

**ATTRIBUTES OF TEST PAPERS AND LEVEL OF PERFORMANCE OF STUDENTS:
BASIS FOR THE FORMULATION OF THE GUIDELINES ON TEST PAPER
CONSTRUCTION**

Amadeo P. Cristobal, Jr. Ed. D.

ABSTRACT

Aiming to establish standards in Saint Paul School of Business and Law, the system of Test Constructions was given attention. Doing this in research, entails the validation of test papers used for the preliminary term examinations, school year 2014-2015 which was submitted to the office of the Academic and Research. This was randomly selected from the total of 115 test papers, 50% was subjected to validation of four (4) experts. The level of validation was then related to the level of performance of students and this served as the basis for the formulation of the guidelines in test construction.

An overall average weighted mean of 3.65 and verbally interpreted quite valid is the result of the level of validation of the test papers. The Mean of 66.26; Median of 67.00; and Mode of 60.00 reflect the level of performance of students. The level of validity of test papers has moderately small positively correlation (0.050) to the level of performance of students of Saint Paul School of Business and Law. It is therefore highly recommended to implement the formulated guidelines on test paper construction.

Article II, Quality Assurance Framework, Section 7 of CHED Memorandum Order (CMO) No.46, series of 2012, Policy-Standard to enhance Quality Assurance (QA) in Philippine Higher Education Through an Outcomes-Based and Typology-Based Quality Assurance, states "that Quality Assurance (QA) for CHED does not mean merely specifying the standards or specifications against which to measure or control quality, rather, Quality Assurance is about ensuring that there are mechanisms, procedures and processes in place to ensure that the desired quality however, defined and measured is delivered. These mechanisms, procedures and processes must be reflected in all elements of the entire organization, thus in higher education this is translated to its trifocal roles: instruction, research, and extension.

Instruction, an important role of higher education institutions (HEIs) must be given enough attention for education to be effective and efficient. The ultimate measures of which are indicated by the performance of HEIs in the different national board examinations and the outcomes in the employment of the graduates. These should be actualized in the various day to day activities in the classroom. Classroom instruction and classroom management must be attended into.

The assessment process must be considered and be given due attention as here lies the kind of outcomes higher education aspires to have. Assessment as defined is the process of judging the performance of students and it serves so many functions such as

measures of student's achievement, motivates students' learning, predicts students' success, diagnose difficulty and evaluates instruction (Raganit, 2010; Santos, 2007; Calderon, 1993; and Asaad, 2004). This can be in a form of summative, formative, placement or diagnostic assessment (Santos, 2007).

The principles of good assessment and evaluation must not be left behind. According to Navarro (2012) and Asaad (2004), assessment and evaluation should start with the vision, mission and goals of the schools. They should be based from the clear instructional objectives; they must be also comprehensive, continuous, diagnostic and functional. They must be conducted through the cooperative efforts of the instructors the students, school administrators, parents and the stakeholders in the educational community.

Assessment in the classroom is usually conducted through a series of tests which are in various forms (Rico, 2011; Santos, 2011; Navarro, 2011; Raganit, 2010; Barosse, 2005) such as subjective, objective, written works, portfolio assessment and the use of rubrics. This can be presented in a quiz, recitation, long exams, and other forms. There are many types of test to consider and choose from (Calderon, 1993; Buendicho, 2011; Corpuz, 2009): recall or supply types (simple recall, fill-in, identification, labelling, enumeration); recognition or selection type (alternative response, multiple choice, multiple response, matching type); and rearrangements of item (pictures, words).

In this paper, the test papers used during the preliminary period will be subjected to a validation process. Though there are different types of validation such as face validation, content validation, criterion validation and construct validation, it is for practical purposes that only the content validation will be utilized to propose an institutionalized format of test papers.

One important aspect in the validation is the preparation of the test papers during final examinations. The validity of the test papers must be ensured to have an accurate result of performance evaluation. Validity, one of the qualities of the good examination is defined as the degree to which an instrument measures what it intends to measure (Rico, 2011; Buendicho, 2011; Navarro, 2012; Raganit, 2010; Miller, 2005; Padua 1997). This is done by submitting test papers to at least three (3) experts to look into the appearance, directions, readability, correctness of grammar, spacing, and suitability of words (Rico, 2011). According to Gabuyo (2012); Asaad (2004); and Del Socorro (2011) the following factors affect the validity of the test papers: appropriateness of test items; directions; construction of test items; arrangement of items; difficulty of items; reading vocabulary and sentence structures; length of the test; and patterns of answers. Also, according to Santos (2007), the adequacy, objectivity, testing condition administrative procedures, usability, administrability, scorability, economy, utility, interpretability are to be checked to establish a good test.

