Development and Validation of Engineering Aptitude Test: Using Rasch Model Analysis

Leila A. Ramos
Technological University of the Philippines – Taguig Campus

An aptitude test for engineering was developed to be able to determine the ability of the students who will pursue engineering programs. The purpose of the test is to assess students' probability of success to finish the course. This instrument served as helpful tool for universities and colleges admission test in selecting and admitting students in engineering programs that only those students who are capable to pursue an engineering courses will be admitted to be able to minimize labor/skills employment mismatch. The engineering aptitude test was also useful to the business and industry human resource personnel in order to identify applicants' and to discriminate between those individuals who have the abilities to carry out a specific task and those who do not. The findings generally support the test validity with forty six items demonstrated acceptable model fit. Rasch Model analysis showed the results of mean scores in the content areas, test range from 2.7 to 18.2. Majority of the students got correct scores on the Number Ability specifically in Number Computation than in Technical Ability. The internal consistency of the test was measured using Kuder Richardson Formula 20, test range from .00 to .61 which indicates low to moderate test reliability. Results show that person reliability ranges from .00 to .50 which indicates low to moderate reliability however the results of the item reliability for all the content areas was high with item reliability of.83 to .97.

Keywords: engineering aptitude, item fit, person fit

The institutionalization of ladderized interface between technical vocational and higher Education in the Philippines was conceptualized through an Executive Order 358 last September 15, 2004. This was created in response to the country's major socio-economic concerns namely poverty, unemployment and labor/skills mismatch. This interface intends to create a borderless education and training system that would allow mobility in terms of flexible entry and exit into the educational system of students and workers. The institution of technical education was mandated to provide higher and advanced vocational, technical, industrial, professional education and training leading to certificates, diplomas and degrees.

With the increasing demand for knowledgeable and skilled workers in this dynamic society, industries have developed linkages with academic institutions for their pool of workers. In developing countries like the Philippines, the role of technician workers and engineers in the economy is significantly important as the country becomes more industrialized using modern production techniques. For this reason, it is the duty of the technician and engineering education system acting on behalf of its students and graduates to anticipate employers' preference for the services of technicians and engineers in terms of numbers, varieties and quality to provide the economy with the right type of workforce (Tadeo, 1988). Blair (2005) mentioned that colleges and universities are being driven by the government's desire to create a more egalitarian higher education system, there has also been mounting pressure on their part to adopt admissions tests to distinguish between the best candidates as record numbers of A-level students are forecast to gain top grades (Stringer, 2008).

With this situation, institution of higher learning in order to produce a right type of workforce in the field of engineering has to acquire an accurate admission and selection tools to measure and assess students' aptitude for engineering students.

According to Magno (2009), aptitude tests are commonly used in the educational, clinical, and industrial settings to determine the potential of individuals for a wide variety of purpose. However, for some experts in measurement and assessment, available aptitudes tests in the market do not provide them with certain needs such as measuring domain specific-variables like potentials of nurses and engineers (Magno, 2009).

Cronbach and Snow (1977) defined aptitude as any characteristic of a person that forecasts his/her probability of success under a given treatment. Psychologically, this aptitude is whatever makes a person ready to learn in a particular situation and is therefore a predictor or forecaster (Russo, 2011). Aptitude testing is based firmly upon the assumption that a large numbers and varieties of mental tests can be parsimoniously described by a limited number of factors. These factors are then taken as representations of underlying aptitude (Hammond, 1984). According to Hammond (1984) there were varieties of aptitude tests such as: Differential aptitude batteries which are designed to measure aptitude profiles in order to discriminate between those individuals who have abilities to carry out a specific task and those who do not.

Several research issues were raised on aptitude testing. Whether aptitude tests are more valid predictor of achievement in college and if the aptitudes tests are fairer predictors of achievement in higher education? The questions raised by Stringer (2008) on aptitude tests would be a great challenge to the other researchers and expert in educational measurement and evaluation to conduct more research studies to prove whether aptitude tests are more valid predictors of achievement in higher education and if the aptitudes tests fairer predictors of achievement in higher education.

