The Conformity of Test Construction of the Achievement Test Papers of College Teachers: A Case Study

Pablito A. dela Rama
Office of Instruction and Evaluation
Silliman University

This paper examines the rules in test construction that are commonly observed and violated by the college teachers of Silliman University and measures if significant relationship exists between the profile of teachers—sex, number of years in teaching, number of education units taken, number of seminars in test construction attended, and academic background—and their ability to observe the rules of test construction. Although majority of the rules of the five types of tests investigated were observed by teachers, particularly the True or False type, there are still rules where violations were greater than compliance, particularly in the Matching and Enumeration types. Moreover, the sex of teachers and number of seminars in test construction attended by these teachers are significantly related to their adherence to the rules of test construction. Therefore, more credits or time to courses in teacher training and seminars in test construction among in-service teachers are recommended to address the need, particularly of male teachers, to enhance their skills in constructing quality test questions.

KEYWORDS: conformity, rules of test construction, achievement test, types of test, Silliman University

INTRODUCTION

An achievement test is a systematic procedure for measuring a representative sample of learning tasks which are of two types: standardized and teacher-made (Salkind, 2003, p. 129) and done by every teacher since the teacher plays an important role in the instructional programs. As Gronlund (1993) puts it, the evaluations teachers make can have a tremendous influence on

the lives of their students; hence, they should not be lightly made. He further stressed that the role of evaluation is so intrinsic to the teaching-learning situation that even hasty consideration seems to indicate the advantages of a systematic use of planned evaluation procedures. This is so because decisions about the amount of learning that students have done, emanates from achievement test results. Therefore, there is a need to investigate this matter to generate data useful for developing a re-orientation program for faculty about the rules of testing and why these are necessary for achieving realistic measurement and evaluation of the learning of their students (e.g., Linn & Gronlund, 2000).

In this paper, the ability in complying with the rules in test construction of full time college faculty in the preparation of their final examinations was examined using Silliman University as a case study. The findings and recommendations in the end may be found relevant to other colleges and universities in the Philippines having similar conditions. What these rules in test construction relative to the type of test that are commonly observed and violated were identified using the final examination test papers of the sampled teachers from different academic units. It further examined any relationship between the teachers' sex, number of years in teaching, number of education units taken, and number of seminars in test construction attended and their ability to observe the rules of test construction. The similarities and differences in the observance of the rules of test construction among teachers from Humanities; Mathematics, Science and Technology; and Social Sciences are likewise looked into.

REVIEW OF RELATED LITERATURE

Achievement testing is frequently viewed as an end-of-unit or end-of-course activity that is done primarily for the purpose of assigning grades or certifying mastery (see also Popham, 2002). Because of this commonly held view about the utility of test, the teachers have acquired a great deal of power over the lives of the students, for they decide who pass or fail, or proceed to higher course and finish degrees. Moreover, teachers may also find themselves wanting to make rational decisions that will help to improve achievement in the students' performance in the whole program (e.g., Mamhot, Mamhot & Kilat, 2007). Or they may find a need to make and justify changes in materials, facilities, and teaching strategy. Such decisions most often

should be made with the aid of achievement test scores (Salkind, 2003, p. 129).

In coming up with sound decisions about the achievement of students and about ways to improve that achievement usually involves testing to find out how much each person has learned within the program. Hence, an achievement test should be constructed in the context of the particular course. This requirement necessitates that the achievement test be directly based on course objectives. Achievement tests must not only be designed to measure the objectives of a given course but also be flexible enough to help teachers readily respond to what they learn from the test about the students' abilities, the students' needs, and the students' learning of the course objectives. In other words, a good achievement test can tell teachers a great deal about their students' achievements and about the adequacy of the course (Brown, 1996).