The construction of a test is not an easy task. The teacher should develop that skill of constructing tests considering the qualities of a valid and good tests. The steps in test constructions are to identify objectives, decide the type of objective test, formulate the table of specifications, draft the test items and try out and validate (Rico, 2011; Gabuyo, 2012). Moreover, Buendicho (2011) added the following steps: determine the purpose of the test, select appropriate assessment tools, prepare them, assemble and appraise them and use the results.

Likewise the following were suggested by most of the experts mentioned above: the use of the table of specifications; write more items than

needed; write items well in advance of the testing date; write items so that they call for the performance described in the behavioural objectives; specify the tasks to be performed clearly; write items for appropriate reading; provide no clues to answer; and recheck items when revised for relevance.

Considering the dissertations conducted by Rivera (2007) and Abimbola (2012), Rivera found out that items developed were directly linked to specific content domains or objectives. After utilizing the norm-referenced test design, there was a direct alignment between the agriculture and natural resources core curriculum and the items developed. Item writing is a skill that with practice one can learn to master, but it was very difficult to find agriculture teachers with the skills to write good items, and it is equally difficult to find test specialists with expertise of the specific content. The team of item writers made up of teachers and extension agents might not have been the best group to design and write questions. They were knowledgeable in content, but lacked the skills in generating well constructed test items. Though this dissertation was focused on teaching agriculture, the procedures conducted in validation are the same with this study.

The latter recommends that examination bodies should make effort to improve the validity of their examination; need for them to hire appropriate external personnel to help in item construction; screen the items before and after administration to establish the gender equity; take interest in how teachers teach their recommended syllabi; and publish this information as books and training cassette and videotapes and also mount workshops that students and teachers will attend.

This proposal is anchored on the scientific management theory (Frederick Taylor, 1911) and system theory (Chester Barnard, 1938) cited by Robbins (2012). The former theory states that in order to develop effectiveness and efficiency the organization should develop a science in each element of the organization. Likewise, education as a system is composed of sub-systems in it. All parts of the system must be established in the bound of science to function well. One of these systems is how the performance of the students being assessed.

The paradigm on the next page shows three boxes, the criterion variable which is presented on the first box, the attributes of the test-papers which has the following sub-variables appearance; grammar and structure; variety of test types; accuracy of direction; and hierarchy of taxonomy of objectives. The variable, Level of performance of students presented on the second box. These two boxes are connected by a line which means that there is a predicted relationship between the two variables. The third box below is the proposed guidelines in the construction of the test papers for the Saint Paul School of Business and Law which serves as the offshoot of the validation process represented by an arrow.

It is the overall objective of this proposal to validate the test papers currently used during the preliminary examinations, 1st semester AY 2014-2015 and relate it to the level of performance of students towards proposing a guidelines to standardize the test paper format of the St. Paul School of Business and Law. Through this the institution will have

the mechanics and science in the preparations of test papers as an indicator of quality and excellence.

This proposal sought answers to the general statement of the problem: What relationship exists in the level of test paper's attributes to the level of performance of students of the Saint Paul School of Business and Law.

Specifically, the following questions have answered:

1. What is the level of validity of the attributes of test papers for preliminary examinations in terms of:
 - 1.1 appearance;
 - 1.2 grammar and structure;
 - 1.3 Characteristics of test ;
 - 1.4 hierarchy of taxonomy; and
 - 1.5 content level?
2. What is the level of performance of the students in different courses based from the result of the prelim examinations?
3. What is the significant relationship of the test paper's attributes to the level of performance of students?
4. What policy guidelines in test construction for theory-based and skill-development based courses is to be formulated based from the findings of the study?

Method

Research Method

This proposal utilized a quantitative-descriptive type of study. This is a scientific method which refers to a general set of orderly, disciplined procedures to acquire information which is usually based from the data which are treated statistically. Also, this aims to gather more information about characteristics within a peculiar field of study which purports to provide a picture of a situation as it naturally happens (Cristobal, 2013).

