Development and Validation of a Philippine Music Achievement Test in Addressing the K to 12 Music Curriculum Learning Competencies

Almighty C. Tabuena
Teacher, High School Department,
Espiritu Santo Parochial School of Manila, Inc., Manila, Philippines

Abstract:
The study aimed to develop and validate an achievement test in Philippine Music for Grade 7 junior high school students that can measure the learning competencies in the K to 12 Music Curriculum. The researcher, during the inter-judge consistency process, aligned the learning competencies of Music, the result of which became the basis in preparing the first draft of the test. Based on the test, a set of 115 questions with four options was formulated. The draft of the instrument was submitted to a panel of music experts for content and face validation. Also, the test underwent two tryouts. After the item analysis and the final test administration, a total of 80 items were included in the final form of the test. To ensure reliability, the achievement test underwent the process of the split-half reliability method, the use of the Pearson Product-Moment Correlation Coefficient and the correction of the Spearman-Brown Formula applied to the coefficient. The results of the test were reliable at 0.82. Similarly, the data gathered through predictive validity determined by correlating the scores with the subjects' final grades in the Philippine Music Achievement Test (PMAT). The researcher concluded that the achievement test is valid and reliable for it covers the areas of Music, and the items were responsive to the K to 12 Music learning competencies.

Keywords: Achievement test, music curriculum, music learning competencies, philippine music, test development, test validation

1. Introduction

In line with the present curriculum – the K to12 Curriculum – in terms of the emergence of the existing curriculum in education, there is a need to prepare instructional materials that would cater to the needs and interests of the students from the context of musical learning, depending on their knowledge and skills. As curriculum sets new standards in terms of learning (content) and performance, the outcomes must be aligned to its new sets of competencies – particular to cognitive domain of learning as the learner demonstrates an understanding of basic concepts and processes in music through appreciation, analysis and performance for his/her self-development, celebration of his/her Filipino cultural identity and diversity, and expansion of his/her world vision (Department of Education [DepEd] K to 12 Music Curriculum Guide, 2016).

As stated in the Philosophy and Rationale for Music Education in K to 12 Basic Education Curriculum (DepEd, 2016), a keen sensitivity to environmental and musical sounds needs to be developed, in which the student must learn to – hear, –speak, and –think in the medium of music. Simultaneously, growth and development in the skills that enable the application of the learner's knowledge should be encouraged, through active involvement in the various musical processes. One good example of assessing and evaluating musical learning through an objective test item is the development and validation of a test material particular to an achievement test.

The main goal of an achievement test is to obtain valid, reliable, and useful information concerning student learning. This suggests a means in determining what are to be measured and defining objectives precisely so that the test items constructed can evoke the desired performance of the learner. To obtain validity, reliability and other useful information, there's a need to specify the achievement domain such that the sample test items will represent the achievement tasks and that the results will be appropriate for the intended instructional use. However, few teachers in the grand arena of teaching are aware of and knowledgeable about the important considerations in the construction of achievement tests regardless of academic disciplines. That at times lead to confusing and conflicting views and ideas which results in a meaningless endeavor. Classroom tests and assessments can be used for a variety of instructional purposes. They can be best described in terms of their purpose and use in the instructional process parallel to the types of

assessment, such a fixed-choice test, a form of assessment that is efficient in measuring the knowledge and skills of an individual.

According to the Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program, classroom assessment is an integral part of curriculum implementation. It allows teachers to track and measures learners' progress and to adjust instruction accordingly. Classroom assessment informs the learners, as well as their parents and guardians, of their progress. By the progress of development, this study will ensure pieces of evidence in the analysis and evaluation of the validation process concerning the assessment and evaluation of musical learning.

This research paper is an attempt to furnish the teachers to provide background knowledge of the characteristics of proper evaluative measures, assisting teachers in developing skills necessary to construct, validate, interpret results, and determine the advisability of using one of the assessment instruments for evaluation, which is the achievement test.

2. Statement of the Purpose

This study aimed to develop and validate an achievement test in Philippine Music for junior high school students, particularly to grade level 7. Furthermore, this study aimed to accomplish the following:

- Construct test items aligned to the K to 12 Music curriculum learning competencies.
- Determine the difficulty and discrimination indices of the test items through item analysis.
- Evaluate the achievement test in Philippine Music in terms of reliability and validity.

