Journal of Applied Sciences Research, 8(3): 1797-1801, 2012
ISSN 1819-544X
This is a refereed journal and all articles are professionally screened and reviewed

ORIGINAL ARTICLES

Rasch Modeling Analysis in Assessing Student's Ability and Questions Reliability in Architecture Environmental Science Examination

Nik Lukman Nik Ibrahim, Nangkula Utaberta and Badiossadat Hassanpour

Department of Architecture, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia.

ABSTRACT

Assessment and measurement in teaching and learning process is one of the most vital processes in order to enable the continuous quality improvement. However, the over use of Cumulative Grade Point Average (CGPA) which is a computation process is purely the mean of raw scores, lacks precision and linearity hence validity required to meet the fundamental criteria of measurement warrants a review. Rasch modeling was used to assess students' ability and the quality of questions in the Environmental Science examination at an architectural school. The modeling demonstrates that the questions are reliable tool for assessing the students. However, more than half of the questions are beyond the average capability of the students.

Key words: Rash Modeling, Architecture Course, Examination.

Introduction

Some believe that when students are taught with good learning style, the student's academic performance will improve significantly. Learning is a complex process. It entails not only what students know but what they can do with, what they know; it involves not only knowledge and abilities but values, attitudes, and habits of mind that affect both academic success and performance beyond the classroom. According to Hussein (Hussein Mahmood. 1993), academic excellence is student's achievements which are based on university's assessments such as test, assignment, presentation, final exam and etc. Assessment should reflect these understandings by employing a diverse array of methods, including those that call for actual performance, using them over time so as to reveal change, growth, and increasing degrees of integration. Such an approach aims for a more complete and accurate picture of learning, and therefore firmer bases for improving our students' educational experience (Astin. A.W, *et al*., 2005).

Responses from the students in an examination, test or quizzes is normally marked against a marking scheme comprising keywords; where when there is a match then the student would be given a mark or otherwise. This is the traditional mark system. In theory, at this stage truly the assessors are only counting the number of correct answers which is then added up to give a total raw score. The raw score only give a ranking order which is deemed an ordinal scale that is continuum in nature (Sick, J., 1999).

It is not linear and do not have equal intervals which contradicts the nature of data fit for the due statistical analysis. It does not meet the fundamentals of sufficient statistics for evaluation (Wright B.D., 1989). Present students' evaluation measurement generally practiced in Institutions of Higher Learning (IHL) is largely dependent on students' performance in carrying out tasks such as a series of tests or quizzes, final examination and submission of assignments. Analysis must be based on valid data and suitably interpreted to generate a reliable report with meaningful information for prudent decision making towards continuous improvement of teaching and learning. And better management to improve their achievement in meeting the targeted learning outcomes.

A good assessment recognizes the value of information for the process of improvement. Assessment approaches should produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made The point of assessment is not to gather data and return "results"; it is a process that starts with the questions of decision-maker that involves them in the data gathering and subsequent analysis;

1. How do you assure the correct instrument is used for purpose? And subsequently;
2. What is the correct method of such data analysis?

It is of utmost importance on the onset these fundamentals of measurement must be correct. (Saidfudin, M, *et al*. 2007).

Corresponding author: Dr. Nangkula Utabera, Department of Architecture, Faculty of Engineering and Built Environment, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.
E-mail: nangkula_arch@yahoo.com, Ph: 60192423816

In IHL, the theory and practice of classical test theory, the traditional approach of students' evaluation must be re-assessed (Saidfudin, M, *et al*., 2009).

Lack of such measurement in architecture education has made the necessary corrective action in the form of skills development, education and competency training different to framing (Saidfudin, M, *et al*., 2009).