The test-papers used for the preliminary examinations across courses were described to arrive at the common characteristics.

Participants of the Study

The instructors of Saint Paul College of Business and Law comprised the population of the study which has a total of 65 instructors which was primarily based from the submitted preliminary examinations. The total of test papers submitted is one hundred fourteen (114) sets from different courses. Fifty percent (50%) of these instructors, full time and part-time were randomly chosen and about fifty percent (50%) of the test papers they have used were subjected for documentary analysis, a technique used to analyse primary or secondary sources of data which are available.

Using a stratified sampling procedures, the final documents were selected. The first stratum is the selection of the instructors, second stratum is the test papers and then the last the classification of the subjects.

The fifty (50) percent of the total instructors who submitted their test papers, about 33 of them were randomly chosen. Their names in even number as listed served as the participants of the study. The fifty percent of their test papers totalling to 57 sets from different courses were randomly chosen. Using a stratified sampling procedures, the final document was selected. First, these names were written in strips of paper then was put inside a bowl to randomly select the fifty percent (50%) of the population to serve as samples. Next, the fifty percent (50%) of the test papers constructed and utilized for the preliminary examinations by each selected sample was further randomly selected.

Sources of Data

The main source of data are the test papers actually utilized during the preliminary examinations conducted on July 8-12, 2014.

Instrument of the Study

A checklist was used as the main instrument for content validation. After reading related literatures the researcher arrived at a proposed checklist.

This checklist is composed of the variables of the study, the attributes of a test-paper. Various indicators per variable were included to fully analyse the documents and arrived at a thorough analysis of the said attributes. Each indicator was measured by the level of appropriateness through a Likert Scale.

EV - Extremely Valid (The indicator is 81% to 100% valid)

QV – Quite Valid (The indicator is 61%-80% valid)

FV – Fairly Valid (The indicator is 41%-60% valid)

SV – Slightly Valid (The indicator is 21%-40% valid)

NV – Not at all Valid (The indicator is 0% - 20% valid)

This checklist was submitted to three (3) experts in the field of education for validation purposes. Some of the comments noted were: inclusion of parallelism and use of tenses in the sub-variable grammar and structure; and the change of the likert scales from very good to needs improvement to excellently valid to not at all valid. This result was incorporated in the final checklist.

Procedures

After conducting the stratified sampling for samples and for the documents to validate, a letter was submitted to the Assistant Vice - President of the Office of the Academic and Research to gather the test papers which were submitted.

A letter was also sent to the four (4)- selected qualified validators who have the expertise on test paper construction, two(2) internal validators from SPSBL and two (2) external validators to conduct the content validation using the validated instrument checklist.

The checklists were consolidated and subjected to appropriate statistical formulae for analyses and interpretation purposes.

The result of the content validation and statistical analyses were the bases for the formulation of the guidelines for the standardized test papers for the SPSBL.

On the Development of the Guidelines for Test Construction

The development of the guidelines for test construction is developed with the use of the model, Organizational Behaviour Modification developed by Drucker (Robbins, 2000). This model is composed of six(6) primary steps:

Step 1-Identify Performance – Related Behaviour Events. This where the researcher identified the level of validation of the test papers used during the preliminary term of the SPSBL instructors.

Step 2- Measure: Baseline the Frequency of Responses. Tabulate the level of validity of test papers and measure the gap. According to Yauch (cited by Navarro, 2000) excellence can be achieved in an organization if efforts are directed at the different aspects of school operations. This will only happen if the school managers and instructors are equipped with needed competencies to assure quality. In this study, the competency in test construction must be possessed by the school managers and instructors and quality assurance is expected when competencies have value of perfect 5.0.

Step 3-After measuring the gap of the variables of the attributes of test papers, each variable was analysed with its indicators that need to be enhanced or improved.

Step 4-Develop Intervention Strategy. Based on the analysis of the attributes of test papers, the researcher developed a guideline on test construction dividing the subjects into two (2): theory-based subjects and skill-development based subjects. The guideline is developed based from intensive readings of literatures related to test construction.