In the statement made of Atkinson (2001) in achievement versus aptitude in college admissions. Atkinsons (2001) cited that "students should be selected on the basis of their demonstrated success in learning, not some ill-defined notion of aptitude. Atkinson argued that admitting students to a college or university should be based on three principles. First, students should be judged based on their actual achievement, not on ill-defined notions of aptitude. Second, standardized tests should have a demonstrable relationship to the specific subjects taught in high school, so that students can use the tests to assess their mastery of those subjects. Third, universities in the U.S.A. should employ admission processes than look at individual applicants in their full complexity and take special pains to ensure that standardized tests are used properly in admission decisions.

The research conducted by Stringer (2008) on aptitude tests versus school exams as selection tools for higher education and the case for assessing educational achievement in context. Stringer (2008) argued if there is any evidence that aptitude tests are the solution to either the problem of widening participation or the problem of discriminating between the best applicants. Stringer (2008) also raised the questions as to whether aptitude tests are better predictors than school exams for higher education was broken down into two main issues: The first issue whether the aptitude tests are more valid predictors of achievement in higher education? And the second issue, if the aptitudes tests fairer predictors of achievement in higher education? There is no evidence to suggest that aptitude tests alone provide any more predictive power than curriculum-based tests alone (Stringer, 2008).

There are several studies that support the use of aptitude tests as to whether aptitude tests are better predictors than school exams for higher education and if the aptitudes tests are fairer predictors of achievement in higher education.

West and Gibbs (2004) had argued that test of "potential" in addition to achievement is needed. An aptitude test is needed to support the results of the achievement tests to fully assess the student potentials. According to the United States Commission on Civil Rights (USCCR), the SAT is used by the majority of the American higher education institutions in the admissions process (West & Gibbs, 2004). The American SAT, like General Certificate of Education Advanced levels measures achievement although unlike the latter, it is not designed to be closely related to curriculum content. Universities are left in the dilemma that though they may wish to assess the potential, it is difficult to separate potential from actual test performance since tests are supposedly measuring potential but in fact measure achievement at a particular point in time (West & Gibbs, 2004).

The study on engineering persistence on the past, present, and future factors and gender differences examined the different factors that are related to persistence in engineering for women and men. The research focused on both the academic and non-academic factors in the individual's past, present, and future that might influence persistence in engineering and whether different factors were important to engineering persistence for women and men. More recent research, using a sample of over 2400 women and men undergraduate engineers, found that freshmen grade point average (GPA), Scholastic Aptitude Test scores (SAT), and self-perceptions of math and science abilities predicted persistence for both sexes (Jackson, et al., 1993).

Magno (2010) proposed to integrate both the quality and abilities of nurses in assessing nursing potential. According to Magno (2010) nurses who provide services with care needs to be aligned with their ability to handle the technical aspects of their job. The study tested a model that shows the direction from qualities of nursing students (caring, compassion, commitment, and connectedness) directly affecting their aptitude and academic achievement. This model implies that there should be an integration of the knowledge, skills, and behaviours of nurses to optimize their potential for success in the nursing profession. The results of the study show that integrating nursing quality with both aptitude and achievement show a considerable good fit for the proposed model. Based on the results nursing quality traits do not affect ability measures, but aptitude affect achievement in nursing Magno (2010).

The ground on which the use of aptitude tests is justified that it measures something that is school-proof: Something fundamental about the ability of the individual that is not affected by their level and quality of education. The idea of aptitude testing is based on the noble aim of levelling the playing field at school-leaving age so that access to higher education is not limited to those who have had access to greater educational resources during childhood. Stringer (2008) the fundamental problem for aptitude testing for admissions to higher education is that the more they are grounded in a particular discipline or body of knowledge, the more they become like achievement test and therefore become increasingly redundant. The less they are grounded in knowledge and experience, the more school proofed they become, and the less useful they become at predicting future achievement.

According to Stringer (2008) that existing aptitude tests are simply not environment-proof and, even if they were, there should be great concern over both the fairness of rewarding the outcomes of such tests and the validity of using them for admissions to higher education. Stringer (2008) stated that it is unlikely that universities would gladly relinquish control of their admissions systems, and the proposed system would be not encourage or require this; it is not an issue of allocating students to universities on the basis of their academic rankings but that would be the basis of the initial sift of applications.