Effective teachers are responsible to their students because of their broad range of impact on their lives. Among their multifarious responsibilities is proper evaluation of the learning of students, therefore, the teachers need to ascertain that the rules governing test construction are observed to ensure fair evaluation. In other words, the test items should be of the type found in the recommendations of educational communities. Teachers have at their disposal a great variety of sources and methods for gathering information about their students. Their decisions pertaining to each of the students should be built around reliable indicators and sources of evidence. This needs to be done in order to come up with decisions which are fair and just to each of the students (Brady & Kennedy, 2001).

However, Gronlund (1993, see also Kubiszyn & Borich, 1999) said that despite the widespread use of achievement testing and the important role it plays in the instructional programs, many teachers, particularly in college level, do not have education units, receive little or no instruction how to construct good achievement tests. Moreover, accrediting agencies such as the Philippine Association of Accrediting Schools Colleges and Universities (PAASCU), where the author had the opportunity to deal with, has noted that there are some test papers included in exhibits of colleges that violate some of the rules in constructing good examinations.

Silliman University, which aims to provide quality education, is undoubtedly composed of dedicated and committed teachers. They aim for excellence in their teaching function, however, toward this end two basic requirements should be met which include willingness

and capability. It is in this perspective that this study had been conducted. As mentioned earlier, the first requirement is already a given, but the second needs verification. A teacher might be so willing to come up with an appropriate and reliable achievement test, but if he or she does not have the skills in doing it, his or her aspiration remains as is, since he or she cannot achieve it. As Protagoras (cited in Lorber, 1996) said, "Art without practice, and practice without art, are nothing." It can be drawn from the preceding passage that in order for the teacher's evaluation of students' learning to be meaningful, his or her tests should conform to the rules governing test construction (see also Linn & Gronlund, 2000).

Generally, there are two types of test to measure the learning of students in the cognitive domain: objective and essay or subjective (Linn & Gronlund, 2000). An objective test is a kind of test wherein there is only one correct answer to each item. On the other hand, an essay test is one wherein the test taker has the freedom to respond to a question based on how he feels it should be answered. Moreover, there are generally two types of objective tests namely: selection and supply. In the selection type, the student chooses the right answer to each item. Conversely, the student constructs his or her own answer in the supply type. Included in the selection category are the following: arrangement type; grouping type; matching type; multiple choice type; alternative response type; key list test and interpretive exercise. Supply type, on the other hand, includes the following: completion drawing type; completion statement type; correction type; identification type; simple recall type; and short answer type.

METHODS

There are three types of assessing the learning of students which include formative, summative and diagnostic testing (Oosterhof, 1996, p. 5). For this study, the final achievement test papers for the first semester of school year 2009-2010 of fulltime college teachers of Silliman University were used because they were summative and contained a variety of test types. The test papers were classified according to the disciplines of teachers as listed here with the corresponding sample sizes identified through cluster random sampling: Humanities (17); Math, Science and Technology (33), and Social Sciences (42). Only these numbers of test papers per classification which total to 92 were finally included in the study since some teachers did not give written

final examination and some gave problem solving or computations which were not covered in this study.

To facilitate the collection of data, the researcher communicated with the unit heads through the Vice President for Academic Affairs for their administrative support. A self-administered questionnaire was employed to obtain the profile of the faculty in terms of sex, number of years in teaching, number of education units taken, and number of seminars in test construction attended. The collection of the final examination test papers of teachers was done by a research assistant. In order to determine the conformity of the college faculty, the sample achievement test papers were analyzed using the rules adopted from Gronlund (1993). The specific rules and to what extent they were observed or violated in the construction of the sample test papers are listed in Tables 1 to 5. Meanwhile, the types of test covered in this study only include True or False, Matching Items, Multiple Choice, Short Answer and Enumeration because these were commonly employed by teachers.

The data gathered were statistically analyzed using percentage distribution, chi square, Pearson Product Moment Correlation Coefficient and analysis of variance. Percentage was used in determining the distribution of teachers who demonstrated conformity to the rules of test construction. To test whether or not a significant relationship existed between the teachers' sex and their test construction ability, chi square was utilized while Pearson r was employed in determining whether or not a significant relationship existed between the teachers' number of years in teaching, number of education units taken, number of seminars in test construction attended and their test construction ability. Finally, the analysis of variance was used to find out if the teachers' ability to conform to the rules in test construction significantly differed when they were grouped according to their disciplines: social sciences; math, science and technology; and humanities.

