International Journal of Engineering & Technology, *7 (4.33) (2018) 32-36*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Evaluating Quality and Reliability of Final Exam Questions for Probability and Statistics Course Using Rasch Model

Faiz Zulkifli*, Rozaimah Zainal Abidin, Nor Faezah Mohamad Razi, Nor Hazlina Mohammad, Rusliza Ahmad, Anis Zafirah Azmi

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Tapah Road, 35400 Perak, Malaysia
**Corresponding author E-mail: faiz@tmsk.uitm.edu.my*

Abstract

Evaluation of the questions' level of complexity for the statistical course was proposed using the revised version of Bloom's taxonomy. The use of Bloom's taxonomy in statistical examination papers allows the degree of difficulty to be pseudo-objectively assessed. Well-constructed questions in the final examination will help in measuring students' abilities based on comprehensive cognitive skills. Therefore, this study used Rasch Model to evaluate the quality and reliability of final exam questions for probability and statistics course. According to research findings, five out of 30 questions are considered as misfit items. It is therefore recommended that these items be removed or rephrased to better suit the students' ability level in a course. Whereas, nine questions have significant differences between taxonomy level and Rasch level that require further analysis. Overall, students view the set of exam questions as simple due to the unavailability of difficult items. Based on this result, it is suggested that the exam questions should undergo verification process from the expert and students should be exposed early to various types of questions with different level of difficulty.

Keywords: *Probability and Statistics Reliability; Students' ability; Quality; Questions' difficulty.*

1. Introduction

Statistics is an important element of the curriculum for students in a wide field of study. Unfortunately, for many of these students, statistics is often considered as the most difficult course to learn in their program of study. Previous works have explored the students' challenges in learning statistics, in which the statistics anxiety is identified as an important factor that affects the student's performance. The researchers indicated that many university students have high level of statistics anxiety when facing with statistical ideas, problems, issues, instructional situation and evaluate situation [1]. As a result, students often delay to an extent in enrolling for statistics courses and may lead to students' low academic achievement in statistics course.
Students' difficulties in understanding and learning have been reported. According to [2], students have underlying difficulty with ideas of probability concepts, which they need to construct their own understanding. Studies have also shown that students have less understanding concepts of statistical variation, sampling distribution and certain reasoning about graphical representation [1]. In a recent study, the finding shows that the students perceive little understanding on counting rule and basic probability concept [3]. Meanwhile, some researchers suggested that the lack of students' understanding in learning statistics is not the only factor that causes failure in academic performance. Students' low academic performance in the examination is one of the most challenging problems that students face as well as lecturers. This problem has many causes such as student attributes, exam questions construction and teaching approach. Teaching and learning methodology without problem-solving approach does not involve with a variety of cognitive thinking skills in a process of answering different difficulty level of questions [1]. Moreover, the questions provided often lack adequate context and not in line with the difficulty level of the questions. Consequently, the students are having difficulty in assessing the level of questions. Hence, this may lead to frequent repetition of failure that affects the students' performance. Word problems may be the most frequently used format for exam questions in statistics courses [4]. Commonly, with suitable method of teaching and learning statistics, students with accurate understanding are able to find ways to correctly answer this kind of word problems.
Due to this matter, the improvement in teaching and learning process should be emphasized. As a matter of fact, educators should use appropriate educational procedures that are common within the scope of in-depth knowledge understanding and development of thinking skills. Assessment tools, especially the examination papers modelled using Bloom's taxonomy are introduced to compare the difficulty of questions and students' performance in accessing the students' understanding in learning statistics. Bloom's taxonomy is a model of hierarchy proposed by [5], a team of educational psychologists in 1956. The research is to promote ways of thinking according to levels of complexity and inculcate higher forms of thinking in education sector. It is also being used by educators to control the level of questions' difficulty when designing assessment. Bloom's model can be divided into two, which are lower level and higher level. Knowledge, comprehension and application are categorized as lower level since they are more basic than the higher level. The higher level consists of analysis, synthesis and evaluation. Bloom's model has been updated since and a new version is used until now to which was divided into six categories: remembering, understanding, applying, analysing, evaluating and creating; a lower order to higher order of hierarchy [6]. In statistics course, the main task is to interpret the Bloom's

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

taxonomy levels in the context of the cognitive processes required to complete exam questions. Well-organized exam questions modelled using Bloom's taxonomy should be able to measure the level of students' ability, thus contribute to the improvement in students' performance.