Results

Table 1 -5 present the weighted mean of the level of validity of the attributes of the test papers in terms of the different sub-variables such as appearance, grammar and structure, characteristics of test, hierarchy of taxonomy and higher ordered thinking skills. This is to answer the specific statement of the problem, "What is the level of validity of the attributes of test papers for preliminary examinations in terms of the different sub-variables mentioned above.

Font style with the highest average weighted mean of 4.33 and size of fonts are rated EV, excellently valid while attractiveness with the lowest average weighted mean of 3.79 with spacing and uniformity of elements are rated QV, quite valid. The variable, appearance has an overall weighted mean

Table I

Weighted Mean of the Level of Validity of the Attributes of Test Papers in Terms of Appearance

N=57

Indicators	E1	E2	E3	E4	AWM	VI
Fonts Style	3.90	4.26	4.75	4.42	4.33	EV
Size of Fonts	3.86	4.07	4.57	4.42	4.23	EV
Spacing	3.63	4.39	4.50	3.30	3.96	QV
Attractiveness	3.75	3.54	4.40	3.45	3.79	QV
Uniformity of Elements	3.79	3.63	4.58	3.63	3.91	QV
AWM	3.79	3.98	4.56	3.84	4.04	QV

SD= .35425

Var.= .125

of 4.04 which is equivalent to QV, quite valid. The ratings of the evaluators has a standard deviation of .35425 and a variance of .125.

Table 2 shows the ratings of the evaluators on the sub-variable, grammar and structure.

Three indicators, capitalization, pluralisation with an average weighted means of 4.25 and spelling have a rating of EV, excellently valid while the remaining indicators, subject-verb agreement, use of tenses, parallelism, use of punctuations and sentence construction have similar ratings of QV, quite valid. This sub-variable, grammar and construction has an overall average weighted mean of 4.17, verbally interpreted quite valid.

The ratings of evaluators show a standard deviation of .41524 and a variance of .172.

Table 2

Weighted Mean of the Level of Validity of the Attributes of Test Papers in Terms of Grammar and Structure

N=57

Indicators	E1	E2	E3	E4	AWM	VI
Capitalization	4.01	4.14	4.93	3.93	4.25	EV
Pluralisation	4.06	4.02	4.93	4.00	4.25	EV
Subject-Verb Agreement	4.02	4.00	4.72	3.98	4.18	QV
Use of Tenses	4.07	3.90	4.75	3.96	4.17	QV
Parallelism	3.96	4.04	4.79	3.84	4.16	QV
Spelling	4.07	4.05	4.91	3.89	4.23	EV
Use of Punctuations	4.02	4.02	4.89	3.68	4.00	QV
Sentence Construction	3.89	4.02	4.49	3.95	4.09	QV
Average Weighted Mean	4.01	4.02	4.80	3.90	4.18	QV
Sd =		.41524		Var.=		.172

Table 3 shows the weighted mean of the level of validity of the attributes of test papers in terms of characteristics of test.

Table 3

Weighted Mean of the Level of Validity of the Attributes of Test Papers in Terms of Characteristics of Test

Indicators	E1	E2	E3	E4	AWM	VI
Variety of Test Types	2.96	2.82	4.49	2.54	3.20	FV
Arrangement of Items	3.16	2.80	4.60	2.73	3.32	FV
Points System	1.54	1.03	4.61	2.12	2.33	SV
Accuracy of Directions	3.05	2.72	4.05	2.49	3.08	FV
Mechanisms	3.02	1.32	4.35	3.07	2.94	FV
Average Weighted Mean	2.78	2.14	4.42	2.59	2.98	FV
Sd =		.99520		Var. =		.990

Four out of five indicators, variety of test types, arrangement of items, accuracy of directions and mechanisms have average weighted means ranging from 2.94 to 3.32 in which arrangement of items has the highest average weighted mean which are verbally interpreted fairly valid. The last indicator, points system has an average weighted mean of 2.33, interpreted SV, slightly valid. The sub-variable, characteristics of test has an overall average weighted mean of 2.94 and verbally interpreted FV, fairly valid. The ratings have a standard deviation of with a variance of .990.

The sub-variable hierarchy of taxonomy has three indicators. These are arrangement, balance of difficulty and higher ordered thinking skills. These have average weighted means ranging from 3.47 to 3.54 which are verbally interpreted quite valid. This sub-variable has an overall weighted mean of 3.50 which is verbally interpreted QV, quite valid.

