

Validity, Reliability and Item Analysis of AMAIUB Admission Test

Dr. Lina S. Calucag¹ and Dr. Danilo A. Tabalan²

Mathematics & Science Department, College of Engineering, AMA International University – Bahrain¹

Social Science Department, Center for General Education, AMA International University – Bahrain²

Abstract: The main objective of the study was to determine the validity, reliability and item analysis of the AMAIUB Admission test. A multiple-choice test of 150 items was finalized after pilot study. The test was administered to a sample of incoming college freshmen. Separate sheets were provided to students included in the sample. Several work sheets were prepared to perform item analysis. Classical test theory method was used in the analyses of the data. Item analysis, difficulty index, and discriminatory index were calculated. The reliability coefficient was 0.86 and 0.82 for English and Mathematics tests, respectively indicating high reliability of the whole instrument.

Keywords: development, validation, admission test, item analysis.

1. INTRODUCTION

College admissions tests also called college entrance exams, are designed to measure students' skills and help colleges evaluate how ready students are for college-level work. Admission tests apply a common standard to everyone. This helps colleges evaluate and compare the preparation of students who go to different high schools. Not all schools offer the same academic programs, learning environments or even expectations. Colleges look at your test scores, along with your high school grades and courses, to see how well prepared you are for college-level work, (Brenzel, 2015).

There is a need to standardize college entrance exams, (Peterson, 2013). Not all high schools have the same standards for giving out grades. Schools vary, countries vary, and states vary. Therefore, a standardized gauge – a college entrance exam – is necessary to measure critical reading, thinking and the problem solving skills that are necessary for college, (Harriet, 2014).

Researchers must demonstrate instruments that are reliable since without reliability research results using the instrument are not reliable, and reliability is fundamental to the scientific method. Reliability is the correlation of an item, scale, or instrument with a hypothetical one that truly measures what it is supposed to measure. The AMA Admission Test measures achievement in the core curriculum areas and is based on academic knowledge and skills typically taught in high school college-preparatory courses.

2. OBJECTIVES OF THE STUDY

The study developed a valid and reliable instrument that could be used as a measure to assess high school graduates who want to earn college degree in AMA International University-Bahrain. The study tried to come up with items that will measure the aptitude of the students along Mathematics and English, determined the index of difficulty and discrimination of the items and reliability of the developed aptitude test. The specific objectives of this study were to:

1. Validate items through expert judgment.
2. Pilot test items.
3. Use classical test theory and item response theory to analyze items based on statistical measures providing further evidences of validity and reliability.

2.1 Item validation

Writing the test item does not produce an item ready to be tested until it is validated (Rivera, 2007). Messick (1989) as cited by Rivera (2007) emphasized the importance of item reviews, strengthening validity and Hambleton (1994) provided a summary of methods used to review and validate items. There have been multiple techniques established for reviewing item-objective congruence based on large-scale assessments to small classroom assessments, such as the use of empirical techniques similar to norm-referenced testing, expert judgment used to calculate the index of item objective congruence, a rating of item-objective match on a 5-point scale conducted by experts, and the use of a matching task (Hambleton, 1984).

2.2 Analyzing Item Responses

Common statistics are used to describe classical test theory parameters including p-values, item discrimination, point-biserial correlation coefficient, alpha coefficient, and variance (Rivera, 2007). Classical Test Theory statistics can also be used to determine values of reliability through the use of internal consistency methods such as split halves (Rulon, 1939; Guttman, 1945; Cronbach, 1951) and item co-variance. Kuder Richardson 20 (KR 20) and Kuder Richardson 21 (KR 21) (Kuder & Richardson, 1937), along with Cronbach's alpha (Cronbach, 1951), and Hoyt's method (Hoyt, 1941) are other methods used to circumscribe problems inherent in split half techniques Rivera (2007).

