

Development of Diagnostic Test in Reading for Grade 7

ART V. MANGUBAT¹

¹Bula National High School, Bula, Camarines Sur, Philippines
<https://orcid.org/0009-0004-5697-758X>

Corresponding Author: art.mangubat@deped.gov.ph

Originality: 100%

Grammarly: 99%

Plagiarism: 0%

ABSTRACT

Article history:

Received: 10 Jun 2022

Revised: 13 Dec 2022

Accepted: 31 Mar 2023

Published: 26 Jul 2023

Keywords — Language arts and reading, reading test development, mixed-method developmental research design, Philippines

Reading is unquestionably a crucial component of any educational system because it not only underlies it but also permeates its aspects. Therefore, reading should be appropriately assessed. This mixed-methods developmental study developed a test package for assessing the reading competency of Grade 7 learners. The researcher used a variety of scales for validation, attribute assessment, assessment adherence principles, and indicators of a good reading diagnostic test. Analysis methods included mean, weighted mean, difficulty-discrimination indices, and Cronbach's Alpha. It was revealed that the comprehension test has a strong reliability score for a classroom test ($\alpha=.700$).

In contrast, the vocabulary test has a somewhat low-reliability score ($\alpha=.600$). The validity results indicated that the fluency test was extremely valid (4.535), while both vocabulary and comprehension tests were very valid (4.322, 4.486). It was described as very good (4.33), and the experts concurred that it adheres to the indicators of a good reading diagnostic test (fluency: 1.00; vocabulary: 0.97; and comprehension: 0.97). Generally, the reading diagnostic test



© Art V. Mangubat (2023). Open Access. This article published by JPAIR Multidisciplinary Research is licensed under a Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0). You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material). Under the following terms, you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes. To view a copy of this license, visit: <https://creativecommons.org/licenses/by-nc/4.0/>

is a valid and reliable instrument for assessing students' reading competence levels, particularly in terms of fluency, vocabulary, and comprehension, but further validation is required.

INTRODUCTION

With the Education for All (EFA) goal in mind, the Department of Education (DepEd) exerted efforts to improve its quality by implementing Republic Act No. 10533, otherwise known as the “Enhanced Basic Education Act of 2013,” which aims to have Filipino graduates who are holistically developed with 21st-century skills prepared for higher education, middle-level skills development, employment, and entrepreneurship (SEAMEO INNOTECH, 2012). More specifically, they envision Filipino graduates to be proficient in the essential competencies (i.e., literacy, numeracy, and problem-solving) to develop themselves fully. However, this would be difficult to realize and materialize since one of the fundamental skills they need to succeed is at stake – literacy. With low literacy comes low reading proficiency. Reading test results from Elementary Schools within the Division of Manila show that only one-sixth to one-third of students can read independently at the appropriate grade level. If Metro Manila's reading proficiency is poor, one can only anticipate much worse reading results in other parts of the nation with fewer school resources (Luz, 2017). Considering the pressing issue, the researcher focused on one literacy skill that is apparently on a problematic scale, i.e., reading.

Using a developed teacher-made tool on the SY 2017-2018, six are Highly Competent, 55 are Moderately Competent, 50 are Competent, 30 are Slightly Competent, and 11 are Not Competent readers. This SY 2018-2019, using the same teacher-made tool, there is no Highly Competent reader recorded, 9 of which are Moderately Competent, 47 are Competent, 45 are Slightly Competent, and 15 are identified as Not Competent readers. At the same time as this SY, the revised Phil-IRI was conducted on the students using the leveled passages for Grades 4 and 5 and found out that only 46 are Independent, 17 are Instructional, 34 are in the Frustration level, and 18 are Non-readers.

From the data presented above, it is alarming that the level of Frustration between readers and Non-readers is increasing over time, and the Slightly Competent and Not Competent readers are doing the same. Apart from these statistics, the school used several reading assessment tools, which caused the fluctuation in the number of frustrated, non-readers, slightly competent, and not competent readers. Since there is no established or standard assessment tool for reading in Junior High School (JHS), it is difficult to discover the literacy trend of the school.