Table 4

Weighted Mean of the Level of Validity of the Attributes of Test Papers in Terms of Hierarchy of Taxonomy

N=57

Indicators	E1	E2	E3	E4	AWM	VI
Arrangement	3.33	2.88	4.72	3.07	3.50	QV
Balance of Difficulty	3.35	2.95	4.68	2.89	3.47	QV
HOTS	3.42	2.82	4.74	3.19	3.54	QV
Average Weighted Mean	3.37	2.88	4.71	3.05	3.50	QV

Sd = .83024 Var.=.689

The ratings given have a standard deviation of .83024 and a variance of .689.

The sub-variable, content level has five(5) indicators. Four of it are scope, depth, clarity ad relevance. They have average weighted means ranging from 3.46 to 3.72 with relevance has the highest average weighted

Table 5

Weighted Mean of the Level of Validity of the Attributes of Test Papers in Terms of Content Level

N=57

Indicators	E1	E2	E3	E4	AWM	VI
Scope	3.42	3.28	4.72	2.42	3.46	QV
Depth	3.35	3.28	4.67	2.96	3.57	QV
Adequacy	3.28	3.21	4.16	2.68	3.33	FV
Clarity	3.49	3.21	4.70	3.02	3.61	QV
Relevance	3.49	3.21	4.65	3.51	3.72	QV
Average Weighted Mean	3.40	3.28	4.58	2.92	3.54	QV

Sd = .76879 Var.=591

mean. They are verbally interpreted, QV or quite valid. The last indicator, adequacy has an average weighted mean of 3.33 and is verbally interpreted FV, fairly valid. This sub-variable has an overall average weighted mean of 3.54, interpreted QV, quite valid.

The evaluators' ratings have a standard deviation of .76879 and a variance of .591.

Table 6 shows the answer for the stated question, "What is the level of performance of the students in different courses based from the result of the prelim examinations?"

Table 6

Measures of Central Tendencies of the Level of Performance

	Mean	Median	Mode	Cases
57	66.26	67.00	60.00	

The level of performance of students in the preliminary examinations based from the result of their final examinations has a mean of 66.26, median of 67.00 and a mode of 60.00

Table 7 on the next page shows the statistical data to answer the posted problem, "What is the significant relationship of the test paper's attributes to the level of performance of students?"

Only one evaluator's ratings on the level of validity of the test papers show a moderately small positive correlation with a correlation coefficient of 0.263 to the level of performance of students, their relationship is significant. Two evaluators' ratings on the level of validity of test papers show a very small negative correlation with correlation coefficients of -0.047 and -0.141 to the level of performance of students and the other has a moderately small

Table 7

Significant Relationship of the Test Paper's Attributes to the Level of Performance of Students

Test Paper's Attributes	r coefficient	Relationship	Significance	Decision
Evaluator 1	0.263	Moderately small positive correlation	0.480	Significant
Evaluator 2	-0.141	Very small negative correlation	0.296	Not significant
Evaluator 3	-0.047	Very small negative correlation	0.727	Not significant
Evaluator 4	0.076	Very small positive correlation	0.574	Not significant
Overall	0.050	Moderately small positive correlation	0.355	Not significant

positive correlation, 0.076. The relationship of the two (2) variables is not significant. As a whole the relationship of the level of validity of the test papers and the level of performance of students for the preliminary term is moderately small positive correlation and this relationship is not significant since the computed value of r, 0.050 is lower than the significance value of 0.355.

Discussion

The level of the validity of the test paper's attributes is assessed using the five sub-variables. These are the appearance, grammar and structure, characteristics of test, hierarchy of taxonomy and content level.

Based from the ratings, the indicators of appearance of test papers: fonts style and size of fonts are 80% to 100% properly observed while the other indicators such as spacing, attractiveness, uniformity are 60% to 80% observed. The attribute, appearance has an average weighted mean 4.04 is quite valid. The difference in the ratings of the evaluators is so low which means that their evaluation is almost unanimous. This means that the different criteria cited were 60% to 80% satisfied. There is a deficient of not more than 40% . The instructors must be careful in the construction of their test papers with respect to the various elements of appearance. They should choose the best fonts style, size of fonts and spacing. This will make the test papers more attractive and encouraging to the students especially when all the test papers have uniformity in all elements. This is in consonance to what was said by Gabuyo (2012); Asaad (2004); and Del Socorro (2011) that appearance will affect the validity of the test papers. As when the test papers are not valid they will hinder the good performance of students.

