZStd), and point measure correlation (PtMea Cor). Tolerance limit for the three index to determine credibility are Outfit MnSq (0.5 – 1.5), Outfit ZStd (-2.0 – 2.0), and PtMea Corr (0.4 – 0.85) (Bond & Fox, 2015; Boone, 2016; Linacre, 2018; Sumintono & Widhiarso, 2015).

Results and Discussion

Result analysis using multifacet on assessor, practitioner, instrument, and scale used in MT can be described in a logit scale map that has an equal interval. A logit scale map can be seen in Figure 1.

Figure 1. Logit scale map in practitioner's item, assessor and microteaching instrument scale

```
Vertical = (2A,3A,1A,S) Yardstick (columns lines low high extreme)= 160,4,-2,1,End
+-------------------------------------------------------------------------------------------------------+
|Measr|+Praktika|-Item                                                             |-Penilai |Scale|
|-----+---------+---------------------------------------------------------------------+---------+-----|
|  1 +         +                                                             + C       + (5) |
|     | 6       | Item 14  Item 18   Item 4                                  |         |     |
|     | 2       | Item 11  Item 12                                           |         |     |
|     | 1  3  9 | Item 1   Item 22   Item 6                                  |         |     |
| *  0 *       * Item 10  Item 13  Item 17  Item 19  Item 2   Item 21  Item 5   Item 7   Item 9 * A  H  * 3 *
|     | 7       | Item 23  Item 3   Item 8                                   |         |     |
|     | 5  8    | Item 16  Item 24                                           | E       |     |
|     | 4       | Item 15  Item 20                                           | B  F  G |     |
| -1 +         +                                                             +         +     |
|     |         |                                                           | D       | --- |
|     |         |                                                           |         |     |
|     |         |                                                           |         |     |
| -2 +         +                                                             + I       + (1) |
|-----+---------+---------------------------------------------------------------------+---------+-----|
|Measr|+Praktika|-Item                                                             |-Penilai |Scale|
+-------------------------------------------------------------------------------------------------------+
```

The Figure I explains that the logit scale ranges from -2 until 1. This logit scale was used as mapping measurement toward the practitioner, assessor, and scoring scale used. Based on Figure 1 and Table 1, practitioner 6 derives the best assessment from peer practitioner (assessor), conversely practitioner 4 derives the poorest assessment from peer practitioner (assessor).

Figure 1 also explains that instrument item used to assess the practitioner who generates the least assessment score are instrument item number 14, 18, and 4. Whereas, the instrument items that generate the highest score are the instrument items number 15 and 20. Meanwhile, it is reviewed from the assessor, Figure 1 explains that the stingiest assessor in giving grade is assessor C. Otherwise, the most generous assessor in giving grade is assessor I.

Practitioner's responses toward MT assessment instrument

Practitioner's responses toward MT assessment instruments were shown by MT practice score gain. MT practitioner's performance total score gain becomes a basis in determining logit value which will be placed in a logit ruler. Besides performance total score gain, determining the logit score also can be seen based on the assessment probability of every instrument's item. Therefore, it does not always that a practitioner who gains a high score will get a higher score as well.

Score gain pattern tendency was assessed as a practitioner's probability in obtaining a score that will be taken into account in determining the logit scale. Data summary on score gain, logit value, statistic, and point measure correlation is presented in Table 1.

Based on Table 1, it is derived information that the practitioner who gains the highest practice score is practitioner 6 with 682 total score that places a logit scale on 0.8. While a practitioner who obtains the lowest practice, value is practitioner 4 with a total score of 550 and places a logit scale -0.82. Therefore, MT practitioner's responses toward MT assessment instrument were on logit ruler between -0.82 until 0.8 which was illustrated by

practitioner 4 until practitioner 6 in Figure 1. This indicates that there were assessment distributions in MT practitioner's student performance. In other words, the assessor can distinct participants (practitioners) in giving an assessment.

Table 1. Summary of data on score acquisition, logit value, statistical fit, and point measure correlation

Total Score	Total Count	Obsvd Average	Fair (M) Average	+ Measure	Model S.E	Infit		Outfit		Estim. Discrm	Correlation		N	Practical
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp		
682	192	3.55	3.57	.80	.12	1.28	2.5	1.27	2.5	.71	.43	.50	6	6
657	190	3.46	3.48	.56	.12	1.06	.6	1.05	.5	.93	.38	.50	2	2
619	189	3.28	3.35	.24	.11	.84	-1.7	.83	-1.7	1.18	.46	.44	9	9
642	191	3.36	3.34	.21	.11	1.08	.8	1.09	.8	.92	.56	.50	1	1
651	192	3.39	3.32	.18	.11	.88	-1.1	.89	-1.1	1.13	.38	.42	3	3
602	190	3.17	3.19	-.15	.11	.90	-1.0	.90	-1.0	1.11	.55	.51	7	7
590	190	3.11	3.08	-.40	.11	1.03	.3	1.03	.3	.97	.56	.51	8	8
569	190	2.99	2.99	-.65	.11	.87	-1.3	.87	-1.3	1.14	.61	.51	5	5
550	192	2.86	2.91	-.82	.11	1.03	.3	1.03	.3	.96	.40	.49	4	4
618.0	190.7	3.24	3.25	.00	.11	1.00	-.1	1.00	-.1		.48			Mean (Count:9)
41.3	1.1	.21	.21	.51	.00	.13	1.3	.13	1.3		.08			S.D. (Population)
43.8	1.1	.23	.21	.54	.00	.14	1.4	.14	1.4		.09			S.D. (Sample)

