

Prototype Qualifying Examination for Filipino Majors

Alvin Rom De Mesa
Leyte Normal University
ar_demesa@ymail.com

Abstract

This study was on Qualifying Examination. It aimed to construct and developed a prototype qualifying examination for students who wish to major Filipino. The study followed the Product Development Cycle as suggested by Popham and Baker. These are: Instructional Specifications, Item Validation, Product Development, Product Try-out, and Product Revision.

Using the descriptive developmental research design, the following were identified: for the content of the final form of the qualifying examination for Filipino majors were on the following: Batayang Kaalaman sa Komunikasyon has twenty-three (23) items or 21.90%, mga Batayang Kaalaman sa Pagsulat at Pagbasa has two (2) items or 1.90%, Pagkilala sa Batayang Istruktura at Hulwaran ng Organisasyon sa Iba't ibang Genre ng Teksto has ten (10) items or 9.52%, Iba't ibang paraan ng Pagpapalawak ng Bokabularyo has fourteen (14) items or 13.33%, mga Kasanayan sa Pag-unawa has seventeen (17) items or 16.20%, and, Pagbasa at Pagsulat sa Iba't ibang Disiplina has thirteen (13) items or 12.39% with a total number of items of one hundred five (105) or 100%. On the basis of item distribution, the researcher followed Bloom's Taxonomy of Objectives in formulating the questions. These are the following: Remembering (Knowledge) has twenty-five (25) items or 23.81%, Understanding (Comprehension) has twelve (12) items or 11.43%, Applying (Application) has twelve (12) items or 11.43%, Analyzing (Analysis) has twenty-five (25) items or 23.81%, Evaluating (Evaluation) has twenty (20) items or 19.04%, and lastly, Creating (Synthesis) has eleven (11) items or 10.48%.

The researcher used Multiple Choice Type of test as the format for the qualifying examinations. Further, the 3-point rating scale was used in validating the content and face validity of the qualifying examination. The test established a high content validity of 2.36 and face validity at 2.71.

Keywords: Prototype, Qualifying Examination, Filipino majors, Product Development Cycle, and multiple choice

Dr. Alvin Rom De Mesa is a Doctor of Philosophy in Education major in Educational Programs Management graduate at Eastern Visayas State University (EVSU), Tacloban City, Leyte, Philippines. In the same country and province, he finished also his Master of Arts in Teaching Filipino and Bachelor of Secondary Education major in Filipino at the Leyte Normal University (LNU). Currently, he is working from 1998 to present at LNU. He is also a former Head in the Filipino Department and at present is the Director of the office of the Quality Management System for Accreditation and Sentro ng Wika at Kultura. Further, since 2010 up to the present, he is an active member and Accreditor of Accrediting Agency of Chartered Colleges and Universities in the Philippines, Inc. or AACCUP. His research studies focus on the development of the instructional materials in Mother Tongue and Filipino, teaching strategies in the Filipino courses, and, on the Waray and Filipino Language and Literature.

INTRODUCTION

Selecting Filipino majors in the Bachelor of Secondary Education (BSEd) program needs some kind of evaluation in order to select competent Filipino teacher for the future. Testing is one of the evaluation measures commonly used in the different units at Leyte Normal University (LNU). Educators and administrators of the school are responsible to make education relevant to the needs of every citizen.

According to Sanchez (1989) in his study, education has been blamed for an imbalanced education outputs and manpower requirements. It must be made clear however, that education has no power in its hand to limit or direct college enrollments, but this became a big problem in Philippine education. Educators were discovered to have an entrance examination, the National College Entrance Examination (NCEE). On November 25, 2973 the first NCEE was conducted but it was discontinued on September 2, 1993 and changed to National Secondary Assessment Test or NSAT. The said test was the basis in all schools including state colleges and universities to select students who are qualified to study in their respective schools. Moreover, teacher and educators are concerned in measuring and evaluating the level of their students' performance. This is necessary in understanding the student's knowledge and to guide the teachers themselves as to whether they have obtained their objective or not.

Meanwhile, Oriondo and Antonio (1984), believe that test is the best tool in evaluation. Based on the results, the teachers know what skills they have focused in their instruction. According to Thorndike and Hagen (1961), a good test helps to motivate the interest of students in order to pass and to choose a particular subject as his major. Therefore, there is a need to have a qualifying examination to be given by the Filipino unit to those incoming first year college students who wish to major Filipino subject in the BSED program. This qualifying examination will be administered during the First-Year college screening. Since there is no qualifying examination to those aspiring students to become a Filipino major, the Filipino unit must have a qualifying examination in order to identify who among the applicants are competent and qualified to be a Filipino major. Further, this study made a prototype qualifying examination for Filipino majors where the Filipino Unit of the Leyte Normal University adapt or use.