RESULTS

Table 1 shows that rule numbers 1, 5, 6, 7, and 10 are the ones commonly observed by the teachers; in fact, 100% of the 51 samples who included True or False type in their final examination followed these rules. This means that all of them used declarative sentences; used negative statements sparingly and did not use double negative;

attributed to some source statements of opinion; and used true propositions in statements with cause-effect relationship. Rule number 9, which requires not providing any pattern in the arrangement of the answers, was used by 96% of the teachers. Closely following next is rule number 3 with 82% of the teachers demonstrating it.

Moreover, Table 1 shows that rule number 8 is the most frequently violated which was committed by 45% of the respondents. They missed to observe it and had test items that contained words which provided clues to the answers. The use of words considered as determiners such as: always, never, all, none, only which tend to be false and usually, may, sometimes which tend to be true was also violated. The other rules which are commonly violated by the teachers are numbers 4 and 2. Number 4 is closely related to number 8. Using any of the determiners makes the statement difficult to judge whether it is true or false. The percentage manifested in rule number 2 indicates that the teachers have violated such by having more than one central idea in an item. Though these percentages are not so high, but any violation to the rule puts students at a disadvantage.

Table 1.

True or False Type of Test.

Rules in Test Construction	Observed (%)	Not Observed (%)
1. Declarative sentences should be used	51 (100.00)	—
2. Include only one central idea in each statement	33 (64.71)	18 (35.29)
3. Keep the statement short and use simple vocabulary and sentence structure	42 (82.35)	9 (17.65)
4. Word the statement so precisely that it can unequivocally be judged true or false	32 (62.75)	19 (37.75)
5. Use negative statement sparingly and avoid double negative	51 (100.00)	—
6. Statement of opinion should be attributed to some source unless used to distinguished between facts from opinion	51 (100.00)	—
7. When cause-effect relationships are being measured, use only true propositions	51 (100.00)	—
8. Avoid extraneous clues to the answer	28 (54.90)	23 (45.10)
9. In arranging the items avoid the regular recurrence of "true" and "false" statements	49 (96.08)	2 (3.92)
10. Score is number of correct answers (This holds true to all objective types of tests)	51 (100.00)	—

Meanwhile, Table 2 shows that in the matching type items, four of the six rules are commonly observed by the 36 teachers who included this type of test in their examination papers. These rules are the following: include only homogeneous material in each matching item, put all the matching items on the same page, use a larger or smaller number of responses than premises and permit the response to be used more than once, and there should only be two columns. These were observed by 94%, 83%, 72% and 69% of the faculty, respectively. On the other hand, two of the rules are commonly violated namely: place the responses in alphabetical, numerical, or chronological order (83.33%), and specify in the directions the basis for matching and indicate that each response may be used once or more than once (77.78%).

Table 2.

Matching Type of Test.

Rules in Test Construction	**Observed (%)**	**Not Observed (%)**
1. There should be two columns. Under column "A"are the stimuli which should be longer and more descriptive than the responses under column "B"	25 (69.44)	11(30.56)
2. Include only homogeneous material in each matching item	34 (94.44)	2 (5.56)
3. Use a larger or smaller number of responses than premises, and permit the responses to be used more than once	26 (72.22)	10 (27.78)
4. Place the responses in alphabetical, numerical, or chronological order	6 (16.67)	30 (83.33)
5. Specify in the directions the basis for matching, and indicate that each response may be used once, or more than once	8 (22.22)	28 (77.78)
6. Put all the matching items on the same page	30 (83.33)	6 (16.67)

As delineated in Table 3, all the thirteen rules are commonly observed by the 63 teachers who used this type of test. The teachers complying ranged from 70 to 100%. Rule number 12 being the highest with 100% of the teachers' compliance, followed by rules number 3, 5, 9, 11, and 13 with 98% of the teachers' compliance. Rule number 8 followed having 95% of the teachers' compliance then by rules 1 and 2 with 89%, then by rule number 4 with 84%. Rule numbers 6 and 7 have been observed by 79% of the teachers while rule number 10 comes last in the order with 70% of the teachers complying. While all the rules are observed by the teachers, there are also some violations committed. The rules which are considerably violated are number 10 wherein 30% of the teachers violated, 6 and 7 with 21%, rule number 4 with 16% and rules 1 and 2 with 11% each.