3. METHODOLOGY

The present AMA Admission Test was developed after an extensive review of the literature. Textbooks for secondary

level students were also studied and the concepts were drawn on the base of the books. After getting the concept from the books, various aptitude tests were also studied to construct the items. After examining the related literature, the researchers prepared the items for the aptitude test to be administered to the freshmen students. During the construction of the multiple choice items, the mental capacity of the students were considered. The researchers initially developed 150 test items under two areas – 100 items for English and 50 items for Mathematics. There were four to five options in each item. Incoming freshmen students to AMAIUB were selected as the population of this study. Sample for the study was selected based on simple random sampling technique. Student-participants were selected from the different degree programmes: Bachelor of Science in Computer Science, Bachelor of Science in Informatics Engineering, Bachelor of Science in Mechatronics Engineering, Bachelor of Science in Business Informatics and Bachelor of Science in International Studies.

The researchers personally administered the test. The instruction stated in the test was followed strictly and answer sheets were collected after the allotted time. The students were informed that the test result would not affect academic grades. They were asked to mark the correct answer on the answer sheet only which were separately provided to every students. Relevant information was collected through answer sheets and a master sheet was prepared in order to tabulate results. The test papers were checked, scored and arranged from highest to the lowest. Item analysis was employed to determine the difficulty index and discriminating index. Selection and improvement/ revision of the items needed for final form of the test was done by the use of U-L Index Method. Items that have levels of difficulty index ranging from 20%-70% were retained after the tryout test. Thus item with discrimination index of 0.20 and above were considered acceptable.

There were two sets of tabulation made for the data collected. One set was intended for internal consistency analysis. Actual responses for each item were recorded. Another set of tabulation was made by marking the correct responses with “1” while “0” for each incorrect response. The two major classical test theory properties were used in item analysis i.e. difficulty index and discrimination index. To establish reliability of the instrument, Kuder-Richardson was used. The binary coding (2nd set of data) was used for the analysis using Microsoft Excel. To determine the internal consistency of the instrument, the 1st set of data was used using SPSS.

The **Kuder and Richardson Formula 20** test checks the internal consistency of measurements with dichotomous choices. It is equivalent to performing the split half methodology on all combinations of questions and is applicable when each question is either right or wrong. A correct question scores 1 and an incorrect question scores 0.

Guttman split-half reliability coefficient is an adaptation of the Spearman-Brown coefficient, but one which does

not require equal variances between the two split forms. This is an alternative split-half model which computes Guttman's lower bounds for true reliability.

Alpha (Cronbach) This models internal consistency based on average correlation among items. Cronbach's alpha is superior to Kuder and Richardson Formula 20 since it can be used with continuous and non-dichotomous data.

4. RESULTS AND DISCUSSION

A. Difficulty Index

A.1. The value of difficulty index for all items lies between 0.16 and 0.73. Item number 42 had the largest value (0.73) and item number 29 has the smallest value (0.16). On the other hand, 45 and 46 had the greatest discrimination power (0.80) and item number 96 got the least discrimination power (0.15). The reliability coefficient for the whole instrument in English was calculated to be 0.86.

Table 1: Index Difficulty of English Items

Difficulty Index		English Test				Item Placement for the Revised Test
		Initial		Final		
		f	%	f	%	
0.86 - 1.0	Very easy	9	9	0	0	
0.71 - 0.85	Easy item	30	30	1	1	42
0.40 - 0.70	Moderately Difficult	54	54	72	72	1-5, 7, 9-12, 15-17, 21, 35, 36, 39-41, 43-69, 71, 74-85, 87, 88, 90-100
0.15 - 0.39	Difficult	7	7	27	27	6, 8, 13, 14, 18-20, 22-34, 37, 38, 70, 72, 73, 86, 95, 96, 98
0 - 0.14	Very Difficult	0	0	0	0	
	Total	100		100		

Item analysis in English AMAIUB Admission test shows that 72 or 72 percent of the prepared items had an average difficulty. Only 1 item or 1 percent was found to be easy while 27 or 27 percent of the items were difficult.

A.2. The value of difficulty index for all items lies between 0.19 and 0.63. Item number 1 had the largest value (0.63) and item number 16 had the smallest value (0.19). On the other hand, item number 22 got the greatest discrimination power (0.85) and item numbers 11, 13, 25 and 32 had the least discrimination power (0.22). The reliability coefficient for the whole instrument in Mathematics was calculated to be 0.82.