3. Literature Review

The theoretical basis of this study is anchored in the provisions and policy guidelines on classroom assessment for the K to 12 Basic Education Program issued by DepEd, Order No. 8, s. 2015; at the heart of this assessment framework is the recognition and deliberate consideration of the learners' zone of proximal development (Vygotsky, 1978). Appropriate assessment is committed to ensure learners' success in moving from guided to an independent display of knowledge, understanding, and skills, and to enable them to transfer this successfully in future situations. From this point of view, assessment facilitates the development of learners' higher-order thinking and 21st-century skills. The view of assessment, therefore, acknowledges the unity of instruction and assessment.

In addition, the Department of Education had released DepEd Order No. 79, s. 2003, in which Assessment and Evaluation of Learning and Reporting of Students Progress in Public Elementary and Secondary Schools were promulgated in emphasizing in the process of Assessment and Evaluation of Learning in all of the general subjects taken at a primary and secondary level. Even Thorndike (1977) supports that the achievement test is typically given to find out how much a student has profited from the past instruction or what level of knowledge or skill a person possessed.

In 1972, Stanley and Hopkins claimed that logical considerations and research have shown that skillfully prepared informal test can be as reliable as some standardized test, and often more valid for a particular class or student. Standardized test tends to be a focus upon broad, general objective that cover a wide range of content. A teacher needs to evaluate frequently so he can identify the specific learning difficulties of individual children and the class as a whole. Needless to say to consider the general principle of constructing informal tests; the following the four fundamental steps: (1) planning the test; (2) preparing the test; (3) trying out the test; and (4) evaluating the test. High-quality assessment can produce valid information about students' learning outcomes and provide insights into the effectiveness of teachers' instruction. Research indicates that teachers who introduce formative assessment into their classroom practice can affect substantial achievement gains. Stronge (2010) cited a research review: Black & William (1998) examined multiple empirical studies to determine whether improvement in classroom assessments can lead to an improvement in learning. They found that informative assessment has substantial positive effects on student achievement, with effect size raging 0.3 to 0.7 standard deviation. In particular, they found that formative assessment is more effective for low achievers than for other students, thus reducing an achievement gap while raising achievement overall at the same time.

In the final form of achievement test developed and validated by Simeon (2003), it consisted of 94 items. The achievement test in Araling Panlipunan for First Year High School Students has an average difficulty index of 0.55 (moderately difficult) and a discrimination index of 0.39 (good). The test-retest method was used and the reliability coefficient of the test was 0.96. The Split-Half method was also used for the reliability of the achievement test. The Pearson Product-Moment Correlation Coefficient was computed and the obtained value of r was 0.92. This was corrected by the Spearman-Brown Formula, which revealed a value of r equal to 0.96.

Parrotina (1994) developed and validated a test in musical knowledge, understanding, and aural discrimination for grade six pupils. The findings revealed that the test is valid, reliable and neither easy nor difficult. It also showed that it can serve as a fair measure of Grade 7 pupils' achievements in the knowledge of music theory. In addition, Flores (2002) developed modules in music. This was conducted primarily in support of the mission and objectives of the BSU Distance Education Program. Generally, the modules obtained a very high rating ranging from 3.4 (Good) to 5.0 (Excellent), based on teachers' evaluation. Modules 1 to 5 got excellent ratings while Module 6 had a very good rating.

From these reviews, the researcher drew much of methodology and processes, that is, from the steps in test construction, to item analysis and to the methods of establishing validity and reliability.

4. Methodology

4.1. Research Design

This study used the descriptive-developmental design in the development of the achievement test in Philippine Music. Studies with a descriptive-developmental design would tend to focus on the present prevailing conditions.

Descriptive-developmental studies are valuable for the following reasons: (1) providing facts on which scientific judgments may be based, (2) providing essential knowledge about the nature of objects and persons, (3) for closer observation into practices, behaviors, methods, and procedures, (4) in playing a large part in the development of instruments for the measurement of many things – instruments that would be employed in all types of descriptive-developmental research such as data-gathering instruments, and (5) in the formulation of policies in the local, national, or international level (Calmorin, 2010).

4.2. Sample and Sampling Technique

The first tryout of the achievement test in Philippine Music was administered to 131 junior high school students selected randomly from three (3) different schools considering the following variables: (a) time-constraint, (b) proximity, and (c) schedule availability. Out of 131 junior high school students from the three different secondary schools, the researcher used the stratified proportional random sampling technique, where the number of students in each school was proportioned as to the computed given population in each sample distribution. Table 1 shows the stratified proportional random sampling in the first tryout (pilot testing).

