Cognitive Test Construction Skills of Prospective Teachers: Ensuring Quality Student Learning Outcomes

Jenifer Raymond R. Tallungan[1] and Demetria A. Corpuz[2]
jenz2319@gmail.com
[1] Department of Teacher Education, Nueva Vizcaya State University, 3702 Bambang, Nueva Vizcaya, Philippines
[2] Department of Teacher Education, Nueva Vizcaya State University, 3700 Bayombong, Nueva Vizcaya, Philippines

Keywords:

cognitive test construction skills, higher order thinking skills, prospective teachers, validity, usability

ABSTRACT

The success of a learning community lies on how the students may benefit from the curriculum which is fundamentally anchored on constellation of human needs. This study delved on the cognitive test construction skills of prospective teachers in a state university of Cagayan Valley who will eventually partake enormously in such success through assessment. Through a writing simulative activity, the level of cognitive test construction skills of 103 randomly selected respondents are unveiled to be high along multiple choice, matching, alternate response, and simple recall types, however average in completion and essay types. Further, their level of proficiency in establishing general validity and usability of cognitive tests is high while the level of their proficiency in preparing cognitive tests that measure higher order thinking skills (HOTS) is average. At 0.05 level, significant correlations were disclosed between their proficiency in preparing cognitive tests and their proficiency in establishing general validity and usability while otherwise with their proficiency in writing tests that measure HOTS. The results prompted better perspectives on the teacher education students' enhancement experiences in preparing quality assessment tools.

INTRODUCTION

The yardstick of student learning is assessment. Its residue is the learning outcomes that gear students in a highly technologically and scientifically propelled society. Effective assessment brings about change through any effort of an educational entity of fortifying skills in realizing objectives aligned to the goals of global education. The meaning of learning lies hence on what is retained in the head, heart and hands of the students and how they can use such learning in surmounting the challenges of modern life.

Assessment is a systematic process of gathering and creating full range of information through informal and formal ways such as observation or verbal exchange, assignments, tests, written reports or outputs, portfolio, rubrics, essay journals, rating scales, and checklist among others. According to Garcia (2008), it looks into account how much change has occurred on the students' acquisition of knowledge, skills and values before and after a given learning activity.

An educational activity that will never be faltering in meeting the goals of the teaching-learning process is preparation of pen and paper tests whose form is authentic in licensure or board examinations. Through the formative and summative tests rendered by the teacher, he is able extract important information which serves as reference for educational decisions thereby boosting learning experiences.

Before embarking into student teaching, the prospective teachers are offered

with a laboratory where they could acquire professional skills in preparing tests that optimize student learning outcomes. Tests are prepared to measure the holistic development of an individual – cognitive, affective and psychomotor learning.

Giving focus to cognitive learning in this light, one of the specific competencies wherein prospective teachers are honed is writing objective and subjective tests like multiple choice, matching, alternate response, completion, simple recall and essay types. Moreover, skills in establishing qualities in assessment tools like validity, reliability and usability prove equal foothold in writing effective examinations. On another face, the skills in preparing the blueprint of test likewise share importance in putting emphasis to competencies needing more weight. In the revised Bloom's taxonomy, the first three hierarchical levels could be regarded under the lower order thinking skills (LOTS) and the three remaining, higher order thinking skills (HOTS).

In this perspective, a thorough study of assessment skills of junior teacher education students was carried out with the title: "*Cognitive Test Construction Skills of Prospective Teachers: Ensuring Quality Student Learning Outcomes*" to highlight the importance of test construction skills in enhancing student learning outcomes. The nose of the prospective teachers for validity, usability and prepping for test that measures higher order thinking skills were likewise taken into consideration.

The findings of this study ensures quality student learning outcomes as it investigates the level of cognitive learning the test items that the prospective teachers prepare. It is expected that learning is optimized by offering the learners experiences they can relate with real-life situations. This could be materialized as well by considering items in assessment tools which measure higher order thinking skills (HOTS).

This study in sought to determine skills of prospective teachers in preparing supply and selection examinations and how these skills may correlate with skills in establishing validity and usability, as well as how they may relate with skills in preparing items that measure HOTS.