Table 2: Index of Difficulty of Math Items

Difficulty Index		Mathematics Test				Item Placement for the Revised Test
		Initial		Final		
		f	%	f	%	
0.86 - 1.0	Very easy	0	0	0	0	
0.71- 0.85	Easy item	0	0	0	0	
0.40- 0.70	Moderately Difficult	19	38	27	54	1, 3, 4, 6, 9, 11, 12, 14, 15, 17, 18, 22, 30-36, 38-41, 44-46, 50
0.15- 0.39	Difficult	30	60	23	46	2, 5, 7, 8, 10, 13, 19-21, 23-29, 37, 42, 43, 47-49
0 - 0.14	Very Difficult	1	2	0	0	
Total		50	100	50	100	

Item analysis in Mathematics AMAIUB Admission test shows that 27 or 54 percent of the prepared items had an average difficulty and 23 or 46 percent of the items were difficult.

B. Index of Discrimination

B.1 Result of analysis shows that 20 or 20 percent were found to have marginal index while 24 or 24 percent had good discrimination level but subject for improvement. Fifty-six or 56 percent of the items were found to be “very well item” in discriminating the high performing from slow learner students in AMA Admission test – English category.

Table 3: Index of Discrimination of English Items

Discrimination Index		English Test				Item Placement for the Revised Test
		Initial		Final		
		f	%	f	%	
0.40 - above	Very well item	40	40	56	56	1-7, 9-16, 19, 27, 30, 33, 35, 36, 39-50, 53, 62, 67-70, 74-77, 79, 81-85, 87, 89-92, 97, 99, 100
0.30 - 0.39	Good but subject for	31	31	24	24	26, 31, 34, 37, 38, 51, 55-57, 59-61,

	improvement					63-66, 71-73, 78, 93, 94
0.20 - 0.29	Marginal item, major revision	16	16	20	20	8, 17, 18, 20-25, 29, 32, 52, 54, 58, 80, 86, 88, 95, 96, 98
0.19 - below	Very poor item to be rejected	13	13	0	0	
Total		100	100	100	100	

B.2. Result of analysis shows that 7 or 14 percent were found to have marginal index while 9 or 18 percent had good discrimination level but subject for improvement. Thirty-four or 68 percent of the items were found to be “very well item” in discriminating the high performing from slow learner students in AMA Admission test – Mathematics category.

Table 4: Index of Discrimination of Math Items

Discrimination Index		Mathematics Test				Item Placement for the Revised Test
		Initial		Final		
		F	%	f	%	
0.40 - above	Very well item	19	38	34	68	1-4, 6, 8, 9, 12, 15, 17-19, 22-24, 26, 27, 29-31, 33-38, 40-47
0.30 - 0.39	Good but subject for improvement	15	30	9	18	5, 10, 14, 20, 28, 39, 48-50
0.20 - 0.29	Marginal item, major revision	4	8	7	14	7, 11, 13, 16, 21, 25, 32
0.19 - below	Very poor item to be rejected	12	24	0	0	
Total		50	100	50	100	

C. Accepted Items

C.1 Item analysis shows that 55 or 55 percent of the English items were accepted as items of the admission test. This means that English items were appropriate for the admission test and 45 or 45 percent was to be improved or revised.

Table 5: Acceptability of English Items

Decision	f	%	Item Placement in English test
Accept	55	55	1-7, 9-16, 19, 27, 35, 36, 39-50, 53, 62, 67-70, 74-77, 79, 81-85, 87, 89-92, 97, 99, 100
Improve /Revise	45	45	8, 18, 19, 20-26, 28-34, 37, 38, 51, 52, 54-66, 71-73, 78, 80, 86, 88, 93-96, 98
Reject	0	0	
Total	100	100	

C.2 Item analysis shows that 34 or 68 percent of the Math items were accepted as items of the admission test. This means that Math items were appropriate for the admission test. Only 16 or 32 percent was to be improved or revised.