Students' performance has been essentially dependent on the grades scored from quizzes, assignments, tests, projects and final examinations. Measuring students' performance is essential in education to ensure that they achieved the expected learning outcomes. In order to increase students' performance, construction of a good assessment based on Bloom's cognitive thinking skills and the degree of students' ability should be taken into consideration [7]. A discussion on reliability is also needed when evaluating the quality of the questions. In order to assess the understanding and ability of the students, a reliable evaluation instrument in teaching and learning activity is required.

Rasch measurement model has been widely used to examine, validate and analyse the students' ability and quality of exam questions [7]. Most of the researchers found that the quality of the final assessment questions set up were incredibly satisfactory and also were calibrated with the learning ability of the students. Rasch model is also able to categorize grades into respective learning outcomes, especially when the number of sampling unit is small [8]. According to [7], Rasch model is considered as a better evaluation model for evaluating course outcomes performance as compared to standard evaluation method. Therefore, validation of instruments constructed is important especially to control the measurement error.

Thus, the aims of this study are to access the quality of exam questions via Bloom's Index and evaluate whether the exam questions calibrate with students' learning abilities on prior lesson, by using Rasch measurement model. In this study, the final exam questions for the statistics program which employ the Integrated Cumulative Grade Point Average (iCGPA) system is taken into account as assessment tools. It is part of the study to improve students' understanding and ability in learning statistics based on Bloom's cognitive thinking skills. Hence, it will encourage educators to think more broadly about cognitive measures and act as guidance in determining the appropriate Bloom's level for the item tests.

2. Methodology

The information was gathered from the final assessment results of Probability and Statistics (STA150) course, taken by part two CS111 (Diploma in Statistics) and third semester CS143 (Diploma in Mathematical Sciences) students from Machang, Perak and Raub campuses, both wherein used iCGPA system in the results. The final examination paper contained 10 questions, where each question contains several items. All the questions were developed according to Test Specification Tables or *Jadual Spesifikasi Ujian* (JSU) that had been set up by a resource person. Items were labelled as Questions, Learning Topic, Cognitive Level, and Marks which involved past semester final examination papers as described in Table 1. The final examination papers were from March 2017 and Jan 2018. All students were required to answer all questions.

Table 1: Examination Questions Specification for Course Code STA150

Question			Learning Topic	Cognitive Level	Marks
1	a	i	Set, Sample space and Events	C1	2
		ii	Set, Sample space and Events	C1	2
		iii	Set, Sample space and Events	C1	2
	b	i	Permutation and Combination	C3	2
		ii	Permutation and Combination	C3	2
2	a		Conditional Probability	C3	3
	b		Bayes Theorem	C2	3
	c		Conditional Probability	C2	3
3	a		Continuous Random Variable - Mathematical Expectations	C5	5
	b		Continuous Random Variable - Mathematical Expectations	C2	3
	c		Continuous Random Variable - Probability Distribution	C5	5
4	a		Continuous Bivariate Probability Distributions - Marginal Distributions	C4	3
	b		Continuous Bivariate Probability Distributions - Marginal Distributions	C4	4
	c		Continuous Bivariate Probability Distributions - Marginal Distributions	C2	2
	d		Continuous Bivariate Probability Distributions - Conditional Distributions	C5	4
5	a		Discrete Distribution - Binomial Distribution	C4	3
	b		Discrete Distribution - Binomial Distribution	C2	3
	c		Discrete Distribution - Binomial Distribution	C4	3
6	a		Continuous Distribution - Uniform Distribution	C4	2
	b		Continuous Distribution - Uniform Distribution	C3	2
	c		Continuous Distribution - Uniform Distribution	C3	2
7	a		Discrete Bivariate Probability Distributions - Marginal Distribution	C2	2
	b		Discrete Bivariate Probability Distributions - Conditional Distributions	C2	2
	c		Discrete Bivariate Probability Distributions - Mathematical Expectations	C5	2
8	a		Continuous Distribution - Binomial Distribution	C3	3
	b		Continuous Distribution - Binomial Distribution	C4	3
	c		Continuous Distribution - Normal Approximation to Binomial Distribution	C3	5
9	a		Continuous Distribution - Poisson Distribution	C4	2
	b		Continuous Distribution - Poisson Distribution	C5	2
	c		Continuous Distribution - Poisson Distribution	C4	3
10	a		Continuous Distribution - Normal Distribution	C3	3
	b		Continuous Distribution - Normal Distribution	C3	3
	c		Continuous Distribution - Normal Distribution	C4	4

