MEASURING THE QUALITY OF MEAL TRAINING FORMATIVE TEST USING THE RASCH MODEL

Asep Sapa'at

Dompet Dhuafa (INDONESIA)

asepsapaat@dompetdhuafa.org

Abstract

Assessment for learning is an essential topic in the educational assessment context. To ensure quality learning, valid and reliable assessments should be developed that show the learning progress and needs of students as well as the effectiveness of teaching. The basic Rasch model (Ming Ho et. al, 2013) for dichotomous items (item responses are either right or wrong) is a probabilistic model that describes the probability of getting an item correct in terms of a simple logistic function of the difference between the person's ability and the item difficulty: the higher the ability of the person compared with the item difficulty, the higher the probability of getting the item correct. The applications of the Rasch model in formative exams of students with Rasch have numerous advantages because it utilizes measurement accuracy. This can be for problem quality detection, as well as for the detection of individual abilities and the identification of assistance to their learning needs. This study aimed to evaluate the participants' understanding of the concept of Monitoring, Evaluation, Accountability, and Learning (MEAL). Data were acquired by 10 multiple-choice written test items distributed to 25 participants of training MEAL. The ability of participants MEAL to answer formative tests, the item difficulty, and the analysis of the wright map will be discussed in this study.

Keywords: *RASCH Model, Assessment for Learning, Assessment of Education, MEAL Training*

1 INTRODUCTION

Educational assessment is a process that cannot be separated from educational activities. Learning as the core of education requires assessment to determine the learning progress of learners. Schellekens et al., (2021) state that assessment has a major impact on students' learning. Assessment influences what students regard as important; it affects students' understanding of learning tasks and impacts the quality of students' involvement in these tasks; and it influences the transfer of these insights to future learning.

Assessment is meant to contribute to students learning and thus need to be integrated into daily teaching and learning process in the classroom. The quality of learning is determined by the quality of assessment practice in the classroom. Assessment is a critical aspect of the teaching and learning process that aims to collect, interpret, and analyze students' performance (Khan,

2012). Effective and well-planned assessment strategies have a great impact on students' learning because assessment provides an opportunity for teachers to place students in a situation where they exhibit their true potential. This way students and teachers can be in a better position to decide where they are and where they need to go.

In general, there are two objectives of assessment, namely formative assessment and summative assessment. The two have different goals. Summative assessment is also known as assessment of learning. While summative assessment is known as assessment for learning. Smith (2013) explains that the assessment of learning aims to consider student placement, grade promotion, and awarding certificates. The assessment for learning aims to provide information to teachers about the achievement of student competencies so that teachers can make decisions about the next learning. On the other hand, assessment as learning aims to monitor and self-correct students independently by students. Judging from the benchmarks that are used as comparisons, in the assessment of learning the comparisons are other students.

ARG, 1999; Crooks, 2011; Earl, 2003 (Schellekens et al., 2021) state that assessment with a formative function became an Assessment for Learning, which in general emphasizes the purpose of assessment to improve the learning and teaching process. Assessment with a summative function became an Assessment of Learning, which in general was used to judge performance and measure outcomes after a formal learning activity.

In the same context, formative assessment was attached to the improvement of learning in progress, whereas summative assessment was attached to making judgments about achievement at the end of learning (Houston & Thompson, 2017). Formative assessment is carried out throughout the learning time, while summative assessment is carried out at the end of learning.

To improve student learning and the quality of education, the use of formative assessment is emphasized, stressing that teachers need to facilitate changes in students' attitudes toward learning. Cakane (Gabdullina, 2023) states that an essential part of formative assessments is the provision of feedback between students and the teacher or between students. This type of assessment helps to better understand what the students know or do not know and what needs to be improved. Formative assessments provide rich information and judgments about student learning that are mainly fed back into the central dialogue between teachers and learners to inform future student learning (Houston & Thompson, 2017).

The approach that is widely used today in the analysis of exam results is the test theory approach (Classical Test Theory or CTT). Classical test theory can be used to make predictions about the results of an exam (test). This prediction is done by considering several parameters such as the student's ability and the level of difficulty of the problem. Classical test theory (CTT) emphasizes only the apparent score of one exam, which is usually inferred as a person's ability (ability) from the exam taken.

Georg Rasch developed an analytical model from Item Response Theory (or Item Theory, IRT) in the 1960s commonly called 1PL (one logistical parameter). This mathematical model was later popularized by Benjamin Wright (Linacre, 2013). With raw data in the form of dichotomous data (in the form of right and wrong) that indicate student ability, Rasch formulated this into a model that connects students and items.

