How to Cite:
Lohgheswary, N., Leelavathi, . R., & Lun, A. W. (2022). Evaluating students performance in advanced programming final exam questions. *International Journal of Health Sciences*, *6*(S2), 1040–1049. https://doi.org/10.53730/ijhs.v6nS2.5074

Evaluating Students Performance in Advanced Programming Final Exam Questions

N. Lohgheswary
Department of Electrical and Electronics Engineering, XIAMEN University Malaysia, Sepang, 43900, Selangor Darul Ehsan, Malaysia

R. Leelavathi
School of Information Technology, SEGi University, Kota Damansara, 47810 Selangor darul ehsan, Malaysia

A. Wei Lun
Centre for Sustainable Process Technology, Faculty of Engineering and Built Environment Universiti Kebangsaan Malaysia, Bangi, 43600 Selangor Darul Ehsan, Malaysia

***Abstract*---**Students grade in a subject is determine mainly from the final examination marks. The final exam questions need to be reliable in order to measure the performance of students in the particular subject. Rasch model able to measure the reliability of an instrument, the final exam questions. A total of 60 students from the School of Information Technology sat for the Advanced Programming final examination. Marks were entered in excel and transferred as *prn format. Marks were analyzed against the Rasch model, WINSTEPS. The summary statistics for person, summary statistics for item, item dimensionality, item statistics, item distribution map and person distribution map are among the output being analyzed. Two items were identified as misfit items. The Advanced Programming exam questios was very reliable and able to give insight view of the questions. A corrective action is taken to review the misfit item and rephrase the item.

***Keywords*---**advanced programming, evaluate, final exam, questions, students performance.

Introduction

Students grade in a subject is determine mainly from the final examination marks. This is because final examination contributes 60% of the total course

***Corresponding author**: Lohgheswary, N.; Email: lohgheswary.nagarethinam@xmu.edu.my*
Manuscript submitted: 27 Nov 2021, Manuscript revised: 09 Feb 2022, Accepted for publication: 18 March 2022

grade and the remaining 40% is determine from the coursework. The coursework includes quizes, assignment and mid term test. The exam questions must be reliable as that will ensure the Course Outcome of a subject is achieved. The selection for final exam questions will be base on the Course Outcomes. Thus the Course Outcomes will be achieved if the questions are reliable enough to measure the performance of students.

Advanced Programming has been a foundation subject for all the undergraduate Information Technology programme. This is because the strong foundation in understanding the programming knowledge and applying those knowledge to other application of Information Technology subjects is the aim of incorporating Advanced Programming subject in undergraduate Bachelor of Information Technology courses.

Rasch model has been used widely in education. Reference [1] focused on the procedure to adapt Mathematical test items using the Rasch model. Content validation is based on expert evaluation of the development of test questions or items. Item statistics from the study revealed that only 86 of the test questions met the Rasch model assumption which was unidimensionality of the Mathematical test items. The study concluded that 14 items need to be dropped, ignored or reviewed because they are not compatible with Rasch model. An item reliability value of 0.96 indicates that the item is reliable for administering.

Reference [2] described the use of the Rasch model to analyze the content of college placement tests in Mathematics subjects. The Rasch model was used to determine whether placement test could be used as a predictor of success for all 4 courses. The study used a partial credit model known as item response and belonged to the Rasch model. Reference [3] describes the relationship between individual ability and item complexity on a scale to that measured in 'logits' or odd log units.

Methodology

The final examination for Advanced Programming was conducted in session 2018 / 2019. After the students undergo a teaching of 14 weeks in a semester, the final exam was conducted. The total percentage of the mid semester test, assignment, quiz and final examination will determines the grade of a student for the Advanced Programming subject. A total of 4 subjective questions were designed and validated by a group of lecturers. The duration of the final examination is 3 hours and the total marks are 100. A total of 60 students from Information Technology department sat for the final exam of Advanced Programming.

Table 1 list the Course Outcome for the Advanced Programming subject. There are a total of 4 Course Outcomes. They are declare and use interface and generic types, build a multi-threaded graphical user interface in accordance with a detailed software application, implement network client and server, and describe the concept of software agents and multi-agents systems. Table 2 shows the details entry of each of the final exam questions. This includes the details of Course Outcome, the level of Bloom Taxonomy and the description of Bloom Taxonomy for each question.

