**INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY: APPLIED BUSINESS AND EDUCATION RESEARCH**
2024, Vol. 5, No. 7, 2459 – 2473
http://dx.doi.org/10.11594/ijmaber.05.07.08

**Research Article**

# An Achievement Test in Mathematics in the Modern World Course: The Standardization Process

Jericho Y. Baybayan*, Michelle R. Lacia

Notre Dame University, Cotabato City, Philippines

*Article history*:
Submission 28 April 2024
Revised 07 July 2024
Accepted 23 July 2024

**Corresponding author:*
E-mail:
baybayan.jy@ndu.edu.ph

**ABSTRACT**

This study aimed to construct and standardize an achievement test in Mathematics in the Modern World (MMW) course. Achievement test evaluates a student's performance after a given period of instruction. Using the instrumentation research design, the data were gathered among 273 tertiary students of Notre Dame University enrolled during the second semester of Academic Year 2020-2021. Among these, 38 pilot tested and 213 tested the reliability of the instrument. This study further utilized the following: a 4-point Likert scale for the validation of the instrument, difficulty value and discrimination index formulas for the analysis of each item, and KR20 reliability for internal consistency. Furthermore, a test manual for the MMW achievement test was also established. The preliminary total number of Items before MMW teachers' validation was 42 items covering the topics for the first 12 hours of the MMW Syllabus. These items were validated through Aiken's Item Validity method which resulted in the acceptance of 42 items. After rigorous processes, it resulted in the acceptance of 34 items for the MMW achievement test. The instrument's average validity is 0.935, and the reliability is 0.701 which means that this tool has high validity and acceptable reliability.

***Keywords****: Construct and standardize, Achievement test, Instrumentation, Mathematics in the Modern World (MMW)*

## Introduction

Mathematics in the Modern World (MMW) is a three-unit course offered to all Tertiary students in the Philippines. Through the Commission on Higher Education Memorandum Order Number 20, series of 2013 (CMO No. 20, s. 2013), the Commission on Higher Education (CHED) with the experts in the Philippines created this new course considering the new and enhanced Basic Education curriculum or commonly known as the K to 12 curriculum.

This new general education course is now in its third year of implementation since it started in the Academic Year 2018 - 2019 when the first batch of K to 12 graduates became First Year College students. Within the two and a half years of implementation, there were still no published standardized achievements made to

*How to cite:*
Baybayan, J. Y. & Lacia, M. R. (2024). An Achievement Test in Mathematics in The Modern World Course: The Standardization Process. *International Journal of Multidisciplinary: Applied Business and Education Research*. *5*(7), 2459 – 2473. doi: 10.11594/ijmaber.05.07.08

test how students do well as they move on with the other topics or as they finish the desired topics and number of hours in this course. With this gap, the researchers decided to address this concern by aiming to construct, validate, and standardize a Multiple-Choice type of test that will help the Commission on Higher Education (CHED) and universities or colleges, especially Notre Dame University to measure students' achievement as they end the first 12 hours of stay in MMW and proceed to the next topics.

Moreover, the researchers developed an achievement test encompassing only Section 1: The Nature of Mathematics, which is good for the first 12 hours so that the validation, data gathering procedure, analysis, and standardization will be done on time and since the target respondents of the researchers are the students of Notre Dame University who are currently having their Mathematics in the Modern World classes for this Second Semester of Academic Year 2020-2021.

Achievement tests are designed to measure the learned skills, knowledge, and abilities of a student in the subject in which the student has received training or instruction (Frey, 2018). This type of test is done with the use of "backward-looking" which determines a student's academic progress by measuring the skills and knowledge in school over time (Great Schools Partnership, 2015).

According to Cerezo (2021), standardized testing involves rigorous development, trials, and revision processes to determine the measurement properties of a test. In addition, there are three reasons for standardized testing as narrated by Churchill (2015), namely objectivity, comparability, and accountability. Objectivity takes place with the use of similar questions given to students with identical testing conditions to provide accurate and unfiltered knowledge and a much clearer view of academic mastery. Comparability is another reason, especially for parents who select the best school for their child. The results obtained from a standardized test will help parents to inspect and compare a range of schools of where they think their child best fits in. School practitioners use standardized results to benchmark students' achievement across school and district lines. Finally, accountability is attached to the school for schools are accountable to the students' growth in terms of their academic performance.

Thus, the researchers envisioned that they would Construct and Standardize an Achievement Test for Mathematics in the Modern World.

## Objectives

The study aimed to Construct and Standardize an Achievement test in Mathematics in the Modern World (MMW). Specifically, it sought to find answers to the following questions: (i.) What is the level of difficulty and discrimination power of each item in the developed Mathematics in the Modern World (MMW) Achievement Test? (ii.) What is the effectiveness of the distracters in each item in the developed Mathematics in the Modern World (MMW) Achievement Test? (iii.) What is the reliability and validity of the developed Mathematics in the Modern World (MMW) Achievement Test?

## Methods

### *Research Design*

The instrumentation research design was used in this study since the primary aim is to Construct and Standardize an Achievement Test in Mathematics in the Modern World. As defined by Kpolovie (2016), instrumentation research design deals with the psychometric principles for the actual test development, the establishment of validity and reliability, providing standard instructions, writing suitable detailed test manuals, elimination of biases in testing, scaling methods, and norm or criterion, based form some test theories like Classical Test Theory and Item Response Theory.

### *Population and Sampling*

The data were gathered via census from 273 student respondents enrolled this second semester of the academic year 2020-2021 at Notre Dame University, particularly in Mathematics in the Modern World (MMW) subject. Through the non-probability purposive sampling, one section with 42 students was specifically chosen and requested to be the respondents for Pilot Testing, and the remaining seven

sections with a total of 231 students were requested for the Standardizing Phase. Due to unavoidable circumstances such as Slow to No internet connection and sickness, only 38 out of 42 students accepted the request of the researchers to answer the questions for the Pilot Testing Phase, while only 213 out of 231 answered for the Standardizing Phase.

### *Instrument*

The main tool used in this study was a developed test instrument for Mathematics in the Modern World (MMW) Achievement Test, which was Content Validated by Mathematics in the Modern World Teachers, Pilot tested and Reliability tested by researchers to student respondents, and Established through a Detailed Test Manual encompassing the Purpose of the Test Manual, The Objectives of the Test, CHED Mathematics in the Modern World Course Syllabus Major Topics, Sub Topics, Specific Topics, and Learning Outcomes, Test Blueprint, The Testing Committee, Preliminaries Before the Conduct of the Test, Instruction after the Conduct of the test, Interpreting Test Scores, Reporting Test Scores, Appendices, and References.