Hence, a total of 213 students' results were analyzed. Marks from final assessment results were obtained and compiled. A standardization method was used since this raw score had different total marks for a question [7]. The standardization was given in (1):

$$z_{ij} = \frac{x_{ij} - min\left(x_j\right)}{max\left(x_j\right)} \tag{1}$$

where

$i = \text{the } i^{th} \text{ students } (i = 1,2,....,213),$

$j = \text{the } j^{th} \text{ questions } (j = 1,2,3,...,30),$

$z_{ij} = \text{standardized marks for } i^{th} \text{ student and } j^{th} \text{ question},$

$x_{ij} = \text{marks for } i^{th} \text{ student and } j^{th} \text{ question},$

$\min\left(x_j\right) = \text{minimum marks for } j^{th} \text{ question, and}$

$\max\left(x_j\right) = \text{maximum marks for } j^{th} \text{ question.}$

Students were ranked according to their accomplishment. Responses from the students' results were analyzed using rating scale:

$$z_{ij} \times 5 = A$$

Henceforth, the marks obtained by each student was organized in Excel *prn. Rasch software, called Winstep was used in this study to evaluate the examination questions for STA150.

3. Results and Discussion

This study investigated students' examination marks in the subject of Probability and Statistics offered by the Faculty of Computer and Mathematical Sciences, MARA University of Technology. To examine the students' ability and questions' difficulty, the exam questions were prepared according to Bloom's revised taxonomy. The exam questions consisted of 30 structured questions were then administered to 213 students.

Table 2 shows the reliability and separation index for 30 items and 213 students measured. The reliability index of esteemed person is at 0.84 (practically equivalent to the traditional Cronbach's alpha), thus it shows that the items have high reliability in imitating a student's score. Meanwhile, with the item reliability of 0.98, it shows that a comparative item along the variable is very reproducible in a comparable sample from the population. This index value implies that the scores of each question are consistent and steady [9].

The person separation index is measured at 2.32, with three levels of student ability indicated low, moderate and excellent. Next, the separation index of item is measured at a high 8.05 which contributed by the large item spread value of 1.77 logit. This index value is reasonable for the five levels of revised Bloom's Taxonomy which are remembering, understanding, applying, analyzing and evaluating. Both separation index values are acceptable as they are higher than 2.

Table 2: Summary of 213 measured persons and 30 measured items

SUMMARY OF 213 MEASURED Persons

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	107.3	30.0	.36	.15	1.02	.0	1.03	.0
S.D.	18.8	.0	.42	.03	.32	1.2	.56	1.2
MAX.	145.0	30.0	1.61	.32	2.33	4.9	4.51	6.3
MIN.	40.0	30.0	-1.21	.13	.36	-2.3	.31	-1.8

REAL RMSE .17 ADJ.SD .38 SEPARATION 2.32 Person RELIABILITY .84
MODEL RMSE .15 ADJ.SD .39 SEPARATION 2.50 Person RELIABILITY .86
S.E. OF Person MEAN = .03

Person RAW SCORE-TO-MEASURE CORRELATION = .93
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .87

SUMMARY OF 30 MEASURED Items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	762.7	214.0	.00	.06	1.03	-.1	1.03	.0
S.D.	190.6	.0	.51	.01	.30	3.2	.36	2.3
MAX.	1004.0	214.0	.91	.09	1.73	6.2	2.27	6.3
MIN.	411.0	214.0	-.86	.05	.51	-8.0	.56	-4.9