Model, Populn: RMSE .11 Adj (True) S.D. .50 Separation 4.34 Strata 6.12 Reliability .95
Model, Sample: RMSE .11 Adj (True) S.D. .53 Separation 4.61 Strata 6.49 Reliability .96
Model, Fixed (all same) chi-square: 178.3 d.f.: 8 significance (probability): .00
Model, Random (normal) chi-square: 7.7 d.f.: 7 significance (probability): .36

A more obvious description of the practitioner's response toward MT assessment instrument based on its response precision was exposed through its statistics fit score and combined by outfit statistics and point measure correlation. Practitioners' performances in MT were responded varied by the assessor through 24 instrument assessment items.

Generally, every practitioner will be assessed by 8 assessors, so that, it will collect a score of 24 x 8 = 192 times. This value of 192 times will occur if every assessor gives their complete assessment. Based on infit and outfit value both in MnSq or ZStd as well as point measure correlation value, practitioners' position is described in Table 2.

Table 2. Infit value and statistical outfit value and point measure correlation value for practical position

Total Score	Total Count	Obsvd Average	Fair (M) Average	+ Measure	Mode l S.E	Infit		Outfit		Estim. Discrm	Correlation		N	Practical
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp		
619	189	3.28	3.35	.24	.11	.84	-1.7	.83	-1.7	1.18	.46	.44	9	9
569	190	2.99	2.99	-.62	.11	.87	-1.3	.87	-1.3	1.14	.61	.51	5	5
651	192	3.39	3.32	.18	.11	.84	-1.1	.89	-1.1	1.13	.38	.42	3	3
602	190	3.17	3.19	-.15	.11	.90	-1.0	.90	-1.0	1.11	.55	.51	7	7
590	190	3.11	3.08	-.40	.11	1.03	.3	1.03	.3	.97	.56	.51	8	8
550	192	2.86	2.91	-.82	.11	1.03	.3	1.03	.3	.96	.40	.49	4	4
657	190	3.46	3.48	.56	.12	1.06	.6	1.05	.5	.93	.38	.50	2	2
642	191	3.36	3.34	.21	.11	1.08	.8	1.09	.8	.92	.56	.50	1	1
682	192	3.55	3.57	.80	.12	1.28	2.5	1.27	2.5	.71	.43	.50	6	6
618.0	190.7	3.24	3.25	.00	.11	1.00	-.1	1.00	-.1		.48			Mean (Count:9)
41.3	1.1	.21	.21	.51	.00	.13	1.3	.13	1.3		.08			S.D. (Population)
43.8	1.1	.23	.21	.54	.00	.14	1.4	.14	1.4		.09			S.D. (Sample)

Model, Populn: RMSE .11 Adj (True) S.D. .50 Separation 4.34 Strata 6.12 Reliability .95
Model, Sample: RMSE .11 Adj (True) S.D. .53 Separation 4.61 Strata 6.49 Reliability .96
Model, Fixed (all same) chi-square: 178.3 d.f.: 8 significance (probability): .00
Model, Random (normal) chi-square: 7.7 d.f.: 7 significance (probability): .36

Table 2 shows that the lowest statistics fit value is practitioner 9 with infit MnSq value = 0.89, while the highest statistics fit value is practitioner 6 with MnSq Infit value of 1.28, but it is observed using three parameters which are: outfit MnSq, outfit ZStd and point measure correlation of both practitioners are still reasonable. Thus, the Infit MnSq score ranges between 0.84 until 1.28. The

conformity of MT practitioners' responses based on the score obtained can be seen from the statistics fit score. Limitation used is amount of mean + SD = 1.00 + 0.13 = 1.13. Based on these criteria, there is response (assessment result) toward practitioner which is not inappropriate, specifically practitioner number 6. It gains Infit MnSq score as much as 1.28 > 1.13. This was seen clearly if it was observed on the total score attained, which is 682 (much higher from the other practitioners' score gain). However, this must be confirmed by seeing MnSq outfit, outfit ZStd, and point measure correlation value.

Afterward, the examination of practitioners' precision responses can be carried out from a statistic outfit score. Through this examination practitioner, number 6 has an Outfit MnSq value of 1.27 (within acceptable ranges). However, when it was seen from outfit ZStd, practitioner number 6 has Outfit ZStd value = +2.5 higher than ZStd acceptance limit which is +2.0. Based on the analysis, there was one practitioner who derived a mistaken assessment that is practitioner number 6. This condition could be occurred because the practitioner's performance number 6 was confusing the assessor in giving proper assessment on the performance.

The practitioners gave varied responses to the MT assessment instrument used. This appears in the logit map of practitioners which ranged from -0.82 until 0.80. Practitioners' logit scores were obtained based on the performance and scored based on instrument items used. In this case, it meant that the best practitioner in obtaining score was practitioner number 6 with a 682 total score. While a practitioner who gained the lowest score is practitioner number 4 with a 550 total score.