The constructed test items before the Pilot Testing phase were 42 multiple-choice items with five options encompassing Section 1: The Nature of Mathematics with topics I. Mathematics in our World, II. Mathematical Language and Symbols, and III. Problem Solving and Reasoning. All the constructed items were Content Validated resulting in the acceptance of 42 Multiple Choice Items with five options. After the analysis in the Pilot Testing Phase, 35 Multiple Choice Items with Two, Three, Four, or Five options were accepted for the Standardizing Phase. After checking the reliability and item deletion to improve the reliability, 34 Multiple Choice Items with Two, Three, Four, or Five options were accepted and became the Mathematics in the Modern World Achievement Test Items. These remaining 34 Multiple Choice Test Items were plotted through a Detailed Test Manual.

### *Data Collection and Analysis*

The researchers used an online platform Google Classroom, specifically Google form to gather all the data needed for the study in the Pilot Testing Phase and Standardizing Phase.

In the conduct of the Pilot Testing Phase, preliminaries were conducted such as inviting Student Respondents to enter the Google Classroom, sending the prayer, greetings, recommendation, and approval of the Dean for the Conduct of the Study, and Instructions for the Test. A total of 15 minutes was used in the Preliminary Part of the Data Collection. After the preliminaries or during the test proper, open-ended time within the day was given to students who had answered the 42 Multiple Choice Items with five options for the Pilot Testing Phase. After the student respondents had sent their answers, the data gathered were analyzed through Item Difficulty, Discrimination Power, and Distracter Analysis. Accepted items were arranged through a Test Blueprint, and the Standardizing Phase followed.

In the conduct of the Standardizing Phase, the same preliminaries as in the Pilot Testing were done. After the preliminaries or during the test proper, as recommended by Clay (2001) in his short guide to writing effective test questions, one minute per item or a total of 35 minutes was given to student respondents who were requested to take part in the Standardizing Phase. After the student respondents had sent their answers, the data gathered were analyzed through Kuder Richardson 20 (KR 20) for the internal consistency reliability of the instrument with the help of the Statistical Package for the Social Sciences (SPSS). Item deletion was also done after the calculation of internal consistency to ensure that the instrument had a good reliability result.

### *Phases and Steps in Constructing and Standardizing an Achievement Test in Mathematics in the Modern World*

This study Construction and Standardization of an Achievement Test in Mathematics in the Modern World has three phases, namely the Planning and Constructing Phase, Pilot Testing Phase, and the Standardizing Phase.

In the planning and constructing phase, the objectives and content of the developed MMW Achievement Test were determined through the aid of the Commission on Higher Education Mathematics in the Modern World Course

Syllabus. The desired size and type of test is 45 Multiple Choice Items. However, due to calculation, it resulted in 42 Multiple Choice Items in the created Table of Specifications before the Pilot Testing Phase. Consultation of the researchers with the adviser was done and resulted in the approval of 42 Multiple Choice Test Items in the Table of Specifications. After the approval of the number of test items, the researchers constructed each of the questions with five options. After the approval of the Test questions, the creation of Test Instructions was done and was followed by the creation of letters addressed to the MMW Teacher Validators. They were then asked to rate the items through the use of a 4-point Likert Scale with 1 – The item is Not Relevant to the Learning Outcome, 2 – The item is Somehow Relevant to the Learning Outcome, 3- The item is Relevant to the Learning Outcome, and 4 – The item is Very Relevant to the Learning Outcome. After that the MMW Teacher Validators rated each item, and the calculation using the content validity method, specifically Aiken's Validity Index was done which resulted in the acceptance and rejection of items. Aiken's Validity Index as narrated by Irawan and Wiluyeng (2020) was calculated through the formula, wherein V = Validity of Aiken's Index, S = The score awarded by the rater minus the lowest possible rating a rater can award, C = The highest possible rating a rater can award, and N = The number of raters rating the test questionnaire. The Validity of Aiken's Index or V obtained in each item is then classified by Irawan and Wilujeng (2020) into different categories with respective decisions, namely $0 \leq V \leq 0.4$ as Invalid, $0.4 < V \leq 0.8$ Medium Validity, and $0.8 < V \leq 1$ Very Valid. Accepted items described as Medium and Very Valid Items are arranged and then plotted to a Final Table of Specifications ready for the Pilot Testing Phase. Final Instructions for the Pilot Testing were also made.

In the pilot testing phase, the accepted test items were then administered to 38 respondents. Difficulty Value and Discrimination Power were computed for each item after all respondents had submitted their answers. Difficulty Value was calculated through the formula of Sharma and Sarita (2018) which states that $DV = R_U + R_L$, where DV is the Difficulty Value, RU is the number of students in the upper group who responded correctly, RL is the number of students in the lower group who responded correctly, NU is the total number of students in the upper group, and NL is the total number of students in the lower group and was interpreted through the Table for Difficulty Value adopted from Lacia (2007) in which 0.81 - 1.00 means Very easy item, 0.61 – 0.80 means Easy item, 0.41 – 0.60 means Moderately Difficult item, 0.21 – 0.40 means Difficult item, and 0.00 – 0.20 means Very Difficult item. Discrimination Power was then calculated through the formula of Kumar (2016) which is $DP = (R_U - R_L) / 0.5\ N$, where DP is the Discrimination Power, RU is the number of right responses of the upper group, RL is the right responses of the lower group, and N is the total number of students in both groups. The results were then interpreted through the Table for Discrimination Power adopted from Ebel and Frisbie (1986) in which 0.40 and above is interpreted as an Excellent item and should be Retained, 0.30 – 0.39 means Good and there are Possibilities for Improvement 0.20 – 0.29 means Mediocre and Needs to be checked or reviewed, 0.00 - 0.19 means Poor and should be discarded or reviewed in depth, and Negative Values are Worst items which should be Discarded. Moreover, the decision from Lacia (2007) to Revise an item due to its importance and that it is the only item left that can measure a specific area of content was also used. A table of acceptance and rejection of items based on the Difficulty Value and Discrimination Power was reported. Distracter Analysis was also done and resulted in the rejection of confusing and ineffective distracters. Items that have a decision of Accept, May Need Revision, or Revise were then arranged and plotted to a Final Table of Specifications ready for the Standardizing Phase. Creation of the Final Test Instructions, which includes the Time Limit and Scoring System for the Standardizing Phase was also done. In general, the pilot testing phase resulted in the acceptance and rejection of items based on the respondents' answers.