Specifically, this study aimed to find answers to the following research questions:

1. What is the level of proficiency of the prospective teachers of NVSU in preparing cognitive tests along multiple choice, matching, alternate response, completion, simple recall and essay types?
2. What is the level of proficiency of the respondents in establishing the general validity and usability of cognitive tests?
3. What is the level of proficiency of the respondents in preparing cognitive test items that measure higher order thinking skills (HOTS)?
4. Does the proficiency of the respondents in preparing cognitive tests correlate significantly with:
   - their proficiency in establishing general validity of their prepared cognitive tests;
   - their proficiency in establishing the usability of their prepared cognitive tests; and
   - their proficiency in preparing test items that measure higher order thinking skills (HOTS)?

This study hence tested the null hypotheses: there is no significant relationship between the respondents' proficiency in preparing cognitive tests and their proficiency in establishing general validity; there is no significant relationship between the respondents' proficiency in preparing cognitive tests and their proficiency in establishing usability; and there is no significant relationship between the respondents' proficiency in preparing cognitive tests and their proficiency in

preparing test items that measure higher order thinking skills (HOTS).

METHODOLOGY

The study employed the descriptive research design to appropriately characterize the cognitive test construction skills of prospective teachers as well as their proficiency in establishing general validity and usability of tests and proficiency in preparing cognitive tests that stimulate higher order thinking skills (HOTS). The correlational approach was used to divulge extent and significance of relationship between the dependent and independent variables of the study.

The study was conducted toward the end of the first semester, SY 2015-2016 involving 103 randomly selected third year teacher education students of the Nueva Vizcaya State University-Bambang Campus, taking 52.28% of the entire junior population. Such number is acceptable for descriptive researches which may consider at least 20% of a small population (Cudia *et. al.*, 2015).

The student respondents wee initially asked to prepare a Table of Specifications (TOS) from which percentages of items under HOTS or LOTS were classified. They were then asked to write a one-hundred (100) items test with six sections wherein all the objective types of test (multiple choice, matching, alternate response, completion and simple recall types) as well as the essay type were incorporated. They were given a common chapter as reference in constructing the test.

The standards of evaluating the prospective teachers' prepared examinations in terms of validity and usability as well as in evaluating the individual objective tests and the essay test, are taken from the stipulations of assessment specialists as follows: Raagas (2010), Garcia (2008), Calmorin (1994), Oriondo & Antonio (1989), and Maranang & Maranang (1984). The examinations prepared by the respondents were evaluated by assessment instructors and students.

The following scale was used to interpret the respondents' level of proficiency along preparing cognitive test, preparing items that measure higher order thinking skills, establishing general validity, and establishing usability:

Preparing cognitive tests and preparing items that measure higher order thinking skills	Establishing general validity and usability	Level
81-100	4.20-5.00	Very High
61-80	3.40-4.19	High
41-60	2.60-3.39	Average
21-40	1.80-2.59	Low
1-20	1.00-1.79	Very Low

Statistical tools used to derive answers to the research questions were means, percentages and correlational procedures using 0.05 level of significance.

RESULTS AND DISCUSSION

Level of proficiency of the prospective teachers of NVSU in preparing cognitive tests along multiple choice, matching, alternate response, completion, simple recall and essay types

Table 1 divulges the summary of respondents' proficiency in test preparation along the six types: multiple choice, matching, alternate response, completion, simple recall and essay types. The first three are selection types while the three latter are supply tests.

Observing further the data in the table, supply tests are areas in which the respondents are less proficient wherein they garnered overall means of 53.98 (average) for essay, 60.18 (average) for completion and 65.13 (high) for simple recall.

In the selection types of test, the respondents obtained overall means of 76.44

Table 1. Level of Proficiency of Respondents in Preparing Cognitive Test

Cognitive Test Type	Mean	Level
Multiple Choice	71.73	High
Matching	76.44	High
Alternate Response	70.28	High
Completion	60.18	Average
Simple Recall	65.13	High
Essay	53.98	Average
Overall Mean	**66.29**	**High**

(high) for matching type, 71.73 (high) for multiple choice type and 70.28 (high) for alternate response type. The data show that the respondents manifest more proficiency in preparing selection type tests like matching, multiple and alternate response types.