Schools	Sample Distribution N = 131	Proportional Sample	Total Sample (Pilot Testing)
Timoteo Paez Integrated School	46	42	
Florentino Torres High School	43	39	120
Jose Abad Santos High School	42	39	

Table 1: Stratified Proportional Random Sampling in the First Tryout (Pilot Testing)

The final form of the test was administered to 156 Grade 8 students, where the researcher purposively chose the following schools for the final testing of the Philippine Music Achievement Test. Out of 156 junior high school students from the four different secondary schools, the researcher used the stratified proportional random sampling technique. The distributions of the proportional samples are indicated below where the researcher limited the extent of the sample size in a total of 100 students in estimating the reliability and validity of the achievement test. Table 2 shows the stratified proportional random sampling in the second tryout (final testing).

Schools	Sample Distribution N = 156	Proportional Sample	Total Sample (Final Testing)
Timoteo Paez Integrated School	45	29	
Jose Abad Santos High School	41	26	100
Manila Science High School	36	22	
Victorino Mapa High School	34	23	

Table 2: Stratified Proportional Random Sampling in the Second Tryout (FinalTesting)

4.3. Locale

This descriptive-developmental type of research aimed to develop and validate the achievement test in Music learning to junior high school students was administered in the Division of City Schools, Manila. Schools were limited in the locale area of Division of City Schools, Manila, with the participating schools from District 1, 2, 3, 5, and 6 as shown in Table 3.

Tryout	District	School
First Tryout	1	Timoteo Paez Integrated School
	2	Florentino Torres High School
	3	Jose Abad Santos High School
Second Tryout	1	Timoteo Paez Integrated School
	3	Jose Abad Santos High School
	5	Manila Science High School
	6	Victorino Mapa High School

Table 3: Participating Schools in the Administration of the Achievement Test in Philippine Music

4.4. Instrument

The instrument used in validating the achievement test in Philippine Music was designed and adopted based on the suggestions, patterns, and ideas of the validation process of Flores (2002) and Narzo (2005) who intended also to validate an instrument. Some of the questions were adapted from the book of Linn and Gronlund (2000), Measurement and Assessment in Teaching, and the format of the scale from Fox (1969) which presented in a numerical rating scale as adapted in the form of a Likert Scale. In judging the face validity (Asaad & Hailaya, 2004), there should be at least three knowledgeable persons who are qualified to pass judgment on the appropriateness, suitability, and mechanics in the construction of the tests. In this case, the researcher sought the help of five (5) experts/experienced music teachers

validating the test items before the pilot testing stage took place in the selected schools of Division of City Schools, Manila. The validation instrument is composed of 14 items which are arranged in terms of (a) format and design, (b) objectives, (c) content, and (d) language. Table 4 shows the rating scale used in the validation of the instrument.

Scale	Range	Interpretation
5	4.21 – 5.00	Very Acceptable
4	3.41 – 4.20	Acceptable
3	2.61 – 3.40	Moderately Acceptable
2	1.81 – 2.60	Fairly Acceptable
1	1.00 – 1.80	Not Acceptable

Table 4: Rating Scale Used in the Validation of the Instrument

4.5. Data Gathering Procedure
        The stages used in the test development and validation process were based from the stages developed by Sevilla et. al. (1984), and the instructional design developed by Aguirre, Jr. & de Cadiz (2013): Stage I, Planning Stage; Stage II, Test Construction Stage; Stage III, Tryout Stage; and Stage IV, Evaluation Stage, as shown below:
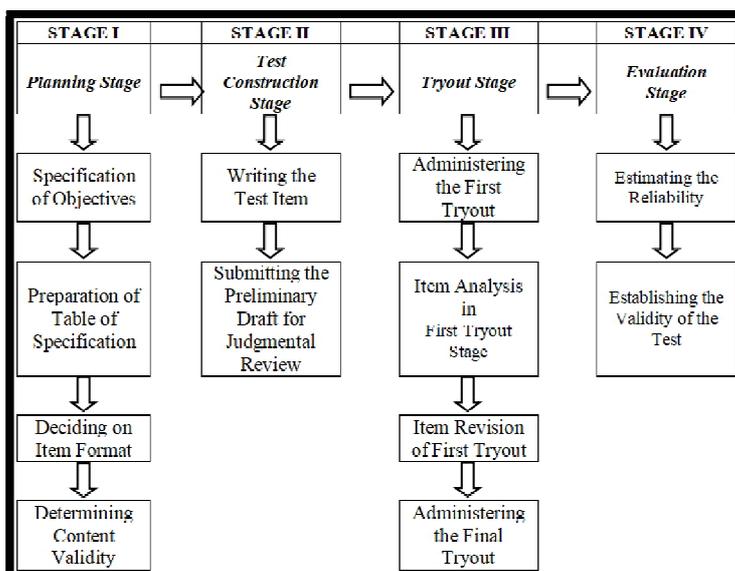