Broadly, reading can be considered an activity characterized by translating letters into words and sentences that have meaning to the individual (Musa & Balami, 2016). Undoubtedly, reading is an indispensable aspect of the educational system because it is not only superimposed on it but also crisscrosses all its facets. Thus, teachers need to properly diagnose the reading skills that need more focus. Teachers could not pinpoint students and regard them as non-readers or struggling readers. However, teachers can usually distinguish between students who struggle with reading and those who are proficient by observing and noting reading behaviors. These behaviors may be evident throughout the reading. They will be determined by the nature of the student's reading difficulty as well as the content and context of the reading task. Hence, we cannot verify the veracity of one's judgment without a proper tool to assess one's reading levels and behaviors. The Phil-IRI is the assessment part of ECARP and the most familiar reading assessment tool for educators, especially elementary teachers.

The Phil-IRI was recently updated through DepEd Order No. 14 s. 2018 or the "Policy Guidelines on the Administration of the Revised Phil-IRI." As clearly stated in Paragraph 4, Scope of the Policy of DO No. 14, s. 2018, "this tool is administered for Grades 3 to 6 learners in public elementary schools nationwide (DepED, 2018a)." Furthermore, in the Phil-IRI Manual (p. 3), it is reiterated that "all students in Grades 3 to 6 will undergo the Phil-IRI Group Screening Test (GST) in Filipino while students in Grades 4 to 6 will undergo GST in English" (DepED, 2018b). From these guidelines, Junior High School (JHS) is again left with no tool to assess reading. Using the argument that foreign tests are not culturally sound to Philippine culture, the DepEd has opposed using these tests. The tendency to go with an "easier yet effective" diagnostic test is not an option, defeating the purpose of measuring results (Luz, 2017). Formal group assessments (like standardized achievement and competency tests) can flag problems; however, individual assessments can provide valuable information. However, the Department lacks a developed assessment tool for the JHS to diagnose or assess students regarding reading proficiency.

With the unavailability of a reading diagnostic test for Grade 7 students, as mentioned above, and considering reading as a complex cognitive process of different interrelated skills, the researcher deemed it necessary to develop a reading diagnostic test package intended to evaluate their reading competence level concerning the skills in reading development and instruction as identified by Shanahan (2005) of the National Reading Panel and the National Literacy Panel (2006) i.e., phonics, phonemic awareness, fluency, vocabulary, and comprehension, by which the result may be used as basis for a correct and appropriately designed reading intervention program in respective schools. This

was supported by Kooij (2013), who said that diagnostic assessment provides a detailed performance analysis. Diagnostic assessment aims to improve the overall quality and accuracy of interpretation by collecting performance data. Furthermore, Dunkley (2012) believed that results could guide skill development by examining the impact of various patterns on the overall quality and accuracy of the interpretation and determining priorities accordingly. He also suggested that a remediation plan can be developed when severe skill gaps are noted.

FRAMEWORK

Theories and principles always support research. This research used three theories to support its claims. The Classical Test Theory (CTT) served as a Measurement Theory, the Modified Cognitive Model as a Reading Theory, and Test Development as a Theory for Test Construction.

Measurement theory is a theoretical model with an underlying statistical basis for measuring or analyzing data. It further defines measurement theory as a loosely integrated conceptual framework within which rigorously formulated statistical models of estimation and inferences about the properties of measurements or scores. It is believed that reading tests are affected by measurement theory and reading theory, and in turn, they should interact for mutual development (Sainsbury et al., 2006).

The mathematics used for developing and responding to tests and measuring scales is called Classical Test Theory (CTT), commonly called the true score model. (Statisticshowto, n.d.). It evaluates two things: 1) the test and 2) the items. On the test level, the reliability is computed. Reliability is a classical concept that seeks to quantify the consistency or repeatability of the measurement. If a test produces consistent scores, we say it is reliable (Zeng & Wyse, 2009). Reliability cannot be estimated directly; however, reliability estimates can be obtained by various means. The most common estimate is the internal consistency using Cronbach's Alpha, which is used in this research to determine the reliability of the whole test. The higher internal consistency reliability coefficient indicates that the items on the test are similar to each other regarding construct. Also, a high-reliability result often yields high validity.

Classical test theory evaluates individual items from a quantitative perspective on the item level. The goal of item analysis is to use detailed statistics to determine possible flaws in the item and decide whether to retain, revise, or reject the item. Two indexes are computed in an item analysis: the index of difficulty and discrimination. The item measures the difficulty index associated with every test question. It is the percentage of test takers who correctly answered out of

all test takers (Statisticshowto, n. d.). On the other hand, item discrimination refers to the item's ability to distinguish between examinees with high and low knowledge or skill levels. These computations were also done in the study to arrive at a refined test.