Multiple Choice Type. This is a type of test which is commonly administered to licensure or professional examinations because of its capacity to measure learning up to the highest cognitive level. Following the guidelines of multiple test preparation will ensure considerable validity, reliability and usability of the test.

The junior students obtained an overall mean of *71.73* or qualitatively categorized as *high*. This could be attributed to some advanced skills of the teacher education students under study in maintaining constant number of options in all items of the test, constructing the main stem of the test item using statement form, direction form, completion form or question form, using 4-5 options in each item to avoid chances of guessing and of obtaining correct answers by logical elimination, constructing the stem or question such that there is only one correct answer, not several possible answers, avoiding the use of "none of these" or "all of these" as one of the options, arranging correct answers following any pattern and avoiding articles "an" and "a" as last word in an incomplete sentence to avoid clues to expected answer.

Number of items which are constructed based on the foregoing guidelines could be qualitatively described as very high. Generally, the strength of the student test writers lies on the rules underlying construction of options or choices.

Some of the rules in writing the options which were observed by the prospective teachers run parallel with the provisions of Gronlund and Linn (1990) that test constructors should write the distracters to be plausible yet clearly wrong. An important, and sometimes difficult to achieve is ensuring that the incorrect choices (distracters) appear to be possibly correct. Distracters are best created using common errors or misunderstandings about the concept being assessed, and making them homogeneous in content and parallel in form and grammar.

On the other hand, the respondents tend to write questions that could be addressed through rote learning and memory which should be avoided. In similar sense, the respondents also write test items or statements borrowed directly from the books or other reference materials. Also, the prospective teachers find difficulty in structuring all options in an item in parallel language, maintaining grammatical consistency and formulating statement that tap single ability only.

Synthesizing their weaknesses from low to average levels, the respondents find difficulty in preparing items that are rephrased not to encourage memorization but understanding and even higher level.

Matching Type. This is another type of cognitive test through which test takers select from logically arranged choices written opposite the array of stems. This test can measure from not only remembering but evaluation as well of concepts in the lesson.
The respondents registered a performance under this test type with a mean of *76.44* or *high*. This rating could be characterized by remarkable evaluation given to the respondents in the following specific guidelines: score is the number of correct answers which should have one point each; the stimuli under column

A should be numbered and responses under column B should be lettered; the test contains column B which represents the options from which the taker selects the correct answer; the item column must be placed at the left and the option column at the right; make sure that each item has a pair in the option column; and there should be only one correct response in each item. All of which are qualitatively described as very high. The strength of the respondents lies on assuring placement and format of the options in the correct columns.

The strength of the prospective teachers is in conformity with the rules set by Wiggins (1998), and McMillan (2001) who asserted that in matching type test, brief portion should be kept in the left column, and shorter responses should be placed on the right, responses should be in logical order and placed in alphabetical order or sequence.

Conversely, the respondents pegged average as qualitative description of their level of proficiency in constructing a matching type test in terms of following guidelines stipulating that the number of options should be more than the items in column A or there should be at most three (3) distractors/distracters; and options in column B expressed in words or statements must be arranged alphabetically and dates in chronological order to facilitate the selection of correct answer. Both indicators are qualitatively described as average.

Among the selection and supply types of test considered in this study, matching type could be considered the area wherein the respondents displayed their optimum test preparation skills.

Alternate Response Type. This is a selection type of test making use of only two options, e.g. natural dichotomy as in true or false, yes or no; or any other paired responses or artificial dichotomy as in X or M and O or Y.

In preparing such a test, the respondents, recorded an overall mean of *70.28* or *high*. Detailing the evaluation, the respondents garnered very high marks for this type because of avoiding qualitative language like several and many whenever possible; avoiding unfamiliar, figurative or literary language; avoiding using negative statements especially double negatives; avoiding using commands in a true-false type of test- commands cannot be true or false; and using declarative sentences for true-false type while interrogative sentences for yes-no type. Ratings are qualitatively described as very high.