The development of Rasch Measurement Model in social science educational measurement has rapidly expanded to other areas of education including technical and engineering fields. And the problem can be solved with use of Rasch measurement model in architecture too. Rasch moves the concept of reliability from establishing 'best fit line' of the data into producing reliable repeatable measurement instrument (Azrila, A, *et al*.2008) This measurement model uses empirical data directly from the lecturer's assessment on student for a given task and transformed them into logic scale which have equal interval (Saidfudin, M, *et al*., 2008).

Rash modeling was conducted to check students' ability in answering the Environmental Science course examination questions at the Department of Architecture, Faculty of Engineering and Built Environment, National University of Malaysia (UKM). The candidates consist of 20 first year architectural students who took the subject (code: KKSB1263) and sat for the final semester examination in April 2011. The examination questions consist of two major parts; Part A and Part B, each has three main questions from which a student has to choose two. Part A provides questions on climatic factors influencing architecture, thermal comfort, green buildings and the Environmental Impact Assessment (EIA). Part B of the examination covers questions on daylighting and natural ventilation in building, passive cooling and solar shading strategies. The results of the examination serve as an input for Rasch Modeling to test students' capability and the exam questions reliability as an assessment tool.

Descriptions Of The Examination Question:

Part A of the examination consists of Question 1, 2 and 3. Question 1 tests student on his or her knowledge about the world major climatic zones, instruments for weather forecast, definition of thermal comfort and factors influencing it. In Question 2, knowledge on factors influencing room temperature and relative humidity are tested as well as knowledge on human metabolic heat gain. Question 3 tests students on the purpose of the Environmental Impact Assessment (EIA) and the criteria of a green building. In general, all questions in Part A can be categorized as knowledge acquisition in Bloom Taxonomy (Woolfolk, A.E., 1995).

Part B of the examination consists of Questions 4, 5 and 6. Question 4 test students on their knowledge about daylight factor (DF) and methods of calculation if a scenario is given. Question 5 requires students to calculate sun paths and the dimensions of two shading devices given several scenarios for them to consider. Question 6 tests students on their knowledge about natural ventilation strategies as well as methods of calculation given certain scenarios. Question 4, 5 and 6 therefore, can be classified as application knowledge in Bloom Taxonomy (Woolfolk, A.E., 1995).

Table 1: The examination questions and categories in Bloom Taxonomy.

Part A	Bloom Taxonomy
Q1	Knowledge Acquisition
Q2	Knowledge Acquisition
Q3	Knowledge Acquisition
Part B	Bloom Taxonomy
Q4	Knowledge Application
Q5	Knowledge Application
Q6	Knowledge Application

Rasch Modeling: Fit Statistic Calculation:

Rasch Modeling of the students' examination results generates the following output summary or fit statistic calculations (Figure 1).

From the output summary or fit statistic calculations above, the Item separation of 2.33 shows that there are two categories of questions; those which are easy for the students to answer and those which are difficult. Separation of 1.53 under the Persons list indicates that there are two categories of students; those who are capable and those who are relatively weak. Persons Reliability of 0.7 shows that generally most students could answer the questions correctly. Reliability of 0.84 under the Item list also indicates that the examination questions are the right instrument for testing the students' ability.

```
 Calculating Fit Statistics
>==================================<
Standardized Residuals N(0,1)  Mean: .00 S.D.: .99
ENVSC
-------------------------------------------------------------------------------
| Persons      20 INPUT      20 MEASURED              INFIT         OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR     IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN       9.1       3.8       -1.59     .89       .94    -.1    .98     .0|
| S.D.       3.1        .5        1.67     .21       .65    1.1    .71    1.0|
| REAL RMSE    .91  ADJ.SD    1.40  SEPARATION  1.53  Person RELIABILITY  .70|
|-----------------------------------------------------------------------------|
| Items       6 INPUT       6 MEASURED               INFIT         OUTFIT    |
| MEAN      30.3      12.7         .00     .45       .92    -.1    .95     .0|
| S.D.      10.5       2.9        1.16     .06       .23     .6    .20     .5|
| REAL RMSE    .46  ADJ.SD    1.07  SEPARATION  2.33  Item   RELIABILITY  .84|
-------------------------------------------------------------------------------
Output written to C:\Users\sony\Desktop\ZOU805WS.TXT
CODES= 12345
Measures constructed: use "Diagnosis" and "Output Tables" menus
```

Fig. 1: Fit Statistic Calculation.