Table 3.

Multiple Choice Type of Test.

Rules in Test Construction	**Observed (%)**	**Not Observed (%)**
1. Construct the stem of the item in question, completion, or direction form	56 (88.89)	7 (11.11)
2. Present a single clearly formulated problem in the stem of the item	56 (88.89)	7 (11.11)
3. State the stem of the item in simple, clear language	62 (98.41)	1 (1.59)
4. Put as much of the wording as possible in the stem of the item	53 (84.13)	10 (15.87)
5. Use a negatively stated item stem only when significant learning outcomes require it	62 (98.41)	1 (1.59)
6. Emphasize negative wording whenever it is used in the stem of an item	50 (79.37)	13 (20.63)
7. Make all alternatives grammatically consistent with the stem of the item and parallel in form	50 (79.37)	13 (20.63)

Table 3.

Multiple Choice Type of Test.

Rules in Test Construction	Observed (%)	Not Observed (%)
8. Verbal associations between the stem and the correct answer should be avoided	60 (95.24)	3 (4.76)
9. Avoid "always" and "never"	62 (98.41)	1 (1.59)
10. Avoid using the alternative "all of the above" and use "none of the above" with extreme caution	44 (69.84)	19 (30.16)
11. Vary the relative length of the correct answer to eliminate length as a clue	62 (98.41)	1 (1.59)
12. Random occurrence of responses should be employed	63 (100.00)	—
13. Make certain that each item is independent of the other items in the test	62 (98.41)	1 (1.59)

Table 4 shows that among the seven rules governing short answer items, number 7 came out to be the one having no violation among the 63 teachers who used it, or not one of them formulated an item which requires the numerical answer. Rule number 1 registered a 90% compliance among the teachers. This indicates that they construct the items in such a manner that only a single, brief answer is required. Rule number 3 came third with 83% which indicates that a good number of the teachers provided only with one blank in each item.

Rule number 2 registered 77% compliance indicating that most of the items in this test are stated in interrogative form. The same table, on the other hand, delineates that there are violations in all the rules. Rule number 4 got the highest violation (41%) which indicates that a considerable number of teachers did not have the blanks in the same length. Following are rules 6 and 5 with 35% and 32%, respectively. This means that these teachers provided some clues to the correct answer and at the same time failed to put the blanks near or at the end of the sentence. In terms of rule number 6, one instance wherein

a clue is provided is by having the blanks at different lengths as well as having articles *a* or *an* right before the blank.

Table 4.

Short Answer Type of Test.

Rules of Test Construction	Observed (%)	Not Observed (%)
1. State the item so that only a single, brief answer is possible	28 (90.32)	3 (9.68)
2. Start with a direct question and switch to an incomplete statement only when greater conciseness is possible by doing so	24 (77.42)	7 (22.58)
3. Leave only one blank and it should relate to the main point of the statement	26 (83.87)	5 (16.13)
4. Blanks should be of equal lengths	18 (58.06)	13 (41.94)
5. Place the blanks near or at the end of the sentence	21 (67.74)	10 (32.26)
6. Avoid extraneous clues to the answer	20 (64.52)	11 (35.48)
7.For numerical answer, indicate the degree of precision expected and the units in which they are to be expressed	31 (100.00)	—

As shown in Table 5, enumeration is not popular among the teachers. Of the 92 teachers only 12 have used it. Nevertheless, information pertaining to how the teachers observe the rules governing this type of test is manifested. Of the three rules, number 3 has the highest percentage (75%) which indicates that a good number of the teachers did not use the phrase "at least." This is followed by rule number 1 with 67% indicating that the teachers used letters to designate the stem of the item. It is also evident that all the three rules are being violated by some of the teachers. The rule most violated is number 2 with 58%, followed by numbers 1 and 3 with 33% and 25%, respectively. Meaning to say some of the teachers were redundant. This can happen when the teacher still provides instruction in the stem or the specific item. For

example, in an item the teacher will say give/list/enumerate. For this type of test, further instruction in each item is unnecessary since the type of test already serves as an instruction.