Figure 1: Stages in Test Development and Validation Process

        In the planning stage, objectives are aligned to the learning competencies (outcomes) of the K to 12 Music Curriculum Guide in the development of the achievement test in Philippine Music; specifically in the cognitive domain aspect in terms of knowledge, comprehension, application, analysis, synthesis, and evaluation level. The test items were based from a table of specifications, prepared by the researcher in accordance to the Content Standard and Learning Competencies as indicated in the K to 12 Music Curriculum Guide (Department of Education, 2016) and to the Taxonomy of Educational Objectives by Bloom (1956) to make the initial draft of the achievement test of 115 items. Multiple-choice test items were constructed as the main instrument for the achievement test as generally recognized as the most widely applicable and useful type of objective test item (Linn & Gronlund, 2000). The initial number of items was 35 items more than the considered number in the final form of the instrument.
        In the test construction stage, Bloom's taxonomy of cognitive objectives may seem cumbersome and irrelevant to the music teacher, who will be able to see the importance of the categories into which the various aspects of cognition have been placed. Bloom and his co-authors believe that the taxonomy does progress from simple to complex so that examination of the order in which the categories of learning have been placed can be very helpful for planning learning experiences (Colwell, 1970).
        In the tryout stage, the researcher wrote a letter of request to the Schools Division Superintendent and concurrent Officer-in-charge Office of the Assistant Regional Director, Dr. Wilfredo E. Cabral, to allow the researcher to conduct the research. Request letters were also prepared for the secondary school principals of the selected public high schools. After the test administration, the researcher checked the papers and subjected them to item analysis. Deciding whether to retain or discard an item was based on two ranges. In deciding the test items to be retained, revised or rejected, the Difficulty Index within 0.26–0.75 and the Discrimination Index of 0.21 and above were retained. Items that passed only one of the given ranges were considered for revision, while those items that failed in both given range of the indices were discarded. Table 5 shows the difficulty and discrimination indices of an item.

Index Range	Difficulty Level	Discrimination Level
0.00 – 0.10	Very Difficult	Questionable Item
0.11 – 0.20		Not Discriminating
0.21 – 0.30	Difficult	Moderately Discriminating
0.31 – 0.40		Discriminating
0.41 – 0.60	Optimum Difficult	Very Discriminating
0.61 – 0.80	Easy	
0.81 – 1.00	Very Easy	

Table 5: Difficulty and Discrimination Indices of an Item

In the evaluation stage to ensure reliability, the achievement test underwent the process of split-half reliability method, the use of the Pearson Product-Moment Coefficient of Correlation and correction of Spearman-Brown Formula applied to the coefficient. Similarly, the data gathered through predictive validity determined by correlating the scores with the subjects' final grades in the Philippine Music Achievement Test (PMAT). Furthermore, the researcher initiated the content-related evidence of validity which refers to the content and format of the instrument which was done through content and face validity.

4.6. Statistical Treatment of Data

The following statistical tools were used to analyze and interpret the data:
- Percentage - a proportion in relation to a whole to describe the data vividly.
- Weighted Mean - used to compute and to determine the intensity of the response of the respondents for each behavioral indicators in music learning.
- Average Weighted Mean - the average (mean) of the weighted mean scores (ratings).
- Pearson's r Coefficient of Correlation – a measure of correlation appropriate when the data represent either interval or rational scales; it takes into account each score and procedures a coefficient between 0.00 and +1.00.
- Split-Half Method of Reliability – a method of estimating reliability.
- Spearman-Brown Formula - makes a clear useful principle of test reliability.
- Difficulty Index – refers to the proportion of the number of students in the upper and lower groups who answered an item correctly.
- Discrimination Index – refers to the proportion of the students in the upper group who got an item right minus the proportion of students in the lower group who got an item right.