Although, the Filipino unit which is with the College of Arts and Sciences (CAS) found difficulty in making qualifying examination for aspiring Filipino majors because it takes time to make the said validated examination, they can use the prototype qualifying examination made by the researcher aside from the College Admission Test (CAT) given by the university to those First-Year college applicants. The selection of the prospective Filipino majors will be done in two ways: a qualifying examination and an interview. Student applicants who passed in the qualifying examination are the ones who will be interviewed. This study assumes that one of the admission requirements for the Filipino majors is the qualifying examination will reveal the ability of students to achieve in the course as their major.

To ensure the best possible test, one must plan for it. A good test does not just happen, it results from careful planning. Above all, planning helps to ensure that a test has balance and comprehensiveness (Furst, 1958). While, Kubizyn and Borich (2007) supports Furst, that planning takes in virtually all phases of test development, administration and use. Its scope is broad. Anti-test advocates suggest that testing should be done away with. Unfortunately, decisions will still have to be made. Teachers as human beings are subject

to *good and bad days*, biases, student and parent pressures, faculty perception, and a variety of other influences. This means that relying solely on a teacher's judgement on a subjective decision-making process is not perfectly dependable. Of course, no teacher would intentionally make mistakes.

Further, Klenowski & Wyatt-Smith (2014), briefly explained that in 1905 Alfred Binet and his assistant, Theo Simon, were commissioned to develop an intelligence test. The aim of their test was to measure a trait that would predict school achievement. They revised their test in 1908 in 1911. The concept of mental age as an index of mental development was introduced in the first revision and defined in the second. Since Binet was commissioned to develop a test that would predict school achievement, he was concerned with the predictive validity of his scale and repeatedly studied its validity for use in the public school.

The researcher followed the steps of Binet and his assistant. This study constructed a prototype qualifying examination for Filipino majors. Klenowski & Wyatt-Smith (2014) found out other literature which said that teachers are also required at least once a year to administer standard tests, and interpret the result to interested and concerned student applicants. Thus, it is necessary that teachers spend time getting to know the advantage and uses of standardized test.

Furthermore, Brookhart & Nitko (2015) suggest that a good testing programs must be designed to cover as many important outcomes as possible. Test must be progressive, practical, economical, usable and flexible. It must also supplement rather than supplant the informal tests given by classroom teachers. In addition, Padua and Santos (1997) contend that standardized tests are norm-referenced tests which have been administered to a large group of potential users of the test. The process of standardization of a test could take a very long time before finally arriving at acceptable standardized test from the time that the test developer prepared the table of specifications up to the point of validation and testing. This was already proven by many researchers in which their studies were about standardized test.

On the other hand, Adina (1980), constructed and developed a prototype achievement test in Mathematics VI for Taft District on the need assessment on the curricular relevance of the periodical examination currently used by the different schools in the district. Hus study followed the descriptive developmental research composed of two stages, first was on determining the curricular relevance of the periodical tests currently used by the different schools in the district of Taft, and second was the constructing and developing the proposed prototype achievement test. In fact, Agner (1981) also attempted to conceptualize, construct, administer, administer, and validate a communicative proficiency test in Filipino for finishing elementary school learners in the language skills, listening, speaking, reading and writing. The intention of her study was to come up with a 200-item valid and reliable communicative proficiency test from a pool of 405 communication items tried and responded to a four hundred (400) native Tagalog grade six children enrolled at Isabela Delos Reyes Elementary School year 1980 to 1981. Her study was founded on the integrated psycho-sociolinguistic theory/communicative competence theory of language testing/learning/testing and the four hundred and five (405) communication items constructed were meant to measure globally and integratively, both linguistic and psycho-linguistic competence which together compose with what so called *Communicative Competence*.

The common information gathered by the researcher from the related literature is to prepare a table of specifications to make sure that there is an equal distribution of items in the test and have an item analysis of the test in order to find out the validity and reliability of the test. Because of this, the main purpose of this study is to develop a prototype qualifying examination in the Filipino Unit of Leyte Normal University. It sought to accomplish the following objectives: 1. To identify the skills and competencies that shall be measured in the prototype qualifying examination; 2. To establish content validity and face validity of the test; and, 3. To establish the index of reliability.

METHODOLOGY

This focuses on the activities that were used in this study. This includes explanation of the research design, research locals, instruments, the respondents and sampling method, and the research procedure done in the study.