Mok and Wright (B. Sumintono, 2016) state that the concept of objective measurement in the social sciences and educational assessment according to must have five criteria, namely: 1. Give a linear size with equal intervals; 2. Carry out a precise estimation process; 3. Finding inappropriate items (misfits) or uncommon (outliers); 4. Resolve lost data; 5. Produce replicable measurements (independent of the parameters studied).

Instrument testing and determination of student ability in educational assessment are essential. Analysis that can produce more precise measurements (because it is an equal interval) will determine the quality of the analysis results and efforts to improve the educational process to be able to help students with learning difficulties. The Rasch model can help teachers, lecturers, and educational assessment researchers a lot in improving the quality of the analysis carried out, because the basic principles are appropriate and data processing models are suitable for the analysis of exam results, especially in ordinal data processing. This is because the Rasch model conforms to five objective measurement requirements (B. Sumintono, 2016). The applications of the Rasch model in formative exams of students with Rasch have numerous advantages because it utilizes measurement accuracy. This can be for problem quality detection, as well as for the detection of individual abilities and the identification of assistance to their learning needs.

In this study, we will measure the quality of the MEAL training formative test. Monitoring Evaluation Accountability and Learning (MEAL) is an approach or framework used by organizations or institutions to monitor, evaluate, be accountable, and study the results and

impacts of programs or projects implemented. The MEAL Dompet Dhuafa program is an effort to ensure the success and effectiveness of DD's programs. The program includes various components designed to conduct systematic monitoring, quality evaluation, transparent accountability, and continuous learning. In the MEAL training, there were 25 Dompet Dhuafa employees experiencing a learning process to be able to understand and apply MEAL knowledge in program implementation practices.

This study seeks to provide a clearer understanding of the quality of MEAL training formative test instruments. To pursue that, quantitative methods using formative tests for data collection were adopted where the approach of the Rasch Model and Winsteps software version 3.73 was to assess problems related to the understanding of MEAL trainees and the quality of formative test instruments.

2 METHODOLOGY

This study utilized a quantitative approach by conducting a survey to measure the understanding level of MEAL trainees with formative test instruments. The following subsections provide details of the survey.

2.1 Participants

Twenty-five MEAL trainees were participants in the study. Data was collected from participants of the MEAL training held from May 23, 2023, to May 25, 2023. The demographic profiles of participants are shown in Table 1.

Table 1. Demographic Data of MEAL Trainees

Demographic		Frequency	Percentage (%)
Gender	Male	19	76
	Female	6	24
Entity	DD Head Office	18	72
	Organ	7	28

2.2 Instrument

MEAL trainees completed a 10-question multiple choice Quizizz test to assess their understanding of the MEAL training material.

2.3 Measurement Model and Data Analysis

Formative test instruments in the form of 10 multiple-choice questions will be analyzed using the Rasch model approach and Winstep software version 3.73. To find out the level of difficulty of the question, the main menu of the output tables in the program appears in **table 13 (item measure)** to detail the logit information of each question item. A high logit score indicates a high level of difficulty in the question.

Rasch modeling can also detect if there are individuals whose response patterns are not appropriate. What is meant by different response patterns is the mismatch of answers given based on their ability compared to the ideal model. On the main menu output tables, **table 6 (person fit order)** is selected to appear. Person fit order that displays sequentially the question items that have no fit criteria at the top. Information on unusual response patterns can be further identified by looking at the scalogram which can be called the Guttman matrix. The characteristic is that each item has a systematic order that can be ranked from low to higher. The goal is to make it easier to analyze, provide explanations, and predict individual abilities and the level of difficulty of question items at once (Sumintono, B. dan Widihiarso W. 2015).

Table 2. Summary Person and Item Statistics

	Person	Item
N	25	10
Measures		
Mean	0.25	0.00
SD	1.00	0.04
SE	0.20	0.40
Outfit Mean Square		
Mean	1.06	1.06
SD	0.82	0.66
Strata	1.05	2.94
Reliability	0.22	0.79
Cronbach's Alpha	0.29	

Based on the information in table 2, we can see the interaction between person and item as a whole, person measure, person reliability value, and item reliability value. The person measure value is 0.25 logit. An average score of more than 0.0 logit indicates a tendency for the ability of MEAL participants to be greater than the difficulty level of the question. Cronbach alpha value (measures reliability, i.e. the interaction between person and item as a whole). A Cronbach alpha value of 0.29 means that the interaction between a person and an item is bad.

The person reliability value is 0.22 and the item reliability value is 0.79. That is, the consistency of answers from respondents is weak and the quality of items in the instrument is sufficient (B. Sumintono & Widhiarso, 2014).