Rasch Modeling: Item Measurement:

From the Item Measure summary below, Question 4 and Question 6 are the easiest questions in the examinations. However, Question 6 was not attempted by two highest students (see figure 2). Questions 5, 2, 1 and 3 are the more difficult questions in this hierarchical order.

```
INPUT: 20 Persons  6 Items  MEASURED: 20 Persons  6 Items  5 CATS         3.68.2
-------------------------------------------------------------------------------

GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |465213
       |------
     8 +4 53 4  20
     7 +5 443   20
    10 +33  33  19
     9 +34 3 1  19
    11 +33 3 2  20
    14 + 42 22  21
    12 +4 3 21  20
    15 +3 22 3  20
     3 +43  21  20
    13 + 22 32  21
     2 +33  21  19
    17 +3 2     19
     4 +4 2 11  21
    18 +4  1 1  21
     1 + 3131   21
     6 +1 223   19
    20 +3 211   20
     5 +2 1 21  20
    19 +3 111   21
    16 +1 2 1   20
       |------
       |465213
```

Fig. 2: Guttman Scalogram.

It is also interesting to notice that Question 4 which is the easiest had been attempted by most students as they had became familiar with it when they did day lighting experiments in the laboratory for one of their small projects. This seems to support the principle of 'learning by doing' as an effective mode of education (McInerney, *et al*., 1994).

Point-Measure Correlations in Figure 3 are all in positive values, therefore the quality of the examination questions are generally good and are suitable instruments for testing the students' capability.

```
INPUT: 20 Persons  6 Items  MEASURED: 20 Persons  6 Items  5 CATS         3.68.2
-------------------------------------------------------------------------------
Person: REAL SEP.: 1.53  REL.: .70 ... Item: REAL SEP.: 2.33  REL.: .84

         Item STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------
|ENTRY   TOTAL                    MODEL|    INFIT    |   OUTFIT    |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE   S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|-----------------------------------+----------+----------+-----------+-----------+------|
|     3     23     13     1.63    .42| .86   -.3| .77   -.5|  .79   .65| 53.8  52.9| I0003|
|     1     27     14      .68    .41| .92   -.1| .93    .0|  .65   .71| 57.1  55.1| I0001|
|     2     23     10      .49    .47| .91   -.1| .88   -.1|  .77   .79| 50.0  50.4| I0002|
|     5     31     14      .23    .40| .63  -1.1| .89   -.2|  .87   .78| 64.3  55.3| I0005|
|     6     25      8    -1.35    .57| .84   -.1| .87   -.1|  .31   .34| 62.5  60.5| I0006|
|     4     53     17    -1.68    .38|1.38   1.1|1.39   1.1|  .66   .74| 47.1  56.9| I0004|
|-----------------------------------+----------+----------+-----------+-----------+------|
| MEAN    30.3   12.7      .00    .44| .92   -.1| .95    .0|           | 55.8  55.2|      |
| S.D.    10.5    2.9     1.16    .06| .23    .6| .20    .5|           |  6.2   3.1|      |
-------------------------------------------------------------------------------------------
```

Fig. 3: Item Statistic.