Research Design

This study followed a descriptive developmental research. It involves the development of the prototype qualifying examinations which followed the Product Developmental Cycle as suggested by Popham and Baker (1975). The steps are: Instructional Specifications; Test Item Validation; Product Development; Product Try-out; and, Product Revision.

Instructional Specifications. The first step in the development of the product. It includes the following: *Determining the content of the test.* The intention of the test should evaluate the acquired knowledges, concepts and skills of the students after taking the course. The content of the test must be based on the present curriculum. *Preparing the table of specifications.* There are different type ways of preparing a table of specifications, depending on the area being tested. Generally, table of specifications have some commonalities. Among them are course content, behavior, number of test items, placement and percentage. *Selecting the appropriate format.* Based on the review on the review of related literature, multiple choice type of test is the most widely used. It is versatile and can measure both complex as well as simpler learning objectives. In this study, multiple choice type of test was used. And, *Constructing of the test items.* There were two parallel items constructed. Each competency was included in the preliminary version of the test to give allowance for discarding the undesirable items. Each item was constructed in such a way that it would measure directly one of the specific objectives of the course.

Qualifying Examination Item Validation. The constructed qualifying examination with the copy of syllabi, the 3-point rating scale and the table of specifications were submitted to the validity panel for content and face validation and for suggestions on the improvement of the test.

Product Development. This is the preparation of the items for the final run. This will be determined by the difficulty level of the test, the discrimination index, the effectiveness of the distractors, and the establishment of test reliability.

Product Try-out. This was for item analysis and the first and second dry-run was tried out to the respondents' schools. This step includes item analysis, difficulty and discrimination indices of each item, determination of the effectiveness of the distractors and determination the test reliability. *Item Analysis.* This study followed the method as illustrated by Oriondo and Antonio (1983). First, the papers were arranged properly from

highest to lowest. Second, the answers of the students who got high and low scores were taken down. Garrett (1965) has the same suggestions as of Oriondo and Antonio that U-L method in the item analysis be used. After checking the papers, they were arranged from highest to lowest were 27% of the highest scores was separated and considered as upper group, and likewise the lower group. *Difficulty Index of Each Item*. The following formula was used and defined by Garrett (1965) in order to determine the difficulty index of each item.

$$D = \frac{U - L}{N}$$

Where:

- D - Level of difficulty index of the item
- U - Total number of the students who got the correct answer above 27%
- L - Total number of the students who got the correct answer below 27%
- N - Total number of students used in the item analysis

Garrett also interpreted the difficulty level of each item.

Level of Difficulty	Interpretation
85% - 100%	Very Easy
50% - 84%	Easy
15% - 49%	Difficult
0% - 14%	Very Difficult

Item difficulty means the percentage who got the correct answers, in which if the percentage is low, the items are considered very difficult. Lindemen (1976) believes that a test is considered to be a good test if it is between easy and difficult.

Discrimination Index of Each Item. To determine the discrimination index, the following formula was used:

$$D_i = \frac{N_u - N_r}{N_c}$$

Where:

- N_u - Total number of students in the upper group who answered the item correctly
- N_r - Total number of the students in the lower group who answered the item correctly
- N_c - Total number of the students who got the answers correctly

Ebel (1965), interpreted the computed index of discrimination.

Index of Discrimination	Interpretation
0.40 and up	Very Good Items
0.30 - 0.39	Reasonably Good Items but possibly subject to improvement
0.20 - 0.29	Marginal Items, usually needing and being subject to improvement
Below 0.19	Poor items, to be rejected or improved by revision

He added that items with discrimination indices of 0.20 and above are considered for inclusion in the final form of the test. *Determining the Effectiveness of Distractors*. This is a phase in analyzing the items in a multiple-choice type of test. According to Downie & Heath (1974), one procedure of estimating the effectiveness of distractors is through inspection. Best (1981) emphasizes that distractors not selected by at least 3% of the

examinees are not functional, this means they have to be discarded or revised. Item analysis was employed after the dry-run. Items that were selected and considered for inclusion were arranged from to difficult. This was done for the final form of the test. *Determining the Test Reliability*. The Kuder Richardson formula 21 was used in this study of determining the reliability of the test for the final form. The formula was:

The Kuder Richardson Formula 21 was used in this study. The formula was:

$$r_{KR21} = \frac{K}{K - 1} \left(1 - \frac{M \left[1 - \frac{M}{K} \right]}{\sigma^2} \right)$$

Where:

r_{KR21}

- The Reliability index of the Test

K - The Total Number of Items of the Test

(\bar{X}) - The Mean Score of all examinees who took the test

σ^2 - The Variance

It resulted to a 0.86 which means the final form of the test has a high reliability.