McKenna and Stahl's Modified Cognitive Reading Model portrays reading as integrating three strands: automatic word recognition, language comprehension, and strategic knowledge (2009). Automatic Word Recognition allows the student to focus his/her attention on comprehension. Strategic Knowledge includes the general and specific purposes for reading and knowledge of strategies to use when reading (Stahl et al., 2019, p. 23). Language Comprehension is the understanding of written and spoken language, which relies on correctly processing word and phrase meanings, sentence grammar, and discourse or text structure (Garnham, 2005). This model was based on the premise that reading comprehension is the primary purpose of reading (Assessment for Reading Instruction, 2003). This theory, as applied in this study, testifies to the process of reading, its goal, and how it may be assessed, considering that reading is a complex, interrelated process. Hence, each component must be individually and adequately assessed to generally arrive at an assessment result: reading as a whole.

Test development comprises the full array of activities associated with implementing an assessment (LaDuca et al., 1995). In developing any language test, the cycle of construction of developing and using tests by Carroll and Hall (1985) must be used. They divide the cycle into the following four phases: (1) Design, (2) Development, (3) Operation, and (4) Monitoring (Yuko, 2005). The Design Phase involves three tasks: description of testees, specification of settings and needs, and statement of test tasks and topics. In the Development Phase, a test developer must construct the draft test. Afterward, the test will undergo trial testing followed by the analysis of test trials and test revision. Phase 3, or the Operation Phase, begins with introducing a test for practical use, and with test information gathered, the developer must make necessary decisions. The last step is Monitoring, and in this step, a survey of test administration must be conducted, test measurement characteristics will be established, and the test revision schedule will be prepared.

The Design Phase, as applied in this research, started with determining the target respondents which served as a pre-survey on the number of teachers and student respondents out of the fifteen (15) secondary schools of Tinambac North and South District of the Division of Camarines Sur, Bicol. Afterwards, specifications are set. In this step, the researcher wrote the objectives, which determine the test's reading and development goals. The next step is to counter-

check the alignment to the DepEd K-12 Competencies and the five areas of reading and development. Crafting the Table of Specifications is the last part of the design phase. The Classical Test Theory (CTT) was used to design the test as it evaluates the level of the whole test (reliability) and its items (item analysis).

On the other hand, the Development Phase commences from the construction of the draft test. This step involves searching compatible information/materials that might fit the competencies set for the test; selecting the format and parts of the test, forms to accomplish, font style, font size, general appearance, and layout; and benchmarking on other reading assessments or diagnostic tests. The next step is the validation of the test. In this part, three (3) jurors, i.e., the reading experts, validate and revise the instrument before the trial testing. In the trial testing of the draft test, the researcher sought permission from its respondents, wherein a letter of request and permission were duly signed by the adviser and given to the school principals of the target schools for approval. When the approval was granted, the researcher proceeded to administer the test. Besides the trial testing, survey questionnaires were also given to teacher respondents to evaluate the test package. An analysis of the survey questionnaires from the teacher respondents is done. Test revision and refinement were done using the results of the analyses from the trial testing the survey questionnaires, including the indicators of a good test as recommended by experts.

This research will only undergo the design and development phases, and the other phases are subjected to further research. The test development theory serves as the underlying principle on how the test must be crafted and summarizes the test construction process, connecting the other two theories in this research. It can be explained that to achieve the goal of this research i.e., the Reading Diagnostic Test for Grade 7 is encompassed with the Classical Test Theory (CTT) wherein item level and test level analysis and evaluation must be observed as the test is being developed. With its development, two other theories interact the Modified Cognitive Reading Model and Test Development. As it is designed and developed, the cognitive reading process must be considered, meaning that the test should start with the sub-skill of reading, and cumulatively, they would obtain the level of reading competence. These two are also observed during the item level and test level analysis and evaluation. With this framework, the researcher believes in the acceptability of its output in the academe.

OBJECTIVES OF THE STUDY

This study aimed to design and develop a diagnostic test for reading intended for Grade 7 learners to assess the students' reading competence level.

It is anchored on the following objectives: (1) develop a teacher-made reading diagnostic test for Grade 7 students along fluency, vocabulary, and comprehension; (2) determine its index of discrimination, index of difficulty, and test validity; and 3) identify the indicators of a good reading diagnostic test based on reading experts; reading teachers, and education program specialist.