INPUT: 20 Persons 6 Items MEASURED: 20 Persons 6 Items 5 CATS 3.68.2

Person: REAL SEP.: 1.53 REL.: .70 ... Item: REAL SEP.: 2.33 REL.: .84

Person STATISTICS: MEASURE 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
8	16	4	2.55	.86	1.82	1.2	1.81	1.2	.08	.57	.0	60.3	20
7	16	4	2.29	.85	.57	-.5	.58	-.5	.88	.48	75.0	60.5	20
10	12	4	-.38	.75	.76	-.2	.74	-.2	.00	.66	50.0	51.1	19
9	11	4	-.98	.74	.92	.1	.92	.1	.82	.68	25.0	48.7	19
11	11	4	-.98	.74	.28	-1.4	.28	-1.3	.79	.68	75.0	48.7	20
14	10	4	-1.05	.72	.55	-.6	.64	-.4	.88	.59	50.0	50.1	21
12	10	4	-1.11	.71	.69	-.3	.73	-.3	.96	.62	50.0	48.7	20
15	10	4	-1.17	.72	1.07	.3	1.02	.3	.16	.62	50.0	48.1	20
3	10	4	-1.48	.75	.57	-.5	.63	-.4	.97	.71	50.0	52.9	20
13	9	4	-1.56	.72	1.16	.5	1.25	.6	-.20	.61	50.0	47.6	21
2	9	4	-2.05	.78	.13	-1.8	.14	-1.6	.98	.73	100.0	60.1	19
17	5	2	-2.07	1.02	.00	-2.6	.00	-2.4	1.00	.56	100.0	55.7	19
4	8	4	-2.16	.75	1.00	.2	1.07	.3	.96	.66	50.0	55.7	21
18	6	3	-2.25	.91	1.57	.9	1.49	.8	.94	.72	33.3	60.0	21
1	8	4	-2.33	.71	1.43	.8	1.38	.8	.55	.52	25.0	47.3	21
6	8	4	-2.40	.72	2.81	2.2	3.12	2.4	-.88	.58	50.0	47.6	19
20	7	4	-2.96	.78	.44	-.8	.45	-.8	.93	.61	100.0	54.5	20
5	6	4	-3.50	.91	.68	-.2	.81	.2	.59	.64	75.0	69.3	20
19	6	4	-3.64	.88	.63	-.3	.48	-.4	.99	.59	75.0	67.8	21
16	4	3	-4.47	1.15	1.73	.9	2.08	1.1	-.34	.51	33.3	73.2	20
MEAN	9.1	3.8	-1.59	.81	.94	-.1	.98	.0			55.8	55.4	
S.D.	3.1	.5	1.67	.11	.65	1.1	.71	1.0			26.1	7.7	

Fig. 4: Person Statistic.

The Scalogram in Figure 5 demonstrates that most of the examination questions are beyond the average capability of the students. Most of the students could answer Question 4 and 6 very well but are less capable in answering Questions 5, 2, 1 and 3.