Product Revision. This is the necessary process for the development of a product. The purpose of this is to take out those items that are irrelevant, to identify and revise weak items, to make functional distractors, to determine the validity and reliability of the test, and to make corrections of the instruments in order to come up with a prototype qualifying examination for Filipino majors.

Research Locale

This study was conducted in selected private higher education institution in Samar and Leyte. On the first dry-run, the researcher used one school in Leyte represented by Asian Development Foundation College (ADFC) in Tacloban City and one respondent school in Samar represented by Samar College in Catbalogan, Samar. Both schools were chosen because they are under the supervision of the Commission on Higher Education (CHED). On the second dry-run, the respondent schools utilized were Leyte Colleges in Tacloban City and Saint Mary's College of Catbalogan, in Catbalogan Samar.

Instruments

The instruments that were used in this study were based on the syllabi of the two basic Filipino 101 now General Education 122 (GE 122) and Filipino 102 now is General Education 123 (GE 123) used by the instructors, table of specifications of the cognitive processes of the items for the constructed test items and three rating scale for determining face and content validation of the test. *The Syllabi*. This is commonly used by the teachers in a particular course. It contains the objectives and the coverage of the course to be covered the whole semester. *Table of Specifications*. This contains the six cognitive processes under Bloom's Taxonomy of Objectives. These are remembering (knowledge), understanding (comprehension), applying (application), analyzing (analysis), evaluating (evaluation), and creating (synthesis). This was used in order to have a balanced distribution of items in the

test. It tells the kinds of distribution of items in the test and the kind of information in which the Filipino unit wants to obtain. *The -point rating scale*. This was used for the content and face validity of the preliminary version of the test items which is composed of two parts. The first contain the criteria for evaluating the content validity of the test. Some of these criteria were taken from the criteria in evaluating tests given by Ebel (1965) as cited by Lumen (1983) and Adina (1980) but modifications were made to suit the purpose. The second contain the criteria for evaluating the face validity of the test. These instruments were submitted to the Filipino teachers for editing, comments, suggestions, and improvements.

The respondents and the Sampling Method

The number of students who took the test was taken from 20% of the population of Bachelor of Secondary Education (BSEd) sophomore students. Before administering the test, the researcher asked permission from the president/director of the said schools. The schools involved in this study are private schools which are not under State Colleges and universities (SUC's). The school chosen were two from Samar and from Leyte respectively.

Table 1. Distribution of Students in Each School who took the Test

	School	Number of Students	Percentage
A. First Dry-run			
	Samar College (SC)	50	50.51%
	Leyte Colleges (LC)	49	49.49%
	Total	99	100%
B. Second Dry-run			
	Saint Mary's College of Catbalogan	48	51.06
	Asian Development Foundation College	46	48.94%
	Total	94	100%

For every dry-run of the examination, one school in Samar and one from Leyte were utilized. The researcher was able to select two schools in Samar and in Leyte. Random Sampling was used in choosing the examinees which followed the steps done by Caharop (1991) in her study. Twenty percent (20%) from the population of the BSED sophomore students were chosen regardless of their major course but had already taken Filipino 101 or now GE 122 and Filipino 102 or now GE 123.

Research Procedure

In this study the following steps in constructing and developing a prototype qualifying examination for Filipino majors were followed: The researcher constructed a 200-multiple choice items test for the first dry-run. After administering the test, the item analysis followed. It was found out that there were 148 items left for the second dry-run. The researcher determined the difficulty and discrimination indices and established the reliability of the test. The results of the second dry-run was analyzed and only 105 items and two essay questions in the test were included in the final form of the test as a whole. The 105 multiple choice items and two essay questions were given to thirty (30) BSED sophomore students of Leyte Normal University in order to establish the reliability of the test. Based on the results of the dry-run, the researcher was guided in revising the test. The items that were taken out were those are found to be very difficult and with poor discriminatory power including those with negative indices of discrimination. Those items

with low discriminatory power but with high level of difficulty and discriminatory power were improved and included in the final form of the test.

Trying out of the Final Form of the Test. In any developmental study, the finished product is always subjected to another try out to establish the face value or reliability of the product. To maintain this practice, the final form of the test was again conducted in another dry run at Leyte Normal University to a thirty (30) BSED sophomore students regardless of their major course. The purpose of this was to get the reliability of the final product. There was no item analysis here.

Safeguarding the Constructed Test. The final product is now confidential. This must be kept in order to safeguard its validity and reliability. The final form of the test was not appended in this manuscript but the researcher is the one keeping a copy to make it available during the selection of the aspiring Filipino majors. Further, the test is intended to evaluate the student applicants who wish to major Filipino course. The test should only be used in selecting Filipino majors and not for any other purpose in order to preserve its validity and reliability. A test loses its reliability if it is always given to the same students, but since this a qualifying examination it is understood not the same students will take the test.