METHODOLOGY

Research Design

The study employed a mixed method design; specifically, descriptive surveys, document analysis, and abstraction methods were employed to attain further the data necessary to address the objectives. In survey research, a group of people or items is examined by gathering and examining data from a small number of individuals or items representing the entire group. The recursive abstraction method is a simple method mainly based on summarizing the data. It starts by summarizing a set of data, then summarizing the summarized portion, and so on, until we end up with a focused and very compact summary that is both accurate and distinct (Hardcastle et al., 2006). Comparatively, qualitative document analysis is qualitative research in which the researcher interprets documents to provide context and voice for an evaluation issue. Document analysis involves categorizing information into themes, like focus groups or interview transcript analysis (Smulowitz, 2017). Such methods will be used since this study only describes specific variables about the target respondents and analyzes the recommendations. This is concerned with a description of the material as it is. These methods yield better necessary information as to what teacher-made reading diagnostic test may be developed to assess the reading levels of Grade 7 students along fluency, vocabulary, and comprehension; the classification of the test items in terms of index of discrimination, index of difficulty and test validity; and the indicators of a good reading diagnostic test as recommended by teachers, experts and education program specialist.

Research Site

Tinambac is a first-class Municipality in the Province of Camarines Sur. By geographical characteristic, it is located south of Siruma, north of Calabanga, and east of Goa in level terrain. Located on the slopes of Mt. Isarog, Tinambac virtually retained its original topography despite the expansion of its neighboring municipalities. A large portion of the municipality's land is predominantly used for agriculture, while its hydrographic features favor fishing, which is the sustainable livelihood of the coastal barangays. It is divided into three (3) sites: 1) Coastal, composed of 16 barangays; 2) Upland, comprised of 19 barangays; and

3) Lowland/ Poblacion having eight barangays. Of these 44 barangays, only 15 have Public Secondary Schools (The Saga of a Town, n.d.).

Respondents

Four groups of respondents in this study provide data triangulation: 1) the student respondents, 2) the reading teachers, 3) the education program specialist, and 4) the reading experts.

In identifying the student respondents, the fifteen (15) Schools of Tinambac North and South District, the researcher used both the quota and multi-stage sampling techniques. The purposive selection is made through the geographical category of the upland, coastal, and Poblacion barangays. The School Category was also considered, classified as a small school or a big school. Then, it was subdivided into the number of students per section and further clustered into male and female. Five (5) schools were identified: one (1) big school in the Poblacion with 67 respondents; two (2) small schools in the upland with seven (7) and two (2) respondents, respectively; and two (2) small schools in the coastal area with 15 and nine (9) respondents with a total of 100 student respondents as quota.

This research used the convenience and homogeneous purposive sampling technique for the reading teachers, education program specialists, and reading experts. They are the Grade 7 English teachers who give the reading assessment to entrants, the education program specialist in English of the Division of Camarines Sur, and three reading experts. A homogeneous purposive sample is selected for having a shared characteristic or set of attributes (Crossman, 2017), while convenience sampling is used when the availability of a sample is rare and costly (Singh, 2018). The researcher purposively selected them since they have firsthand and adept knowledge of reading, the English language, and its facets. At the same time, they are based on convenience since the reading teachers are limited in the locale of the study. The education program specialist, including reading experts, considers that the qualifications of these people are also limited and cannot be found in the research's locale.

Instrumentation

Measurement instruments are various methods through which a researcher obtains data from respondents for his research work. This study used survey questionnaires as its research instrument to ensure the validity and reliability of the results. It is said that questionnaires are very practical, that any number of individuals can respond to them, and that the findings can be quickly measured. It is more scientifically accurate than other quantitative research instruments (William, 2018).

The first questionnaire used is the Validation Sheets for the draft test. It is a 5-point Likert-type rating scale containing three parts for each test. Part I would be for the Fluency test, Part II for the Vocabulary test, and Part III is dedicated to the Reading Comprehension test. This questionnaire shall determine the test validity regarding face, construct, and content.

Aside from the validation sheets, the test package is pilot tested to five (5) school respondents wherein 100 students answered and underwent the three-part 100-item test to see if it suits the Grade 7 learners. The draft test package comprises three separate tests, each with different measures. The first part is the 20-item equivalent fluency test to be done orally, one per student for one minute. The second is the 40-item multiple-choice vocabulary test done in groups for 40 minutes. Lastly, the other 40-item multiple choice reading comprehension test was done in groups, too, and also good for 40 minutes. The draft test is administered to determine if it suits the learners and that all necessary revisions are made based on the test results.