REAL RMSE .06 ADJ.SD .50 SEPARATION 8.05 Item RELIABILITY .98
MODEL RMSE .06 ADJ.SD .50 SEPARATION 8.65 Item RELIABILITY .99
S.E. OF Item MEAN = .09

UMEAN=.000 USCALE=1.000
Item RAW SCORE-TO-MEASURE CORRELATION = -.99
6390 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 13298.43 with 6145 d.f. p=.0000

Fig. 1 shows that item difficulty is distributed on the right side and person ability is distributed on the left side with the same logit ruler. The ruler comprised of samples extending from 1.61 to -1.21, where the most excellent student and the most difficult question are laid out over the scale. The students' identity field were recoded based on their respective campus location (M - Machang, P - Perak, R - Raub). Meanwhile, question items were recoded as per the question number and cognitive level of Bloom's revised taxonomy. For example, 4b_A represents the question 4b with "Analysing" cognitive level.

Generally, the students thought that the given exam question set was easy since the person mean (0.36 logit) had fallen over the item mean (0.00 logit). There were no difficult items since all items were located inside the person range (-1.21 to 1.61 logit). Items "4a and 10a" from the "Applying" and "Analyzing" levels were viewed as the easiest by most students. In addition, 15 students were considered as excellent students while two students were found unable to answer even one question. Apart from that, determining the difficulty level of a question cannot rely solely on the intended cognitive process. Based on the map, some questions were assumed to be at a difficult level, but most students could answer the questions (example: 4a_A, 10a_P).

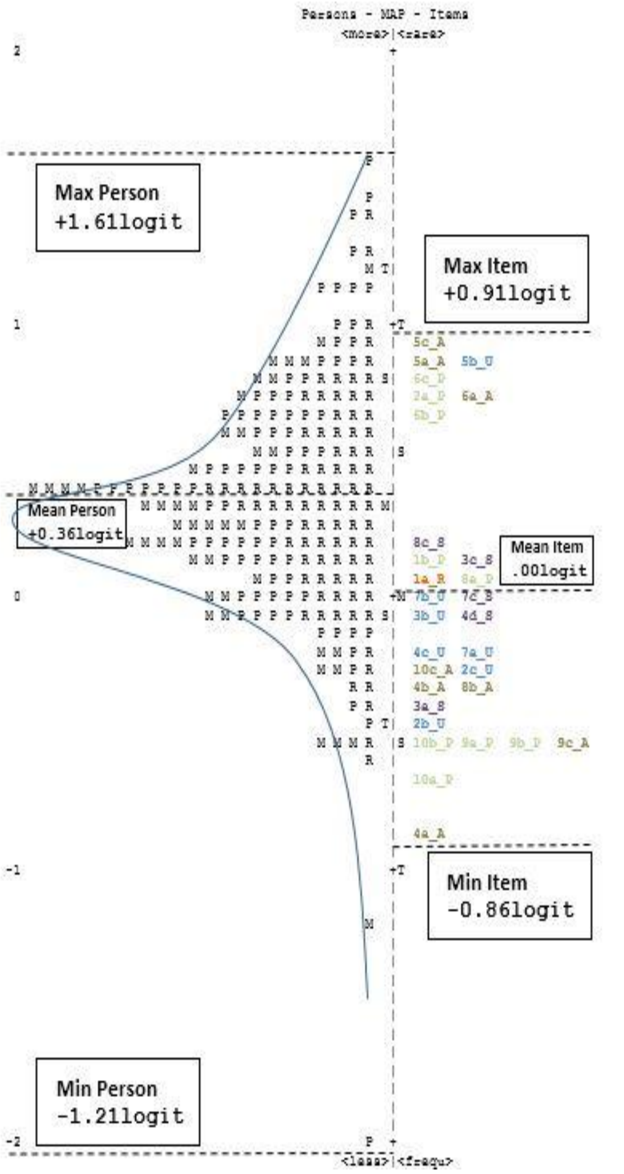

Fig. 1: Person-item distribution map

The Rasch model expects the response pattern to move from the easy item to the left with a score of 5 and up towards the tough item on the right with a score of 1. Based on scalogram in Figure 2, the simplest level of questions to the hardest level can be expressed in Table 3.