RESULTS AND DISCUSSION

The first illustration here begins with the combined judgement of teachers teaching Filipino courses with more than five years of teaching experience in the teaching of the course in terms of content and face validity of the prototype qualifying examination for Filipino majors. The second, illustrates the analysis of the results of the final dry-run of the qualifying examination which include item analysis, determining the difficulty and discrimination indices of the items and the effectiveness of distractors. The last part discusses the reliability of the prototype qualifying examination and its final form.

The Validity of the Prototype Qualifying Examinations

Qualifying Examination. The instrument used by the college Filipino teachers was a 3-point rating scale. This was provided by the researcher as guide for validation in order to ascertain the content and face validity of the qualifying examination.

Its Content Validity. The mean score for every criterion was computed based on the validation guide which in the 3-point rating scale.

Table 2. Mean Indicators of the Content Validity of the Qualifying Examination for Filipino majors

Criteria	Mean	Interpretation
1. The items are based on the Filipino 101 now GE 122 and Filipino 102 now GE 123	2.26	Very Satisfactory
2. The items present clear problems	2.24	Very Satisfactory
3. The Correct responses are better than the other responses	2.39	Very Satisfactory
4. The distractors are likely to attract mainly students of low achievement	2.54	Very Satisfactory
Combined Mean	2.36	Very Satisfactory

Table 2 illustrates the summary of the mean indicators of the content validity of the qualifying examination for Filipino majors. The mean rating of criterion number 1 is 2.26, number 2 is 2.24, number 3 is 2.39, and the last criterion is 2.54 interpreted as very satisfactory. The evaluators suggested that the copy for the examinees must be clear because they believed that this might be the cause for failing in the test. The last line of the

table shows the combined mean which is very satisfactory. This shows that the test questions possess high content validity.

Its Face Validity. This is part II of the rating scale which was administered by the researcher to the evaluators of the test. The mean score for every criterion was computed.

Table 3. Mean Indicators of the Face Validity of the Qualifying Examination for Filipino majors

Criteria	Mean	Interpretation
1. The standardized qualifying examination is appropriate in terms of:		
a. Format	2.75	Very Satisfactory
b. Page Lay-out	3.00	Very Satisfactory
c. Language	2.75	Very Satisfactory
2. The items are clearly stated	2.50	Very Satisfactory
3. The words are clearly printed and legible	2.50	Very Satisfactory
4. The directions are concise and clear	2.75	Very Satisfactory
Combined Mean	2.71	Very Satisfactory

Table 3 shows the mean indicators of the face validity prototype qualifying examination for Filipino majors. As illustrated in the table, the Filipino teachers have a mean score of 2.75, very satisfactory. The page lay-out has a mean score of 3.0 very satisfactory. It also shows that the mean score of language is 2.75 very satisfactory. They believed that the language use in the test was simple and relevant. In addition, the table shows that the validity panel agreed that the items were clearly stated and the words were clearly printed and legible although there were some items which need improvement. The mean score is 2.5, satisfactory. The college Filipino teachers agreed that the directions were concise and clear because their mean rating was 2.75, very satisfactory. The combined mean score of the face validity of the test is 2.71, very satisfactory. Therefore, the test has a high face validity.

Analysis of the Test Results

The examination was conducted in two dry-runs. These were 200 multiple choice items, in the first dry-run. The purpose of the first dry-run was to determine which items would be included in the second dry-run. Out of 200 multiple choice items, only 148 items were included in the second dry-run. During item analysis, 66 items were revised and 54 items were rejected. The researcher did not include the rejected items. Only those items belonging to the average discrimination and difficulty indices were revised. The first dry-run took three hours and thirty minutes to finish. The second dry-run was considered as the final dry run in this study. Prior to that, determining the difficulty and discrimination indices was followed from 148 multiple choice items. This also included the determining of the effectiveness of distractors and establishing of test reliability.

First Dry-run. The test was composed of 200 multiple choice items. This was conducted at Samar College of Catbalogan and at Leyte Colleges of Tacloban City. The examinees were BSEd second year college students who took Filipino 101, Komunikasyon sa Aakdemikong Filipino but now is GE 122, Kontekstwalisadong Komunikasyon and Filipino 102, Pagbasa at pagsulat sa Iba't ibang Disiplina which now GE 123, Filipino sa Iba't ibang Disiplina. After checking the test, the scores were arranged from highest to lowest. The researcher separated the upper and lower and lower 27%, that is 27% from the higher score and lower score. Items that were rejected, retained, and revised were based on

the difficulty and discrimination indices. Table 4 shows the results of the item analysis of the First Dry-run.