Moreover, two adapted survey questionnaires, namely: "Assessment of the Attributes of the Test Package" and "Adherence to the Principles of Assessment of the Test Package," are administered to the English Grade 7 Teachers to verify if the test package is representative of a Diagnostic Test. The first questionnaire is a 5-point Likert-type survey questionnaire. It consists of two parts, though not labeled. Part 1 consists of several profiling information and questions. The second part is where the teacher respondents rated by ticking or checking the quality of the attributes present in the test package. The attributes are categorized into five with sub-features each.

On the other hand, the last questionnaire is a dichotomous type survey questionnaire. Like the first one, it consists of two parts with the same contents as Part 1. The second part contains the principles of assessments, also categorized into five with sub-statements to which the respondents checked its presence or absence in the text package.

Moreover, the researcher considered the books and articles of foreign experts as instruments, for they are the main indicators of a good reading diagnostic test. Lastly, a survey questionnaire is given to the three reading experts to affirm the indicators of a good reading diagnostic test by the foreign experts. Though unlabeled, the first part of the questionnaire asks for the profile of the reading experts. Furthermore, the questionnaires contain three major parts: 1) indicators of a good fluency test, 2) indicators of a good vocabulary test, and 3) indicators of a good reading comprehension test.

Data Analysis

A careful data analysis plan was used to answer the research objectives. Calculating variable frequencies and variances is one method of analyzing quantitative data. (Mertens, 2017). On the other hand, as a rule of thumb, there is no single “right” way for qualitative data analysis. A researcher begins with a large body of knowledge and information and must deploy inductive reasoning, sorting, and categorization and make it precise with key themes (Sofaer, 2002).

The first problem is answered by discussing how the test came to be. It details the process of test design and development. This part talks about the draft of the test package from its conception until its refinement.

The second problem is answered through careful statistical and item analysis wherein each item’s difficulty and discrimination index is computed. The data from the draft test results are used to compute the mentioned computations. Moreover, Cronbach’s Alpha is used to find the reliability coefficient, strengthening the claims of validity derived from the experts’ validation of the test package.

The third problem is answered through the recommendations of the reading teachers, education program specialists, and reading experts obtained by analyzing the results of the three survey questionnaires. A data table is presented, and the mean scores for each survey are shown to reveal their overall judgment of the reading package. Moreover, according to local experts, the indicators of a good test are abstracted. This is supported by the document analysis on established, tested, and validated reading assessments, diagnostic tests, and assessment frameworks, including articles of theories and principles of reading and assessment where authors, considered experts, have presented their viewpoints on the characteristics or indicators of a good reading assessment tool. Moreover, the comments, suggestions, and overall feedback of the Reading Teachers and Education Program Specialist are considered if they do not contradict the principles and guidelines in test construction on the crafting of the refined test.

The data from the three problems, together with the theories of learning, reading and instruction and principles of assessment that back up the study, including the features of a good assessment tool as recommended by experts, reading teachers, and education program specialists, are used as bases for the crafting of the refined test. This refined test is the output of this research.

RESULTS AND DISCUSSION

A teacher-made reading diagnostic test that is valid and reliable, consisting of items for fluency, vocabulary, and comprehension, can be developed to assess the reading competence level of Grade 7 students. It was successfully

made by the researcher and was subjected to validation.

Under the CTT, finding out the indices of discrimination and difficulty and its cross-sectioning, as well as its test validity and reliability, is needed.

For the discrimination index, 10 items of the 40-item Vocabulary test under Excellent Item/ Very Discriminating. Five items are classified as Very Good Items., 11 are Good Items/ Moderately Discriminating, eight are Fair Items/ Not Discriminating, and six are classified as Questionable/ Poor Items. Furthermore, in the Reading Comprehension test, 14 items are very discriminating/excellent, with six items classified as discriminating/very good. Nine items were considered Moderately Discriminating/ Good Item, and five were interpreted as Not Discriminating/ Fair Item. Finally, six items are classified as Questionable/ Poor Item. The item distribution of the difficulty index for the Vocabulary test is as follows: a) seven items are classified as Easy; b) 23 items are classified as Moderate, and 10 items are regarded as Difficult. The item distribution of the difficulty index for the Reading Comprehension test is as follows: a) only one item is classified as Easy; b) 30 items are classified as Moderate; and 9 items are regarded as Difficult. From the cross-sectioning of the index of discrimination and index difficulty of the Vocabulary test, 15 items are to be retained, 19 items need revision, and six items are to be rejected. From the cross-sectioning of the index of discrimination and index difficulty in the Reading Comprehension test, 25 items are to be retained. Nine items need revision, and six items are to be rejected.