This result is in conformance with the provision of Gronlund and Linn (1990) that while constructing true-false items, attempts should be made to avoid trivial, broad, general and negative statements.

On the other hand, the weak points of the test writers lies on indicating by short line or by ( ) where the response is to be recorded - the responses may be arranged in a column preferably at the right of the items (average); arranging the items in the test into clusters, with five statements in each cluster (low); and avoiding lifting statements directly from books (low). The same weakest point could be observed in the succeeding types wherein the student writers could hardly deviate from the words of the reference materials or textbooks.

Completion Type. One of the more challenging test types that prospective teachers prepare is the completion type. The respondents obtained an overall mean of *60.18* or *average* along this area of evaluation. This type of test is commonly known as "fill-in-the-blanks". This type of test is rarely included in licensure or professional board examinations because it tends to measure rote memorization which is not inclined to outcomes-based approach. Further, it appeals to learners who have flair of lower order thinking skills (LOTS).

The overall performance of the respondents in this domain could be attributed to their adherence to the rules that the required response should be a simple word or a brief phrase (very high), allow one point for each correctly filled blank (very high), make the

blanks in uniform length - do not indicate the expected answer by varying the length of blanks or by using a dot to represent each letter in the correct word (high).

However, the respondents find it hard to avoid lifting statements directly from the text. This finding is opposed to the standards set by Wiggins (1998) and McMillan (2001) that direct statements from textbooks should not be taken as an item.

Other areas of difficulty are writing items in the form a(n), so that the examinees must decide whether the correct answer begins with a consonant sound or with a vowel sound; arranging the test so that the answers are in a column at the right or left of the sentences; avoiding statements which do not clearly specify the expected answer so that there is no more than one possible answer to only one blank; avoiding statements which will create confusion among the examinees relative to the subject area included in the item and to prepare scoring key that contains all acceptable answers. All indicators referring to their weaknesses are classified either as low or average.

The multiplicity of responses to a completion type of test is one consideration in crafting the test alongside its challenge of elevating the level of cognitive learning not just a memory test. Rewording, rephrasing or rewriting sentences can help in making the test gauge higher thinking skills.

Simple Recall Type. This type of test is commonly known as identification type either in clause, question or sentence forms. It may also come in the form of questions or statements asking for a list of objects or the more commonly known as enumeration type. Again, this type of test is rarely included in licensure or professional board examinations because it tends to measure rote memorization which is classified under lower order thinking skills (LOTS). There are some techniques however which could reduce the possibility of a simple recall test to be used as a test memory alone. Just like in the completion type, rewording, rephrasing or rewriting sentences can help in making the test more useful.

For the simple recall test, the respondents obtained an overall mean of *65.13* described as *high* though ranked as 3rd among their weak test types. The respondents obtained a high level in areas like making the test item so worded that the response is brief preferably a simple word, number, symbol or a very brief or simple phrase, preferring the direct question form over the statement form, making the question so worded that there is only one correct answer, including all acceptable answers in the scoring key, and putting the blanks for the responses in a column at the left or right of the items. All evaluation for this area are qualitatively described as high.

The only item falling under average level is the respondents' making a minimum use of textbook language. Because of this, unfamiliar phrasing may reduce the possibility of correct responses that represent more meaningless verbal association.

Same weakness could be noted in completion type that test items could best serve their purpose if they are not taken directly from printed resources but from they are presented in a manner that higher order thinking skills are motivated.

Essay Type. This is a test type with low validity, reliability and usability because of its nature of lending its scoring to subjectivity if there are no specific standards through which a written response to a stimulus may be evaluated (Cashin, 1987; Worthen *et. al.*, 1993). Preparation of such a test likewise is challenging considering elements like time, space, level of students proficiency in the language and in the subject per se, directions, and even the level of learning that the essay item intends to measure. Scoring in the same manner is challenging due to the tendency of answers to be divergent. More personalities mean more approaches to answering essay items, thus requiring thorough reading of responses and establishing evaluation through

rubrics.

