CONSTRUCTION AND EVALUATION OF A DIAGNOSTIC EXAMINATION IN COLLEGE ALGEBRA FOR FRESHMEN OF THE COLLEGE OF SCIENCE, UNIVERSITY OF SANTO TOMAS

Mark Louie F. Ramos, MMATHED

Proceedings of the Mathematics Teachers Association of the Philippines Tertiary Level Annual Convention, West Visayas State University, Iloilo City, Philippines, August 30-31, 2008

Narrative Abstract

The purpose of this study was to construct and evaluate an instrument for determining student preparedness in College Algebra. A 73-item instrument covering prerequisite arithmetic and high school Algebra knowledge for College Algebra was constructed. The instrument was pilot-tested on a freshman population of 595 students. Results of reliability testing using the split odd-even test showed that the instrument is reliable at a 0.05 significance level. Concurrent and predictive validity testing likewise showed that the instrument is valid (α=0.05). Students who scored at least 56 out of 73 in the instrument were found to be much more likely to pass a College Algebra course than those who scored less than 56 ($p \approx 0.00$). Item analysis showed that 71% of the items have acceptable discriminatory indices. The remaining 29% were considered for revision or removal. It was recommended that the instrument be refined based on the outcomes of the evaluation and that it be subsequently administered to other freshman populations prior to taking their respective College Algebra courses. (Contains 1 figure)

Introduction

Competency in basic algebra is of prime importance to most if not all tertiary courses. This remains to be a nationwide concern as separate studies on performance in Mathematics of pre-service and current Mathematics teachers reveal a substandard quality of Mathematics education in the elementary and secondary levels (Leongson & Limjap, 2003). In the latest results of the Trends in Mathematics and Science Study (TIMSS) which was administered in 2003, the Philippines lagged behind other participating countries, placing 24th out of 25 countries in Grade 4 Mathematics and 41st out of 45 participating countries in second year high school Mathematics. These rankings were noted to be very similar to TIMSS' assessment of the country in 1999. (Cristobal, 2005).

In a roundtable conference held in 2004 at the SEAMEO-INNOTECH in Diliman, Quezon City, Dr Allan Benedict Bernardo, Director of the Lasallian Institute for Development and Education Research (LIDER) of the De La Salle University suggested four specific research directions and studies that could be focused on in response to the TIMSS findings. These were, 1) seeking explanations for the findings of TIMSS including studies on good and bad practices; 2) understanding constraints and enabling factors for improvement; 3) evaluating interventions; and 4) rationalizing options at different levels of the educational bureaucracy. (Cristobal, 2005)

Diagnostic exams in Mathematics have long been used in other countries for determining course placements. For example, the University of Colorado at Colorado Springs administers an algebra diagnostic exam to determine a student's capacity to take a particular math subject. The topics of the said exam encompass all of the topics covered

by typical College Algebra courses in the country. Other diagnostic examinations such as the one used by the University of Texas at Austin include more basic topics which are similar to the one constructed for this paper. However, one clear distinction is the use of multiple choice questions by the former against the traditional short response type of the latter.

This paper serves as the beginning of a task which seeks to address the second mentioned research direction of Dr. Bernardo, by creating a customized diagnostic exam in Algebra for college freshmen that can predict how students would fare in their College Algebra course.

The constructed diagnostic exam is composed of 73 short response type questions with topics ranging from operations with real numbers to solving different algebraic equations. It was administered for evaluation purposes to a population of 595 college freshmen from the University of Santo Tomas, College of Science last July 2006.

This diagnostic tool aims to understand the constraints that high school graduates have with regard to their competencies in Algebra which would in turn suggest enabling factors for improvement. Specifically, this paper aims to assess the quality of the constructed diagnostic examination based on acceptable measures which are: (1) item analysis, (2) reliability testing and (3) validity testing.

With reference to recommendations obtained in this paper, the test developer intends to propose the operational use of the constructed diagnostic examination on the freshmen of the College of Science, University of Santo Tomas for S.Y. 2007 – 2008. Further development of the tool will also be sought based on data gathered.