In terms of the result of Test Validity, the Fluency test has a Face Validity mean of 4.600 and is interpreted as Extremely Valid. The construct Validity mean is 4.338, which means Very Valid, and the Content Validity mean is 4.667, interpreted as Extremely Valid. The Vocabulary test's Face Validity is 4.333. Construct Validity is 4.234, and the Content Validity mean is 4.400. All of them are interpreted as Very valid. The face validity mean in the Reading Comprehension test is 4.600, interpreted as extremely valid. The construct Validity mean is 4.325, which means Very Valid, and the Content Validity mean is 4.533, interpreted as Extremely Valid.

In terms of the inter-rater reliability, the Fleiss' Kappa value of the Fluency test is 0.034 or Slight Agreement. The Vocabulary and the Reading Comprehension tests are negative in value of -0.751 and -0.142, respectively, which means Poor Agreement. However, regarding internal consistency identified through Cronbach's Alpha, the Vocabulary test, an α of .600 is computed from the test scores. It is interpreted as Somewhat Low. Hence, the test needs to be supplemented by other measures, and some items could be improved. For the Reading Comprehension test, an α of .700 is computed from the test scores. It is

interpreted as Good for a Classroom Test. This means there are a few items that could still be improved.

According to foreign and local reading experts, these are the indicators of a good reading diagnostic test. A diagnostic test should measure the students' skills, weaknesses, and strengths (Local Expert 2). A fluency test must measure rate, accuracy, and prosody. Also, administration, scoring, and interpretation must be quick and easy because teaching and progress monitoring time must be maximized (Snellings et al., 2010). Accuracy could be assessed by counting the number of errors per 100 words. Assessing rate must consider both automaticity and reading speed, and prosody can be assessed through observation. It may be quantified through the Multi-Dimensional Fluency Rubrics developed by Zutell and Rasinski (1991) (Hudson et al., 2005).

In terms of the selection of material or text, the passage may be taken from any different reading materials for as long as it matches the student's reading level. Expository texts that the students have not previously read must be used. The passage must be long enough that students cannot completely read the entire passage within a minute (Linan-Thompson et al., 2001). Furthermore, text choice is crucial in determining what competencies the students are good at and which ones they find difficult (Expert 2).

With regards to the vocabulary and comprehension test, a multiple-choice type of test is an ideal format that can express the levels of cognitive complexity (Clay, 2011). It is one of the most common formats for testing students' reading abilities (Alderson, 2000). Four or five choices per question are optimal for an item (Cohen & Wollack, 2015). Sentences should not be vague, and choices for vocabulary should be simplified and not more difficult than the unfamiliar word/s (Local Expert 2). Grammatical clues should be avoided in the stem. Distractors should be the same form of the word as the correct answer or the concept of parallelism among the choices should be observed. If nouns or verbs are used, these should be consistent for each number (Local Expert 2). Apart from that, all options should be on nearly the same level of difficulty and avoid needless redundancy. Also, distractors with similar meanings and using a pair of words with opposite meanings should be avoided.

Most importantly, each item should contain only one correct or best answer. The correct answer should not match the exact words in the passage. A test should also avoid using "All of the above" as an option (Local Expert 2). For comprehension, items should require comprehension of the passage. It should not be feasible to answer items correctly purely from general knowledge, and all the options should be roughly the same length (Yu & Lin, 2002).

As evaluated by the reading teacher, the mean scores of the reading

teachers regarding the categories of the test attributes revealed that the Appearance and Characteristics of the Test tied at the first rank with 4.92 and is interpreted as Excellent. Grammar and Structure ranked third with 4.80, followed by Content Level with 4.76, and last is the Hierarchy of Taxonomy with 4.71. However, all of the interpretations of the other attribute categories are still Excellent. Nevertheless, the total mean score is 4.82, also interpreted as Excellent. Moreover, the reading teachers all agreed with validity, reliability, flexibility, fairness, and currency. For the betterment of the test, a reading teacher added that it would be better if there were consistency in the number of choices in every item with four options of the same level. Regarding the overall feedback, the reading teachers gave the researcher positive feedback and encouragement.