Table 1
Course outcome for advanced programming subject

Course Outcome	Description
1	Declare and use interface and generic types.
2	Build a multi-threaded graphical user interface in accordance with a detailed software specification.
3	Implement network client and server.
4	Describe the concept of software agents and multi-agents systems.

Table 2
Entry for final exam question for advanced programming subject

Question	Course Outcome	Bloom Taxonomy Level	Description	Marks
1.a.i.	CO1	1	Knowledge	2
1.a.ii.	CO1	2	Comprehension	2
1.a.iii.	CO1	3	Application	2
1.b.i.	CO1	1	Knowledge	2
1.b.ii.	CO1	2	Comprehension	3
1.b.iii.	CO1	2	Comprehension	3
1.b.iv.	CO1	2	Comprehension	3
1.b.v.	CO1	2	Comprehension	3
1.b.vi.	CO1	3	Application	2
1.b.vii.	CO1	3	Application	2
1.c.	CO1	1	Knowledge	1
2.a.	CO1	3	Application	3
2.b.	CO1	4	Analysis	4
2.c.	CO1	4	Analysis	4
2.d.	CO2	3	Application	2
2.e.	CO1	3	Application	8
2.f.	CO1	3	Application	4
3.a.i.	CO1	3	Application	3
3.a.ii.	CO1	4	Analysis	2
3.a.iii.	CO1	2	Comprehension	4
3.b.	CO1	5	Synthesis	4
3.c.i.	CO1	1	Knowledge	2
3.c.ii.	CO1	1	Knowledge	2
3.d.	CO1	4	Analysis	4
3.e.i.	CO1	5	Synthesis	2
3.e.ii.	CO1	5	Synthesis	2
4.a.i.	CO2	1	Knowledge	2
4.a.ii.	CO2	1	Knowledge	2
4.b.	CO2	3	Application	8
4.c.	CO2	5	Synthesis	4
4.d.i.	CO3	5	Synthesis	2
4.d.ii.	CO3	5	Synthesis	2

Question	Course Outcome	Bloom Taxonomy Level	Description	Marks
4.e.	CO3	2	Comprehension	3
4.f.	CO4	2	Comprehension	2

Results and Discussion

As a first step, grades were compiled in the EXCEL *prn format. The grades were transferred using Bond and Fox [4] known as WINSTEPS. It is a Rasch analysis software used to obtain the logit values. Figure 1 shows the summary statistics for person. Person represents the students who took Advanced Programming final examination. The person summary reveals a strong reliability of Cronbach Alpha = 0.86 and person reliability = 0.80. The result of separation was 2.00 indicating that the students can be divided into two groups.

Figure 2 shows the summary statistics for the 34 items involved in this study. "Items" represents the questions tested on the final examination. The item summary summarises very high reliability of 0.89 and item separation = 2.79. This means the items can be divided into 3 groups. The value for the mean item is 0.

Fig. 3 shows the item dimensionality for Advanced Programming subject. The raw variance explained by measures from Fig. 4 is 39.27%. This value fulfill the need for unexplained raw variance more than 20% [5]. There is a sufficient reliability in measuring problem solving skills in Advanced Programming subject. The measured noise level for item or known as unexplained variance in 1st contrast is 6.8% whereby the value is in a good control as it is within the range of 5% - 15% [6] and [7]. These values show that the items or the final exam questions are unidimensionality.

Fig. 4 shows the item statistics for Advanced Programming subject. There are 3 steps of examination is done in order to find the misfit item. Thery are Point-Measure Correlation, the Outfit MNSQ and Outfit z-Standard. The standard range for Point-Measure Correlation is [8] and [9]. Item 1ai, item 1aiii, item 1bi, item 1bii, item 1biii, item 1biv, item 1bvi, item 2b, item 2d, item 2e, item 3ci, item 3eii, item 4ai, item 4aii, item 4di and item 4dii are out of Point-Measure Correlation range. Other items are within the range.