In this study, the respondents pegged an overall mean of *53.98* or *average*, the least performance among all areas of evaluation. Among the strengths though of the respondents along this type include that examinees are not made to choose which among a number of items they would work on- optional question(s) are not given (very high); each question is clear and indicates the task for students to perform; questions are phrased or carefully designed to elicit the particular aspects of behavior stated in the learning outcomes; items are not merely asked for repetition of memorized facts; test takers are made to use the higher level skills of analysis and evaluation by use of specific verbs, e.g. defend, compare, evaluate and facilitate. All levels for the latter indicators are high.

Since essay tests could measure higher order thinking skills, it should necessitate critical and creative thinking through *why* and *how* questions. For the respondents, the challenge of preparing this type of test lies on giving the relative scoring weights of the key points to be discussed (average); constructing a detailed key for each item especially if the scoring procedure to be used is analytic (average); stating the scope and length of the required sentence(s) in each item (low); and indicating the appropriate time limit for each question so that the students can pace their answering accordingly for a series of essay questions (very low).

In summary, the greatest concern of the student teachers is allotting conditions or restrictions for time and space/length when they construct essay questions. This can make essay examinations efficient and beneficial to the teacher.

Level of proficiency of the respondents in establishing the general validity and usability of cognitive tests

In order to come up with a true measure of learners' achievement in the classroom, teachers may exert effort in preparing measuring instruments that possess good qualities like validity, reliability and usability. One of the objectives of teacher is to collect data that will describe and determine the level of performance of students through these measuring instruments. The concern of this study is to look into the proficiency of prospective teachers in establishing general validity and usability of cognitive tests.

Validity. This refers to the extent to which the test serves the purpose for which it is constructed or the efficiency with which the test measures what it intends to measure. It is the most important criterion of a good examination. A valid test is always reliable. (Calmorin, 1994)

To evaluate the skills of the respondents in establishing general validity, the factors affecting validity according to Gronlund as cited by Oriondo and Antonio (1984) served as the standards of evaluating the tests prepared by the student teachers. These factors include appropriateness of test items, directions, reading vocabulary and sentence structures, difficulty of items, construction of test items, length of the test, arrangement of items and patterns of answers as displayed in table 2.

The respondents' proficiency in establishing general validity could be described with an overall mean of *3.72* or *high*. For the respondents, the easiest procedure through which validity may be established is observing no patterns of answers (very high). When examinees obtain the correct answer because of patterns, then the result is not valid.

This is followed by preparing a test with acceptable length (high). A test may be of sufficient length to measure what it is supposed to measure. A test that is too short cannot adequately sample the performance or behavior of the learners. Thirdly, the respondents prepared tests with appropriate test items (high). This refers to the extent to which the test items were constructed to serve the purpose for which it is designed.

Table 2. Respondents' level of proficiency in establishing general validity in cognitive tests

Indicators	Mean	Level
Appropriateness of Test Items	3.70	High
Directions	3.66	High
Reading Vocabulary and Sentence Structures	3.65	High
Difficulty of Items	3.52	High
Construction of Test Items	3.60	High
Length of the Test	3.80	High
Arrangement of Items	3.52	High
Patterns of Answers	4.32	Very High
Overall Mean	**3.72**	**High**

Table 3. Respondents' level of proficiency in establishing general usability in cognitive tests

Indicators	Mean	Level
Ease of Administration or Administrability	3.49	High
Ease of Scoring or Scorability	3.82	High
Economy	3.67	High
Interpretability	3.63	High
Objectivity	4.04	High
Adequacy	3.53	High
Comparability	3.70	High
Overall Mean	**3.69**	**High**

On the other hand, the areas in which the respondents considered establishing validity more challenging is in the arrangement of items and difficulty of items though both are still qualitatively categorized as high.

Usability. This refers to the degree to which the measuring instrument can be used by teachers, supervisors, and school administrators without undue expenditure of time, money, and effort. It means practicability. (Calmorin, 1994)

To evaluate the skills of the respondents in establishing general usability, the factors determining usability of tests according to Maranang & Maranang (1983), Garcia (2008), Raagas (2010) served as the standards of evaluating the tests prepared by the student teachers. These factors include ease of administration or administrability, ease of scoring or scorability, economy, interpretability, objectivity, adequacy and comparability as displayed in table 3.