For the Education Program Specialist (EPS), the mean score of the Test Attributes survey shows that 3.84 is still interpreted as Very Good. With the categories, Content Level ranked first for her with 4.40, followed by the Hierarchy of Taxonomy with 4.33, and Grammar and Structure in its third spot with 3.88. The three categories are interpreted as Very Good, while Appearance 3.40 and Characteristics of the Test with 3.20 ranked fourth and fifth, respectively, and interpreted as Good. She suggested it would be better if the number of options is consistent, as expressed in attribute A.5 (Uniformity of Elements). She also said, “There are letters omitted due to typographical errors,” as expressed in attribute B.6 (Spelling).

Moreover, she also suggested that “Vocabulary must be taken/ administered first, by Reading Comprehension then, Fluency” as expressed in attribute C.1 (Arrangement of Items). The EPS generally agreed with the test package’s adherence to the principles of assessment, with a mean of 0.85 interpreted as Agree. She pointed out factors about the test that were representative of the principle of assessment. They are as follows: a) the assessment approach can be adapted to meet the needs of all students; b) the assessment strategy adequately covers both intellectual and behavioral components of reading; and c) the texts and sentences in the test remain reflective of the current time.

CONCLUSIONS

The following conclusions were drawn based on the findings derived from this study. Based on the first problem’s findings, it can be concluded that a teacher-made reading diagnostic test is a good tool for assessing fluency, vocabulary, comprehension, and the reading competence levels of students in general. Based on the findings in the second problem, it can be concluded that the test is valid, considering that the indicators of a good reading diagnostic

test, as recommended by foreign and local reading experts, were used as guiding principles in crafting the test. Since the problem concentrated on the statistical computations derived from students' answers to the test, it can also be concluded that the test is somehow a reliable test considering the Cronbach's Alpha result. Experts, whether foreign or local, have a universal view of the indicators of a good reading diagnostic test that they adhere to. The reading teachers and EPS are also saying the same. With that, it can be concluded that the teacher-made reading diagnostic test for Grade 7 complies with the criteria of a good reading diagnostic test.

With the first conclusion, the researcher recommends that the test package be utilized for Grade 7 English teachers to have a basis for an appropriate reading remediation program since this test can specifically determine what reading competency is problematic to students. Apart from that, the researcher strongly recommends that this test be subjected to further refinement, leading to the standardization of the items to achieve its finality for it to be disseminated to its beneficiaries and attain its acceptability in the academe and the discipline.

Considering that there are 12 items to be rejected out of the 100-item test, it is recommended that these items be replaced to complete the number of items and that the 28 items be dutifully revised. Considering the reliability coefficients of 0.600 and 0.700, the researcher recommends that continuous improvement be made in the test package to attain the acceptable reliability of a very good classroom test so that it may be used by and distributed to its intended beneficiaries. Since the whole test is already valid, the researcher recommends constantly monitoring the validity as the process of improving reliability is made. Regarding the low inter-rater reliability, the researcher sees it fit to be re-validated by more than three experts to improve the level of agreement. It should comprise at least five reading experts, three assessment experts, three master teachers, three grade 7 teachers, and one education program supervisor.

The foreign and local experts, including the reading teachers and EPS, view reading assessment as a unified concept; therefore, the guidelines and indicators set by these experts and people must be adapted, and none of those should be neglected. However, considering that we have different settings, texts, passages, and questions should be contextualized and/or to suit the Philippine learners.

TRANSLATIONAL RESEARCH

Additionally, this study may help teachers plan for their various remedial reading efforts. Since it is easier to conduct and there are specific topics

in each test, they could concentrate more on developing their remedial reading interventions. It will also lead students to learn reading without fearing rigorous assessment. This will also help reading teachers in improving various activities. Finally, the developed test was very useful to teachers and future researchers venturing into reading education.