The educational institutions should help their students in assessing the person capabilities in choosing a career choice specifically in engineering programs. In order to ensure that college students who will enrol in engineering programs posses the abilities to do the specific tasks and competencies required in the engineering course. It is the duty of every technical instructor to equip their students with the necessary knowledge and skills to be immediately productive in their job. The role of educational higher learning is to assess the competencies of their incoming students in order to develop their potentials to the fullest, prepare them for work and equip them for occupations to be able to provide the right type of workers for the job.

In the Philippine education system, the National College Entrance Examination (NCEE) in the Department of Education was abolished in year 1994. The NCEE was used to for all high school graduates to be able to enter college and have a chance of a better career in the future. However, the removal of the NCEE resulted to a less accurate screening of the high school graduates in entering college. College admission and selection rests on the decisions of the universities and colleges which they decide on the assessment tools they administer. Recently, the Commission of Higher education (CHED) is requiring for tighter screening of students who want to enter college, and there will be a plan to revive the abolished NCEE. But the department clarified that the introduction of an examination for entering college students should be more of an aptitude test. This provision stated that the students graduating from high school who would want to go to college will take an Aptitude Test and the results of which will be used to classify them into two groups: the first group will include those who will go to the pre-college or the Technical Vocational track and the second are those who will go to the university college track. The reestablishment of the NCEE or an aptitude test is one of the reform measures that the commission proposed in order to improve the quality of higher education in the country. The proposal to re-introduce the NCEE was aired by several sectors saying that it would upgrade the quality of tertiary education in the country while at the same addressing the growing problem of job mismatch among college graduates.

The present study focused on the development of aptitude test that are needed to support the results of the achievement tests in order to fully assess the student potentials. The construction and validation of engineering aptitude test involves the use of Rasch Model. The study was conducted for the following purpose: First, to be able to address the issue of labor/skills employment mismatch as cited in Executive Order 358 and will serve as supporting tool in CHED programs in reviving the NCEE but will focus in measuring the aptitude of the students, specifically in the field of engineering. An aptitude test for engineering was constructed to be able to measure students' capabilities to pursue and finish engineering courses. This will be helpful to supplement other universities and colleges admission test in selecting and admitting students in engineering programs.

Second, the engineering aptitude tests will be useful for the business and human resource industry personnel to identify applicants' aptitude and assess if this person will be capable to do the job in a particular situation specifically in the technical and engineering works. The test was designed to measure aptitude profiles in order to discriminate between those individuals who have abilities to carry out a specific task and those who do not. The issue of defining aptitude and ways of reliably testing it, however, is not only in the area of academic, but also in the business and industry settings.

Method

Participants

A total of one hundred ten (N=110) third year technology engineering students from the different courses were selected as respondents. The participants age ranged from 17-20 years are in their major field of specialization and presently involved in acquiring the expected competencies in their field. The purpose of the Engineering Aptitude Test (EAT) was explained to the students before the test was administered.

Instruments

The first draft of the engineering aptitude test contained 90 items. The test was content validated and reviewed by the experts in advance psychometrics. The suggestions were considered and the test was revised based on the recommendations. The test was organized into two schemes the verbal and non-verbal. A total of 60 items were retained in the Engineering Aptitude Test. The following items were divided based on the set table of specification. The Table of Specifications contains four content areas with sub-factors: (1) numerical ability with two sub factors: number sequence and numerical computation; (2) abstract reasoning which measures the student figure sequence and relationship ability; (3) technical ability with two sub factors: mechanical, electronics and electrical ability; (4) word analogy with three sub-factors: cause and effect analogy, object and relationship analogy and opposites analogy. The content areas are based on the description of taxonomy of aptitude items devised by Magno (2009). The 25% of the items were placed for the knowledge and application and 16.66% were placed for comprehension, analysis, and synthesis. Most of the items were concentrated on knowledge and application since the main purpose of the test is to measure the aptitude for engineering students which focus on the application of technical knowledge.