The proficiency of the respondents in establishing general usability of tests could be characterized by the overall mean of *3.69* or *high*. All of the areas are qualitatively categorized as high. Among the seven standards included in evaluating the tests prepared by the respondents, the prospective teachers observed well the objectivity of questions (high). Scorability of test items is another strength of the tests that the respondents have prepared (high).

Conversely, in this area they manifested least skills in ease of administration or administrability of the tests (high) and adequacy (high). To make administration of the test easy, the test may have included clear, simple and direct instruction to the examinee, to the examiner, and to the scorer. Sample test exercises may be illustrated to clarify the instructions for performing the test.

Moreover, adequacy is the degree to which a test contains a fairly wide sampling of items to determine the educational outcomes or abilities so that the resulting scores are representative of the total relative performance of the students in the areas measured.

Level of proficiency of the respondents in preparing cognitive tests that measure higher order thinking skills (HOTS)

The category of learning objectives under the cognitive domain of learning according to Raagas (2010) is connected with the intellectual component of behavior. The revised Bloom's taxonomy considers

the following hierarchy of cognitive learning: remembering, comprehending, applying, analysis, evaluating and creating. Effective examinations do not only include LOTS items but more importantly, HOTS items which develop critical and creative thinking, thereby making learning more meaningful.

The respondents were tested along this area by letting them prepare TOS through which the level of cognitive learning is inspected. The percentages of HOTS were observed and summarized in table 4.

Most of the respondents (46 or 44.66% of the total respondents) obtained evaluation between 41-60 or qualitatively described as average; followed by 27 respondents or 26.21% who garnered evaluation between 61-80 or high; closely next are 23 respondents or 22.33% who were rated between 21-40 or low. Seven respondents or 6.80% on other hand are classified as either very low or very high.

Overall, the respondents registered a mean performance of *52.63* or *average*. In this area, the respondents' general proficiency in preparing test items that measure HOTS is considerably unremarkable which means that this area is needing improvement. In this light, the respondents could be trained how to prepare items categorized under analyzing, evaluating and creating, in other words test that develop higher order thinking skills.

Analysis of correlation of proficiency of the respondents in preparing cognitive tests with proficiency in establishing the general validity and usability of cognitive tests and proficiency in preparing cognitive tests that measure higher order thinking skills (HOTS)

Correlational procedures were employed to point out significant relationship that may be derived between the proficiency of the respondents in preparing cognitive tests and proficiency in establishing validity, usability and preparing test that measure HOTS.

Table 5 provides that all correlation coefficients computed for validity and usability exceeded the critical r-values thus corresponding to p-values less than the level of significance which is equivalent to 0.05. All such values lead to the rejection of the null hypotheses correspondingly. This implies that the skills of the respondents in preparing cognitive tests - the supply and selection types of tests - are significantly correlated with their proficiency in establishing general validity and usability. This further means that observing the specific rules of writing objective and subjective tests ensures general validity and usability of the tests.

Nevertheless, looking into the probability values under the correlational procedure carried out for test types and HOTS, all p-values are greater than the 0.05 level of significance which means that the null hypotheses are accepted correspondingly in this domain. This implies that the proficiency of the respondents in writing cognitive tests is not significantly correlated with their proficiency in writing tests that measure HOTS. Such finding is indicative of the respondents' weaknesses in deciphering items that measure critical and creative thinking.

Table 4. Respondents' level of proficiency in in preparing cognitive tests that measure higher order thinking skills (HOTS)

Level	Range	Frequency	Percentage
Very High	81-100	6	5.83
High	61-80	27	26.21
Average	41-60	46	44.66
Low	21-40	23	22.33
Very Low	1-20	1	0.97
Total	103		100.00
Mean	52.63		Average