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       | 22222  12 311 1222   21 11111
       |987956460350197210212847368435
       |------------------------------
   49 +555555555555555555555453355555  P49F
   22 +555555555555555555555454325555  P22F
   77 +555555555555555555553555555355515  P77F
  159 +555555555555555555554555215555 5  R77F
   59 +555555555555555555555553313555  P59M
  163 +555555555555555555554555215355 5  R81F
  177 +555555554555515555525455535555 5  M8 F
   24 +555555555555455555555453313555  P24F
   28 +555555555555555555535545535333 3  P28M
   35 +555555555555555555535545531155 5  P35F
   50 +555555555553555554155445535355 5  P50F
   14 +555555555555555555455444533115 55  P14F
   81 +555555555555525455535545531155 5  P81F
  130 +555555555555554553554554313155 5  R48F
    2 +555555155555555551555555553155 15  P2 M
   16 +555555555555455555353453311555  P16F
  113 +555555555525555555544552131555  R31M
    . ..............................  ....
    . ..............................  ....
    . ..............................  ....
```

Fig. 2: Scalograms

Taxonomy level is tied with revised Blooms' taxonomy proposed by [4]. Meanwhile, the Rasch level is based on the order of the question from the scalogram diagram. Through comparative analysis, the difficulty level of questions "2b, 4a and 8c" is the same according to both levels. Eighteen questions have slight differences between two levels. Whereas, nine questions of "1a, 2a, 3a, 5b, 6b, 6c, 7b, 9c, and 10a" have significant differences which require further study. To improve the quality of the exam questions, it is important to first have these exam questions vetted by experts from the respective field, so that the questions are appropriate to students' level and ability. In addition, lecturers should also emphasize on the Bloom's taxonomy level in questions during teaching and learning sessions [11, 12].

Table 3: Level of questions

Question	4a	10a	9c	10b	9a	9b	2b	3a	4b	8b	2c	10c	4c	7a	3b
Taxonomy Level	1	3	4	3	3	3	2	5	4	4	2	4	2	2	2
Rasch level	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3
Question	4d	7c	7b	8a	1a	1b	3c	8c	6b	2a	6a	6c	5b	5a	5c
Taxonomy Level	5	5	2	3	1	3	5	5	3	3	4	3	2	4	4
Rasch level	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5

*Note: Taxonomy level (1-remembering, 2-understanding, 3-applying, 4-analyzing, 5- evaluating)

To identify the item misfits using the Rasch Measurement Model, three item fit statistics were performed; infit MNSQ, outfit standardized value and point measure correlation (PMC) value. Measurement on the item fit started with MNSQ. Basically, MNSQ is the relationship or ratio of an observation compared with expectation. The value of MNSQ that equals 1 is said to be perceptible as indicated by desires. In the meantime, MNSQ is calculated to be out of desire when the value of MNSQ Infit is out of range from 0.73 to 1.33. Indicators that fortify the Misfit Item are the value of the Z-Standard Items that drop out of range from -2 to 2. Constraints of satisfactory PMC varying range is dependent on the purpose of the instrument. Final exam questions generally require more precise items inside the range of 0.4 to 0.8. However, a negative value PMC item is found to not measure what should be measured and thus, should be dropped. From Table 4, items "7a_U, 2b_U, 6c_P, 7c_S and 6b_P" could be viewed as misfit items because they could not meet all the three measurements. These misfit items should be considered for further analysis, for example, eliminating or rephrasing.