The range for Outfit MNSQ is [10]. Item 1aiii, item 1bi, item 2b, item 3ci, item 3ei, item 3eii, item 4ai and item 4dii are out of range of Outfit MNSQ. Other items are within the range. The range for Outfit z-Standard is [6]. Only item 3ci and item 4dii are out of the range of Outfit z-Standard. Other items are within the range. Since item 3ci and item 4dii is out of range of Point-Measure Correlation, Outfit MNSQ and Outfit z-Standard therefore item 3ci and item 4dii are misfit items for the final examination.

Fig. 5 shows the item distribution map for the Advanced Programming subject. From the PIDM, the items can be divided into 3 groups. They are difficult, mediocre and easy. Fig. 5 shows the classification of items into groups. Item 3ci, item 4c, item 1ai, item 4f and item 2c belongs to difficult group. There are 14

items in mediocre category. They are item 2a, item 3ai, item 3aii, item 3b, item 1bv, item 2f, item 3aiii, item 4di, item 2d, item 3cii, item 3d, item 1c, item 4aii and item 1bvii. The easy group comprises of 15 items. They are item 1bvi, item 4b, item 4dii, item 1bvi, item 2e, item 4e, item 1aii, item 3ei, item 4ai, item 1biii, item 3eii, item 1bi, item 1aiii, item 1bii and item 2b. Table III summarises the results of item distribution map from Fig. 5.

Question 3ci was difficult as the students do not understand the need of the question. Hence the question can de rephrase as' Define the following: i. throw'. Question 4c was difficult for students as they could not manage to find the need of the question. The question can be written as 'i. Identify the use of method A ii. Identify the use of method B'. Question 1.a.i was difficult for students as they might be confused on how to answer the question. The question was 'What is the return type of a constructor?'. The answer is 'There is no return type of a constructor'. Perhaps students was thinking too long for an answer for question.

SUMMARY OF 60 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	131.8	33.5	.51	.14	1.00	.02	1.07	.14
SEM	2.7	.1	.04	.00	.04	.14	.14	.13
P.SD	20.8	.5	.34	.03	.31	1.07	1.11	.98
S.SD	20.9	.5	.34	.03	.31	1.08	1.12	.98
MAX.	161.0	34.0	1.11	.21	1.93	3.83	8.42	3.78
MIN.	79.0	33.0	-.27	.11	.43	-2.88	.31	-1.45

REAL RMSE	.15	TRUE SD	.30	SEPARATION	2.00	Person RELIABILITY .80
MODEL RMSE	.14	TRUE SD	.31	SEPARATION	2.15	Person RELIABILITY .82
S.E. OF Person MEAN = .04						

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .86 SEM = 7.89

Figure 1. Summary statistics for person

SUMMARY OF 34 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	232.6	59.1	.00	.13	1.07	.10	1.07	.18
SEM	9.1	.9	.09	.01	.05	.17	.10	.18
P.SD	52.1	5.1	.52	.08	.29	.97	.60	1.05
S.SD	52.9	5.1	.53	.08	.30	.99	.60	1.07
MAX.	297.0	60.0	.97	.42	1.81	2.12	2.82	3.40
MIN.	89.0	30.0	-1.20	.08	.46	-2.62	.34	-1.64

REAL RMSE	.18	TRUE SD	.49	SEPARATION	2.79	Item RELIABILITY .89
MODEL RMSE	.15	TRUE SD	.50	SEPARATION	3.24	Item RELIABILITY .91
S.E. OF Item MEAN = .09						

Item RAW SCORE-TO-MEASURE CORRELATION = -.89
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 2. Summary statistics for item

```
Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units
                                          Eigenvalue   Observed   Expected
Total raw variance in observations     =     55.8899 100.0%         100.0%
  Raw variance explained by measures   =     21.8899  39.2%          39.5%
    Raw variance explained by persons  =      7.8052  14.0%          14.1%
    Raw Variance explained by items    =     14.0848  25.2%          25.4%
  Raw unexplained variance (total)     =     34.0000  60.8% 100.0%   60.5%
    Unexplned variance in 1st contrast =      3.8229   6.8%  11.2%
    Unexplned variance in 2nd contrast =      2.8620   5.1%   8.4%
    Unexplned variance in 3rd contrast =      2.5982   4.6%   7.6%
    Unexplned variance in 4th contrast =      2.5580   4.6%   7.5%
    Unexplned variance in 5th contrast =      2.0587   3.7%   6.1%
```