5. Results and Discussion

5.1. Construction of Test Items in the K to 12 Music Curriculum Learning Competencies

Cognitive Domain of the Taxonomy of Educational Objectives	Music Learning Competencies (Philippine Music)	Unit
Knowledge, Synthesis	Identifies musical characteristics of representative music selections.	I, II, III, IV
Knowledge, Synthesis	Narrates the origins and cultural background of selected music selections.	IV
Knowledge, Synthesis	Describes how a specific idea or story is communicated through music in a particular music selection or musical production.	IV
Comprehension, Synthesis	Explains the distinguishing characteristics of representative Philippine music selections in relation to its culture and geography.	I, II, III, IV
Analysis	Analyzes the musical elements of some vocal and instrumental music selections.	I, II, III
Synthesis, Application	Describes how music contributes to the performance of the musical production.	IV
Application, Evaluation	Evaluates music and music performances applying knowledge of musical elements and styles.	I, II, III

Table 6: Cognitive Domain of the Taxonomy of Educational Objectives in the Music Learning Competencies

Cognitive Instructional Objectives	Item Placement				Total	%
	Unit I	Unit II	Unit III	Unit IV		
Knowledge	1, 8, 11, 17, 20, 26, 28	31, 43, 44, 48, 61, 64	68, 78, 81, 87, 90, 93, 95	96, 99, 103, 104, 108, 111, 113, 114	28	24%
Comprehension	2, 4, 9, 15, 21	33, 35, 36, 39, 41, 45, 46, 50, 59	66, 67, 69, 76, 83, 84, 86, 88, 89, 92	100, 105, 107, 112	30	26%
Application	5, 10, 18, 23, 25, 29	37, 53, 54, 57, 65	70, 75, 80, 85	106	13	11%
Analysis	6, 7, 13, 30	32, 34, 42, 47, 51, 56, 60	72, 77, 79, 82, 91, 94	102, 110	20	18%
Synthesis	3, 12, 14, 19, 22, 24, 27	38, 40, 49, 52, 58, 62, 63	71, 73, 74	97, 98, 101, 109, 115	22	19%
Evaluation	16	55	-	-	2	2%
Total	30	35	30	20	115	100%
Percentage	26%	30%	26%	18%	100%	-

Table 7: Placement, Frequency, and Percentage Distribution of Test Items in the Cognitive Instructional Objectives

Table 6 shows the cognitive domain of the taxonomy of educational objectives aligned to the music learning competencies. Through the item inspection on the two-way table of specifications in the first tryout, in terms of the cognitive instructional objectives, as shown in Table 7, there were 28 or 24% in the Knowledge Level, 30 or 26% in the Comprehension Level, 13 or 11% in the Application Level, 20 or 18% in the Analysis Level, 22 or 19% in the Synthesis Level, and 2 or 2% in the Evaluation Level. According to the Taxonomy of Educational Objectives, Knowledge and Comprehension Level are classified in the Low Order Thinking Skills (LOTS) and the remaining (Application, Analysis, Synthesis and Evaluation Level) are classified in the High Order Thinking Skills (HOTS), therefore, in this case, 58 items are classified in the Low Order Thinking Skills (LOTS) and 57 items in the High Order Thinking Skills (HOTS), a total of 115 test items.

In addition, according to the content in the first tryout of test administration as shown in Table 7: Unit I which is concerned with the Music of Luzon (Lowlands) has 30 items prevailing 26% in the whole achievement test; 35 items or 30% in Unit II, Music of Cordillera, Mindoro, Palawan, and the Visayas; 30 items or 26% in Unit III, Music of Mindanao; and 20 items or 18% in Unit IV, Philippine Festivals and Theatrical Forms.

5.2. Test Items in the Cognitive Domain of the Taxonomy of Educational Objectives Aligned to the Music Learning Competencies

The sample questions in line with the taxonomy of educational objectives, cognitive domain are as follows:

A. Knowledge
        Rondalla was first started in _____.
                A. the Philippines   C. South America
                B. Portugal   D. Spain

In this example, it is presented through the knowledge of specific facts. A fact may be simple or complex, a specific is an isolated bit of information, meaningful in itself. In this particular question, it is in the form of an incomplete-statement which is more concise to present a well-defined problem.

B. Comprehension
        Various redefined compositions of Nicanor Abelardo in a more artistic way such as Mutya
        ng Pasig, Nasaan Ka Irog and Bituing Marikit are all examples of _____.
                A. Folk Song   C. Sacred Music
                B. Kundiman   D. Sarsuela

In this example, it is presented under interpretation which includes a translation of various parts of the problem and comprehension of the relationships that exist between the parts. In other words, it involves the ability to locate the major elements or ideas and the knowledge of how these fit together. It is in the form of an incomplete-statement which is more concise to present a well-defined problem.

C. Application
        The Kulintang is played by _____.
                A. anyone in the group.
                B. a young man without looking at the gongs.