Even though there is a tendency that practitioner number 6 was given the highest score and practitioner number 4 obtained the lowest assessment by the assessor, but it was still on tolerance limitation. This means that measurement toward practitioners using the instrument used was still reasonable. This normality indicates that practitioners' responses toward the instrument became one of the proofs that the instrument has shown its credibility in measuring, if it was seen from the practitioners' point of view.

Based on data in Table 2, the practitioners' grade was supported by the fact that practitioner number 6 was the most diligent practitioners during MT practice and always respond task assignment seriously. Conversely, practitioner number 4 was the least took seriously on responding the assignment given. The attitude in implementing MT mostly looked carelessly and ignored to complete the assignment. This condition was also affirmed by the peer in the group, based on assessment result data that has been conducted.

Practitioners' performances, assessed based on instrument items, were seen its pattern precision based on statistics fit score particularly using Outfit MnSq, Outfit ZStd, and point measure correlation score limitation (Boone, 2016; Boone, Staver, & Yale, 2014; Engelhard & Wind, 2017). Based on those criteria, there was only one practitioner who was confusing the assessor in giving a score, which was practitioner number 6, who has ZStd = 2.5 value.

Practitioner number 6 made assessor confused in giving their score. Further analysis can be used to clarify this phenomenon. Practitioner number 6 was a coordinator or responsible in the MT group. As a person in charge of the group, practitioner number 6 was responsible for giving all the information in MT practice instruction from the lecturer to the students and became liaison in communication with the MT supervisor lecturer. This role demanded practitioner number 6 became more diligent and responsive.

The role was quite strategic and make other assessors who were also co-practitioner gave their appreciation for practitioner number 6. The impact was the practitioner number 6 gain the highest total score compared to the other practitioners' score gain. Generally, it can be concluded that MT practitioner assessment by peer practitioner was quite objective, it's just that the role as coordinator has influenced psychological condition among the participants, so that, practitioner number 6 was given the highest score.

Assessor responses toward MT assessment instrument

Assessor responses in assessing practitioners were actualized by giving scores in every instrument items. The summary of responses based on the logit score is presented in Table 3.

Based on Table 3, it was obtained information that the stingiest assessor in

the grading practitioner was assessor C with logit value (measure) as much as 0.94. In this case, assessor C gave grade as much as 188 times with the total score given was 492. Conversely, the most generous assessor in grading was assessor I with a logit score of -1.99. In this context, the assessor I gave grade as much as 191 times with the total score given was as much as 726. Grading pattern was a probability that can be calculated, so that, it did not solely

determine based on total score given. For example, assessor B who assessed with total score 629 has higher logit score (parsimonious) compared to F who gave a total score of 622. This was understandable if it was viewed from the number of grading, which is assessor B gave the grade as much as 192 times while assessor F gave the grade as much as 190 times.

Table 3. Summary of rater (Assessor) responses based on logit values

Total Score	Total Count	Obsvd Average	Fair (M) Average	- Measure	Model S.E	Infit		Outfit		Estim. Discrm	Correlation		Exact Obs %	Agree Obs %	N	Assessor
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp				
492	188	2.62	2.63	.94	.11	.89	-1.0	.89	-1.1	1.12	.40	.39	35.4	32.1	3	C
569	192	2.96	2.98	.12	.11	1.62	5.3	1.62	5.3	.29	.51	.39	36.2	39.0	1	A
585	192	3.05	3.03	-.01	.11	.89	-1.1	.89	-1.1	1.13	.38	.38	42.4	39.6	8	H
625	192	3.26	3.23	-.48	.11	.95	-.5	.95	-.4	1.06	.14	.37	41.4	40.9	5	E
629	192	3.28	3.31	-.68	.11	1.17	1.6	1.16	1.6	.86	.47	.37	38.9	40.9	2	B
622	190	2.27	2.32	-.71	.11	.78	-2.3	.78	-2.3	1.24	.42	.35	43.6	40.7	6	F
638	191	3.34	3.34	-.75	.11	.74	-2.8	.74	-2.8	1.28	.30	.39	43.8	40.9	7	G
676	188	3.60	3.56	-1.31	.12	.63	-4.2	.62	-4.2	1.39	.27	.34	41.9	39.1	4	D
726	191	3.80	3.82	-1.99	.12	1.28	2.5	1.30	2.6	.69	.39	.37	32.4	34.0	9	I
618.0	190.7	3.24	3.24	-.54	.11	.99	-.3	.99	-.3		.36		Mean (Count:9)			
62.3	1.6	.33	.32	.80	.00	.29	2.8	.29	2.8		.10		S.D. (Population)			
66.1	1.7	.35	.34	.84	.00	.31	3.0	.31	3.0		.11		S.D. (Sample)			

Model, Populn: RMSE .11 Adj (True) S.D. .79 Separation 6.90 Strata 9.53 Reliability (not inter-rater) .98
Model, Sample: RMSE .11 Adj (True) S.D. .84 Separation 7.32 Strata 10.10 Reliability (not inter-rater) .98
Model, Fixed (all same) chi-square: 424.7 d.f.: 8 significance (probability): .00
Model, Random (normal) chi-square: 7.9 d.f.: 7 significance (probability): .35
Inter-Rater agreement opportunities: 5964 Exact agreements: 2359 = 39.6% Expected: 2301.9 = 38.6%

A provision in giving the grade toward practitioners was also reported as seen in Table 4. Information derived from Table 4 is that the lowest statistics fit score was assessor D with MnSq Infit score = 0.63 and the highest was assessor A with MnSq infit score = 1.62. The assessor precision range

in assessing MT was between 0.63 until 1.62. Assessor responses conformity based on the score obtained can be observed from statistics fit score and discrepancy grading pattern can be seen from its outfit value.