CONCLUSIONS AND RECOMMENDATIONS

After careful investigation and

Table 5. Summary of Correlational Analysis of the Study Variables

Test Type	Statistic	Validity	Usability	HOTS
Multiple Choice	r	0.3237	0.4236	0.0373
	p	0.0009*	0.0000*	0.7083
Matching Type	r	0.3456	0.3283	0.0793
	p	0.0004*	0.0007*	0.4258
Completion Type	r	0.2721	0.3285	0.0164
	p	0.0054*	0.0007*	0.8691
Simple Recall	r	0.2685	0.2888	0.0249
	p	0.0061*	0.0031*	0.8029
Alternate Response Type	r	0.2010	0.2032	0.0480
	p	0.0417*	0.0395*	0.6300
Essay	r	0.3049	0.3229	0.1827
	p	0.0017*	0.0009*	0.0647

presentation of data collected to address the purpose of this research piece, the following conclusions were derived:

1. The level of proficiency of junior prospective teachers of Nueva Vizcaya State University- Bambang Campus in preparing selection and supply tests is high. Specifically, their level of proficiency in preparing multiple choice, matching, alternate response, and simple recall types of test is high while in preparing completion and essay types, average.
2. The level of proficiency of the respondents in establishing general validity and usability of cognitive tests is high.
3. The level of proficiency of the respondents in preparing cognitive tests that measure higher order thinking skills (HOTS) is average.
4. The proficiency of the respondents in preparing cognitive tests is significantly correlated with their proficiency in establishing general validity and usability while their proficiency in writing cognitive tests is not significantly correlated with their proficiency in writing tests that measure HOTS.

These findings suggest more learning experiences for the prospective teachers of the university in terms of test construction specifically along completion and essay types. Likewise, emphasis may be given on improving proficiency of the prospective teachers in preparing cognitive tests that measure higher order thinking skills, After all, learning is optimized by offering experiences to their future students which are assessed through cognitive tests based on evaluating and creating levels. This as well urges the prospective teachers to utilize authentic assessment aside from the pen-and-paper test to bring learning into the real world.

In the light of the above salient findings, the following are recommended:

Educational Training. That an in-house seminar may be designed and implemented before deployment to equip pre-service teachers with competencies pertinent to their future profession, e.g. cognitive test construction, preparation of Table of Specifications with emphasis on higher order thinking skills, the lesson objective and qualities of a good measuring instrument;

Establishing Quality of Tests. That skills of student teachers along establishing validity, reliability and usability may be

strengthened to assure effective evaluation of test content and format, consistency of responses of test takers and efficiency of measuring instruments;

Other Researches. That other researches may conducted to disclose correlation of test construction skills of teacher education students among the three domains of learning (cognitive, affective and psychomotor) to ensure holistic learning.

REFERENCES

Calmorin, L., Educational Measurement and Evaluation, National Book Store, Philippines, 1994.

Cashin, W. E. (1987). Improving essay tests. Manhattan: Center for Faculty Evaluation and Development.

Cudia, C.M. & Tallungan, J.R.R. (2015). Educational Research Made Simple. Nueva Vizcaya State University

Cudia, C.M. & Tallungan, J.R.R. (2015). Work texts in Assessment of Student Learning.Nueva Vizcaya State University

Garcia, C. Measuring and Evaluating Learning Outcomes: A Textbook in Assessment of Learning 1 and 2. Books Atbp. Publishing Corp., 2008

Gronlund, N.E. & Linn, R.L. (1990). Measurement and evaluation in teaching. 6th edition. New York: Macmillan.

Gronlund, N., Measurement and Evaluation in Teaching, New York: Macmillan Company, Inc., 1981.

Manarang, L., and Manarang, R., Educational Measurement and Evaluation, Progressive Printing Palace, Quezon City, Philippines, 1987.

McMillan, J.H. (2001). Classroom assessment: Principles and practice for effective instruction. Boston: Allyn and Bacon.

Oriondo, L., and Antonio, E., Evaluating Educational Outcomes, REX Printing Company, Inc., 1989.

Raagas, E. Assessment and Evaluation of Student Learning: Concepts and Applications, 3rd Edition, ELR DATStat Analysis Center, 2010.

Wiggins, G.P. (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.

Worthen, B.R, Borg, W.R. & White, K.R. (1993). Measurement and evaluation in the schools. New York: Longman.