There were also some teachers who designated the items with numbers making it difficult to reconcile with the table of specifications. It is to be recalled that the items in the table of specifications are designated with numbers. Another controversial violation is the use of the phrase, *at least*. This is something that should be avoided since this indicated that the teacher sets the minimum number of answers, but not prohibiting the students from giving all the answers. Hence, a student who committed some mistakes in the other types of test in the examination can compensate if he/she can provide all the answers in the enumeration type of test.

Table 5.

Enumeration Type of Test.

Rules of Test Construction	Observed (%)	Not Observed (%)
1. Items should be designated with letters not numbers	8 (66.67)	4 (33.33)
2. Avoid redundancy	5 (41.67)	7 (58.33)
3. Avoid using the phrase "at least"	9 (75.00)	3 (25.00)

In order to determine whether or not a significant relationship existed between the teachers' ability to conform to the rules of test construction and their profile, the percentage of the rules being observed was computed. As shown in Tables 6 to 9, in terms of the profile of teachers and their ability to follow the rules in test construction, the succeeding discussion shows that two variables are significantly related to the latter: sex and seminars attended. Table 6 particularly indicates that the teachers' ability to construct the different types of tests is influenced by their sex. The data suggest that the female teachers are better than the male teachers in terms of observance to the rules in test construction. The present data, however, cannot provide explanation to this and it is decided that a more focused inquiry on the matter has to be done in the subsequent study.

Table 6.

Test of Independence between Test Construction Ability and Sex of Teachers.

Variables	X^2		Decision	Remarks
	Computed	**Tabular**		
Test construction ability and sex of teachers	4.767	3.841	Reject Ho	Significant*

*α=0.05; df=1

It can be gleaned in Table 7 that the computed r is less than the tabular. This indicates that no significant relationship existed between teachers' ability to conform to the rules governing test construction and their number of years in teaching. In other words, irrespective of whether the teacher has been into teaching for few or more years, his or her test construction ability remains the same.

Table 7.

Test of Relationship between Test Construction Ability and Number of Years in Teaching.

Variables	r value		Decision	Remarks
	Computed	**Tabular**		
Test construction ability and number of years in teaching	-0.020	0.203	Accept Ho	Not significant

Similar to the number of years in teaching, test construction ability is not influenced or affected by the number of education units taken by the teacher as shown in Table 8. This may be test construction is

only offered in the College of Education or taken by those who are taking up teacher education and not in other degrees. This result might also indicate that in order for the teacher to acquire skills in test construction, he/she needs to enroll in the subjects which deal on it. In the revised curriculum, there are already subjects or equivalent to six units intended for test construction. In fact these subjects are among those required by the Professional Regulation Commission to be taken in order for the applicant to take the licensure examination for teachers.

Table 8.

Test of Relationship between Test Construction Ability and Number of Education Units Taken.

Variables	r value		Decision	Remarks
	Computed	Tabular		
Test construction ability and number of education units taken	-0.009	0.203	Accept Ho	Not significant

As shown in Table 9, the computed r value is greater than the tabular which indicates that there is a significant relationship between the ability of teachers to comply with the rule of test construction and the number of seminars in this area they attended. This finding delineates the importance of providing seminars on test construction especially among teachers who are not graduates of Teacher Education. Earlier it was shown that the number of education units teachers have do not relate significantly with their ability to observe the rules of test construction which suggests that regular seminars can help check the deficiency of teachers in testing.