Table 4. Data on the accuracy of how to provide an assessment of the practitioner

Total Score	Total Count	Obsvd Average	Fair (M) Average	- Measure	Model S.E	Infit		Outfit		Estim. Discrm	Correlation		Exact Obs %	Agree Obs %	N	Rater (Assessor)
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp				
676	188	3.60	3.56	-1.31	.12	.63	-4.2	.62	-4.2	1.39	.27	.34	41.9	39.1	4	D
638	191	3.34	3.34	-.75	.11	.74	-2.8	.74	-2.8	1.28	.30	.39	43.8	40.9	7	G
622	190	3.27	3.34	-.71	.11	.78	-2.3	.78	-2.4	1.24	.42	.35	43.6	40.7	6	F
585	192	3.05	3.32	-.01	.11	.89	-1.1	.89	-1.1	1.13	.38	.38	42.4	39.6	8	H
492	188	2.62	3.03	.94	.11	.89	-1.0	.89	-1.1	1.12	.40	.39	35.4	32.1	3	C
625	192	3.26	2.63	-.48	.11	.95	-.5	.95	-.4	1.06	.14	.37	41.4	40.9	5	E
629	192	3.28	3.23	-.68	.11	1.17	1.6	1.16	1.6	.86	.47	.37	38.9	40.9	2	B
726	191	3.80	3.82	-1.99	.12	1.28	2.5	1.30	2.6	.69	.39	.37	32.4	34.0	9	I
569	192	2.98	2.98	.12	.11	1.62	5.3	1.62	5.3	.29	.51	.39	36.2	39.0	1	A
618.0	190.7	3.24	3.24	-.54	.11	.99	-.3	.99	-.3		.36		Mean (Count:9)			
62.3	1.6	.33	.32	.80	.00	.29	2.8	.29	2.8		.10		S.D. (Population)			
66.1	1.7	.35	.34	.84	.00	.31	3.0	.31	3.0		.11		S.D. (Sample)			

Model, Populn: RMSE .11 Adj (True) S.D. .79 Separation 6.90 Strata 9.53 Reliability (not inter-rater) .98
Model, Sample: RMSE .11 Adj (True) S.D. .84 Separation 7.32 Strata 10.10 Reliability (not inter-rater) .98
Model, Fixed (all same) chi-square: 424.7 d.f.: 8 significance (probability): .00
Model, Random (normal) chi-square: 7.9 d.f.: 7 significance (probability): .35
Inter-Rater agreement opportunities: 5964 Exact agreements: 2359 = 39.6% Expected: 2301.9 = 38.6%

Further examination can be carried out from statistics outfit score. The accepted outfit mean square (MnSq) score was 0.5 < MnSq < 1.5. The accepted outfit z-standard (ZSTD) score was -2.0 < ZSTD < +2.0. Point measure correlation score was 0.4 < Pt Measure Corr < 0.85. Based on other examination parameter, assessor A had Outfit MnSq score = 1.62 (indicated gave confusing assessment pattern), meanwhile based on Outfit ZStd examination, there were 5 assessors who were outside tolerance limit, they were: assessor A, D, F, G, and I, each of their score were 5.3; -4.2; -2.4; -2.8; and 2.6. Based on point measure correlation examination, there were three assessors whose pattern were not consistent, they were: assessor D, E, G, H, and I, each of them had score of 0.27; 0.14; 0.30; 0.38; and 0.39 (less than 0.4).

Based on the three discrepancy measurement parameters, the assessors in grading all the assessments were still reasonable because none of them met the criteria. Therefore, generally, it can be concluded that the entire practitioner gave proper assessment according to every practitioner who was graded. The assessor gave assessment objectively and orderly according to its pattern in grading.

Assessors' responses in grading the practitioners in MT practice were quite varied. However, grading diversity was within tolerance limitation because it showed an orderly pattern that was still tolerated. Tolerance of fairness was given because it did not exceed three parameters simultaneously, they were: outfit MnSq, Outfit ZStd, and point measure correlation. At most, there were only exceeding the threshold for the two parameters simultaneously.

Analysis toward assessment pattern given, some assessors were indicated to almost make irregular grading. The assessors were A, D, G, and I. Assessor A was exceeding two parameters, which were: Outfit MnSq (1.62) and Outfit ZStd (5.3). Assessor A was indicated giving parsimonious assessment (grading with low score) to practitioners.

Assessor D, G, and I exceed two assessment thresholds, which were Outfit ZStd and point measure correlation simultaneously. The three assessors were indicated to give easy grading (generously gave a high score) to the practitioner or other participants in MT. Grading pattern irregularity has possibly occurred because there was a close relationship between participants or it could be meant they gave grading without considering objectivity.