Misfit person negatively affects item reliability. Responses that cause distortion in actual measurements need to be set aside. Data from these respondents can be categorized as unreliable data. As with the item, the Infit MNSQ values must be within 0.7 to 1.34. Respondents with a negative PMC are assumed to answer an unobtrusive question. In Table 5, the respondent categorized as misfit is P54F.

Table 4: Item Misfit

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	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
19	859	214	-.20	.05	1.73	6.2	2.27	6.3	A .24	.44	23.9	30.7	7a_U
4	933	214	-.46	.06	1.57	3.8	1.67	2.9	B .32	.43	47.9	49.2	2b_U
18	451	214	.79	.05	1.51	4.7	1.56	3.6	C .29	.41	21.1	29.1	6c_P
21	785	214	.00	.05	1.38	4.3	1.37	2.8	D .39	.45	15.5	23.3	7c_S
9	1004	214	-.86	.09	1.38	1.7	.97	.0	E .46	.41	85.9	79.5	4a_A
29	955	214	-.55	.07	1.26	1.7	1.32	1.4	F .40	.42	59.6	58.0	10b_P
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
10	900	214	-.33	.06	.77	-2.2	.65	-2.2	f .49	.43	42.3	38.9	4b_A
17	499	214	.66	.05	.58	-5.9	.74	-2.4	e .37	.43	41.8	23.2	6b_P
2	734	214	.12	.05	.73	-3.9	.73	-2.8	d .43	.45	35.2	19.3	1b_P
7	817	214	-.08	.05	.65	-4.7	.63	-3.2	c .58	.45	33.3	27.6	3b_U
30	876	214	-.25	.06	.59	-4.7	.63	-2.6	b .47	.44	39.9	34.8	10c_A
1	743	214	.10	.05	.51	-8.0	.56	-4.9	a .49	.45	43.7	19.1	1a_R
MEAN	762.7	214.0	.00	.06	1.03	-.1	1.03	.0			37.2	35.8	
S.D.	190.6	.0	.51	.01	.30	3.2	.36	2.3			17.5	15.5	

Table 5: Person Misfit

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	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
54	82	30	-.15	.13	2.33	4.9	4.51	6.3	A -.16	.56	3.3	28.6	P54F
111	119	30	.57	.16	2.19	3.1	3.15	3.3	B .03	.60	43.3	42.9	R29M
63	119	30	.57	.16	1.28	.9	2.98	3.1	C .36	.60	36.7	42.9	P63F
27	124	30	.70	.16	1.70	2.0	2.97	2.8	D .10	.58	40.0	48.8	P27F
2	132	30	.94	.18	1.74	2.0	2.83	2.2	E .29	.50	70.0	67.4	P2 M
177	139	30	1.21	.22	1.57	1.3	2.83	1.8	F .05	.40	63.3	71.6	M8 F
157	113	30	.43	.15	1.67	2.1	2.28	2.6	G .31	.62	23.3	30.3	R75F
191	62	30	-.52	.14	1.08	.4	2.10	2.2	H .36	.46	26.7	32.4	M22F
210	40	30	-1.21	.24	1.37	.8	2.04	1.4	I .08	.27	70.0	68.6	M41F
55	82	30	-.15	.13	1.48	2.1	1.95	2.5	J .37	.56	20.0	28.6	P55F
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
124	116	30	.50	.15	.45	-2.3	.53	-1.3	d .79	.61	53.3	37.6	R42M
141	127	30	.78	.17	.52	-1.8	.31	-1.6	c .81	.55	66.7	54.3	R59F
137	120	30	.60	.16	.46	-2.1	.36	-1.7	b .84	.60	56.7	45.4	R55F
28	137	30	1.12	.20	.36	-2.1	.33	-1.0	a .70	.44	73.3	70.8	P28M
MEAN	106.9	30.0	.33	.16	1.02	.0	1.03	.0			37.2	35.8	
S.D.	19.4	.0	.51	.11	.32	1.2	.56	1.2			16.5	15.5	

4. Conclusion

This study revealed that nine out of 30 exam questions for Probability and Statistics need to be revised in an effort to increase students' ability to answer questions with varying degrees of difficulty. The Rasch model can precisely classify the questions according to students' ability across five cognitive levels of revised Bloom's taxonomy. Rasch measurements can evaluate person and item reliabilities, and subsequent measurements can identify any misfits due to the exceptional response given by the students. To improve the quality of the exam questions, it is necessary to undergo the expert validation process.

Besides looking at the quality of the exam questions, teaching and learning strategies should be tended to. Through the student-centered approach, lecturers can identify what students' most favored way of learning is based on their diverse backgrounds [13, 14]. Additionally, students need to be exposed early on various types of questions with varying degrees of difficulty. This can be done by lecturers throughout the learning session. The use of learning aids can also improve students' thinking skills, particularly towards high-level thinking skills [11, 12, 15].