Table 9.

Test of Relationship between Test Construction Ability and Number of Seminars in Test Construction Attended.

Variables	r value		Decision	Remarks
	Computed	**Tabular**		
Test construction ability and number of seminars in test construction attended	r = 0.636	0.205	Reject Ho	Significant

But as delineated in Table 10, the F value of 2.583 is less than the F critical of 3.099 when the three groups of teachers were compared in terms of ability to follow the rules of test construction. The result indicates that no significant difference existed among the three groups of teachers. This is confirmed by the p-value of 0.081 which is greater than the margin of error or the alpha which is 0.05. So even if a difference is evident between any two of the three mean scores, where the Social Science teachers (including Education teachers) registered the highest mean, such difference is not significant but only suggestive. In other words, the ability to observe the rules of test construction is not inherent in the discipline or degree earned by the teachers but on their attitude and willingness to apply the rules in order to realistically test the amount of learning of their students.

Table 10.

Analysis of Variance Result.

Groups	**Count**	**Sum**	**Average**	**Variance**		
Social Sciences	42	3592.6	85.538	173.98		
Math, Science, and Technology	33	2614.5	79.227	505.8		
Humanities	17	1270	74.706	303.11		
Sources of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F critical*
Between Groups	1635	2	817.53	2.583	0.081	3.099
Within Groups	28168	89	316.5			
Total	**29803**	**91**				

DISCUSSION

Of the five types of test whose rules were the subject of this paper, the rules of the True or False type of test were the most followed or observed by the college teachers included in the study. Five out of the 10 rules were followed by all the teachers while the rules for a statement to be precise so it could "unequivocally be judged true or false" and to "include only one central idea" were violated by a good number, although these were complied with by a majority of the respondents. So True or False test may be easy to prepare but the teachers perhaps failed to closely review the statements to truly measure the learning of students.

Meanwhile, only one each out of the several rules of the construction of Short Answer and Multiple Choice types of test was observed by all the teachers. Respectively, the aforementioned rules include the need to indicate the unit of measures of numerical responses and to randomize the occurrence of responses. Having blanks where students should write their answers that are of unequal length was the leading violation committed by teachers in the Short Answer type of test while the common or indiscriminate uses of choices such as "all of the above" and "none of the above" were noted in the Multiple Choice type. While having unequal length of blanks may offer clues to the answer, the use of "all of the above" and "none of the above" choices suggests that the teacher runs out of possible answers or is in a hurry to finish the test paper.

None of the rules of Matching Type test and Enumeration Type earned 100% compliance from the teachers as compared to the first three types of test discussed earlier although these are seemingly easy to prepare. Nonetheless, majority of the teachers had observed about 67% of the rules in both types of tests as compared to the 33% of the rules being commonly violated. The use of homogenous material or topic in each matching item was observed by the majority, but the requirement to place the responses in alphabetical, numerical or chronological order and to specify in the instruction the basis for matching and how this should be done were the most violated rule under the Matching Type. Meanwhile, the rule of avoiding redundant instruction was violated by the majority of the teachers, for example, the teacher says give/list/enumerate in every item asked.

In general, although the majority of the college teachers included in the study followed or observed the rules in test construction, the

number who violated certain rules in particular types of test demand a closer examination so that appropriate and specific interventions may be designed and introduced to improve testing and rating of the learning of students. The data show that the sex of teachers and the number of seminars they had attended in test construction are significantly related to their observance of the principles of test construction measured by the number of rules in particular types of test they had complied with or violated. Specifically, the female teachers and those who had attended more seminars were able to register higher adherence to the rules of test construction.

Incidentally, the number of years teaching and education units taken in college and the academic units or disciplines of teachers were not significantly related to their observance of the principles of test construction. This means that new or old teachers, those who earned or not baccalaureate degrees in education or earned the mandatory 18 units in education, and those who came from various types of disciplines do not differ with regard to their observance or violations of the rules of test construction. Teacher Education graduates may have greater advantage and familiarity about test construction rules as compared to those from other disciplines. However, Test and Measurement is just one of the courses the former had taken.