Assessor responses toward instrument also can be seen in the comprehension on every meaning scale used. This instrument used 5 scales, which were 1 (very poor), 2 (poor), 3 (enough), 4 (good), and 5 (very good). Based on the analysis result using the Rasch model, it was found that the assessor can distinct the meaning among the 5 scales. This can be proved in Figure 2.
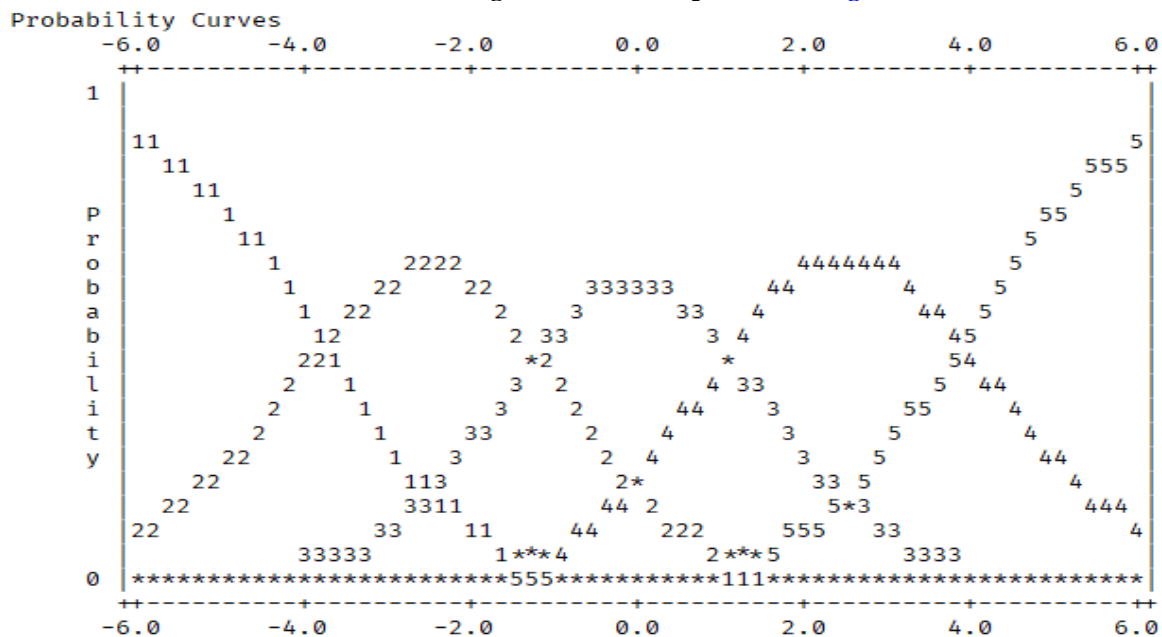

Figure 2. Probability curve scale used by the assessment instrument

Figure 2 shows that the peak of every scale is separated from each other; there is no graphic peak that is coincided. This indicates that the assessors can distinctly meaning of scale 1, 2, 3, 4, and 5. The assessor is aware and able to use the scale as a grading scale (Kudiya et al., 2018). Graphic on Figure 2 gives visual information on the grading scale that can be used to distinguish practitioners according to its performance in MT practice (Van Zile-Tamsen, 2017). The assessor skill in giving grading is also meant as feedback for the practitioner that will be useful to improve its performance in MT practice. Therefore, MT can be function as practitioners' guidance in professional development for teacher candidates (Pekdağ, Dolu, Ürek, & Azizoğlu, 2020).

Analysis of MT assessment instrument

MT assessment instrument consists of 24 items. Assessor gives MT practitioners' performance grading through the 24 items. The position of every instrument in the logit scale ruler was described in Table 5.