Table 1

ENGINEERING APTITUDE TEST

Table of Specification

Content Areas	Knowledge 25%	Application 25%	Comprehension 16.66%	Analysis 16.66%	Synthesis 16.66%
A. Numerical Ability					
Number Sequence					Items 1-10 10 items
Numerical Computation		Items 11-25 15 items			
B. Abstract Reasoning					
Figure Sequences and Relationship			Items 26-35 10 items		
C. Technical Ability					
Mechanical Reasoning				Items 36-40 5 items	
Electronics and Electrical Information				Items 41-45 5 items	
D. Word Analogy					
Cause and Effect Analogy	Items 46-50 5 items				
Object and Relationship Analogy	Items 51-55 5 items				
Opposites Analogy	Items 56-60 5 items				

Data Analysis

The analyses of engineering aptitude test items were conducted using the Rasch Model through the Winsteps Software. The Rasch Model is the only item response theory (IRT) in which the total score across the items characterized a person totally (Magno & Ouano, 2009). In the Rasch analysis, items are ranked or ordered in difficulty and the units of measurement should maintain their size across the entire range of measurement regardless of the exact items used, the instrument used, who is using the instrument or what is being measured. The results using the Rasch Model are the true strong scores, which is more applicable to for the tests with right and wrong (dichotomous) responses. The item polarity may be measure using the PTME correlation results to determine if the items are moving in the same direction. Internal consistencies were generated to determine the reliability of the person and item reliabilities. Difficulty of the level of the items were calculated and if the item fits the Rasch Model were analyze as well.

Results

The Rasch Model was used to analyze the items of the engineering aptitude test. To investigate the functioning of the items in the engineering aptitude test, the one parameter Rasch model was used. The person and item reliability was also obtained separately in the Rasch analysis.

Table 2

Descriptive Statistics of the Content Areas of the Engineering Aptitude Test

N=110	*M*	*SD*	Cronbach's Alpha	Person Reliability	Item Reliability
A. Number Ability	18.2	2.7	.19	.34	.96
1. Number Sequence	7.6	1.5	.00	.11	.97
. 2. Numerical Computation	10.6	2.3	.34	.44	.96
B. Abstract Reasoning	5.8	2.1	.54	.50	.83
C. Technical Ability	6.7	1.4	.28	.34	.95
1. Mechanical Ability	3.4	.7	.35	.00	.88
2. Electronics & Electrical Ability	2.7	.9	.00	.20	.97
D. Word Analogy	9.4	2.6	.61	.26	.96

The descriptive statistics reported the mean scores of the content areas of the instrument, test range from 2.7 to 18.2. In the scoring of 0 and 1, 1 indicates correct answer and 0 as incorrect answer. Results show that majority of the students got correct scores on the Number Ability specifically in Number Computation than in Technical Ability. The internal consistency of the test was obtained with Cronbach's alpha ranging from .00 to .61 which indicates low to moderate test reliability. Person and item reliability was obtained using the Rasch analysis. Results show that person reliability ranges from .00 to .50 which indicates low to moderate reliability however the results of the item reliability for all the content areas was high with item reliability of.83 to .97.

Table 3
Summary of Content Areas for Item Measured

	M of the Total Score	S.D. Measure	S.E of Item Mean	REAL RMSE SEPARATION	Item Reliability
Number Sequence	57.3	1.65	0.55	4.83	0.96
Numerical Computation	52.9	1.39	0.37	4.55	0.95
Abstract Reasoning	43.9	0.66	0.22	2.12	0.82
Mechanical Ability	61.4	1.66	0.83	2.7	0.88
Electronics and Electrical Ability	39.8	1.96	0.98	5.77	0.97
Word Analogy	46.8	1.65	0.44	4.74	0.96

Table 3 shows the results of each content area for item measured with 75 counts. The real RMSE separation scores of 5.77, 4.83, 4.74, and 4.55 indicate that the items in electronics and electrical ability, number sequence, word analogy and number computation highly discriminate the high ability and low ability students. However, real RMSE scores in abstract reasoning (2.12) and mechanical ability (92.7) moderately discriminate high ability and low ability students. Data shows that there is a wide spread of difficulty in the items indicated by the standard deviation of item difficulty estimates (.66 to 1.96) and the separation of 2.7 to 5.77. The standard error item mean range from 0.22 to 0.98 indicating moderate to greater precision of the items. The item reliability of the content areas attained a high item reliability of .82 to .96.