3 FINDINGS AND DISCUSSION

In this section, we will analyze the ability of MEAL participants to answer formative tests, item difficulty, wright map analysis, and scalograms. The results of the study are described in the following sub-sections.

3.1 Item Measure

To analyze the data on the difficulty of the question items, an item measure analysis table (table 3) is used. The item column describes the order of question items from the most difficult to answer to the easiest question items to answer by MEAL training participants. The S4 question item with a logit value of 1.89 is the most difficult question item for MEAL training participants to answer. The S3 question item with a logit value of -4.60 is the easiest question item for MEAL training participants to answer to the easiest question items to answer by MEAL training participants.

Table 3. Item Measure

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
4	5	25	1.89	.54	1.06	.3	1.02	.2	.30	.36	84.0	81.1	S4
9	6	25	1.62	.51	1.52	1.8	2.79	3.0	-.23	.37	72.0	77.6	S9
2	10	25	.73	.45	.77	-1.4	.75	-1.0	.62	.41	84.0	69.1	S2
1	14	25	-.05	.44	1.01	.1	.93	-.2	.42	.42	64.0	68.7	S1
5	14	25	-.05	.44	.79	-1.3	.72	-1.2	.61	.42	80.0	68.7	S5
6	17	25	-.66	.47	.83	-.8	.75	-.7	.55	.39	80.0	72.8	S6
7	18	25	-.89	.48	.90	-.4	.82	-.4	.47	.38	76.0	74.7	S7
10	18	25	-.89	.48	1.40	1.6	1.34	.9	.04	.38	60.0	74.7	S10
8	21	25	-1.71	.58	.65	-1.0	.39	-1.1	.65	.31	84.0	83.9	S8
3	25	25	-4.60	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	S3
MEAN	14.8	25.0	-.46	.62	.99	-.1	1.06	.0			76.0	74.6	
S.D.	6.0	.0	1.75	.41	.28	1.1	.66	1.2			8.4	5.2	

3.2 Person Fit Order

In addition to mapping students' abilities according to the importance of grouping achievement, the Rash Model can detect if there are individuals whose response patterns are not appropriate.

The analysis in Table 4 (person fit order) can be used to determine the consistency of thinking of MEAL trainees and can also be used to determine if there is fraud committed by MEAL trainees.

Table 4. Person Fit Order

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
24	3	10	-1.57	.86	1.43	1.0	3.22	1.9	A .31	.59	66.7	78.4	24P
25	3	10	-1.57	.86	1.43	1.0	3.22	1.9	B .31	.59	66.7	78.4	25P
17	7	10	.86	.81	2.30	2.5	2.53	2.1	C-.13	.46	44.4	75.9	17L
20	5	10	-.32	.76	1.39	1.3	1.80	1.6	D .32	.53	55.6	70.2	20L
9	7	10	.86	.81	1.56	1.3	1.30	.7	E .25	.46	44.4	75.9	09P
18	6	10	.25	.77	1.37	1.1	1.37	.9	F .34	.50	66.7	72.3	18L
12	6	10	.25	.77	1.25	.8	1.35	.9	G .38	.50	66.7	72.3	12L
23	4	10	-.91	.78	1.34	1.1	1.12	.4	H .44	.56	55.6	70.8	23L
13	6	10	.25	.77	1.20	.7	1.20	.6	I .41	.50	66.7	72.3	13L
21	4	10	-.91	.78	1.02	.2	1.04	.3	J .54	.56	77.8	70.8	21L
11	6	10	.25	.77	.99	.1	.90	-.1	K .51	.50	66.7	72.3	11L
5	7	10	.86	.81	.87	-.2	.97	.1	L .50	.46	88.9	75.9	05L
22	4	10	-.91	.78	.92	-.1	.75	-.2	M .61	.56	55.6	70.8	22P
3	9	10	2.56	1.13	.85	.0	.39	-.1	l .40	.29	88.9	88.8	03P
15	5	10	-.32	.76	.77	-.7	.66	-.7	k .64	.53	77.8	70.2	15L
14	5	10	-.32	.76	.71	-1.0	.61	-.9	j .67	.53	77.8	70.2	14L
19	5	10	-.32	.76	.71	-1.0	.61	-.9	i .67	.53	77.8	70.2	19L
6	7	10	.86	.81	.69	-.7	.58	-.7	h .60	.46	88.9	75.9	06P
7	7	10	.86	.81	.69	-.7	.58	-.7	g .60	.46	88.9	75.9	07L
10	6	10	.25	.77	.55	-1.5	.48	-1.4	f .70	.50	88.9	72.3	10L
16	5	10	-.32	.76	.54	-1.8	.47	-1.3	e .74	.53	100.0	70.2	16L
1	8	10	1.57	.90	.47	-1.2	.31	-.8	d .63	.40	100.0	79.7	01L
2	8	10	1.57	.90	.47	-1.2	.31	-.8	c .63	.40	100.0	79.7	02L
4	8	10	1.57	.90	.47	-1.2	.31	-.8	b .63	.40	100.0	79.7	04L
8	7	10	.86	.81	.46	-1.6	.38	-1.3	a .70	.46	88.9	75.9	08L
MEAN	5.9	10.0	.25	.81	.98	-.1	1.06	.0			76.0	74.6	
S.D.	1.6	.0	1.00	.08	.44	1.1	.82	1.0			16.8	4.4	