Table 4. Results of the Item Analysis of the First Dry-run on the Rejected, Retained, and revised items based on their Difficulty and Discrimination Indices

Item Numbers	Difficulty Index	Discrimination Index	Remarks
2, 5, 9, 11, 13, 18, 19, 21, 22, 26, 27, 28, 30, 32, 34, 41, 46, 49, 50, 58, 59, 60, 65, 66, 67, 69, 71, 82, 84, 88, 89, 97, 99, 102, 106, 120, 123, 124, 129, 134, 135, 136, 145, 146, 154, 155, 161, 177, 186, 192, 198, & 199	0% - 14%	Below 0.18	Rejected
7, 4, 8, 10, 12, 14, 15, 20, 25, 29, 38, 40, 44, 45, 47, 48, 51, 52, 54, 55, 62, 73, 75, 81, 83, 85, 91, 92, 98, 103, 104, 107, 108, 109, 112, 115, 116, 121, 126, 130, 131, 140, 141, 142, 143, 144, 148, 149, 150, 156, 157, 158, 159, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 173, 175, 178, 180, 182, 184, 185, 187, 189, 191, 193, 194, 195, 196, 197, & 200	15% - 49%	0.30 – 0.39	Retained
1, 3, 6, 16, 17, 23, 24, 31, 33, 35, 37, 39, 42, 43, 53, 56, 57, 61, 63, 64, 68, 70, 72, 74, 86, 87, 90, 93, 94, 95, 96, 100, 101, 105, 110, 111, 114, 117, 118, 119, 122, 125, 127, 128, 132, 133, 137, 138, 139, 147, 151, 152, 153, 160, 172, 174, 176, 179, 181, 183, 188, & 190	50% - 84%	0.20 – 0.29	Revised

As shown in Table 4 item from 1 to 200 was analyzed. It indicates the upper 27% and the lower 27% of the examinees. This presents also the item numbers, difficulty index, discrimination index, and remarks showing whether the item was accepted or rejected. It was mentioned earlier that those items which are below 0.20 discrimination index but very high in difficulty index are items considered to be revised. As shown in the remarks, 52 items are rejected, 85 items were retained, and 63 items are revised.

Second Dry-run. The researcher used 148 multiple choice items for second dry-run. This was conducted at Saint Mary's College of Catbalogan, and Asian Development Foundation College of Tacloban City. The same program and year level in the first dry-run Was used in the second dry-run. Items that were retained and revised are found in Table 5.

Table 5. Results of the item analysis of the Second and final dry-run on the retained and revised items based on the results of the second dry run

Item Numbers	Difficulty Index	Discrimination Index	Remarks
7, 4, 8, 10, 12, 14, 15, 20, 25, 29, 38, 40, 44, 45, 47, 48, 51, 52, 54, 55, 62, 73, 75, 81, 83, 85, 91, 92, 98, 103, 104, 107, 108, 109, 112, 115, 116, 121, 126, 130, 131, 140, 141, 142, 143, 144, 148, 149, 150, 156, 157, 158, 159, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 173, 175, 178, 180, 182, 184, 185, 187, 189, 191, 193, 194, 195, 196, 197, & 200	15% - 49%	0.30 – 0.39	Retained
1, 3, 6, 16, 17, 23, 24, 31, 33, 35, 37, 39, 42, 43, 53, 56, 57, 61, 63, 64, 68, 70, 72, 74, 86, 87, 90, 93, 94, 95, 96, 100, 101, 105, 110, 111, 114, 117, 118, 119, 122, 125, 127, 128, 132, 133, 137, 138, 139, 147, 151, 152, 153, 160, 172, 174, 176, 179, 181, 183, 188, & 190	50% - 84%	0.20 – 0.29	Revised

Item Difficulty. Items that are easy, receive higher percentage than the difficult items. In constructing, receives all items should be analyzed based on their difficulty so that only those with appropriate difficulty indices are the ones included in the final for of the test. Adina (1980), in his study arranged his final form of the test from easy to difficult so that the examinees will be able to finish answering the test. The same procedure was used in this study. There were 7 items got 4.73% which interpreted as “very easy” with difficulty indices ranging from 85% and above. Ninety items or 60.81% are “easy” with difficulty indices ranging from 50% to 84%. A total of 49 items or 33.11% are “difficult” items ranging from 15% to 49%. Only 2 items or 1.35% are “very difficult” ranging from 0% to 14%. A total of 139 items are easy and difficult. This means that the test has an average difficulty.