Table 5. Position each instrument within the logit scale ruler

Total Score	Total Count	Obsvd Average	Fair (M) Average	- Measure	Model S.E	Infit		Outfit		Estim. Discrm	Correlation		Nu	Item
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp		
214	72	2.97	2.97	.66	.18	.94	-.3	.94	-.3	1.08	.56	.53	4	Item 4
212	71	2.99	2.98	.66	.18	1.05	.3	1.05	.3	.93	.60	.51	14	Item 14
215	72	2.99	2.99	.63	.18	.85	-.9	.85	-.9	1.16	.53	.53	18	Item 18
219	72	3.04	3.05	.49	.18	.93	-.3	.94	-.3	1.06	.54	.53	11	Item 11
219	71	3.08	3.09	.38	.18	1.47	2.5	1.46	2.5	.47	.42	.53	12	Item 12
227	72	3.15	3.16	.22	.18	1.00	0	1.00	0	.50	.50	.52	22	Item 22
229	72	3.18	3.19	.15	.19	.78	-1.3	.78	-1.4	1.24	.58	.52	1	Item 1
226	71	3.18	3.19	.14	.19	1.07	.4	1.07	.4	.91	.49	.53	6	Item 6
228	71	3.21	3.22	.07	.19	.97	-.1	.97	-.1	1.03	.44	.52	19	Item 19
229	71	3.23	3.24	.02	.19	.83	-1.0	.83	-1.0	1.19	.62	.52	9	Item 9
233	72	3.24	3.24	.02	.19	.81	-1.2	.80	-1.2	1.21	.57	.52	2	Item 2
230	71	3.24	3.25	.01	.19	.79	-1.3	.80	-1.2	1.22	.66	.53	5	Item 5
234	72	3.25	3.26	-.02	.19	1.25	1.4	1.24	1.4	.74	.49	.52	7	Item 7
231	71	3.25	3.27	-.04	.19	1.14	.8	1.13	.8	.87	.74	.52	10	Item 10
235	72	3.26	3.27	-.05	.19	1.30	1.7	1.29	1.7	.68	.36	.52	13	Item 13
235	72	3.26	3.27	-.05	.19	1.09	.5	1.09	.6	.90	.53	.52	17	Item 17
236	72	3.28	3.29	-.09	.19	.69	-2.0	.71	-1.9	1.31	.50	.52	21	Item 21
235	71	3.31	3.32	-.27	.19	.63	-2.5	.63	-2.5	1.38	.50	.52	8	Item 8
239	71	3.37	3.37	-.29	.19	1.18	1.0	1.16	.9	.84	.35	.52	23	Item 23
242	72	3.36	3.37	-.29	.19	1.04	.3	1.05	.3	.95	.52	.52	3	Item 3
246	71	3.46	3.47	-.54	.19	.98	.0	1.00	.0	1.01	.60	.52	24	Item 24
249	72	3.46	3.47	-.54	.19	.77	-1.4	.78	-1.4	1.21	.53	.52	16	Item 16
248	71	3.49	3.51	-.64	.19	.95	-.2	.95	-.2	.105	.44	.52	15	Item 15
251	71	3.54	3.54	-.72	.19	1.39	2.1	1.38	2.1	.62	.40	.51	20	Item 20
231.7	71.5	3.24	3.25	.00	.19	1.00	-.1	1.00	-.1		.52			Mean (Count:24)
10.7	.5	.15	.16	.38	.00	.21	1.3	.21	1.3		.09			S.D. (Population)
10.9	.5	.16	.16	.39	.00	.22	1.3	.21	1.3		.09			S.D. (Sample)

Model, Populn: RMSE .19 Adj (True) S.D. .33 Separation 1.77 Strata 2.70 Reliability .76
Model, Sample: RMSE .19 Adj (True) S.D. .34 Separation 1.82 Strata 2.77 Reliability .77
Model, Fixed (all same) chi-square: 99.0 d.f.: 23 significance (probability): .00
Model, Random (normal) chi-square: 18.8 d.f.: 22 significance (probability): .66

Table 5 informs that the most difficult MT assessment instrument to be fulfilled is item 4. It is a practitioner assessment on skill in correlating lesson material with other relevant knowledge. Out of 72 times grading, item 4 only collects 214 total scores with a 0.66 logit score. Meanwhile, the easiest instrument item to be achieved by practitioners is item 20. It is conducting a final assessment according to competency (objective). Out of 71 times in grading, item 20 collects 251 total scores with MT assessment on scale 0.72 until 0.66. This means that every assessment instrument item is quite understandable by the assessor and it gained varied responses.

The entire instrument item used in MT practice assessment can be analyzed its precision in measuring practitioner skills. The precision of every MT assessment instrument item is presented in Table 6.

Based on information from Table 6, the lowest instrument item precision in measuring practitioner is item number 8 with Infit MnSq value of 0.63. While the highest instrument item precision in measuring practitioners' skill is item number 12 with MnSq Infit value of 1.47. Therefore, the whole instrument item has Infit MnSq range value from 0.63 until 1.47, which later will be confirmed through examination based on statistics outfit.