Table 6: Acceptability of Math Items

Decision	f	%	Item Placement in Mathematics test
Accept	34	68	1-4, 6, 8, 9, 12, 15, 17-19, 22-24, 26, 27, 29-31, 33-38, 40=47
Improve/Revise	16	32	5, 7, 10, 11, 13, 14, 16, 20, 21, 25, 28, 32, 39, 48-50
Reject	0	0	
Total	50	100	

D. Content Validity of the AMA Admission Test

Table 7. Summary of Evaluator’s Rating in the Content Validity of the AMA Admission Test

Criteria	Evaluator					M	DI
	1	2	3	4	5		
The AMA Admission test is a tool designed to assess the academic readiness of the students for college education.	4	4	4	4	4	4	A
The AMA Admission test adequately represents the Mathematics and English competency of the students for college education.	4	4	4	4	4	4	A
The AMA Admission test is suitable to the comprehension level of the incoming college freshmen.	4	4	4	4	4	4	A
The AMA Admission test is suited to the behavioral patterns of the incoming college freshmen.	4	3	4	3	4	3.6	A
The AMA Admission test has the GCC and	4	3	4	3	4	3.	A

International standards to assess the ability of the students.						6	
Descriptive Interpretation	A	A	A	A	A	3.84	A

Legend: 1.00-1.49: Strongly Disagree (SD)
1.50-2.49: Disagree (D)
2.50-3.49: Moderately Agree (MA)
3.50-4.49: Agree (A)
4.50-5.00: Strongly Agree (SA)

Table shows the content validity of the AMA Admission test evaluated by English and Mathematics pool of experts. The evaluators agreed that the AMA Admission test “ is a tool designed to assess the academic readiness of the students for college education, adequately represents the Mathematics and English competency of the students for college education, suitable to the comprehension level of the incoming college freshmen, suited to the behavioural patterns of the incoming college freshmen, and test has the GCC and International standards to assess the ability of the students.

E. Reliability:

Table 8: Reliabilities of the Test Variable

Variable	Cronbach’s α	Guttman	KR-20
English Test	0.85	0.82	0.86
Mathematics test	0.81	0.80	0.82

The internal consistency of the English and Mathematics tests were found to be acceptable with a Cronbach’s alpha value of 0.85 and 0.81, respectively. This value indicated accepted reliability for the admission test. Aside from internal consistency, Split-half method was also performed resulting to a Guttman coefficient of 0.82 (English) and 0.80 (Mathematics), both values that indicated high internal consistencies of the responses in the test. Finally, Kuder-Richardson, KR20, was employed to determine internal consistency with values of 0.86 (English) and 0.82 (Mathematics) indicating high reliability of the whole instrument.

5. CONCLUSION

The developed test approximates average item difficulty. Majority had an average difficulty while only few items were found to be easy or difficult. On the other hand, the average discrimination index of the whole instrument was found to be capable in discriminating high performing students from low performers. This reflects that the test items are capable to differentiate the aptitude of the students to a greater extent. In summary, 55 and 34 of the items in English and Mathematics respectively were accepted as items of the admission test. Only 45 and 16 items in English and Mathematics respectively need to be improved or revised. The items to be revised will need changes in its construction. Removing these items would increase the reliability of the test. The reliability of the

whole instrument indicates high reliability of the whole instrument for the incoming freshmen students. This indicates the viability of the instrument to be adapted as a mission test for the prospective college students.

REFERENCES

- [1] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- [2] Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- [3] Hambleton, R.K. (1994). The rise and fall of criterion referenced measurement? *Educational Measurement: Issues and Practice*, 13(4), 21-26.
- [4] Hambleton, R.K. (1984). Validating the test scores. In R.A. Berk (ed.). *A guide to criterion-referenced test construction*. Baltimore, MD: The John Hopkins University Press, pp. 199-230.
- [5] Harriet Broder, (2014). Why are College Entrance Exams Necessary? Breakthrough Test preparation. breakthroughtestprep.com/why-are-college-entrance-exams
- [6] Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, 6, 153-160.
- [7] Jeff Brenzel, (2015) The Real Role of Tests in Your College Application, The College Board. Yale University.
- [8] Kuder, G. F., and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151-160.
- [9] Peterson, (2016). Are Test Scores Still Important College Admission Requirements. www.petersons.com/college-search/college.
- [10] Rivera, J. (2007). Test Item Construction and Validation: Developing A Statewide Assessment for Agricultural Science Education. *International Journal of Interdisciplinary Research and Innovations* ISSN 2348-1226 (online) Vol. 2, Issue 4, pp: (129-133), Month: October - December 2014, Available at: www.researchpublish.com Page | 129 Research Publish Journals
- [11] Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99-103.