Table 6. Difficulty indices of the items in the qualifying examination for Filipino majors

Difficulty Index	Item Numbers	Total No. of items	Percentage	Interpretation
85% - 100%	25, 37, 77, 117, 130, 132, & 136	7	4.73%	Very Easy
50% - 84%	1, 2, 3, 4, 5, 6, 8, 10, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 26, 27, 28, 29, 31, 33, 34, 36, 39, 40, 41, 42, 46, 47, 48, 50, 51, 52, 53, 54, 55, 56, 58, 60, 61, 62, 64, 65, 66, 69, 71, 72, 73, 75, 80, 81, 82, 84, 86, 87, 90, 91, 92, 93, 94, 96, 97, 100, 101, 103, 104, 105, 106, 107, 108, 109, 110, 111, 113, 115, 118, 119, 121, 122, 128, 133, 139, 141, 145, & 147	90	60.81%	Easy
15% - 49%	7, 9, 11, 17, 19, 30, 32, 35, 38, 43, 44, 45, 49, 57, 59, 63, 67, 68, 70, 74, 76, 78, 79, 83, 85, 88, 89, 95, 98, 99, 112, 114, 116, 120, 124, 125, 126, 127, 129, 131, 134, 137, 138, 140, 142, 143, 144, 146, & 148	49	33.11%	Difficult
0% - 14%	24 & 102	2	1.35%	Very Difficult
Total		148	100%	

Index of Discrimination. A good test items identifies good students from the poor students. The examinees in the upper 27% get more correct answers than the students in the lower 27% have more correct answers, the discrimination index is negative. The discrimination indices of the 148 multiple choice items were classified: 11 or 7.43% items classified as very good items, 94 or 63.51% fall under, marginal classifications, 23 or 15.54% considered as poor items, and 20 or 13.52% have negative indices.

Table 6. Difficulty indices of the items in the qualifying examination for Filipino majors

Discrimination Index	Item Numbers	Total No. of items	Percentage	Interpretation
0.40 and up	2, 15, 33, 51, 60, 80, 86, 91, 97, 119, & 121	11	7.43%	Very Good Items
	1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 21, 22, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 38, 39, 42, 43, 46, 48, 50, 54, 56, 59, 61, 62, 64, 65, 66, 67, 68, 71, 73, 75, 79, 81, 84, 87, 92, 93, 94, 98, 99, 101, 103, 104, 105, 108, 110, 111, 113, 114, 116, 118, 120, 122, 123, 124, 127, 128, 129, 126, 131, 133, 137, 138, 139, 140, 142, 143, 144, 145, 146, 147, & 148	94	63.51%	Reasonably Good Items
0.30 to 0.39	37, 40, 41, 45, 47, 49, 52, 53, 55, 57, 58, 69, 70, 72, 74, 78, 82, 83, 88, 90, 100, 106, & 125	23	15.54%	Marginal Good Items
0.20 to 0.29	7, 19, 20, 23, 24, 25, 44, 76, 85, 102, 107, 115, 117, 130, 132, 134, 135, 136, & 141	20	13.52%	Poor Items
0.19 and below				
Total		148	100%	

Effectiveness of Distractors. In a multiple-choice test, the option should be related or similar to one another. The distractors here are option which seem to correct but are not. In order to be functional, it must be attractive and acceptable to the examinees. According to Best (1981), the distractors which are not selected by at least 3% of the examinees are not functional or good. As mentioned earlier, there were 94 students who took the second or final dry-run of the 148-item test. There were 36 students involved in the item analysis, considered to be from the upper and lower group. 3% out 36 students are 1.08. An effective distractor, therefore, must be chosen by at least 1 student. If less than 1 student choose the option, such is not effective one. In the responses of the students, there were 36 students involved in the item analysis. The results were the distractors of the qualifying examination are good except for items number 15, 25, 26, 27, 33, 34, 35, 36, 37, 38, 40, 41, 42, 46, 51, 52, 59, 60, 61, and 62 which means in effective distractors. The easiest item in the final form of the test had difficulty index of 0.64 interpreted as “fairly easy.” The most difficult items that has a difficulty index of 0.33 which is interpreted as “fairly hard.” While, the weighted average of the discrimination index is 0.49 which is “fairly hard.” Therefore, the final items are moderately difficult and the highest computed index of discrimination is 0.47 which means “very good items.”

Test Reliability. The reliability coefficient of the test was determined by using Kuder Richardson Formula 21 because it yields a coefficient of internal consistency. The value of the coefficient of reliability of the final form of the test was 0.86 which means the test is highly reliable.