The MNSQ, ZSTD outfit and PT Measure Corr outfit columns are criteria for assessing individual suitability (person outliers or misfits). To check for inappropriate persons (outliers or misfits) criteria are used:

MNSQ outfit value received: 0.5<MNSQ<1.5

ZSTD outfit value accepted: -2.0<ZSTD<2.0

Pt Mean Corr Value: 0.4<Pt Measure Corr<0.85

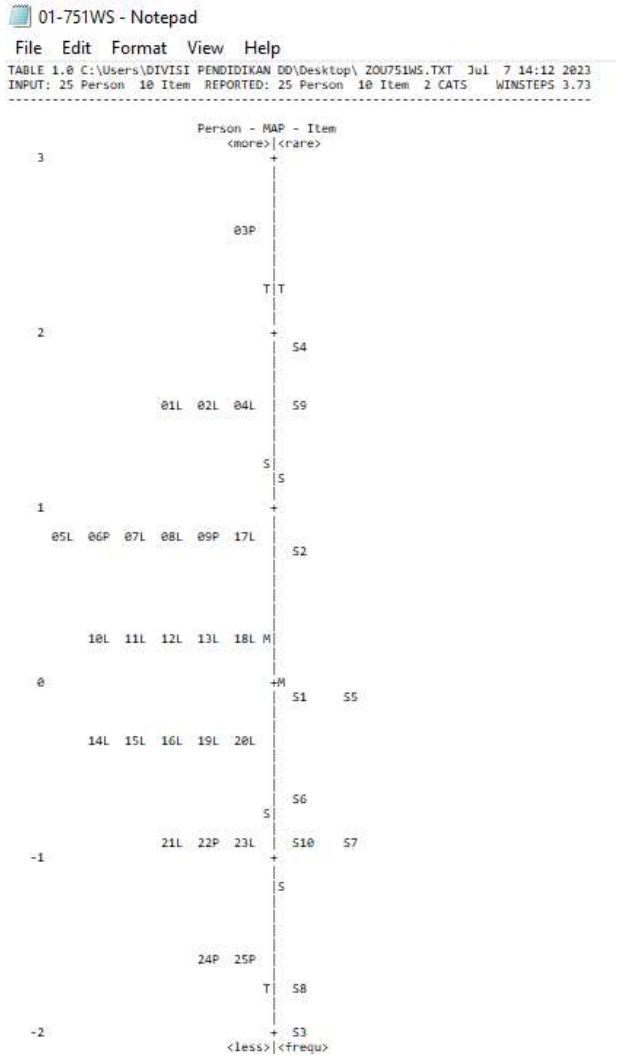
Based on Table 4, to sort the level of incompatibility with the model, there were 11 MEAL participants whose response patterns were rated not fit (outfit MNSQ, namely 24P, 25P, 17L,

20L, 03P, 10L, 16L, 01L, 02L, 04L, 08L. For the ZSTD outfit aspect, only 17L (2.1) qualified. For the Pt Mean Corr value, there are 8 LEAP participants outside the allowed limits, namely 24P, 25P, 17L, 20L, 09P, 18L, 12L, and 03P.

3.3 Wright Map

The map in Figure 1 below illustrates the distribution of the abilities of MEAL participants and the distribution of difficulty levels of questions on the same scale. The map on the left illustrates the ability of MEAL participants who have high ability, namely 03P. Participant 03P lies outside the limit of two standard deviations (T) indicating different high intelligences (outliers). MEAL participants with the lowest ability are 24P and 25P. The question with the highest difficulty is S4 and the question with the lowest difficulty is S3.

Figure 1. Item-Person Map of MEAL Training Formative Test



3.4 Scalogram

Scalograms or Guttman matrices can be used to identify error responses, score loss, and identify guesses.