The process of preparing exam questions needs to be studied in depth. Most importantly, it should take into account the views of lecturers and students. Other than that, the difficulty level of a question should not depend entirely on the verb of cognitive level that has been set. If a question requires a broad knowledge base to answer and involves a lot of work, it should be categorized as a difficult question. Hopefully with this study, the quality of exam questions can be standardized.

Acknowledgement

The authors would like to convey their thanks and gratefulness to Universiti Teknologi MARA Perak Branch and especially to Faculty of Computer and Mathematical Sciences for providing them the opportunity to run this study.

References

[1] Marriott J, Davies N & Gibson L (2009), Teaching, learning and assessing statistical problem solving. *Statistics Education* 17(1), 1-18.

[2] Garfield J & Ben-Zvi D (2014), *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

[3] Baharun N, Mohd Razi NF, Zainal Abidin R, Musa NAC & Mahmud Z (2017), Measuring students' understanding in counting rules and its probability via e-learning mode: A Rasch measurement. Journal of *Fundamental and Applied Sciences* 9(65), 429-441.

[4] Schau C & Mattern N (1997), Assessing students' connected understanding of statistical relationships. In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*. Amsterdam: IOS Press, pp. 91-104.

[5] Bloom BS, Furst MD, Hill EJ, WH & Krathwohl (1956), *Taxonomy of educational objectives: The classification of educational goals, Handbook 1: Cognitive domain.* David McKay Co Inc.
[6] David L (2018), *Bloom's Taxonomy (Bloom).* https://www.learning-theories.com/blooms-taxanomy-bloom.html
[7] Othman H, Asshaari I, Bahaludin H, Nopiah ZM & Ismail NA (2012), Application of rasch measurement model in reliability and quality evaluation of examination paper for engineering mathematics courses. *Procedia Social and Behavioral Sciences* 60, 163-171.
[8] Saidfudin M, Azrilah AA, Rodzo'An NA, Omar MZ, Zaharim A & Basri H (2010), Easier learning outcomes analysis using Rasch Model in engineering education research. *Proceedings of the 7th WSEAS International Conference on Engineering Education*, pp. 442-447.
[9] Creswell JW (2005), *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. Pearson Merrill Prentice Hall.
[10] Anderson LW (Ed.), Krathwohl DR (Ed.), Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, & Wittrock MC (2001), *A Taxonomy for learning, teaching, and assessing: A Revision of bloom's taxonomy of educational objectives.* Longman.
[11] Zulkifli F, Zainal Abidin R, Mansor Z, Mohammad Hamzah MH, & Zulkipli F (2017), A+ Stat. In *Creative Innovation Without Boundaries*, pp. 125-129.
[12] Zulkifli F, Zainal Abidin R, Mansor Z, Mohammad Hamzah MH & Zulkipli F (2018), Use of learning aids on advancing students' thinking ability. *Jurnal Inovasi Malaysia* 1(2), 1-22.
[13] Zulkifli F, Fadhlullah A, Abu Bakar N & Zainal Abidin R (2018), An investigation on multiple intelligence of students from the faculty of Business Management (FPP) and Faculty of Computer and Mathematical Sciences (FSKM), Universiti Teknologi MARA (UiTM). *Insight Journal* 1(1), 41-48.
[14] Zulkifli F, Fadhlullah A, Abu Bakar N, Jahya A, Ismail NF, Hashim H & Ahmad Ridzuan ANA (2014), A comparative analysis on learning preferences of students at Faculty of Computer and Mathematical Sciences (FSKM) and Faculty of Business. Proceedings of the *Kolokium Pengajaran dan Pembelajaran UiTM Ke-2*.
[15] Zulkifli F, Zainal Abidin R, & Abdullah MN (2016), *Mathematical statistics.* Bootstrap Resources.