       C. an aged man and accompanied by the group.

       D. a woman and the melodic instrument in the group.

In the question given, a methodology is used by the performing musician and other methods of how certain things will be processed. It is in the form of an incomplete-statement which is more concise to present a well-defined problem, given an alternative that mainly focuses on the applied knowledge about the topic.

D. Analysis

       Which of these unique groups in the Philippines drew music from unconventional bamboo instruments made of six types of versatile bamboo in various sizes, shapes, and designs?

       A. Angklung Ensemble   C. Pangkat Kawayan

       B. Brass Band   D. Rondalla

As viewed in this example, the analysis of an element is concerned to identify the elements included in communication that gives the ability to recognize unstated assumptions. This particular question is in direct-question form, more likely to present a clearly formulated problem.

E. Synthesis

       Brass band is a group of instruments usually composed of _____.

       A. brass, woodwind, and bamboo instruments

       B. brass, woodwind, and rondalla instruments

       C. brass, woodwind, and string instruments

       D. brass, woodwind, and percussion instruments

In this particular question, it is in the form of an incomplete-statement which is more concise to present a well-defined problem, searching for a variety of sources where the structure of an object corresponds to it. Synthesis is a process of putting together the elements and parts to form a whole.

F. Evaluation

       Which of the following phrases is NOT true about Kundiman?

       A. the used of very fast tempos

       B. the used of triple time signature

       C. the used of minor tonality to a major key

       D. the used of idiomatic expressions

In the example given, the process of critiquing and reviewing a certain topic or idea is observed. Sometimes, it assesses the general probability of accuracy in reporting facts as exactness of statement. Judgments about the value of material and methods for given purposes is the principle of evaluation.

Note: For the complete details of the Philippine Music Achievement Test, you may e-mail the researcher at almighty.tabuena@gmail.com or send a personal message through the Facebook account.

5.3. Difficulty and Discrimination Indices of the Test Items

Tryout	Mean			
	Difficulty Index	Interpretation	Discrimination Index	Interpretation
After the First Tryout	0.357	Difficult	0.150	Not Discriminating
Before the Final Tryout	0.431	Optimum Difficult	0.214	Moderately Discriminating

Table 8: Mean Difficulty and Discrimination Index after the First Tryout and before the Final Tryout of the Philippine Music Achievement Test

From Table 13, the difficulty and discrimination indices of the achievement test in Philippine Music in the first tryout ranged from 0.078 to 0.625 and 0.000 to 0.469, respectively. The mean difficulty index of the test was 0.357 while the discrimination index, 0.150 which means that the test is difficult within the difficultyindex of 0.26 – 0.75 accepted range, and its discriminating power is not discriminating. After the item analysis, rejected items were eliminated. The difficulty and discrimination indices of the achievement test in Philippine Music before the final testing ranged from 0.234 to 0.828 and 0.000 to 0.625, respectively. The mean difficulty index of the test was 0.431 while the discrimination index, 0.214 which means that the test is of optimum difficulty within the difficulty index of 0.26 – 0.75 accepted range, and its discriminating power is of moderately discriminating within the discrimination index of 0.21 and above accepted range. This implies that the test is of optimum difficulty and of moderately discriminating.

Items	Unit				Total
	I (1-30)	II (31-65)	III (66-95)	IV (96-115)	
Retained	4-6, 8, 15-23, 26-27, 30	32, 44, 46, 55 57, 60, 65	68, 71, 73, 76-77, 80, 82, 88- 89, 92	97 -98, 100- 101, 103-104 106, 108-110	43
Revised	1, 3, 10, 13, 14, 24, 25, 28	34-36, 38-39, 41, 50, 52, 56 58-59, 61, 64	66-67, 69, 74- 75, 78, 81, 84-85, 90	96, 99, 102, 107, 112, 115	37
Rejected	2, 7, 9, 11-12, 29	31, 33, 37, 40 42-43, 45, 47- 49, 51, 53-54 62-63	70, 72, 79, 83 86-87, 91, 93-95	105, 111, 113-114	35
Total	30	35	30	20	115

Table 9: Summary of the Items Retained, Revised, and Rejected

Unit	Number of Items							
	Test Construction Stage	%	Revised Tryout Stage	%	Retained Tryout Stage	%	Retained and Revised for Final Testing	%
I	30	26 %	8	22 %	16	37.21 %	24	30 %
II	35	30 %	13	35 %	7	16.27 %	20	25 %
III	30	26 %	10	27 %	10	23.26 %	20	25 %
IV	20	18 %	6	16 %	10	23.26 %	16	20 %
	115	100 %	37	100 %	43	100 %	80	100 %

Table 10: Frequency and Percentage Distribution of Retained and Revised Items in the First Tryout Stage and for Final Testing

The results of the first administration of the achievement test show that out of 115 items 43 items were retained, 37 items were subjected for revision and 35 items were rejected. In the final testing, there is a total of 80 items as shown in Table 9. Among the 80 items from the final set of the test, Unit I composed of 24 items or 30% of the total items, Unit II has 20 items (25%), Unit III has 20 items (25%), and Unit IV has 16 items (20%) as shown in Table 10.