In the standardizing phase, the accepted items were then administered to 213 respondents for the Reliability Testing Sub-Phase. Respondents were then informed about the Set of

Instructions including the Time Limit and Scoring System before the test questionnaire was given to them. After they finished answering the Test within the given allotted time, checking and scoring of answers was done by researchers. The calculation of the instrument's internal consistency through Kuder Richardson 20 (KR20) reliability was also done. The formula KR=N/(N-1) / (V- SUM (pi*qi)/V) where KR is Kuder Richardson 20, N is the Number of items in the test, V is the Variance of the raw scores or standard deviation squared, pi is the proportion of correct answers of question i, or the number of correct answers divided by the total number of responses, and qi is the proportion of incorrect answers of question i, or (i –p) by Patock (2004) was used. Item deletion was also done after the calculation of internal consistency to ensure that the instrument had good reliability results. This resulted in the elimination of one item from the administered 35 items in the Standardizing Phase. After doing the item deletion and ensuring that the test is reliable, the Test Manual which includes the Purpose of the Test Manual, The Objectives of the Test, CHED Mathematics in the Modern World Course Syllabus Major Topics, Sub Topics, Specific Topics, and Learning Outcomes, Test Blueprint, The Testing Committee, Preliminaries Before the Conduct of the Test, Instruction after the Conduct of the test, Interpreting Test Scores, Reporting Test Scores, Appendices, and References was established.

## Results and Discussion

*Table 1. Difficulty Value of Each Item*

| Item No. | $R_U$ | $R_L$ | $R_U + R_L$ | DV | Interpretation |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 0.10 | Very Difficult |
| 2 | 6 | 4 | 10 | 0.50 | Moderately Difficult |
| 3 | 6 | 2 | 8 | 0.40 | Difficult |
| 4 | 2 | 1 | 3 | 0.15 | Very Difficult |
| 5 | 3 | 1 | 4 | 0.20 | Very Difficult |
| 6 | 8 | 0 | 8 | 0.40 | Difficult |
| 7 | 8 | 0 | 8 | 0.40 | Difficult |
| 8 | 7 | 0 | 7 | 0.35 | Difficult |
| 9 | 3 | 5 | 8 | 0.40 | Difficult |
| 10 | 6 | 0 | 6 | 0.30 | Difficult |
| 11 | 9 | 0 | 9 | 0.45 | Moderately Difficult |
| 12 | 10 | 3 | 13 | 0.65 | Easy |
| 13 | 10 | 3 | 13 | 0.65 | Easy |
| 14 | 8 | 5 | 13 | 0.65 | Easy |
| 15 | 7 | 1 | 8 | 0.40 | Difficult |
| 16 | 8 | 1 | 9 | 0.45 | Moderately Difficult |
| 17 | 8 | 1 | 9 | 0.45 | Moderately Difficult |
| 18 | 9 | 1 | 10 | 0.50 | Moderately Difficult |
| 19 | 4 | 6 | 10 | 0.50 | Moderately Difficult |
| 20 | 10 | 6 | 16 | 0.80 | Easy |
| 21 | 0 | 3 | 3 | 0.15 | Very Difficult |
| 22 | 7 | 1 | 8 | 0.40 | Difficult |
| 23 | 5 | 4 | 9 | 0.45 | Moderately Difficult |
| 24 | 7 | 0 | 7 | 0.35 | Difficult |
| 25 | 7 | 2 | 9 | 0.45 | Moderately Difficult |
| 26 | 6 | 3 | 9 | 0.45 | Moderately Difficult |
| 27 | 9 | 5 | 14 | 0.70 | Easy |
| 28 | 4 | 0 | 4 | 0.20 | Very Difficult |
| 29 | 2 | 0 | 2 | 0.10 | Very Difficult |
| 30 | 9 | 5 | 14 | 0.70 | Easy |

| Item No. | $R_U$ | $R_L$ | $R_U + R_L$ | DV | Interpretation |
|---|---|---|---|---|---|
| 31 | 0 | 4 | 4 | 0.20 | Very Difficult |
| 32 | 7 | 4 | 11 | 0.55 | Moderately Difficult |
| 33 | 0 | 1 | 1 | 0.05 | Very Difficult |
| 34 | 9 | 4 | 13 | 0.65 | Easy |
| 35 | 1 | 1 | 2 | 0.10 | Very Difficult |
| 36 | 1 | 2 | 3 | 0.15 | Very Difficult |
| 37 | 8 | 6 | 14 | 0.70 | Easy |
| 38 | 9 | 2 | 11 | 0.55 | Moderately Difficult |
| 39 | 0 | 1 | 1 | 0.05 | Very Difficult |
| 40 | 3 | 2 | 5 | 0.25 | Difficult |
| 41 | 2 | 3 | 5 | 0.25 | Difficult |
| 42 | 7 | 3 | 10 | 0.50 | Moderately Difficult |

***Note.*** $N_U + N_L = 20$

Table 1 reveals the Difficulty Value and Interpretation of each Item. It further revealed that there were 0 Very Easy Items, 8 Easy Items, 12 Moderatety Difficult Items, 11 Difficult Items, and 11 Very Difficult Items.

The test items were divided according to the Mathematics in the Modern World Learning Plan. There were three sub-topics in Section 1: The Nature of Mathematics namely, Mathematics in our World (four hours), Mathematical Language and Symbols (three hours), and Problem-Solving and Reasoning (five hours) a total of 12 teaching hours. Each of the subtopics has six specific topics thus leading to two questions each specific topic for Mathematics in our World and Mathematical Language and Symbols and three questions each for Problem Solving and Reasoning.

In this constructed items, in terms of subtopic Mathematics in our World, items 1 (create) and 42 (remember) are set for specific topic Patterns and Numbers in Nature and the World: the snowflake and honeycomb, tiger's stripes and hyena's spots; the sunflower; the snail's shell, flower petals; the world's pipulation and the weather which is said to be very difficult and difficult respectively, 14 (apply) and 22 (analyze) for topic The Fibonacci Sequence which is said to be Easy and Difficult respectively, 9 (understand) and 18 (evaluate) for topic Mathematics helps organize patterns and regularities in the world which is difficult and moderately difficult respectivelt, 34 (understand) and 5 (evaluate) for topic Mathemetics helps predict the behavior of nature and phenomena in the world which is easy and difficult respectively, 40 (remember) and 33 (create) for topic Mathematics helps control nature and occurences in the world for our own ends which is difficult and very difficult respectively, and 10 (apply) and 39 (analyze) for topic Mathematics has numerous applications in the world making it indispensable which is difficult and very difficult respectively.