Table 4 shows the results of the 25 items that measure number ability: 1-10 items measure the number sequence ability and 11-25 items measure the numerical computation ability. To determine if the items under each content area fits the rasch model, item fit mean square (MNSQ) was computed. MNSQ Infit values within 1.2 and less than 0.8 are acceptable. Items with high MNSQ values indicate a lack of construct homogeneity with other items in a scale, where low values indicate redundancy with the other.

Table 5 shows the result of the items that measure the item difficulty and goodness of fit of the items in abstract reasoning (composed of figure sequence items 26-35). Item fit mean square (MNSQ) was computed, MNSQ Infit values show that items 28 and 32 did not fit the Rasch Model. These items in the abstract reasoning lack construct homogeneity where they do not share a similarity with the items in engineering aptitude.

Table 4
Summary of the Mean Square Measured for Item Difficulty

Number Ability Item	Measure	SE	Infit MSQ	Z	Outfit MSQ	Z	PTME
1	-1.40	0.22	0.23	-.4	0.46	-.3	0.09
2	-1.40	0.22	0.20	-.4	0.15	-.10	0.35
3	-1.44	0.20	0.22	-.6	0.11	1.4	0.25
4	- .98	0.42	0.97	0.4	0.90	0.4	0.20
5	1.45	0.25	6.45	9.9	9.9	9.9	0.36
6	2.70	0.27	1.16	1.4	1.18	1.0	0.21
7	1.25	0.26	0.93	-.7	0.94	-.4	0.46
8	2.42	0.26	0.79	-2.1	0.74	-1.9	0.59
9	-1.34	0.26	0.23	-.3	0.27	-.4	0.33
10	-1.26	0.31	0.40	-.1	0.86	0.4	0.05
11	2.13	0.17	0.26	-0.45	0.66	-0.28	0.37
12	-1.36	0.36	1.16	0.71	1.04	0.2	0.17
13	-1.8	0.41	0.98	0.04	0.88	-0.1	0.28
14	1.76	0.27	1.17	1.43	1.22	1.2	0.17
15	1.48	0.26	1.00	0.04	1.31	1.92	0.31
16	-0.91	0.32	0.90	-0.49	0.71	-1.03	0.46
17	-1.5	0.37	0.86	-0.49	0.62	-0.95	0.45
18	-1.62	0.15	0.93	-0.28	0.78	-0.87	0.44
19	-0.39	0.28	0.96	-0.24	0.86	-0.63	0.43
20	0.52	0.25	0.98	-0.22	0.94	-0.47	0.43
21	2.23	0.29	1.25	1.6	1.28	1.1	0.06
22	0.58	0.25	1.06	0.7	1.14	1.2	0.31
23	-0.31	0.28	0.97	-0.17	0.92	-0.34	0.41
24	-0.63	0.30	0.89	-0.65	1.06	0.3	0.43
25	-1.8	0.41	1.00	0.1	0.68	-0.58	0.31

Four Rasch analyses were conducted separately for each content area. Results show that 8 items in the numerical ability turned out to have a bad fit. The item lacks construct homogeneity and was not acceptable within item fit mean square range. Items 1, 2, 3, 5, 9, 10, 11 21 fitted the Rasch Model. Items in the numerical ability that lacks construct homogeneity do not share similarity with the rest of the items.

Table 5
Summary of Mean Square Measure for Item Difficulty for Abstract Reasoning

Item	Measure	SE	Infit		Outfit		
			MSQ	Z	MSQ	Z	PTME
26	-.05	0.26	1.10	0.90	1.0.	0.5	0.38
27	-.05	0.26	0.86	-1.3	0.83	-1.1	0.56
28	-.73	0.29	0.73	-2.0	0.66	-1.7	0.63
29	-.73	0.29	0.87	-.9	0.77	-1.0	0.54
30	-.73	0.29	1.01	0.1	0.93	-.2	0.43
31	1.17	0.27	1.15	1.4	1.07	0.4	0.32
32	0.56	0.26	1.39	3.6	1.76	4.2	0.09
33	-.34	0.27	0.94	-.5	0.90	-.5	0.49
34	0.96	0.26	1.03	0.4	1.05	0.4	0.41
35	-.05	0.26	0.86	-1.3	0.85	-1.0	0.56