The second sub-variable, grammar and structure is an important ingredient to make the test valid. When the written words and texts are not properly written, the concepts and statements not be properly communicated and of course test takers will be confused. The following indicators: use of capitalization, pluralisation and spelling are 80% to 100% correctly done while the use of subject and verb agreement, tenses, parallelism, punctuations and sentence constructions are 60% to 80% correctly observed. As a whole, this is 60% to 80% properly used. This observation was almost unanimously seen by the evaluators as reflected by the standard deviation and variance of very low value. As emphasized by experts reading vocabulary and sentence structure are very much important in communication, the same will also do in test papers. This shows that not all instructors are good in the construction of test papers in terms of grammar and structure. They should be equipped in these competencies in order for them to construct the test correctly and be able to help students to take test easily concentrating only on the concepts learned not having much concerns on what their instructors want to imply because of the items not correctly expressed.

The third sub-variable, characteristics of test has four (4) out of five (5) indicators which are rated fairly valid. These are variety of test types, arrangement of the items, accuracy of directions and the points system was rated slightly valid. This sub-variable has an average weighted mean of fairly valid. The ratings of the evaluators are almost similar as reflected by the low variance value. This means the instructors have only 20% to 40% correctly construct their test papers in terms of its characteristics. In this element the instructors must have a thorough knowledge for this is so important to guide test takers in passing the test. As mentioned in the introduction, these indicators must be given careful attention (Del Socorro, 2011) the following factors affect the validity of the test papers: appropriateness of test items; directions; construction of test items; arrangement of items; difficulty of items; reading vocabulary and sentence structures; length of the test; and patterns of answers.

The sub-variable, hierarchy of taxonomy with its indicators: arrangement, balance of difficulty and higher ordered-thinking skills has a rating of quite valid. This means that the instructors have 61% to 80% constructed the test properly in terms of the hierarchy of taxonomy. This further shows that the instructors arranged the items not following the principle of from simple to complex or the theory of higher ordered-thinking skills. Items should be arranged from the easiest then accelerated to the most difficult purposely to motivate students to continue taking the test till the last items. Most of the education experts require the use of table of specifications to scientifically assign the number of items per taxonomy dependent upon the number of hours consumed in discussing the different topics or concepts as cited by Rico (2011) and Gabuyo (2012) that the following steps in test constructions must be followed. These are to identify objectives, decide the type of objective test, formulate the table of specifications, draft the test items and try out and validate.

The content of the test is another important sub-variable in test construction. The following indicators: scope, depth, clarity and relevance are rated quite valid while the indicator, adequacy is fairly valid. Summing it all, the sub-variable has an average weighted mean equivalent to quite valid which means that the instructors have 60% to 80% accuracy in satisfying all the concepts and sub-concepts which were discussed during the preliminary period. Although the syllabi was not used to validate the test items, it

can be inferred from the ratings that the items in the test papers did not consider thoroughly the scope and depth of the concepts discussed.

The low validation ratings given by the evaluators show that the testpapers formulated and used by the instructors of Saint Paul School of Business and Law during the preliminary term of the first semester, 2014-2015 needs a lot of improvement and revisions. The result of this study agree to the result of the study of Rivera (2007) and Abimbola (2012) that even their subject is about agriculture subjects, the same methodology and principles were used in test construction as they found out item writing is a skill that with practice one can learn to master, but it was very difficult to find agriculture teachers with the skills to write good items, and it is equally difficult to find test specialists with expertise of the specific content. The team of item writers made up of teachers and extension agents might not have been the best group to design and write questions. They were knowledgeable in content, but lacked the skills in generating well constructed test items. Differently agreed to this findings that items developed were directly linked to specific content domains or objectives.

Lastly, as cited commonly by education experts the following should be used as a guide in the construction of test papers: the use of the table of specifications; write more items than needed; write items well in advance of the testing date; write items so that they call for the performance described in the behavioural objectives; specify the tasks to be performed clearly; write items for appropriate reading; provide no clues to answer; and recheck items when revised for relevance.