Moreover, in terms of the subtopic Mathematical Language and Symbols, items 15 (understand) and 23 evaluate) for the topic Characteristics of mathematical language: precise, concise, powerful which is difficult and moderately difficult respectively, 2 (remember) and 32 (create) for topic expressions vs, sentences which are both moderately difficult, 19 (remember) and 41 (create) for topic Conventions in the mathematical language which is moderately difficult and difficult respectively, 6 (understand) and 27 (evaluate) for topic Four basic concepts: sets, functions, relations, binary operations which is difficult and easy respectively, 24 (apply) and 28 (analyze) for topic Elementary logic: connectives, quantifiers, negation, variables which is difficult and very difficult respectively, and 11 (apply) and 31 (analyze) for topic Formality which is moderately difficult and very difficult respectively.

Finally, in terms of sub-topic Problem Solving and Reasoning, items 38 (remember), 3 (analyze), and 20 (evaluate) for the topic Inductive and Deductive reasoning which is moderately difficult, difficult, and easy respectively, 30 (remember), 12 (analyze), and 37 (create) for topic Intuition, proof, and certainty which are all easy questions, 25 (remember), 16

(apply), and 26 (evaluate) for topic Polya's 4-steps in Problem-solving which are all moderately difficult questions, 21 (understand), 7 (analyze), and 17 (evaluate) for topic problem solving strategies which are very difficult, difficult, and moderately difficult items respectively, 8 (understand), 36 (apply), and 29 (create) for topic Mathematical Problems involving patterns which is difficult, very difficult and very difficult respectively, and 13 (understand), 4 (apply), and 35 (create) for topic Recreational Problems using mathematics which is easy, very difficult, and very difficult respectively.

In general, on average, the respondents found Intuition, proof, and certainty easy while Mathematical Problems involving patterns very difficult.

*Table 2. Discrimination Power of each item*

| Item No. | $R_U$ | $R_L$ | $R_U - R_L$ | DP | Interpretation | Recommendation |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0.10 | Poor | Review in depth |
| 2 | 6 | 4 | 2 | 0.20 | Mediocre | Review |
| 3 | 6 | 2 | 4 | 0.40 | Excellent | Retain |
| 4 | 2 | 1 | 1 | 0.10 | Poor | Review in depth |
| 5 | 3 | 1 | 2 | 0.20 | Mediocre | Review |
| 6 | 8 | 0 | 8 | 0.80 | Excellent | Retain |
| 7 | 8 | 0 | 8 | 0.80 | Excellent | Retain |
| 8 | 7 | 0 | 7 | 0.70 | Excellent | Retain |
| 9 | 3 | 5 | - 2 | - 0.20 | Worst | Definitely Discard |
| 10 | 6 | 0 | 6 | 0.60 | Excellent | Retain |
| 11 | 9 | 0 | 9 | 0.90 | Excellent | Retain |
| 12 | 10 | 3 | 7 | 0.70 | Excellent | Retain |
| 13 | 10 | 3 | 7 | 0.70 | Excellent | Retain |
| 14 | 8 | 5 | 3 | 0.30 | Good | Possibilities for Improvement |
| 15 | 7 | 1 | 6 | 0.60 | Excellent | Retain |
| 16 | 8 | 1 | 7 | 0.70 | Excellent | Retain |
| 17 | 8 | 1 | 7 | 0.70 | Excellent | Retain |
| 18 | 9 | 1 | 8 | 0.80 | Excellent | Retain |
| 19 | 4 | 6 | - 2 | - 0.20 | Worst | Definitely Discard |
| 20 | 10 | 6 | 4 | 0.40 | Excellent | Retain |
| 21 | 0 | 3 | - 3 | - 0.30 | Worst | Definitely Discard |
| 22 | 7 | 1 | 6 | 0.60 | Excellent | Retain |
| 23 | 5 | 4 | 1 | 0.10 | Poor | Review in depth |
| 24 | 7 | 0 | 7 | 0.70 | Excellent | Retain |
| 25 | 7 | 2 | 5 | 0.50 | Excellent | Retain |
| 26 | 6 | 3 | 3 | 0.30 | Good | Possibilities for Improvement |
| 27 | 9 | 5 | 4 | 0.40 | Excellent | Retain |
| 28 | 4 | 0 | 4 | 0.40 | Excellent | Retain |
| 29 | 2 | 0 | 2 | 0.20 | Mediocre | Review |
| 30 | 9 | 5 | 4 | 0.40 | Excellent | Retain |
| 31 | 0 | 4 | - 4 | - 0.40 | Worst | Definitely Discard |
| 32 | 7 | 4 | 3 | 0.30 | Good | Possibilities for Improvement |
| 33 | 0 | 1 | - 1 | - 0.10 | Worst | Definitely Discard |
| 34 | 9 | 4 | 5 | 0.50 | Excellent | Retain |
| 35 | 1 | 1 | 0 | 0.00 | Poor | Review in depth |
| 36 | 1 | 2 | - 1 | - 0.10 | Worst | Definitely Discard |
| 37 | 8 | 6 | 2 | 0.20 | Mediocre | Review |
| 38 | 9 | 2 | 7 | 0.70 | Excellent | Retain |
| 39 | 0 | 1 | - 1 | - 0.10 | Worst | Definitely Discard |

| Item No. | $R_U$ | $R_L$ | $R_U - R_L$ | DP | Interpretation | Recommendation |
|---|---|---|---|---|---|---|
| 40 | 3 | 2 | 1 | 0.10 | Poor | Review in depth |
| 41 | 2 | 3 | - 1 | - 0.10 | Worst | Definitely Discard |
| 42 | 7 | 3 | 4 | 0.40 | Excellent | Retain |

***Note.*** **N (0.50) = 10 since N = 20**

Table 2 reveals the Discrimination Power, Interpretation, and Recommendation in each item. It further revealed that there were 22 Excellent Items to be retained, 3 Good Items which has possibilities for improvement, 4 Mediocre Items to be reviewed or needed to be checked, 5 Poor Items to be discarded or reviewed in depth, and 8 Worst Items which should definitely be discarded.