Table 6
Summary of the Mean Square Measured for Item Difficulty

Technical Ability			Infit		Outfit		
Item	Measure	SE	MSQ	Z	MSQ	Z	PTME
36	0.73	0.35	1.06	0.50	1.07	0.5	0.55
37	-2.87	1.03	0.93	0.2	0.29	-.2	0.24
38	1.53	0.34	0.88	-1.0	0.82	-1.1	0.72
39	-.81	0.47	1.14	0.5	1.57	1.2	0.28
40	1.42	0.34	0.96	-.3	0.98	-.1	0.67
41	2.18	0.35	0.97	-.1	1.08	0.30	0.49
42	2.56	0.37	1.01	0.1	0.98	0.10	0.44
43	-1.29	0.29	0.98	-.1	0.81	-.4	0.45
44	-1.38	0.30	0.90	-.8	1.23	0.70	0.46
45	-2.07	0.34	1.10	0.6	0.95	0.1	0.29

All the items in Technical Ability particularly in mechanical (36-40), electronics and electrical ability (41-45) fit the Rasch Model. These items in the Technical Ability demonstrated construct homogeneity that shared a similarity with the items in engineering aptitude.

Table 7 shows the result of the 25 item measure for Word Analogy: Cause and effect analogy (46-50), object and relationship analogy (51-55), opposite analogy (56-60). The 4 items turned out to have a bad fit, they lack construct homogeneity and were not acceptable within item fit mean square acceptable range. Items in the word analogy that lacks construct homogeneity do not share similarity with the items.

Table 7
Summary of the Mean Square Measured for Item Difficulty

Word Analogy			Infit		Outfit		
Item	Measure	SE	MSQ	Z	MSQ	Z	PTME
46	-3.40	1.02	0.77	0.10	0.09	-.9	0.86
47	1.18	0.25	1.01	0.1	1.17	1.7	0.30
48	0.41	0.26	1.03	0.3	1.16	1.1	0.36
49	2.84	0.21	0.79	-.5	0.84	-.4	0.27
50	1.25	0.25	1.11	1.5	1.11	1.10	0.25
51	-.40	0.31	1.18	1.0	1.18	0.80	0.37
52	-3.40	1.02	0.77	0.1	0.09	-.9	0.86
53	-3.40	1.02	0.77	0.1	0.09	-.9	0.86
54	-1.92	0.53	0.86	-.2	0.55	-.7	0.70
55	1.50	0.25	1.17	2.1	1.20	1.8	0.20
56	1.82	0.26	1.01	0.1	1.01	0.10	0.27
57	0.54	0.26	0.92	-.8	0.94	-.5	0.42
58	0.74	0.25	0.96	-.4	1.05	0.40	0.37
59	0.81	0.25	0.95	-.6	0.92	-.7	0.39
60	1.43	0.25	0.92	-1.0	0.90	-.9	0.35

Overall results show the items with negative measure are easy items: Number sequence (1,2,3,4,9,10); number computation (12,13,16,17,18,19,23,24,25); abstract reasoning ability (26,27,28,29,30,33,35); technical ability (37,39,43,44,45); word analogy (46,51,52,53,54). Items with positive measure engineering aptitude are said to be difficult: number sequence (5,6,7,8); number computation (11,14,15,20,21,22); abstract reasoning ability (31,32,34); technical ability (36,38,40,41,42); word analogy (47,48,49,50,55,56,57,58,59,60). There were 32 items with negative measure which are considered easy and 28 items with positive measure which are considered difficult. All the 60 items in engineering aptitude test attained a positive result in PT-Measure which indicates that most of the items fitted the Rasch Model.

Discussions

The engineering aptitude test was constructed and validated to ensure that college students who will enrol in engineering programs posses the abilities to do the specific tasks and competencies required in the engineering course. The test was also designed to measure aptitude profiles of the person in order to discriminate between those individuals who have abilities to carry out a specific task and those who do not.