LITERATURES CITED

- Clay, B. (2011). *Is This a Is Trick Question? A Short Guide to Writing Effective Test Questions*.
- Cohen, A. S., & Wollack, J. A. (2015). Handbook on test development: Helpful tips for creating reliable and valid classroom tests. *University of Wisconsin-Madison: Madison, WI, USA*.
- Crossman, A. (2017). Understanding Purposive Sampling An Overview of the Method and Its Applications. *Thoughtco*. <https://bit.ly/3T6MMUt>
- Department of Education. (2010). DM 266, s. 2010 Philippine Informal Reading Inventory (Phil- IRI) Reporting and Database System. <https://bit.ly/3t66LIf>
- DepEd. (2018a). DO 14 s.2018. Policy Guidelines on the Administration of the Revised Philippine Informal Reading Inventory. <https://bit.ly/3N8wDdq>
- DepED. (2018b). The Philippine Informal Reading Inventory Manual 2018. <https://bit.ly/3RoRdbP>
- Dunkley, D. (2012). Diagnostic Assessment in Theory and Practice, 55–65.
- Garnham, A. (2005). Language comprehension. *K. LAMBERTS & RL GOLDSTONE The Handbook of cognition, London: Sage*, 241-254.
- Hardcastle, M. A., Usher, K., & Holmes, C. (2006). Carspecken's five-stage critical qualitative research method: An application to nursing research. *Qualitative Health Research*, 16(1), 151-161.

- Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how?. *The Reading Teacher*, 58(8), 702-714.
- Kooij, J. J. S. (2013). Diagnostic Assessment.
- LaDuca, A., Downing, S. M., & Henzel, T. R. (1995). 5. Systematic Item Writing And Test Construction.
- Linan-Thompson, S., Hickman-Davis, P., & Light, M. (2001). Essential Reading Strategies for the Struggling Reader: Activities for an Accelerated Reading Program. Texas Education Agency, 1701 N. Congress Ave., Austin, TX 78701-1494.
- Luz, J. M. (2017). *Brigada Eskwela: Essays on Philippine Education*. Anvil Publishing, Inc..
- Mertens, W. (2017). *Quantitative data analysis*. Springer.
- Musa, A. K. J., & Balami, A. Z. (2016). Effects of phonological awareness training on the reading performance of children with dyslexia in primary schools In Maiduguri metropolis, Borno State, Nigeria. *American Journal of Educational Research*, 4(15), 1078-1085.
- Sainsbury, M., Harrison, C., & Watts, A. (Eds.). (2006). *Assessing Reading: From Theories to Classrooms: an International Multi-disciplinary Investigation of the Theory of Reading Assessment and Its Practical Implications at the Beginning of the 21st Century*. National Foundation for educational research.
- SEAMEO INNOTECH. (2012). K TO 12 TOOLKIT: Resource Guide for Teacher Educators, School Administrators and Teachers. Southeast Asian Ministers of Education Organization (SEAMEO) Regional Center for Educational Innovation and Technology (INNOTECH). <https://bit.ly/3T7mGAv>
- Shanahan, T. (2005). The National Reading Panel Report. Practical Advice for Teachers. *Learning Point Associates/North Central Regional Educational Laboratory (NCREL)*.

- Singh, H. (2018). Understanding Gradient Boosting Machines| by Harshdeep Singh| Towards Data Science.
- Smulowitz, S. (2017). Document analysis. *The international encyclopedia of communication research methods*, 1-8.
- Snellings, P., van der Leij, A., Blok, H., & de Jong, P. F. (2010). Reading fluency and speech perception speed of beginning readers with persistent reading problems: the perception of initial stop consonants and consonant clusters. *Annals of dyslexia*, 60, 151-174.
- Sofaer, S. (2002). Qualitative research methods. *International journal for quality in health care*, 14(4), 329-336.
- Stahl, K. A. D., Flanigan, K., & McKenna, M. C. (2019). *Assessment for reading instruction*. Guilford Publications.
- Statisticshowto. (n.d.). What is Classical Test Theory. <https://bit.ly/415pIr9>
- The Saga of a Town (n.d.). The Official Website of the Local Government of Tinambac. <https://bit.ly/3RtE8OD>
- William, J. (2018). What are Different Types of Questionnaires. *Flatworld Solutions*. Retrieved from <https://bit.ly/3Nf6vgN>
- Yu, M. E., & Lin, Y. C. (2002). Principles for Writing Multiple-Choice Items in Vocabulary, Grammar, and Reading Tests: A Study on the Test Making Process of an Achievement Test at Fooyin Institute of Technology. *宜蘭技術學報*, (8), 107-114.
- Yuko, S. (2005). Analysis of Objective Test Items : Towards a Revision of the Placement Test, 119–127. <https://bit.ly/3NfrffF>
- Zeng, J., & Wyse, A. (2009). Introduction to classical test theory. *Michigan, Washington, US*.
- Zutell, J., & Rasinski, T. V. (1991). Training teachers to attend to their students' oral reading fluency. *Theory into practice*, 30(3), 211-217.