Methodology

Item analysis

Descriptive statistical details were initially obtained from the population followed by computations for the *question difficulty* and *discrimination index* of each question. *Question difficulty* measures how hard a particular test item is. It is determined by obtaining the ratio of the number of respondents who failed to answer a particular item correctly to the total number of students who answered the item. *Discrimination index* measures the ability of a test item to distinguish between the top and bottom performers in the test. It is determined by for each item by the formula below:

Discrimination index = (H– L)/M

Where:

H is the number of respondents in the top 27% of the population who were able to answer the item correctly

L is the number of respondents in the bottom 27% of the population who were able to answer the item correctly

M is 27% of the population

The results obtained from item analysis using both *question difficulty* and *discrimination index* were used to give a basic assessment on the usability of each test item based on existing standards.

Reliability testing

Reliability testing aims to measure the extent to which a test is repeatable and is yielding of consistent results. For this purpose, the *Split Odd-Even test* was selected. A paired t-test for testing significant difference between two means was selected to treat the

data at a .05 significance level which is the acceptable psychometric standard (Michelle, 1999).

Validity testing

Several procedures were considered to provide the strongest possible measure of validity based on the resources available to the test developer. Based on this limitation, the tests for *predictive* and *concurrent criterion validity* were selected. *Predictive validity* determines whether test results correlate to the outcomes of another tool or measure (Wilderdom, 2004). For this paper, a sample of 3 sections collectively composing of 123 students from the population was selected and final grades in their concluded College Algebra course were obtained. An initial One-sample t-test was used to verify whether the sample is an effective representative of the population. Once this was established, two statistical tools were selected to measure *predictive validity* independently from each other. The first is the Pearson test for significant correlation which determines whether there is a significant relationship between the students' scores in the diagnostic examination and their respective final grades in College Algebra. The second is the Chi-square test for independence which determines whether there is an underlying dependence between the numbers of students who pass/fail College Algebra to the numbers of students who obtain a score of at least 56 in the diagnostic examination. The selected benchmark of 56 for the Chi-square test was based on the 75% learning competency standard for basic education (EFA 2000). *Concurrent validity* determines if the constructed diagnostic examination can distinguish subgroups of the population based on the manifestations it seeks to measure (Wilderdom, 2004). For this paper, scores of the different freshman sections divided according to course were treated

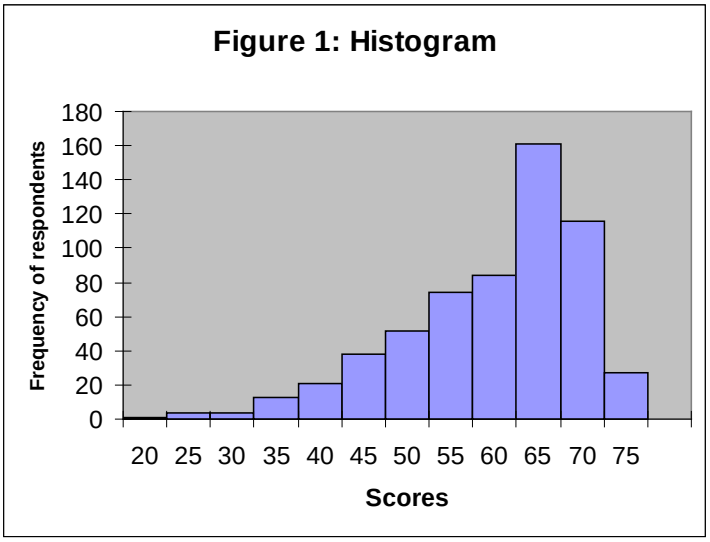
using the Analysis of Variance (ANOVA) test for significant difference, followed by Scheffe's post-hoc test (0.1 confidence level) in order to determine whether students who major in Mathematics obtained a significantly better score than students who are non-Mathematics majors.

The tests for reliability and validity determine if the constructed diagnostic test is of quality and can be used operationally while item analysis provides suggestions of items in the tool that should be revised or omitted.

Results and Discussion

Item Analysis

The data gathered from the 595 respondents who answered the constructed diagnostic tool was negatively skewed with a mean and standard deviation of 57.88±10.32, median of 61 and mode of 65 (see figure 1).