*Table 3. Deleting Ineffective and Confusing Distractors*

| Item No. | Group | Options | | | | | Effective | Confusing | Ineffective |
|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C* | D | E | | | |
| 1 | Upper | 1 | 1 | 1 | 7 | 0 | D, E | A, B | |
| | Lower | 0 | 0 | 1 | 8 | 1 | | | |
| | | A | B | C* | D | E | | | |
| 2 | Upper | 0 | 3 | 6 | 1 | 0 | A, B | D | E |
| | Lower | 1 | 5 | 4 | 0 | 0 | | | |
| | | A | B | C | D* | E | | | |
| 3 | Upper | 3 | 0 | 1 | 6 | 0 | A, B, E | C | |
| | Lower | 5 | 2 | 0 | 2 | 1 | | | |
| | | A | B | C* | D | E | | | |
| 4 | Upper | 6 | 1 | 2 | 0 | 1 | B, D | A, E | |
| | Lower | 5 | 3 | 1 | 1 | 0 | | | |
| | | A | B | C | D | E* | | | |
| 5 | Upper | 0 | 1 | 6 | 0 | 3 | A, D | B, C | |
| | Lower | 2 | 0 | 2 | 5 | 1 | | | |
| | | A* | B | C | D | E | | | |
| 6 | Upper | 8 | 0 | 1 | 1 | 0 | B, C, D, E | | |
| | Lower | 0 | 1 | 6 | 2 | 1 | | | |
| | | A | B | C | D | E* | | | |
| 7 | Upper | 1 | 1 | 0 | 0 | 8 | A, B, D | | C |
| | Lower | 7 | 2 | 0 | 1 | 0 | | | |
| | | A | B | C | D* | E | | | |
| 8 | Upper | 0 | 1 | 1 | 7 | 1 | A, B, E | C | |
| | Lower | 3 | 3 | 0 | 0 | 4 | | | |
| | | A* | B | C | D | E | | | |
| 9 | Upper | 3 | 0 | 0 | 0 | 7 | C | E | B, D |
| | Lower | 5 | 0 | 1 | 0 | 4 | | | |
| | | A | B | C | D | E* | | | |
| 10 | Upper | 1 | 1 | 0 | 2 | 6 | A, B, C | D | |
| | Lower | 6 | 2 | 1 | 1 | 0 | | | |
| | | A | B | C | D | E* | | | |
| 11 | Upper | 1 | 0 | 0 | 0 | 9 | A, B, C | | D |
| | Lower | 7 | 2 | 1 | 0 | 0 | | | |
| | | A* | B | C | D | E | | | |
| 12 | Upper | 10 | 0 | 0 | 0 | 0 | C, D, E | | B |
| | Lower | 3 | 0 | 4 | 1 | 2 | | | |
| 13 | | A* | B | C | D | E | C, D, E | | B |

| Item No. | Group | Options | | | | | Effective | Confusing | Ineffective |
|---|---|---|---|---|---|---|---|---|---|
| | Upper | 10 | 0 | 0 | 0 | 0 | | | |
| | Lower | 3 | 0 | 1 | 4 | 2 | | | |
| | | A | B | C* | D | E | | | |
| 14 | Upper | 0 | 1 | 8 | 0 | 1 | A, D, E | B | |
| | Lower | 2 | 0 | 5 | 1 | 2 | | | |
| | | A* | B | C | D | E | | | |
| 15 | Upper | 7 | 1 | 0 | 2 | 0 | B, C, E | | D |
| | Lower | 1 | 3 | 2 | 2 | 2 | | | |
| | | A* | B | C | D | E | | | |
| 16 | Upper | 8 | 0 | 1 | 1 | 0 | B, C, D, E | | |
| | Lower | 1 | 3 | 2 | 3 | 1 | | | |
| | | A | B | C | D* | E | | | |
| 17 | Upper | 0 | 0 | 0 | 8 | 2 | A, B, C | | E |
| | Lower | 3 | 3 | 1 | 1 | 2 | | | |
| | | A | B | C* | D | E | | | |
| 18 | Upper | 1 | 0 | 9 | 0 | 0 | A, B, D, E | | |
| | Lower | 3 | 3 | 1 | 1 | 2 | | | |
| | | A* | B | C | D | E | | | |
| 19 | Upper | 4 | 2 | 0 | 0 | 4 | B | E | C, D |
| | Lower | 6 | 3 | 0 | 0 | 1 | | | |
| | | A* | B | C | D | E | | | |
| 20 | Upper | 10 | 0 | 0 | 0 | 0 | B, E | | C, D |
| | Lower | 6 | 3 | 0 | 0 | 1 | | | |
| | | A* | B | C | D | E | | | |
| 21 | Upper | 0 | 0 | 7 | 1 | 2 | E | C | B, D |
| | Lower | 3 | 0 | 2 | 1 | 4 | | | |
| | | A | B | C | D* | E | | | |
| 22 | Upper | 0 | 3 | 0 | 7 | 0 | A, C, E | B | |
| | Lower | 1 | 1 | 5 | 1 | 2 | | | |
| | | A | B* | C | D | E | | | |
| 23 | Upper | 5 | 5 | 0 | 0 | 0 | A | | C, D, E |
| | Lower | 6 | 4 | 0 | 0 | 0 | | | |
| | | A | B | C* | D | E | | | |
| 24 | Upper | 1 | 0 | 7 | 1 | 1 | A, B, E | D | |
| | Lower | 6 | 2 | 0 | 0 | 2 | | | |
| | | A | B | C* | D | E | | | |
| 25 | Upper | 0 | 0 | 7 | 2 | 1 | A, B, E | D | |
| | Lower | 1 | 2 | 2 | 1 | 4 | | | |
| | | A* | B | C | D | E | | | |
| 26 | Upper | 6 | 0 | 1 | 2 | 1 | C, E | | B, D |
| | Lower | 3 | 0 | 3 | 2 | 2 | | | |
| | | A | B | C* | D | E | | | |
| 27 | Upper | 0 | 0 | 9 | 1 | 0 | A, B, E | D | |
| | Lower | 2 | 1 | 5 | 0 | 2 | | | |
| | | A | B | C | D | E* | | | |
| 28 | Upper | 1 | 1 | 3 | 1 | 4 | A, B | C, D | |
| | Lower | 4 | 4 | 2 | 0 | 0 | | | |
| | | A | B | C* | D | E | | | |
| 29 | Upper | 8 | 0 | 2 | 0 | 0 | B | A | D, E |
| | Lower | 6 | 4 | 0 | 0 | 0 | | | |