Table 6. Precision in every MT assessment instrument item

Total Score	Total Count	Obsvd Average	Fair (M) Average	-Measure	Model S.E	Infit		Outfit		Estim. Discrm	Correlation		Nu	Item
						MnSq	ZStd	MnSq	ZStd		PtMea	PtExp		
235	71	3.31	3.32	-.17	.19	.63	-2.5	.63	-2.5	1.38	.50	.52	8	Item 8
236	72	3.28	3.29	-.09	.19	.69	-2.0	.71	-1.9	1.31	.50	.52	21	Item 21
249	72	3.46	3.47	-.54	.19	.77	-1.4	.78	-1.4	1.21	.53	.52	16	Item 16
229	72	3.18	3.19	.15	.18	.78	-1.3	.78	-1.4	1.24	.58	.52	1	Item 1
230	71	3.24	3.25	.01	.19	.79	-1.3	.80	-1.2	1.22	.66	.53	5	Item 5
233	72	3.24	3.24	.02	.18	.81	-1.2	.80	-1.2	1.21	.57	.52	2	Item 2
229	71	3.23	3.24	.02	.18	.83	-1.0	.83	-1.0	1.19	.62	.52	9	Item 9
215	72	2.99	2.99	.63	.18	.85	-.9	.85	-.9	1.16	.53	.53	18	Item 18
219	72	3.04	3.05	.49	.19	.93	-.3	.94	-.3	1.06	.54	.53	11	Item 11
214	72	2.97	2.97	.66	.19	.94	-.3	.94	-.3	1.08	.56	.53	4	Item 4
248	71	3.49	3.51	-.64	.19	.95	-.2	.95	-.2	1.05	.44	.52	15	Item 15
228	71	3.21	3.22	.07	.18	.97	-.1	.97	-.1	1.03	.44	.52	19	Item 19
246	71	3.46	3.47	-.54	.19	1.00	0	1.00	0	1.01	.60	.52	24	Item 24
227	72	3.15	3.16	.22	.18	1.04	0	1.00	0	1.00	.50	.52	22	Item 22
242	72	3.36	3.37	-.29	.19	1.05	.3	1.05	.3	.95	.52	.52	3	Item 3
212	71	2.99	2.98	.66	.19	1.07	.3	1.05	.3	.93	.60	.51	14	Item 14
226	71	3.18	3.19	.14	.19	1.09	.4	1.07	.4	.91	.49	.53	6	Item 6
235	72	3.26	3.27	-.05	.19	1.05	.5	1.09	.6	.90	.53	.52	17	Item 17
231	71	3.25	3.27	-.04	.19	1.14	.8	1.13	.8	.87	.74	.52	10	Item 10
239	71	3.37	3.37	-.29	.19	1.18	1.0	1.16	.9	.84	.35	.52	23	Item 23
234	72	3.25	3.26	-.02	.19	1.25	1.4	1.24	1.4	.74	.49	.52	7	Item 7
235	72	3.26	3.27	-.05	.19	1.30	1.7	1.29	1.7	.68	.36	.52	13	Item 13
251	71	3.54	3.54	-.72	.19	1.39	2.1	1.38	2.1	.62	.40	.51	20	Item 20
219	71	3.08	3.09	.38	.19	1.47	2.5	1.46	2.5	.47	.42	.53	12	Item 12
231.7	71.5	3.24	3.25	.00	.19	1.00	-.1	1.00	-.1		.52			Mean (Count:24)
10.7	.5	.15	.16	.38	.00	.21	1.3	.21	1.3		.09			S.D. (Population)
10.9	.5	.16	.16	.39	.00	.22	1.3	.22	1.3		.09			S.D. (Sample)

Model, Populn: RMSE .19 Adj (True) S.D. .33 Separation 1.77 Strata 2.70 Reliability .76
Model, Sample: RMSE .19 Adj (True) S.D. .34 Separation 1.82 Strata 2.77 Reliability .77
Model, Fixed (all same) chi-square: 99.0 d.f.: 23 significance (probability): .00
Model, Random (normal) chi-square: 18.8 d.f.: 22 significance (probability): .66

Further examination can be conducted starting from statistics outfit value. The accepted outfit mean square (MnSq) value was 0.5 < MnSq < 1.5. The accepted outfit z-standard (ZSTD) value was -2.0 < ZSTD < +2.0. Point measure correlation value was 0.4 < Pt Measure Corr < 0.85. Using this examination, all instrument items were still within the range of Outfit MnSq or point measure correlation acceptance limitation. However, if it was seen from outfit ZStd, two instrument items are immoderate which was instrument item number 8 with Outfit ZStd value of 2.5 and item number 12 with Outfit ZStd value of 2.5.

Therefore, all MT assessment instrument items used were not confusing to the assessor in giving practitioner grade. All of the 24 instrument items gave input in assessing practitioner objectively and transparent to fulfill the existing criteria. Thus, using this analysis is derived psychometric description through measurement that connects measuring people and item calibration (Engelhard et al., 2018).

Generally, all of the instrument items were feasible because none of the items that were exceeded three threshold discrepancy parameters. The instrument used in MT assessment was an instrument that was used in the Teacher Profession Training and education activities issued by the Human Resources Development Agency for Cultural Education and Education Quality Assurance in 2013. This instrument was known as the Instrument for Assessing the Learning Implementation (IPPP). This instrument was applied nationally in all regions holding PLPG.

This instrument was considered quite representative in assessing teaching skills which related to aspects in pre-learning, learning approach/strategy, utilizing learning/media resource, skill in involving students, assessing learning process and result, language usage, and closing skill. There for MT assessment instrument that utilizes IPPP is a reliable instrument and able to give accurate information on MT practitioners' performance.

Conclusion

Based on result and discussion, this research concludes that MT assessment instrument in biology education using IPPP is quite reliable that is explained from these following three explanations: 1) MT practitioners' responses based on assessment result from the peer in logit scale ranges between -0.82 until 0.80, which means practitioners students are quite varied and can be differentiated by instrument used; 2) student response as an assessor in grading MT instrument on logit scale unit were on the range -1.99 until 0.94 which means that as assessors, the students gave varied grading based on the instrument used. 3) instrument item

calibration result used in MT assessment shows that the 24 instrument items can give precise information in describing practitioners' performance with logit scale ranges from scale -0.72 until 0.66.