Figure 3. Item dimensionality

```
Item STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL          MODEL|   INFIT  |  OUTFIT  |PTMEASUR-AL|EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+----------+-----------+------|
|    22    115     60      .97     .09|1.17   .99|2.58  3.40|  .16   .39| 23.3  34.6| 3ci  |
|    30    124     60      .89     .09| .72 -1.95| .60 -1.47|  .58   .41| 33.3  29.1| 4c   |
|    34    164     60      .62     .08|1.02   .21|1.08   .47|  .44   .48| 15.0  11.1| 4f   |
|     1     89     30      .59     .11| .98  -.05|1.13   .63|  .07   .48| 26.7   8.3| 1ai  |
|    14    178     60      .53     .08| .90  -.74| .97  -.11|  .42   .50| 10.0  11.1| 2c   |
|    12    192     60      .44     .08| .97  -.16| .88  -.61|  .62   .50|  8.3  12.4| 2a   |
|    18    204     60      .36     .08|1.00   .07| .97  -.11|  .50   .50| 10.0  12.1| 3ai  |
|    21    208     60      .34     .08| .92  -.49| .87  -.63|  .52   .50| 13.3  16.0| 3b   |
|    19    212     60      .31     .08| .83 -1.12| .86  -.63|  .54   .50| 26.7  14.9| 3aii |
|    20    214     60      .30     .08| .75 -1.80| .68 -1.64|  .67   .50|  8.3  15.1| 3aiii|
|    31    214     60      .30     .08|1.25  1.56|1.28  1.27|  .39   .50|  8.3  15.1| 4di  |
|    17    218     60      .27     .08| .99  -.01| .98  -.03|  .53   .50| 21.7  18.9| 2f   |
|     8    222     60      .24     .09| .96  -.19|1.17   .78|  .44   .49| 31.7  22.8| 1bv  |
|    15    224     60      .22     .09|1.09   .60|1.14   .64|  .35   .49| 25.0  25.5| 2d   |
|    23    224     60      .22     .09|1.08   .53| .91  -.31|  .54   .49| 26.7  25.5| 3cii |
|    24    228     60      .19     .09| .99  -.02| .86  -.49|  .58   .49| 26.7  28.4| 3d   |
|    28    234     60      .15     .09|1.18  1.03|1.20   .79|  .38   .48| 33.3  33.9| 4aii |
|    11    238     60      .11     .09|1.01   .11| .91  -.23|  .55   .47| 35.0  34.1| 1c   |
|    10    244     60      .06     .09| .57 -2.62| .89  -.24|  .65   .46| 41.7  35.9| 1bvii|
|     7    256     60     -.05     .10|1.03   .23|1.12   .42|  .26   .43| 50.0  50.7| 1biv |
|    29    258     60     -.07     .10|1.05   .30| .88  -.18|  .50   .42| 51.7  51.2| 4b   |
|    32    260     60     -.10     .11|1.56  2.12|2.53  2.75|  .11   .41| 50.0  53.8| 4dii |
|     9    264     60     -.14     .11|1.20   .82| .86  -.18|  .39   .40| 55.0  54.5| 1bvi |
|    16    264     60     -.14     .11| .89  -.37|1.27   .71|  .27   .40| 48.3  54.5| 2e   |
|    33    268     60     -.19     .12|1.16   .65|1.33   .78|  .41   .38| 61.7  60.1| 4e   |
|     2    272     60     -.25     .12|1.04   .23| .73  -.39|  .50   .36| 68.3  64.6| 1aii |
|    25    280     60     -.39     .14| .64 -1.00| .34 -1.23|  .55   .31| 76.7  73.9| 3ei  |
|    27    284     60     -.48     .16|1.52  1.20|2.21  1.54|  .15   .28| 83.3  80.1| 4ai  |
|     6    287     60     -.56     .18| .98   .13| .56  -.44|  .29   .25| 83.3  86.2| 1biii|
|    26    290     60     -.66     .20|1.36   .75|2.82  1.74|  .12   .22| 91.7  88.7| 3eii |
|     4    292     60     -.76     .23|1.39   .75| .40  -.55|  .36   .19| 93.3  89.7| 1bi  |
|     3    296     60    -1.06     .35|1.74   .94| .40  -.29|  .28   .13| 98.3  96.0| 1aiii|
|     5    296     60    -1.06     .35|1.81   .99| .62   .00|  .22   .13| 98.3  96.0| 1bii |
|    13    297     60    -1.20     .42| .46  -.20| .46  -.13|  .22   .11| 95.0  96.7| 2b   |
|------------------------------------+----------+----------+----------+-----------+------|
| MEAN   232.6   59.1     .00     .13|1.07    .1|1.07    .2|           | 45.0  44.2|      |
| P.SD    52.1    5.1     .52     .08| .29   1.0| .60   1.0|           | 29.4  29.0|      |
-------------------------------------------------------------------------------------------------
```