In posted specific problem no. 2, based from the mean, median and mode of 66.26, 67.00 and 60.00, it could be inferred that the students who took the exam using the test-papers during the preliminary term is low considering that the passing grade is 75%. This further explains that though there are many factors to consider in having a high level of performance, it could not be denied that the manner of constructing test papers is very significant to a level of performance. That when the test papers are valid enough, which measures what it intends to measure, it will result to a higher performance. Likewise, the result of this study shows that most of the test items are not appropriate and wrongly formulated.

The result of the pearson-product moment of correlation which was computed resulted to a moderately small positive correlation. This means that when the test papers used in the final examination has a high validity it will have a small impact to the level of performance of the students or when the test has a low validity it will also have a small impact to the level of performance. The result of this study is negating the result of the study of Rivera (2007) and Abimbola (2012) that though they found out that items developed were directly linked to specific content domains or objectives, they found out that item writing is a skill that with practice one can learn to master, but it was very difficult to find agriculture teachers with the skills to write good items, and it is equally difficult to find test specialists with expertise of the specific content. The team of item writers made up of teachers and extension agents might not have been the best group to design and write questions. Though their study is on agriculture, the principles in test construction are exactly the same.

This study therefore, concluded that though the level of validation on the attributes of test paper does not significantly relate to the level of performance of students of Saint

Paul School of Business and Law, it could not exclude the fact that test papers construction is so important towards a high level of performance. Furthermore, it is recommended that a training on test construction and English proficiency which includes rules in capitalization, pluralization, use of tenses, parallelism, use of punctuation among others must be conducted and the proposed guidelines for test construction be observed for standardization purposes.

SPSBL GUIDELINES FOR TEST CONSTRUCTION

1. Appearance
 - 1.1 Margins: .5 inch in all sides
 - 1.2 Font type and size – Arial; 12 points
 - 1.3 Spacing - double spaced
 - 1.4 Size of paper – 8.5 x 13
 - 1.5 Printing – RISO printed or originally printed
 - 1.6 Pagination – center of the bottom margin
2. Grammar and Structure – Observe and apply the rules for capitalization, pluralization, subject-verb agreement, use of tenses, parallelism, use of punctuations and spell words correctly.
3. Characteristics of Test
 - 3.1 Test Types

For theory-based subjects - Use at least 3 types of objective types (80%): (Identification, Completion, Labelling, Enumeration, Alternative Response (TF), Matching, Multiple Choice, Rearrangement) and a Restricted Essay type (20%).
For skill-development based subjects –cognition (25%) and skill domain (Problem Solving) types (75%)
 - 3.2 Test Types must be arranged according to the level of difficulty.
 - 3.3 Points per item or per test type must be defined. Easy item, 1 point and difficult items, more points (total points for final examination=50 points)
 - 3.4 Directions should be clear on what tasks to do, where to write and other significant requirement
4. Hierarchy of Taxonomy of Objectives
 - 4.1 Arrangement of Items – items must be arranged from simple to complex; from the easiest item to answer to the most difficult
 - 4.2 Difficulty of Items – items should be more on the development of analytical, critical, and decision making skills.
 - 4.3 Higher- Ordered Thinking Skills – utilize the use of table of specifications
5. Content Level – satisfy the standardized syllabi
 - 5.1 Scope – Confine items on the content that is developed in the period based from the approved syllabus.
 - 5.2 Depth – Include the details of the concept in the test.
 - 5.3 Adequacy – Remember the time spent in developing the concept for it is the basis of how many items/points to be given.
 - 5.4 Clarity – Use clear, concise and specific statements in simple sentences to be easily understood by test takers.
 - 5.5 Relevance – Include practical scenario or incidence that is happening locally, nationally or internationally.

6. Others

6.1 The draft of the test papers must be made 10 days before the scheduled final examinations.

6.2 The drafted test papers will be submitted to the dean/academic heads for checking of the content and the format.

6.3 Final copies of the test papers must be signed at the back for printing purposes.

6.4 A training on the following must be scheduled:

6.4.1 test construction

6.4.1 grammar and structure

References

A. Published Materials

Asaad, Abubakar S and Hailaya, Wilham M. (2004). *Measurements and evaluation: concepts and principles*. Quezon City: Rex Bookstore, Inc.