| Item No. | Group | Options | | | | | Effective | Confusing | Ineffective |
|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C* | D | E | | | |
| 30 | Upper | 1 | 0 | 9 | 0 | 0 | A, D, E | | B |
| | Lower | 3 | 0 | 5 | 1 | 1 | | | |
| | | A | B | C | D | E* | | | |
| 31 | Upper | 3 | 7 | 0 | 0 | 0 | C, D | A, B | |
| | Lower | 2 | 2 | 1 | 1 | 4 | | | |
| | | A | B | C | D* | E | | | |
| 32 | Upper | 2 | 1 | 0 | 7 | 0 | A, C, E | B | |
| | Lower | 3 | 0 | 2 | 4 | 1 | | | |
| | | A | B | C | D* | E | | | |
| 33 | Upper | 6 | 1 | 2 | 0 | 1 | C, D | A | B |
| | Lower | 4 | 1 | 4 | 1 | 0 | | | |
| | | A* | B | C | D | E | | | |
| 34 | Upper | 9 | 0 | 1 | 0 | 0 | B, C, E | | D |
| | Lower | 4 | 1 | 4 | 0 | 1 | | | |
| | | A | B | C | D | E* | | | |
| 35 | Upper | 6 | 0 | 1 | 2 | 1 | B, C | A, D | |
| | Lower | 3 | 3 | 2 | 1 | 1 | | | |
| | | A | B | C | D* | E | | | |
| 36 | Upper | 7 | 1 | 0 | 1 | 1 | C | A, E | B |
| | Lower | 6 | 1 | 1 | 2 | 0 | | | |
| | | A* | B | C | D | E | | | |
| 37 | Upper | 8 | 0 | 1 | 0 | 1 | B, D | | C, E |
| | Lower | 6 | 1 | 1 | 1 | 1 | | | |
| | | A | B | C | D | E* | | | |
| 38 | Upper | 0 | 0 | 1 | 0 | 9 | A, B | C | D |
| | Lower | 5 | 3 | 0 | 0 | 2 | | | |
| | | A* | B | C | D | E | | | |
| 39 | Upper | 0 | 1 | 1 | 1 | 7 | B, C, D | E | |
| | Lower | 1 | 5 | 2 | 2 | 0 | | | |
| | | A | B | C* | D | E | | | |
| 40 | Upper | 0 | 1 | 3 | 6 | 0 | A, B | D | E |
| | Lower | 4 | 3 | 2 | 1 | 0 | | | |
| | | A | B | C* | D | E | | | |
| 41 | Upper | 6 | 0 | 2 | 2 | 0 | B, E | A | D |
| | Lower | 1 | 2 | 3 | 2 | 2 | | | |
| | | A | B* | C | D | E | | | |
| 42 | Upper | 0 | 7 | 3 | 0 | 0 | C, E | | A, D |
| | Lower | 0 | 3 | 6 | 0 | 1 | | | |

Table 3 reveals the Distracter Analysis in which Ineffective and Confusing Items were automatically deleted. This resulted in each item having either two, three, four, or five options.

*Table 4. Reliability Test Result*

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|---|---|---|
| .703 | .701 | 34 |

Table 4 reveals the Reliability Test Result of the MMW Achievement Test. Moreover, this revealed that the developed MMW Achievement Test is reliable because the internal consistency value 0.701 is said to be acceptable according to Cortina (1993), Taber (2016), and Metcalf in ResearchGate (2017).

*Table 5. 34 MMW Test questions Item Validity Result*

| Item Number | Raters | | | S1 | S2 | S3 | ∑S | V | Decision |
|---|---|---|---|---|---|---|---|---|---|
| | Aiken's Index Item Validator 1 | Aiken's Index Item Validator 2 | Aiken's Index Item Validator 3 | | | | | | |
| 1 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 2 | 2 | 4 | 4 | 1 | 3 | 3 | 7 | 0.8 | Medium Validity |
| 3 | 4 | 4 | 2 | 3 | 3 | 1 | 7 | 0.8 | Medium Validity |
| 4 | 4 | 4 | 3 | 3 | 3 | 2 | 8 | 0.9 | Very Valid |
| 5 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 6 | 3 | 4 | 3 | 2 | 3 | 2 | 7 | 0.8 | Medium Validity |
| 7 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 8 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 9 | 2 | 4 | 4 | 1 | 3 | 3 | 7 | 0.8 | Medium Validity |
| 10 | 2 | 4 | 4 | 1 | 3 | 3 | 7 | 0.8 | Medium Validity |
| 11 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 12 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 13 | 3 | 2 | 3 | 2 | 1 | 2 | 5 | 0.6 | Medium Validity |
| 14 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 15 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 16 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 17 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 18 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 19 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 20 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 21 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 22 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 23 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 24 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 25 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 26 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 27 | 3 | 4 | 3 | 2 | 3 | 2 | 7 | 0.8 | Medium Validity |
| 28 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 29 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 30 | 4 | 4 | 3 | 3 | 3 | 2 | 8 | 0.9 | Very Valid |
| 31 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| 32 | 4 | 2 | 4 | 3 | 1 | 3 | 7 | 0.8 | Medium Validity |
| 33 | 4 | 4 | 2 | 3 | 3 | 1 | 7 | 0.8 | Medium Validity |
| 34 | 4 | 4 | 4 | 3 | 3 | 3 | 9 | 1.0 | Very Valid |
| **Mean Validity** | | | | | | | | **0.935** | **Very Valid** |

***Note.*** **$L_0$= 1 and C = 4**

Table 5 reveals the Validity of each item in the MMW Test and its average validity of 0.935 which means that the items in the test are highly valid.

In these items, in terms of sub-topic Mathematics in our World, item 13 (remember) is set for the specific topic Patterns and Numbers in Nature and the World: the snowflake and honeycomb, tiger's stripes and hyena's spots; the sunflower; the snail's shell, flower petals; the world's population and the weather which is said to have medium validity, 7 (apply) and 24 (analyze) for topic The Fibonacci Sequence are both very valid, 11 (evaluate) for topic Mathematics helps organize patterns and regularities in the world which is very valid, 6 (understand) and 29 (evaluate) for topic Mathematics helps predict the behavior of nature and phenomena in the world which is medium and very valid respectively, 27 (remember) for topic Mathematics helps control nature and occurrences in the world for our ends which is medium valid, and 28 (apply) for topic Mathematics has numerous applications in the world making it indispensable which is very valid.

Moreover, in terms of sub-topic Mathematical Language and Symbols, items 22 (understand) and 15 (evaluate) for topic Characteristics of mathematical language: precise, concise, powerful which are very valid, 12 (remember) and 10 (create) for topic expressions vs, sentences which is very and medium valid respectively, 34 (remember) for topic Conventions in the mathematical language which is very valid, 21 (understand) and 4 (evaluate) for topic Four basic concepts: sets, functions, relations, binary operations which are very valid, 26 (apply) and 30 (analyze) for topic Elementary logic: connectives, quantifiers, negation, and variables which are very valid, and 17 (apply) for topic Formality which is very valid.