Figure 4. Item statistics

Question 4f was difficult as it required to explain the characteristics of pro-activeness for a software agent. This question was difficult for students as they cannot recall the characteristics of the pro-activeness. Question 2c required to identify what is wrong in the class definitions. The answer was 'Class BClass extends AClass'. Students have to add the word 'extends' between the Blass and AClass. Perhaps the students could not recall the word.

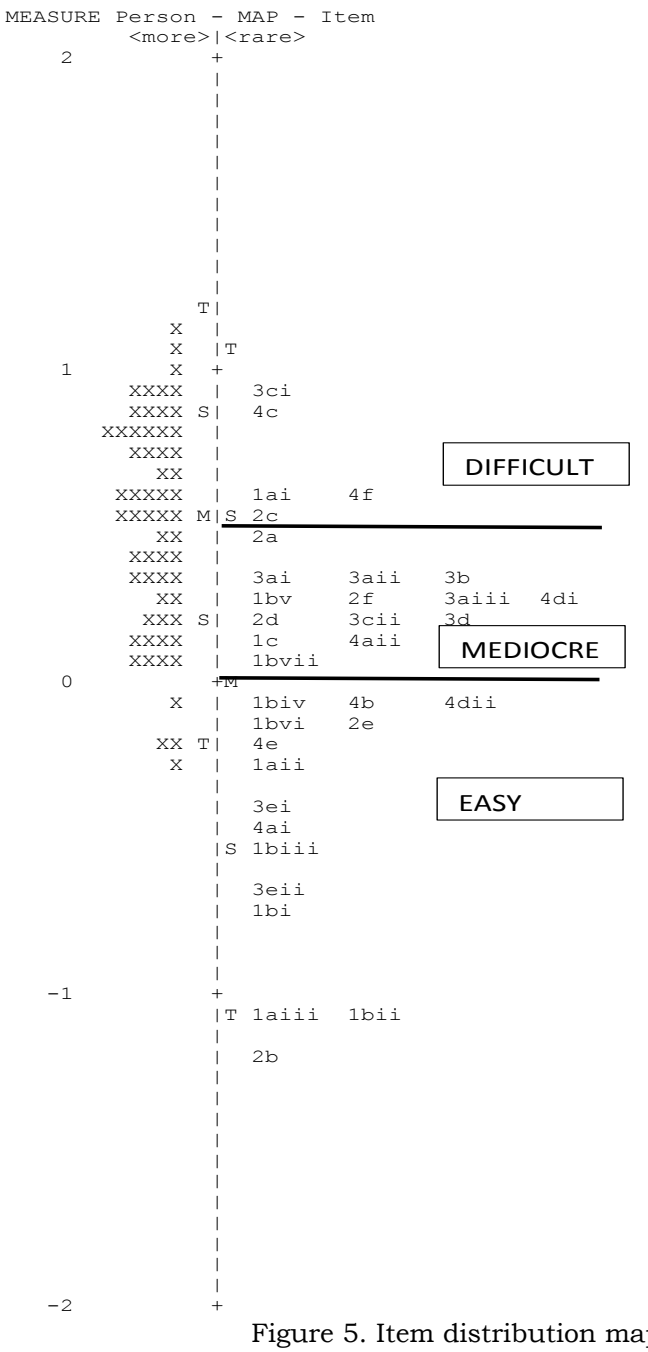

Figure 5. Item distribution map

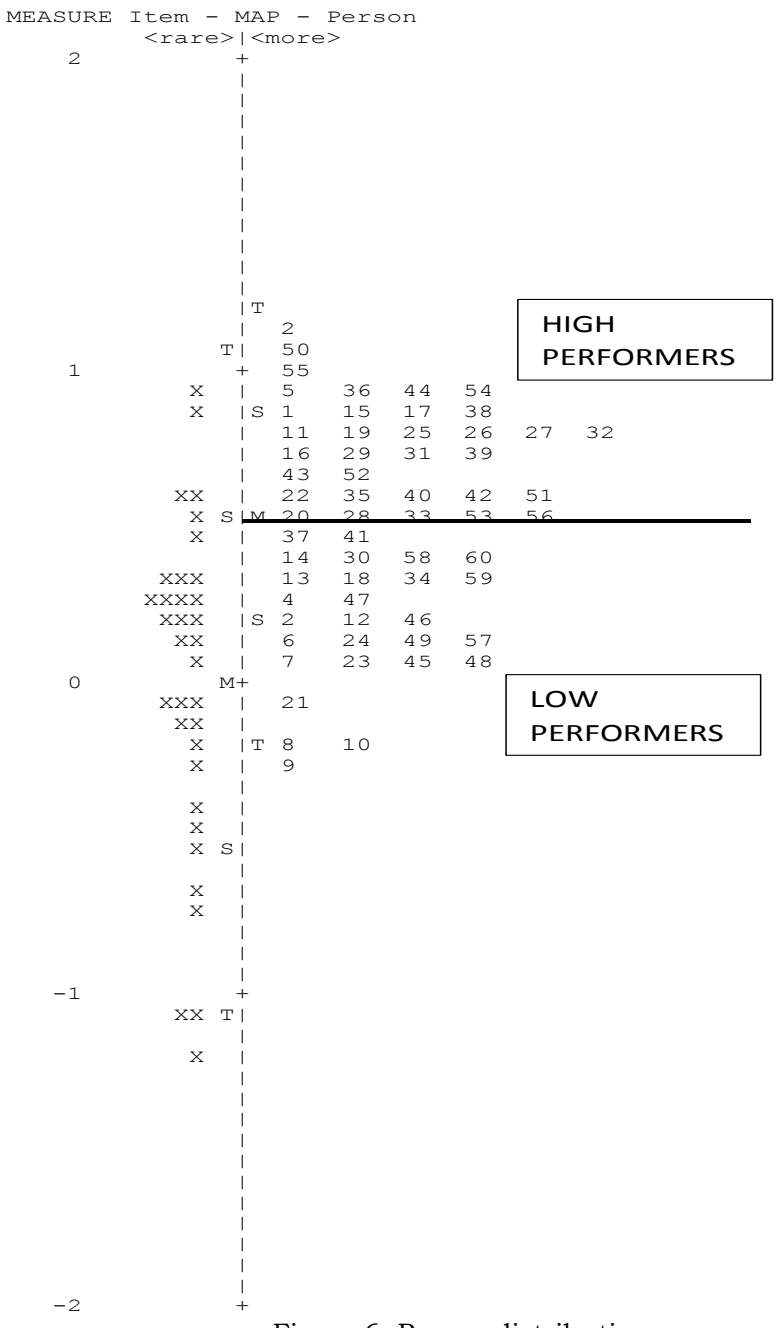

Figure 6. Person distribution map

The students performance divided into 2 groups. Fig. 6 shows the separation of the students. There are more high performers compared to low performers. This means the final exam questions are considered answerable for the students.