The foregoing observation may explain why attendance in a number of seminars in test construction is significantly related to the observance of test construction rules than the number of education units earned. Although an added value, it is not a guarantee that a degree in education means greater ability to develop and implement a valid test; rather the data suggest that it is the regular exposures of teachers to seminars in test construction that sharpen their ability to justly measure the learning of students and to realistically reward them so they can be inspired to pursue more learning encounters than to be frustrated due to dubious testing process. Interestingly, that the female teachers were reportedly more compliant than their male counterparts to the rules of test construction may be due to their inherent or stereotyped nurturing traits, which made them perhaps more careful in formulating test questions.

CONCLUSION

The results of this case study of Silliman University in terms of the observance to test construction rules of its teachers may be unique or

similar with other higher education institutions in the Philippines, but what is important to highlight is the fact that not all teachers are able to satisfy the requirements of good test questions that fairly measure the learning of students and allow them to be realistically rated. For example, although the rules of constructing True or False type of test is the most commonly followed or observed by teachers compared to the other types included in the study, there are still a number of them that unmindfully re-examine the quality of their test questions before administering them. Other problems found in other types of test are related to the format and test instructions that offered hints about the answers, confused the students, or are redundant.

Among the profile of teachers hypothesized to relate with adherence to the rules of test construction, only sex and attendance in seminars in test construction were found out to be significantly linked. Specifically, the female teachers tend to follow more the rules of test construction than their male counterparts which may be due to their inherent qualities and attitudinal differences in the teaching and testing processes. But this finding has to be explored more in future investigation because attendance in seminars was found to improve the quality of test questions prepared by teachers. In fact, number of years teaching and academic preparation related to Teacher Education cannot guarantee that teachers will be adept in test construction. Thus, giving more unit credits or time in pre-training and continuing education program of teachers in test and measurement are needed because the principles and techniques of testing is as important as the art and science of teaching students.

ACKNOWLEDGEMENTS

The study was funded by the Faculty Development Grant for Research of Silliman University through the Research and Development Center. In this regard, I am grateful to the following people who are instrumental in the completion of this study: Dr. Enrique G. Oracion, the Director of Research for his inspiration and expert guidance; Dr. Betsy Joy B. Tan for her administrative support; Dr. Reynaldo Y. Rivera and Dr. Earl Jude Paul L. Cleope for reviewing the initial draft of this paper; Ms. Alma Banabana who helped in the distribution and collection of the final examination papers of teachers; and to the teachers whose participation and cooperation played a significant role in this study. Needless to say, I owe sole responsibility for any opinions, errors and shortcomings this paper has.

REFERENCES

Brady, L., & Kennedy, K. (2001). *Curriculum and assessment*. Sydney, Australia: Pearson Education.

Brown, J.D. (1996). *Testing in language programs*. New Jersey: Merill.

Gronlund, N.E. (1993). *How to make achievement tests and assessments*. Boston, MA: Allyn & Bacon.

Kubiszyn, T. & Borich, G. (1999). *Educational testing and measurement: Classroom application and practice* (6th ed.). New York: John Wiley & Sons.

Linn, R.L. & Gronlund, N.E. (2000). *Measurement and assessment in teaching* (8th ed.). New Jersey: Merill.

Lorber, M.A. (1996). *Objectives, methods, and evaluation for secondary teaching*. Boston, MA: Allyn & Bacon.

Mamhot, M.R., Mamhot A. A., & Kilat, K.S. (2007). The pre-algebra course: A bridge program for mathematically under-prepared college entrants. *Silliman Journal, 48*(1), 101-115.

Oosterhof, A. (1996). *Developing and using classroom assessments*. New Jersey: Prentice-Hall.

Popham, J.W. (2002). *Classroom assessment: What teachers need to know* (3rd ed.). Boston, MA: Allyn & Bacon.

Salkind, N.J. (2003). *Exploring research* (5th ed.). New Jersey: Prentice Hall.