5.4. Reliability of the Test

Total Odd-Scheme Scores (X)	Total Even-Scheme Scores (Y)	Total Squared of Odd-Scheme Scores (X²)	Total Squared of Even-Scheme Scores (Y²)	Total Product of Odd-Even Scheme (XY)	Computed r_{oe} (Pearson r)	Computed r_t (Spearman-Brown Formula)
1699	1664	30029	29046	29147	0.697	0.821

Table 11: Split-Half Method of Reliability Computation Summary of the Relationships of Odd and Even Scores (N = 100)

Computed r–value (Coefficient Correlation)	Interpretation (Degree/Extent of Relationship)
0.00 to + 0.10	No correlation
+0.11 to +0.25	Negligible correlation
+0.26 to +0. 50	Moderate correlation
+0.51 to +0.75	High correlation
+0.76 to 1.00	Very high positive / negative correlation

Table 12: Interpretation of Computed r-value

Table 11 shows the total number of scores in the computation of using the Split. According to Subong, Jr. (2005), to measure the degree or strength of the correlation, Table 12 may be used in interpreting the coefficient of correlation. From the sum of each column in Table 11, the odd-even division scheme, the reliability of one–half of the test was computed using Pearson Product–Moment Correlation Coefficient; the computed value of *r* = 0.697 or 0.70. This value implies a high correlation between the scores of the odd–even items of the whole test. To further establish the reliability of the whole test, the Spearman–Brown Formula was utilized. The coefficient of reliability of the entire test was 0.821 which implies very high reliability. Therefore, the Philippine Music Achievement Test is reliable.

5.5. Face and Content Validity of the Test

Indicators	Weighted Mean	Interpretation
Format and Design	4.47	Very Acceptable
Objective	4.40	Very Acceptable
Content	4.50	Very Acceptable
Language	4.47	Very Acceptable
Average Weighted Mean	4.46	Very Acceptable

Table 13: Face and Content Validity of the Philippine Music Achievement Test by the Validators

The validation instrument is composed of 14 items that are arranged in terms of format and design, objectives, content, and language. Based on Table 13, in terms of the format and design in the initial draft of the achievement test for tryout stage, the obtained value of the weighted mean among the five validators was 4.47 interpreted as acceptable; in terms of the objective, 4.40 interpreted as acceptable; in terms of the content, 4.50 interpreted as very acceptable; and in terms of the language, 4.47 interpreted as acceptable. In this case, the validators' judgment on the appropriateness, suitability, and mechanics in the construction of the test is remarkably considered the face validity in examining the test. All of the five validators interpreted and marked the initial draft of the achievement test as very acceptable. The overall rating of the initial draft of the achievement test for the pilot testing was 4.46 interpreted as very acceptable. Therefore, the achievement test is acceptable and valid.

5.6. Predictive Validity

Total Scores in the Subjects' Final Grade in Music (X)	Total Scores in the Philippine Music Achievement Test (Y)	Total Squared of Scores in the Subjects' Final Grade in Music (X²)	Total Squared of Scores in the PMAT (Y²)	Total Product of Scores in the Final Grade in Music and PMAT (XY)	Computed r-value (Pearson r)
3363	8671	117369	752891	293634	0.9677

Table 14: Predictive Validity of the Music Final Grades in the Philippine Music Achievement Test Scores

Table 14 shows the predictive validity which determined by correlating the scores with the subjects' final grades in the Philippine Music Achievement Test (PMAT). Using Pearson Product–Moment Correlation Coefficient; the computed value of *r* is 0.9677 or 0.97 which implies a high correlation between the scores. Therefore, the Philippine Music Achievement Test is valid.