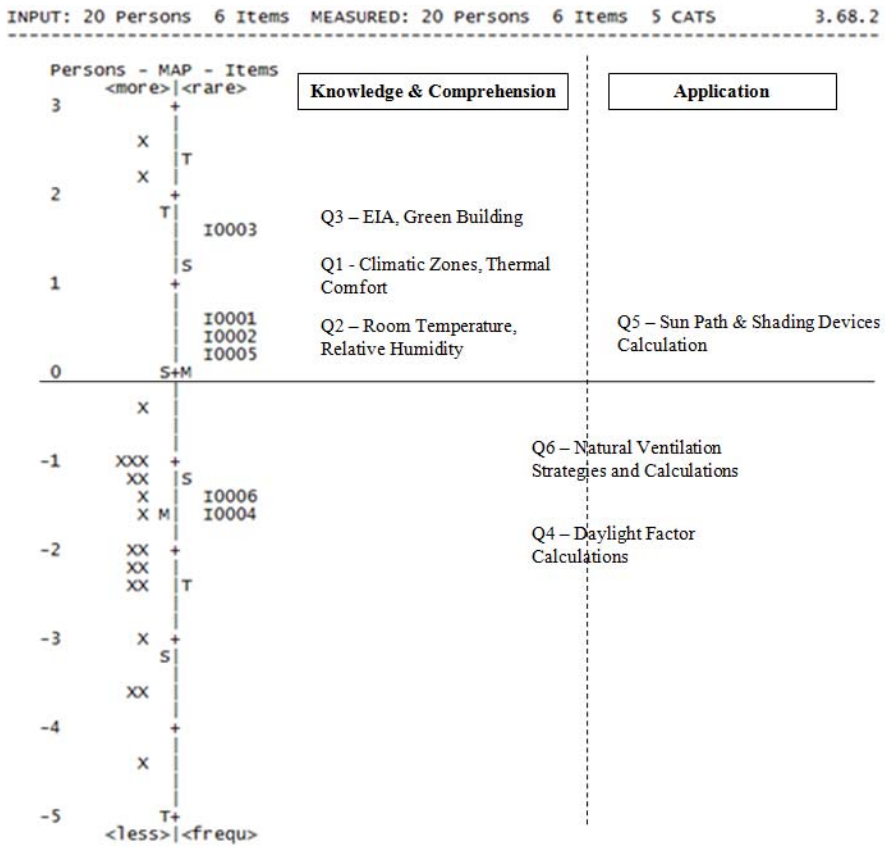

Fig. 5: Scalogram.

However, two of the students have demonstrated capabilities which are beyond average and their capacities even surpass the most difficult question in the examination. The Scalogram also indicates that the easiest questions are the ones which deal with the application of knowledge.

J. Appl. Sci. Res., 8(3): 1797-1801, 2012

Conclusions:

The Rasch modeling conducted proves that the examination questions utilized are reliable tool for assessing the group of students who took the Environmental Science course at the Department of Architecture, UKM. However, most of the questions are rated 'difficult' and exceed the capability of the average students in this group. This indication suggests that examination questions should be revised in the future to ensure that they really match with the students' capabilities. Beside this, architecture is a multidisciplinary field and different types of courses from theory to practice are in universities curriculum to prepare students to enter to practical world. But what is important here is the evaluation and ponder of assessment models and their related criteria in theory based course's exams which are multiple choices or open ended questions and they are more simple and accessible. In fact investigation and evaluation of assessment in practical courses, criteria, objectives and effective rating in them are more complicated and more desired. Revisiting current assessment models with Rasch measurement model and representing new framework for grading and marking sessions is ongoing project in UKM (Universiti Kebangsaan Malaysia).

References

Astin, A.W., *et al*., 2005. 9 Principles of Good Practice for Assessing Student Learning, The American Association for Higher Education, Stylus Publishing, LLC.

Azrila, A.A., M. Saidfudin, 2008. Application of Rasch model in validating the construct of measurement instrument.

Hussein Mahmood, 1993. Kepimpinan dan Keberkesanan sekolah. DewanBahasa dan Pustaka, Kuala Lumpur.

McInerney & McInerney, 1994. *Educational Psychology: Constructing Learning*. Prentice Hall, Sydney.

Saidfudin, M., M. Azlinah, A.A. Azrilah, A. Nor Habibah, & Z. Sohaimi, 2007. Appraisal of Course Learning Outcomes using Rasch measurement: A case study in Information Technology Education, International Journal of Systems Applications, Engineering & Development; University Press, UK., 4(1): 164-172.

Saidfudin, M., Gulman, 2009. Measurement Paradigm in engineering education:Easier to read and better analysis using Rasch-based Approach, ICEED conference.

Saidfudin, M., Gulman, 2008. Application of Rasch-based ESPEGS model in measuring generic skills of engineering students, WSEAS conference.

Sick, J., 1999. Rasch Measurement in Language Education Part 3: The family of Rasch Models, Shiken: JALT Testing & Evaluation SIG Newsletter.

Woolfolk, A.E., 1995. Educational *Psychology* (6^{th} edition). Allyn & Bacon, Boston.

Wright, B.D., 1989. Rasch Model from Counting Right Answers: Raw Scores as Sufficient Statistics. Rasch Measurement Transactions, 3:2 pp: 62.