The Final Form of the Test. Based on the results of the item analysis, the final form of the test is composed of 105 multiple choice items. These items are arranged from easy to difficult.

Safeguarding the Constructed Test. In Filipino 101 or GE 122, these are the two aspects: Batayang Kaalaman sa Komunikasyon, a total of 23 items or 21.90% and Pagtatamo ng Kasanayan sa Komunikasyon, 26 items or 24.76%. The second part of the test in Filipino 102 or GE 123 are the following: Mga Batayang Kaalaman sa Pagsulat at pagbasa, 2 items Or 1.90%, pagkilala sa batayang istruktura at hulwaran ng organisasyon ng iba't ibang genre ng teksto, 10 items or 9.52%, iba't ibang paraan ng pagpapalawak ng bokabularyo, 14 items or 13.33%, mga kasanayan sa pag-unawa, 17 items or 12.39%. Likewise, under the cognitive processes of the test has 25 items or 24% for knowledge, 12 items or 11.43% for comprehension, 12 items or 11.43% for application, 25 items or 24% for analysis, 11 items or 10% for synthesis, and 20 items or 19% for evaluation. To determine its reliability, the final form of the test was tried out to 30 second year BSEd students of Leyte Normal University in which the reliability coefficient is 0.86, which means highly reliable.

CONCLUSION AND RECOMMENDATION

Based on the findings of this study, the following conclusion were drawn: The constructed test has a balanced distribution as shown by the table of specifications, the Filipino expert teachers gave “very satisfactory” rating for the content and face validity of the qualifying examination, the final form of the test has easy and difficult items which belongs to a moderate difficulty level, and the test has high reliability. On the other hand, the following recommendation are advanced with the hope that the administrators of the Filipino unit will be considered for implementation: The Filipino unit of Leyte Normal University will use this test as an evaluation instrument for those aspiring Filipino majors, any qualifying examinations must be relevant to the present curriculum, a further study of this research is encouraged to establish the concurrent and predictive validity, and the future researchers may use this study for guidance in making a similar to this.

REFERENCES

- Adina, B. A. (1980). *A prototype achievement test in mathematics for grade 3*. Leyte State College, Tacloban City, Philippines: Unpublished Thesis.
- Agner, S. B. (1981). *Towards the development of a communicative proficiency test in Pilipino for finishing elementary students: A first attempt*. Philippine Normal College-Della Salle University Consortium: Unpublished Dissertation.
- Brookhart S. M. & Nitko. A. J. (2015). *Educational assessment of students*. 7th Edition. New Jersey: Pearson Education, Inc.
- Caharop, T. B. (1991). *A test of communicative competence for freshmen college students*. Leyte State College, Tacloban City, Philippines: Leyte State College.
- Downie, N. M. & Heath, R. W. (1974). *Basic statistical methods*. New York: Harper and

Row Publishing, Co.

- Ebel, R. L. (1965). *Measuring educational achievement*. New Jersey: Prentice Hall, Inc.
- Furst, E. J. (1958). *Constructing evaluation instrument*. New York: David McKay Company, Inc.
- Garett, H. (1965). *Testing for teachers*. 2nd Edition. New York: American Book Company.
- Klenowski, V. & Wyatt-Smith, C. (2014). *Assessment for education (Standard, judgement and Moderation)*. New Delhi: SAGE Publication, Inc.
- Kubiszyn, T. & Borich G. (2007). *Educational testing and measurement (classroom application and practice)*. 8th Edition. Australia: John Wiley & Sons, Inc.
- Lindemen, R. (1976). *Educational measurement*. Winsconsin: George Banta Publishing Co.
- Lumen, A. C. (1983). *A standardized achievement test in mathematics for high school seniors in region 8*. Devine Word University, Tacloban City, Philippines: Unpublished Dissertation.
- Oriondo, L. L. and Dallo-Antonio, E. M. (1984). *Evaluating educational outcomes*. Manila: Rex Printing Co., Inc.
- Padua, R. N. and Santos. (1997). *Educational evaluating and measurement: Theory, practice and application*. Quezon City, Philippines: Katha Publishing Co.
- Popham, J. W. and Baker, E. L. (1975). *Educational evaluation*. New Jersey: Prentice Hall, Inc.
- Sanchez, N. O. (1989). *Admission requirements and other selected factors or predictors of academic achievements of the college of education students*. University of the Philippines: Unpublished Dissertation.
- Thorndike, R. and Hagen, E. (1961). *Measurement and evaluation in psychology and education*. New York: John Willey and Sons, Inc.