Figure 2. Scalogram of MEAL Training Formative Test

22-673WS - Notepad

File Edit Format View Help

```
TABLE 22.1 C:\Users\DIVISI PENDIDIKAN DD\Desktop ZOU673WS.TXT  Jul 12  9:39 2023
INPUT: 25 Person  10 Item  REPORTED: 25 Person  10 Item  2 CATS    WINSTEPS 3.73
-------------------------------------------------------------------------------

GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |   1
       |3870615294
       |----------
     3 +1111111101  03P
     1 +1111111100  01L
     2 +1111111100  02L
     4 +1111111100  04L
     5 +1110111100  05L
     6 +1111101100  06P
     7 +1111101100  07L
     8 +1111111000  08L
     9 +1111100011  09P
    17 +1100011111  17L
    10 +1111110000  10L
    11 +1110101100  11L
    12 +1110110001  12L
    13 +1110101010  13L
    18 +1101011010  18L
    14 +1111001000  14L
    15 +1101110000  15L
    16 +1111100000  16L
    19 +1111010000  19L
    20 +1100110001  20L
    21 +1101000100  21L
    22 +1011100000  22P
    23 +1010011000  23L
    24 +1001000010  24P
    25 +1001000010  25P
       |----------
       |   1
       |3870615294
```

Pay attention to data 01L, 02L, 04L (has the same logit value of 1.57 logits) and 24P and 25P data (has the same logit value, which is -1.57 logit). These data have indications of mutual cheating because the response pattern is the same. Participant 03P is included in the category of not careful because question number 9 which is much easier than question number 4 cannot be answered correctly. This also happened to MEAL17L training participants.

4 CONCLUSION

The aim of this research was to analyze the responses of MEAL trainees to formative test instruments utilizing the Rasch model approach. The results showed that the easiest question to answer was S3, while the most difficult was S4. Participant 03P displayed the highest level of ability, whereas participants 24P and 25P had the lowest levels of ability. The Cronbach alpha value showed poor formative test instrument reliability, with a person reliability value of 0.22 and an item reliability value of 0.79. This suggests that the consistency of answers from respondents was weak, while the quality of items in the instrument was satisfactory. The weak person's reliability can be attributed to various factors. Firstly, the highest-ability participants were unable to answer the easier questions correctly, leading to inaccuracies. Secondly, there were indications that some participants were cheating. Lastly, correct guesses from several MEAL training participants contributed to the weak reliability of the responses.

The study suggests two recommendations. Firstly, there should be improvements made to the MEAL training formative question instrument, especially in item number 3, so that it can be answered by all the MEAL participants. Secondly, the MEAL training participants who are identified as cheating, not answering easy questions carefully, and guessing should be followed up and their training should be deepened.

ACKNOWLEDGEMENTS

The author thanks Dompet Dhuafa for the opportunity to benefit the community.

REFERENCES

Gabdullina, S. (2023). THE POTENTIAL OF THE ESSAY IN FORMATIVE ASSESSMENT: LITERATURE REVIEW. Education. Innovation. Diversity., 1(6), 48–53. https://doi.org/10.17770/eid2023.1.7176

Houston, D., & Thompson, J. N. (2017). Blending formative and summative assessment in a capstone subject: ‘it’s not your tools, it’s how you use them.’ Journal of University Teaching and Learning Practice, 14(3). https://doi.org/10.53761/1.14.3.2

Khan, B. (2012). Relationship between assessment and students’ learning International Journal of Social Sciences and Education RELATIONSHIP BETWEEN ASSESSMENT AND STUDENTS’ LEARNING. International Journal of Social Sciences and Education2012, 2(1), 576–588. http://ecommons.aku.edu/pakistan_ied_pdcn

Linacre, J. M. (2013). A user's guide to winsteps ministeps Rasch-model computer programs. [version 3.74.0]. winsteps.com (by John M. Linacre). http://www.winsteps.com/index.htm

Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). In Studies in Educational Evaluation (Vol. 71). Elsevier Ltd. https://doi.org/10.1016/j.stueduc.2021.101094

Smith, A. B. (2013). Assessment for Learning. In Understanding Children and Childhood: A New Zealand Perspective (pp. 241–277). Bridget Williams Books. https://doi.org/10.7810/9781927131763_9

Sumintono, B. (2016). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (assessment for learning). http://deceng3.wordpress.com

Sumintono, B. dan W. W. (2015). Aplikasi Pemodelan Rasch pada Assessment Pendidikan. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2014). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (Rev ed.). Trimkom Publishing House.