Finally, in terms of sub-topic Problem Solving and Reasoning, items 9 (remember), 20 (analyze), and 1 (evaluate) for the topic Inductive and Deductive reasoning which is medium, very, and very valid respectively, 3 (remember), 5 (analyze), and 2 (create) for topic Intuition, proof, and certainty which are medium, very, and medium valid respectively, 18 (remember), 19 (apply), and 14 (evaluate) for topic Polya's 4-steps in Problem-solving which are all very valid, 23 (analyze), and 16 (evaluate) for topic problem-solving strategies which are very valid items, 25 (understand) and 32 (create) for topic Mathematical Problems involving patterns which are medium and very valid items, and 8 (understand), 31 (apply), and 33 (create) for topic Recreational Problems using mathematics which is very, very, and medium valid respectively.

## Conclusions and Recommendations

This study was carried out to construct, validate, standardize, and produce a valid and reliable Achievement Test in Mathematics in the Modern World covering all the Specific Topics bound for the first 12 hours. The test was standardized by 213 student respondents of Notre Dame University taking up Mathematics in the Modern World during the second semester of the academic year 2020-2021. The validity of each item in the test calculated through Aiken's Validity Index resulted in the instrument's average validity of 0.935, and the Kuder Richardson 20 (KR 20) Reliability for Internal Consistency was 0.701. With these results, this test can be used by the teachers teaching Mathematics in the Modern World to assess their students' achievement as they finish the first 12 hours of their class.

The researchers had envisioned that future researchers be able to construct and standardize items good for the remaining 10 hours, specifically for Section 2: Mathematics as a Tool (Part 1) as teachers are given discretion to choose three topics to be discussed in class among the eight topics for the remaining 32 hours in Section 3: Mathematics as a Tool (Part 2), namely Geometric Designs, Codes, Linear Programming, The Mathematics of Finance, Apportionment and Voting, Logic, The Mathematics of Graphs, and Mathematical Systems.

## Acknowledgments

The researchers would like to express their deepest gratitude to all of those who shared their precious time and effort in completing the research paper, to:

Dr. Michelle R. Lacia, for the guidance and full support she provided to make the study more meaningful and successful,

Dr. Danilo D. Caburnay, Dr. Edwin L. Apawan, and Dr. Edgard R. Cabales for being the panel of the Oral Defense helped in the technicalities of this paper,

Prof. Donato B. Pidlaon and Dr. Michelle R. Lacia, for teaching the researchers Statistics and Advanced Statistics that helped them to understand the data in this research,

Dr. Gina T. Astillero, Prof. Ma. Annie A. Susaya, Prof. Gregoria D. Fernandez, Prof. Abdul Wahid I. Tocalo, Dr. Diana Jean D. Cadeliña, Prof. Engr. Arnel N. Ben, Mr. Juvelon A. Manonog, Mr. Gavin John Mangulamas, Ms. Meilin Abang, Ms. Ninarita Fe Buot, and Mr. Muhammad Abdullah for sharing their precious time and knowledge which had helped this paper to be fulfilled,

STRAND-Asia (Strategic Research and Development Center, Inc.), PARSSU (Philippine Association of Researchers & Statistical Software Users), AFFIRM Center for Research and Professional Learning, ASTR (Asian Society of Teachers for Research, Inc.), and AAP (Analytics Association of the Philippines) for all the trainings that one of the researchers had attended,

Mr. and Mrs. James, Jocelyn Baybayan, Mrs. Filipinas Baybayan, Mrs. Nerry A. Yu, relatives, students, classmates, and friends for the unending support, inspiration, and care.

and above all, the Almighty God for blessing the researchers with His protection, wisdom, and guidance which increased their confidence to face all the odds and manifested perseverance to make all things possible in completing this study. To God be the glory.

## References

Beck, K. (2020). How to Calcuate a T-Score. Leaf Group Ltd. / Leaf Group Media. https://sciencing.com/calculate-tscore-5135749.html

Bruce, L. (2021). Module 7: Study Skills - Common Types of Tests in College. Lumenwaymaker. https://courses.lumenlearning.com/waymaker-collegesuccess/chapter/text-common-types-of-tests-in-college/#:~:text=Common%20Test%20Types,tests%2C%20and%20physical%20skills%20tests

Cengage Learning (2021). Understanding Standardized Assessment. https://www.shsu.edu/aao004/documents/15_005.pdf

Cerezo (2021). Técnicas de evaluación: Standardized testing. Universidad de Murcia. https://webs.um.es/lourdesc/miwiki/lib/exe/fetch.php?media=standardized_testing.pdf

Churchill, A. (2015). Bless the tests: Three reasons for standardizes testing. The Thomas B. Fordham Institute. https://fordhaminstitute.org/national/commentary/bless-tests-three-reasons-standardizedtesting

Clay, B. (2001). Is This a Trick Question? A Short Guide to Writing Effective Test Questions. Kansas Curriculum Center of Kansas State University. https://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf

Commission on Higher Education [CHED], (2017). KWF- Mathematics in the Modern World. https://ched.gov.ph/wp-content/uploads/2017/10/KWF-Mathematics-in-the-Modern-World.pdf

Cortina, J. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. Journal of Applied Psychology 1993. Vol. 78., No. 1. pp. 98-104. https://www.psycholosphere.com/what%20is%20coefficient%20alpha%20by%20Cortina.pdf

Ebel, R. L., & Frisbie, D. A. (1986). Essentials of education measurement. Englewood Cliffs, NJ: Prentice Hall.