Barosse, Emily (2005). *Classroom assessment: key concepts and applications, 5th edition*. NY; McGraw – Hill.

Buendicho, Flordeliza (2010). *Assessment of Learning I*. Phils: Rex Bookstore, Inc.

Calderon, Jose F. and Gonzales, Expectacion C. (1993) *Tests and measurements*. Pasig City: Capitol Publishing House.

Costa, Arthur L. (2004). *Assessment strategies for self-directed learning*. California: Sage Publications, Ltd.

Cristobal, Amadeo Jr. P. and Cristobal, Maura Consolacion D. (2013). *Research made easier a step by step process*. Quezon City: C & E Publishing Incorporated.

Del Socorro, Felicidad R. et al. (2011). Quezon City: Great Books Publishing.

Gabuyo, Yonardo, A.(2012). *Assessment of learning I textbook and reviewer*. Quezon City: Rex Printing Company, Inc.

Kachru, Upendra (2007). *Production and operations management, text and cases*. New Delhi: Excel books, p.5

Miller, I and David, I.(2005). *Educational tests and measurements*. US: Pearson Prentice Hall.

Navarro, Rosita L and Santos, Rosita G. (2012). *Assessment of Learning Outcomes (Assessment 1), 2nd Edition*. Aurora Blvd: Lorimar Publishing, Inc.

Padua, Roberto N. and Santos, Rosita G (1997). Quezon City. Katha Publishing Co. Inc.

Raganit, Arnulfo R. et al (2010). *Assessment of student learning I (cognitive learning)*. Quezon City: C & E Publishing , Inc.

Rico, Alberto (2011). *Assessment of student's learning: a practical approach*. Philippines: Anvil Publishing, Inc.

Robbins, Stephen R. (2000). *Organizational behaviour*. San Diego, USA: Prentice Hall International, Inc.

Robbins, Stephen R. and Coulter Mary (2012). *Management global edition, 11th edition*. Boston: Pearson Education Limited.

Santos, Rosita (2007). *Advanced methods in educational assessment and evaluation of learning 2*. Aurora Blvd: Lorimar Publishing House.

Santos, Rosita D. (2007). *Assessment of learning I*. Aurora Blvd: Lorimar Publishing Inc.

B. Unpublished Materials

Abimbola, Isaac O. (2012), *Advances in the development and validation of instruments for assessing student's science knowledge*. University of Llorin: Department of Curriculum Studies and Educational Technology

Rivera, Jennifer Ellain (2007). Cornell, University: unpublished dissertation.

APPENDIX A

SAINT PAUL SCHOOL OF BUSINESS AND LAW

Palo, Leyte

PROBLEM: *ATTRIBUTES OF TEST PAPERS AND LEVEL OF PERFORMANCE:
BASIS FOR THE FORMULATION OF THE GUIDELINES ON TEST PAPER
CONSTRUCTION*

DIRECTIONS: Indicate the degree of appropriateness of the different attributes of the test papers. Check the appropriate column.

EV - Extremely Valid (The indicator is 81% to 100% valid)

QV – Quite Valid (The indicator is 61%-80% valid)

FV – Fairly Valid (The indicator is 41%-60% valid)

SV – Slightly Valid (The indicator is 21%-40% valid)

NV – Not at all Valid (The indicator is 0% - 20% valid)

ATTIBUTES OF TEST PAPERS	EV	QV	FV	SV	NV
Appearance					
1. Fonts style					
2. Size of the fonts					
3. Spacing					
4. Attractiveness					
5. Uniformity of elements					
Grammar and Structure					
1. Capitalization					
2. Pluralization					
3. Subject-verb agreements					
4. Use of Tenses					
5. Parallelism					
6. Spelling					
7. Use of Punctuations					
8. Sentence Construction					
Characteristics of Test					
1. Variety of Test Types					
2. Arrangement of test types					
3. Points system					
4. Accuracy of Directions					
5. Mechanisms (Clarity of Directions)					
Hierarchy of Taxonomy					
1. Arrangement of items					
2. Balance of Difficulty					
3. HOTS (Higher Order Thinking Skills)					
Content Level					
1. Scope					
2. Depth					
3. Adequacy					
4. Clarity					
5. Relevance					