6. Conclusions and Recommendations

6.1. Conclusions

Based on the findings, the following conclusions were drawn:

- The Philippine Music Achievement Test, an 80-item test with four options in each item, includes both Low Order Thinking Skills (LOTS) and High Order Thinking Skills (HOTS) which comprised of the indicated learning competencies in the K to 12 Music Curriculum with the areas in Music of Luzon (Lowlands), Music of Cordillera, Mindoro, Palawan, and the Visayas, Music of Mindanao, and the Philippine Festivals and Theatrical Forms. In addition, the test items are consistent with and responsive to the learning competencies of the K to 12 Music curriculum and situated to the cognitive process dimensions of Bloom.
- The difficulty and discrimination indices of the achievement test in Philippine Music before the final testing ranged from 0.234 to 0.828 and 0.000 to 0.625, respectively. The mean difficulty index of the test was 0.431 while the discrimination index, 0.214 which means that the test is of optimum difficulty within the difficulty index of 0.26 – 0.75 accepted range, and its discriminating power is of moderately discriminating within the discrimination index of 0.21 and above accepted range. This implies that the test is of optimum difficulty and of moderately discriminating.
- Test items were very acceptable in terms of format and design, objectives, content, and language. The coefficient of reliability of the entire test implies very high reliability. In addition, the coefficient of the validity of the scores with the subjects' final grades in the Philippine Music Achievement Test implies a very high validity. Therefore, the Philippine Music Achievement Test is reliable and valid.

6.2. Recommendations

In light of the findings of the study, the following recommendations are here offered:

- The Philippine Music Achievement Test for junior high school students may serve as a reliable alternative instrument in assessing the performance of the students in the said subject. The achievement test may be used as a final examination to measure the skills learned in the cognitive domain after a school year of the teaching–learning process.

- The achievement test may serve as a sample of developing a better achievement test for the evaluation of the students' skills in the cognitive domain. The developed test may be administered to a broader range of public high school students in other Division City Schools locale or other regions and/or to private schools throughout the country to ascertain its geographical applicability and to validate the norms established.
- Further validation in other settings by other researchers might still improve the developed achievement test. Parallel studies on the test in music learning may be developed for the other grade/year levels.
- The future music researchers in the development and validation of tests should attend training in test construction, as well as to be in-depth and comprehensively knowledgeable in measurement and evaluation particularly in Music learning.

7. References

i. Aguirre, Jr. D. O. & de Cadiz, G. B. (2013). *Instructional materials development manual.* Eastern Visayas State University, Tacloban City.
ii. Asaad, A. S. & Hailaya, W. M. (2004). *Measurement and evaluation: concepts and principles.* Manila: Rex Book Store, Inc.
iii. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice,* 5(1), 7-74.
iv. Bloom, B. S. (1956). *Taxonomy of educational objectives.* New York: David McKay Company, Inc.
v. Calmorin L. P. (2010). *Research and statistics with computer.* Mandaluyong City: National Books Store.
vi. Colwell, R. (1970). *The evaluation of music teaching and learning.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
vii. Department of Education (2016). *K to 12 music curriculum guide (grade 1 to grade 10).* Pasig City: DepEd.
viii. Department of Education (DepEd) Order No. 8, s. 2015 'Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program.' Pasig City: DepEd.
ix. Department of Education (DepEd) Order No. 79, s. 2003 'Assessment and Evaluation of Learning and Reporting of Students Progress in Public Elementary and Secondary Schools.' Pasig City: DepEd.
x. Flores, Honey Jane M. (2002). *Development of modules in music.* Manila: Philippine Normal University.
xi. Linn, R. L. & Gronlund, N. E. (2000). *Measurement and assessment in teaching.* Upper Saddle River, New Jersey: Prentice-Hall, Inc.
xii. Narzo, P. M. (2005). *Arrangement of selected musical pieces for the angklung ensemble.* Manila: Philippine Normal University.
xiii. Parrotina, Lorna R. (1994). *Development and validation of a music achievement test for grade six.* Manila: Philippine Normal University.
xiv. Sevilla, C. et. al. (1984). *An introduction to research methods.* Manila, Philippines: Rex Book Store.
xv. Simeon, Florisa B. (2003). *Development and validation of an achievement test in araling panlipunan for first year high school students.* Manila: Philippine Normal University.
xvi. Stanley, J. C. & Hopkins, K. D. (1972). *Educational and psychological measurement and evaluation.* New Jersey: Prentice-Hall, Inc.
xvii. Stronge, J. H. (2010). *Evaluating what good teachers do.* New York: Eye on Education, Inc.
xviii. Subong, Jr. P. E. (2005). *Statistics for research.* Manila: Rex Book Store, Inc.
xix. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.