References

Allen, D. W. (1967). *Micro–teaching, a description*. Retrieved from https://eric.ed.gov/?id=ED019224

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory*. https://doi.org/10.1007/978-981-13-7496-8

Badan Pengembangan Sumber Daya Manusia Pendidikan Kebudayaan dan Penjaminan Mutu Pendidikan. (2013). *Buku 4 rambu–rambu pelaksanaan pendidikan dan latihan profesi guru profesi guru (PLPG)*. Jakarta: Kementerian Pendidikan Nasional.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (Third). https://doi.org/10.4324/9781315814698

Boone, W. J. (2016). Rasch analysis for instrument development: Why, When, and How? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., & Staver, J. R. (2020). *Advances in rasch analyses in the human sciences*. https://doi.org/10.1007/978-3-030-43420-5

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. https://doi.org/10.1007/978-94-007-6857-4

Cheng, J. (2017). Learning to attend to precision: the impact of micro-teaching guided by expert secondary mathematics teachers on pre-service teachers' teaching practice. *ZDM*, *49*(2), 279–289. https://doi.org/10.1007/s11858-017-0839-7

Engelhard, G., & Wind, S. A. (2017). Invariant measurement with raters and rating scales. In *Invariant Measurement with Raters and Rating Scales* (pp. 327–336). https://doi.org/10.4324/9781315766829-14

Faculty Development and Instructional Design Center. (1993). *What is microteaching?* Northern Illinois University.

Kilic, A. (2010). Learner-centered micro teaching in teacher education. *International Journal of Instruction2*, *3*(1), 77–100. Retrieved from http://www.e-iji.net/dosyalar/iji_2010_1_5.pdf

Kudiya, K., Sumintono, B., Sabana, S., & Sachari, A. (2018). Batik artisans' judgment of batik wax quality and its criteria: An application of the many-facets rasch model. In *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings* (pp. 27–37). https://doi.org/10.1007/978-981-10-8138-5_3

Linacre, J. M. (2018). A user's guide to FACETS Rasch-model computer programs. Retrieved from Tutorial program website: https://www.winsteps.com/index.htm

Maryati, M., Prasetyo, Z. K., Wilujeng, I., & Sumintono, B. (2019). Measuring teachers' pedagogical content knowledge using many-facet Rasch model. *Jurnal Cakrawala Pendidikan*, *38*(3), 452–464. https://doi.org/10.21831/cp.v38i3.26598

Pauline, R. F. (1993). Microteaching: An integral part of a science methods class. *Journal of Science Teacher Education*, *4*(1), 9–17. https://doi.org/10.1007/BF02628852

Pekdağ, B., Dolu, G., Ürek, H., & Azizoğlu, N. (2020). Exploring on-campus and in real school classroom microteaching practices: the effect on the professional development of preservice teachers. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-020-10109-2

Saban, A., & Çoklar, A. N. (2013). Pre-service teachers' opinions about the micro-teaching method in teaching practise classes. *Turkish Online Journal of Educational Technology*, *12*(2), 234–240. Retrieved from http://www.tojet.net/articles/v12i2/12221.pdf

Seidel, T. (2007). The role of student characteristics in studying micro

teaching–learning environments. *Learning Environments Research, 9*(3), 253–271. https://doi.org/10.1007/s10984-006-9012-x

Sezen-Barrie, A., Tran, M.-D., McDonald, S. P., & Kelly, G. J. (2014). A cultural historical activity theory perspective to understand preservice science teachers' reflections on and tensions during a microteaching experience. *Cultural Studies of Science Education, 9*(3), 675–697.

https://doi.org/10.1007/s11422-013-9503-x

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan RASCH pada assessment pendidikan.* Cimahi: Trim Komunikata.

Van Zile-Tamsen, C. (2017). Using rasch analysis to inform rating scale development. *Research in Higher Education, 58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

supplementary material

MICROTEACHING ASSESSMENT INSTRUMENT/PEER TEACHING

Practitioners' Name : _____
Gender : _____

No	Aspect and Sub-aspect	Score				
		1	2	3	4	5
I	**PRE-LEARNING**					
1.	Preparing the student to study					
2.	Performing apperception activities					
II	**LEARNING STUDYING ACTIVITIES**					
A	Mastering lesson materials					
3.	Showing lesson material mastery					
4.	Relating material with other relevant knowledge					
5.	Delivering material clearly, according to studying hierarchy and student's characteristic					
6.	Relating material with real-life reality					
B	Teaching Strategy/approach					
7.	Performing teaching according to students' characteristic and competency (objective) that will be achieved					
8.	Performing teaching in sequence					
9.	Mastering the class					
10.	Performing contextual learning					
11.	Performing teaching that will likely develop a positive habit					
12.	Performing teaching according to allocation time planned					
C.	Utilizing learning media/resource					
13.	Utilizing media effectively and efficiently					
14.	Generating interesting message					
15.	Involving student in utilizing media					
D.	Learning that triggers and maintains student's involvement					
16.	Building student's active participation in learning					
17.	Showing openness toward the student's response					
18.	Developing student's enjoyment and enthusiasm in learning					
E.	Assessing Studying process and result					
19.	Monitoring student's learning progress during learning					
20.	Performing final assessment according to competency (objective)					
F.	Language Usage					
21.	Using clear, good, and correct written and spoken language					
22.	Delivering the message in appropriate way					
III.	**CLOSING**					
23.	Performing reflection or making a summary by involving student					
24.	Conducting follow-up by giving direction, or activity, or assignment as part of remedies / enrichment					

(Badan Pengembangan Sumber Daya Manusia Pendidikan Kebudayaan dan Penjaminan Mutu Pendidikan [BPSDMPK-PMP], 2013)