Table 3
Summary of the item distribution map for advanced programming subject

Question	Course Outcome	Bloom Taxonomy Level	Summary from PIDM
1.a.i.	CO1	Knowledge	Difficult
1.a.ii.	CO1	Comprehension	Easy
1.a.iii.	CO1	Application	Easy
1.b.i.	CO1	Knowledge	Easy
1.b.ii.	CO1	Comprehension	Easy
1.b.iii.	CO1	Comprehension	Easy
1.b.iv.	CO1	Comprehension	Easy
1.b.v.	CO1	Comprehension	Mediocre
1.b.vi.	CO1	Application	Easy
1.b.vii.	CO1	Application	Mediocre
1.c.	CO1	Knowledge	Mediocre
2.a.	CO1	Application	Mediocre
2.b.	CO1	Analysis	Easy
2.c.	CO1	Analysis	Difficult
2.d.	CO2	Application	Mediocre
2.e.	CO1	Application	Easy
2.f.	CO1	Application	Mediocre
3.a.i.	CO1	Application	Mediocre
3.a.ii.	CO1	Analysis	Mediocre
3.a.iii.	CO1	Comprehension	Mediocre
3.b.	CO1	Synthesis	Mediocre
3.c.i.	CO1	Knowledge	Difficult
3.c.ii.	CO1	Knowledge	Mediocre
3.d.	CO1	Analysis	Mediocre
3.e.i.	CO1	Synthesis	Easy
3.e.ii.	CO1	Synthesis	Easy
4.a.i.	CO2	Knowledge	Easy
4.a.ii.	CO2	Knowledge	Mediocre
4.b.	CO2	Application	Easy
4.c.	CO2	Synthesis	Difficult
4.d.i.	CO3	Synthesis	Mediocre
4.d.ii.	CO3	Synthesis	Easy
4.e.	CO3	Comprehension	Easy
4.f.	CO4	Comprehension	Difficult

Conclusion

Rasch model able to evaluate the reliability of Advanced Programming exam questions. The summary statistics for person, summary statistics for item, item dimensionality, item statistics, item distribution map and person distribution map are the outputs from rasch model. From item statistics, item 3cii and item 4dii are identified as misfit items. The item dimensionality shows that the final exam questions is unidimensional. The item distribution map showed that questions can be group into 3 groups, namely difficult, mediocre and easy. The person distribution map classified the person into 2 groups, namely high

performers and low performers. The final exam questions for Advanced Programming is very reliable.

Acknowledgments

The authors wish to express gratitude towards XIAMEN University Malaysia (XMUMRF/2022-C9/IECE/0026), Universiti Kebangsaan Malaysia (GUP-2018-151) and SEGi University (SEGiIRF/2018-14/FOEBE-21/84) for supporting the research.

References

1. R.T. Aliyu, “Construct validity of mathematics test items using the Rasch model”, International Journal of Social and Humanities Rewsearch”, 3(2), pp. 22-28, 2015.
2. M. Garner, M. Burke and J. Binongo, J. “Analysis of a college placement test in mathematics using the Rasch measurement model”, Electronic sources: http://digitalmeasures.kemesaw.edu, 2016.
3. B. Wright and G.N. Masters, “Rating scale analysis”, Chicago : Mesa Press,1982.
4. T.G. Bond and C. M. Fox, “Applying the Rasch model: Fundamental measurement in the human sciences”, 2nd Edition, New Jersey: Lawrence Elbaum Inc. Publishers, 2007
5. M.D. Reckrase, “Unifactor latenttrait models applied to multifactors tests: Resultsand implications”, Journal of Educational Statistics, 4, pp. 207-230, 1979.
6. J.W.P. Fisher, “Rating scale instrument quality criteria”, Rasch Measurement Transcation, 21(1), pp. 1095, 2007.
7. [10] A.M. Eakman, “Measurement characteristics of the engagement in meaningful activities survey in an age-diverse sample”, American Journal of Occupational Therapy, 66(2), pp. 20-29, 2012.
8. A.A. Azrilah, N. Azlinah, H.A. Noor, A.G. Hamzah, Z. Sohaimi and M. Saifudin, “Application of Rasch model in validating the construct measurement instrument”, International Journal of Education and Information Technologies, 2, pp. 105-112, 2008.
9. A.A. Azrilah, S.M. Mohd and Z. Azami, “Asas model pengukuran Rasch, Bangi, Penerbit Universiti Kebangsaan Malaysia, 2013.
10. J.M. Linacre, A user’s guide to WINSTEPS, Chicago: winsteps.com, 2005.