Descriptive statistics reported the mean scores of the content areas of the instrument range from 2.7 to 18.2. In the scoring of 0 and 1, 1 indicates correct answer and 0 as incorrect answer. Results show that majority of the students got correct scores on the Number Ability specifically in Number Computation than in Technical Ability. The Cronbach's alpha coefficients of the test range from .00 to .61 which indicates low to moderate test reliability. Person and item reliability was obtained using the Rasch analysis. Results show that the person reliability ranges from .00 to .50 which indicates low to moderate reliability however the results of the item reliability for all the content areas was high with item reliability of.83 to .97. Each content area for item measured with 75 counts. The real RMSE separation scores of 5.77, 4.83, 4.74, and 4.55, indicate that the items in electronics and electrical ability, number sequence, word analogy and number computation highly discriminate the high ability and low ability students. However, the real RMSE scores in abstract reasoning (2.12) and mechanical ability (2.7) moderately discriminate students with high ability and low ability. Data showed that that there is a wide spread of difficulty in the items as the standard deviation of item difficulty estimates of .66 to 1.96 and the separation of 2.7 to 5.77. The standard error item mean range from 0.22 to 0.98 indicates moderate to greater precision of the items being measured. The item reliability of the content areas in item measured attained a high item reliability of .82 to .96.

Based on the results, it is evident that with the 60 items measuring engineering aptitude, the 46 items fitted the Rash Model. The item fit mean square (MNSQ) was computed with Infit values within 1.2 and less than 0.8 are acceptable. Items with high MNSQ values indicate a lack of construct homogeneity with other items in a scale, where low values indicate redundancy with the other. Four Rasch analyses were conducted separately for each content area. Results show that 14 items turned out to to have bad fit, it lacks construct homogeneity and were not acceptable within item fit mean square acceptable range: items 1, 2, 3, 5, 9, 10, 11, 21, 28, 32, 46, 49, 52, and 53; did not fit The Rasch Model. The following items do not share a similarity with the items in engineering aptitude can either be removed or revised.

Among the 60 items, 32 items with negative measure are easy items and 28 items with positive measure considered difficult items. Data shows that that there is a wide spread of difficulty in the items as the standard deviation of item difficulty estimates are .66 to 1.96 and the separation of 2.7 to 5.77. The standard error of the item means range from 0.22 to 0.98 indicating moderate to greater precision of the items. The item reliability of the content areas attained a high item reliability of .82 to .96.

However, after considering the item fit mean square (MNSQ), there were 14 items that did not fit the Rasch Model. There were 22 items retained as easy items and 24 items retained as difficult items. The findings still supports the test's validity. The 46 items in engineering aptitude test demonstrated acceptable fit in the Rasch Model suggesting that the construction and validation of the engineering aptitude test is acceptable.

References

Atkinson, R. A. (2001). Achievement versus aptitude in college admissions. *Issues on Science and Technology, 18*(2) 31-36

Grotevant, H.D., & Thorbecke, W. L. (1982). Sex differences in styles of occupational identity information. *Development Psychology, 18*, 396-405.

Hammond, S. M. (1984). An investigation into the factor structure of General Aptitude Test Battery. *Journal of Occupational Psychology, 57*, 43-48.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology 6*, 35-45.

Holland, J. L. (1966a). A psychological classification scheme for vocations and major fields. *Journal of Counselling Psychology, 13*, 278-288.

Jackson, L. A., Gardner, P. D., & Sullivan, L. A. (1993). Engineering persistence: Past, present, and future factors and gender differences. *Higher Education, 26*, 227-246.

Magno, C. P. (2010). Integrating nursing quality with achievement and aptitude towards Assessing nursing potential. *The International Journal of Researcher and Review, 4*, 37-49.

Magno, C. P. (2009). Taxonomy of aptitude test items: A guide for item writers. *The International Journal of Educational and Psychological Assessment, 2*, 39-53.

Magno, C., & Ouano, J. (2010). *Designing written assessment for student learning*. Phoenix House, Inc.

Russo, M. (2011). Aptitude testing over the years. *International Journal of Research and Practice, 13*, 5-30.

Stinger, N. (2008). Aptitude tests versus school exams as selection tools for higher education and the case for assessing educational achievement in context. *Research Paper in Education, 23*, 53-68.

Treumann M. J., & Sullivan, B. A. (1949). Use of the engineering and physical science aptitude test as predictor of academic achievement of freshmen engineering students. *The Journal of Educational Research, 43*(2), 129-133.

West, A., & Gibbs, R. (2004). *Selecting undergraduate students: What can UK learn from the American SAT?*. London: LSE

Wood, D. A. (1972). Moderating the prediction of grades in freshman engineering. *Journal of Educational Measurement, 9*, 311-315.