Educational Research Techniques (2017). Developing Standardized Tests.EducationalResearchTechniques.com. https://educationalresearchtechniques.com/2017/06/21/developing-standardized-tests/

Evroro, E. (2015). Item Analysis of Test of Number Operations. Asian Journal of Educational Research Vol. 3, No. 1, 2015 ISSN 2311-6080. http://www.multidisciplinaryjournals.com/wpcontent/uploads/2015/01/ITEM-ANALYSIS-OF-TEST-OF-NUMBER-OPERATIONS.pdf

Faremi, Y. A. (2016). Reliability Coefficient of Multiple Choice and Short Answer Objective Test Items in Basic Technology: Comparative Approach. Journal of Educational

Policy and Entrepreneurial Research (JEPER), Vol. 3 (No. 3), 60-61. ISSN: 2408-770X (Print), ISSN: 2408-6231 (Online). https://www.zeetarz.com/Jee/wp-content/uploads/2018/01/Reliability-Coefficient-of-Multiple%E2%80%93Choice-and-Short-Answer-Objective-Test-Items-in-Basic-Technology-Comparative-Approach.pdf

Frey, B. B. (2018). Achievement Tests. SAGE Publications. https://dx.doi.org/10.4135/9781506326139.n16. https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i1967.xml

Glen, S. (2017). Classical Test Theory: Definition. StatisticsHowTo.com: Elementary Statistics for the rest of us!. https://www.statisticshowto.com/classical-test-theory/

Glen, S. (2020). Z-Score: Definition, Formula and Calculation. StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/probability-and-statistics/z-score/

Great Schools Partnership (2015). Standardized Test. Edglossary.org. https://www.edglossary.org/standardized-test/

Haladyna, T. M. (2018). Developing Test Items for Course Examinations. IDEA Student Ratings System. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/IDEA_Paper_70.pdf

Ikhsanudin & Subali (2018). Content validity analysis of first semester formative test on biology subject for senior high school. Journal of Physics: Conference Series 1097012039. https://www.researchgate.net/publication/328258459_Content_validity_analysis_of_first_semester_formative_test_on_biology_subject_for_senior_high_school

Irawan, E., & Wilujeng, H. (2020). Development of an online mathematical misconception instrument. Journal of Physics: Conference Series. https://www.researchgate.net/publication/345162423_Development_of_an_online_mathematical_misconception_instrument

Katz, L. J., & Slomka, G. T. (2000). Achievement Testing. Handbook of Psychological Assessment (Third Edition) and Elsevier B. V. https://www.sciencedirect.com/topics/medicine-and-dentistry/achievement-test

Kean, J., & Reilly, J. (2014). Classical Test Theory. https://static1.squarespace.com/static/514fd024e4b0d4d5c3e59e38/t/53bc6138e4b07f64b2496034/1404854584695/Kean+Reilly+%282014+in+press%29+Classical+Test+Design.Pdf

Kpolovie, P. J. (2016). Single-Subject Research Method: The Needed Simplification. British Journal of Education (Vol. 4, No. 6, pp.68-95, June 2016). https://www.eajournals.org/wpcontent/uploads/Kpolovie-Peter-James.pdf

Kumar, N. (2016). Construction and Standardization of an Achievement Test in English Grammar. International Journal of Current Research and Modern Education (Vol. I, Issue II) http://ijcrme.rdmodernresearch.com/wp-content/uploads/2017/01/141.pdf

Lacia, M. (2007). The Notre Dame University Mathematics Qualifying Examination: An Analysis

Logsdon (2020). Different Types of Scores on Standardized Tests. Very Well Family. https://www.verywellfamily.com/what-are-standard-scores-2162891

Math is Fun (2017). Standard Deviation Formulas. MathsIsFun.com. https://www.mathsisfun.com/data/standard-deviation-formulas.html

Metcalf, A. in ResearchGate (2017, October 25). Many sources say above 0.70 is acceptable. 0.80 or greater is preferred. Higher is better. [Comment on the post "What is the acceptable range for Cronbach alpha test of reliability?"]. ResearchGate GmbH. https://www.re-

searchgate.net/post/What-is-the-acceptable-range-for-Cronbach-alpha-test-of-reliability

Mitra N.K., Nagaraja, H.S., Ponnudurai, G., & Judson, J.P. (2009). The Levels Of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. IeJSME 2009: 3(1): 2-7. https://www.researchgate.net/publication/227858882_The_Levels_Of_Difficulty_And_DiscriminationIndices_In_Type_A_Multiple_Choice_Questions_Of_Preclinical_Semester_1_Multidisciplinary_Summative_Tests

Nixon, B. (2021). The Pros and Cons of Standardized Testing. WHITBY School. https://www.whitbyschool.org/passionforlearning/the-pros-and-cons-of-standardized-testing

Pam, M. S. (2013). ITEM VALIDITY. PsychologyDictionary.org. https://psychologydictionary.org/itemvalidity/

Patock, J. (2004). Exam Scores: How to Interpret your Statistical Analysis Reports. University Testing Services Arizona State University. https://jcesom.marshall.edu/media/24101/Interpreting-Statistical-Information.pdf

Schuwirth, L. W., & Van Der Vleuten, C. P. (2011). General overview of the theories used in assessment. T. Gibbs, Penyunt. Association for Medical Education in Europe (AMEE). https://www.tandfonline.com/doi/abs/10.3109/0142159X.2011.611022

Sefcik, D. J., Bice, G., & Prerost, F. (2013). How to Study for Standardized Tests. Burlington, United States of America: Jones & Bartlett Learning. http://samples.jblearning.com/9780763773625/73625_FMXX_Final.pdf

Sharma, H. L., & Sarita (2018). Construction and standardization of an achievement test in Science. International Journal of Research and Analytical Reviews. http://ijrar.com/upload_issue/ijrar_issue_1399.pdf

Sharma, H.L., & Poonam (2017). Construction and Standardization of an achievement test in English grammar. International Journal of Advanced Educational Research (Vol. 2, Issue 5). http://www.educationjournal.org/archives/2017/vol2/issue5/2-5-108

Siegle, D. (2021). Educational Research Basics: Standardized Scores. University of Connecticut Nega School of Education. https://researchbasics.education.uconn.edu/standardized-scores/

Taber, K. (2016) The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ (2018) 48:1273–1296 DOI10.1007/s11165-016-9602-2. https://www.researchgate.net/publication/317777374_The_Use_of_Cronbach's_Alpha_When_Developing_and_Reporting_Research_Instruments_in_Science_Education

The Glossary of Educational Reform (2015). Standardized Test. Great Schools Partnership. https://www.edglossary.org/standardizedtest/#:~:text=To%20evaluate%20whether%20students%20have,student%20learning%20and%20academic%20progress

Tookoian, J. (2016). Deciding the length of a test. https://edulastic.com/blog/the-right-number-of-test-questions/

UTHealth (2021) Lesson 1.6 Standard Scores. University of Texas-Houston Health Science Center. https://www.uth.tmc.edu/uth_orgs/educ_dev/oser/L1_6.HTM

Zucker, S. (2003). Fundamentals of Standardized Testing. Hardcourt Assessment, Inc. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/Fundamentals_of_Standardized_Testing.pdf