Out of 73 items, only 19% (14 items) of the questions have difficulty ratings between 0.3 and 0.7. About 77% (56 items) of the questions have ratings below 0.3 and the remaining 4% (3 items) have ratings above 0.7. Although most of the questions fall outside the general difficulty range of 0.3 - 0.7, items that fall below a difficulty rating of

0.3 are acceptable for diagnostic or preliminary tests (CARnet). The remaining 3 items that have difficulty rating above 0.7 are to be considered for revision. About 71% (52 items) of the questions have discrimination indices ≥ 0.15 which is the acceptable standard of discrimination (CARnet). The remaining 29% (11 items) are to be considered for revision or removal.

Reliability testing

The mean of the scores of the population on the odd numbered items was determined to be 28.94 versus 28.93 on the even numbered items. The absolute value of the obtained value of the paired t-test (0.08) was less than the critical value (1.96). Thus, the difference between the two means was determined to be not significant at a .05 confidence level.

Validity testing

Predictive Validity

The scores of the sample are highly correlated with their final grades in College Algebra with a Pearson correlation coefficient of 0.71. Since the obtained value is greater than the critical value of 0.197, the correlation between the scores of the sample in the diagnostic exam and their final grades in College Algebra is determined to be significant at a .05 confidence level.

The obtained value from the Chi-square test for independence of the numbers of students who pass/fail College Algebra to the numbers of students who obtain a score of at least 56 in the diagnostic examination was 44.83. This value is greater than the critical value of 3.84 which indicates that the numbers of students who pass/fail College Algebra

is significantly dependent on the numbers of students who obtain a score of at least 56 in the diagnostic examination.

Concurrent Validity

The result of ANOVA shows that the computed value of 3.54 is greater than the critical value of 2.23. This means that there is a significant difference at a .05 level of confidence among the means of the scores of students with different majors. It also showed that the mean of the scores of Mathematics majors was the maximum among the different groups. The subsequent Scheffe's test at a 0.1 confidence level reveals that the significant difference can only be located between the means of the scores of Mathematics majors and some non-Mathematics majors. No significant difference was located between the means of any two groups of non-Mathematics majors.

Conclusions and Recommendations

The constructed diagnostic examination in College Algebra was proven to have acceptable quality based on reliability and validity. Reliability testing using the *Split Odd-Even test* showed acceptable reliability of the diagnostic tool while predictive and concurrently validity tests showed acceptable validity. The results of the item analysis suggest that some questions can be considered for revision or omission.

It can be concluded that the diagnostic tool is fit for operational use on incoming freshmen of the College of Science, University of Santo Tomas for S.Y. 2007 – 2008. It is recommended that data gathered from such an operational use be utilized for further development of the diagnostic tool. It is also recommended that provisions be made for further testing of reliability using other methods such as *Test-retest* and *Split-halves test*.

Lastly, other aspects of validity such as *face validity*, *content validity* and *discriminant validity* should also be explored.

References

Cristobal R. (2005). *Educators Talk on Research Implications of Poor Performance of Filipino Students in International Math and Science Assessment Tests.* Retrieved March 5, 2007, from http://www.sei.dost.gov.ph/sep0505.html

Michell, J. (1999). *Measurement in Psychology*. Cambridge: Cambridge University Press.

Leongson J., & Limjap A. (2003) *Assessing the Mathematics Achievement of College Freshmen using Piaget's Logical Operations***,** Hawaii International Conference on Education in Waikiki

Diagnostic Exam. (2006). Retrieved March 5, 2007, from University of Texas at Austin Web site: http://www.utexas.edu/student/utlc/lrnres/mstc/msl-diagnostic/1-math-ic-exam/alg-diag4ml1c.html

Essentials of a Good Psychological Test. (2004). Retrieved March 6, 2007, from http://wilderdom.com/personality/L3-2EssentialsGoodPsychologicalTest.html

Math Placement Test. (2006). Retrieved March 5, 2007, from University of Colorado at Colorado Springs, College of Letters, Arts and Sciences, Department of Mathematics Web site: http://www.uccs.edu/~math/Courses/catalog.php

The EFA 2000 Assesment: Country Reports (n.d.). Retrieved March 6, 2007, from http://www.unesco.org/education/wef/countryreports/philippines/rappport_1.html

Self-assessment and Summative Assessment in E-education. (n.d.). Retrieved March 6, 2007, from http://www.carnet.hr/referalni/obrazovni/en/spzit/theory/analize?CARNetweb=