*Journal of African Interdisciplinary Studies (JAIS)*: **ISSN 2523-6725** (online)
**February 2023 Vol. 7, No. 2**
**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

# Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis

By

Dr. Benson Ngigi Kinuthia
University of Eastern Africa, Baraton, Kenya
Email: kinuthiab@ueab.ac.ke

This study evaluates the quality of 1st semester 2020/2021Educational Psychology examination given in Department of Education at UEAB. A retrospective record review of 125 first year education students' performance was done. Analysis of the difficulty index and discrimination index using Microsoft Excel 2010 was done. Finding of the Fill in Items based on Difficulty Indexes (p) reported; 50%, 45 % and 5% items were too easy, average and too difficult respectively and Discrimination Indexes (d), 55%, 30% and 15% items were excellent, good and acceptable respectively. While the findings of the Matching Items supported by Difficulty Indexes (p) reported; 15%, 80% and 5% items were too easy, average and too difficult respectively and Discrimination Indexes (d), 45%, 20% ,10% and 25% items were excellent, good, acceptable and poor respectively. The study concluded that fill in items were within satisfactory levels of difficulty and discrimination although there were some few items which need to be improved in level of difficulty while matching items were within satisfactory level of difficulty and discrimination however some few items need to be improved in levels of difficulty and/or discrimination. The study recommended that both fill in and matching items with satisfactory levels of difficulty and discrimination should be stored in test item bank for later use, improve test items with near satisfactory level while continually ascertain levels of difficulty and discrimination of new items.

**Key words:** Kenya, Education, Difficulty index, Discrimination Index, Quality of Examination, Item Analysis

*Journal of African Interdisciplinary Studies (JAIS)*: **ISSN 2523-6725** (online)
**February 2023 Vol. 7, No. 2**
**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

# Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis

By

Dr. Benson Ngigi Kinuthia

## Introduction

The quality of any test item is determined by the ability of that item to differentiate between students who understand the subject matter and those who do not. The quality of the test items is ascertained by establishing the difficulty and discrimination indices through a process known as test item analysis. Item analysis is especially valuable in improving items which will be used in later tests and examinations, but it can also be used to eliminate ambiguous or misleading items in a single test administration.

Test item analysis is a powerful procedure for revealing the difficulty level and discriminating value of each item in a test or examination. Some items may not discriminate between students who have good grasp of the subject matter and those who are apparently just sliding by (Mc Daniel, 1914, Kinuthia, 1999). Such items should be improved or completely deleted.

Kocdar, Karadag & Sahin (2016) defined "Item difficulty is the percentage of learners who answered an item correctly and ranges from 0.0 to 1.0. The closer the difficulty of an item approaches to zero, the more difficult that item " (p. 16). The authors further defined "discrimination index of an item is the ability to distinguish high and low scoring learners. The closer this value is to 1, the better the item" (Kocdar, Karadag & Sahin, 2016, p. 16), in helping differentiating the students who have gotten the concepts and those who haven't.

Blerkkom (2009) affirms that item analysis can be done on any type of test items but it is much easier demonstrated with objective test items. This study analyzed 'Fill in Item' and Matching Items types since they the only objective types in the EDPC 106-Educational Psychology Examination.

## Statement of the Problem

Examinations' setting at university level is normally done individually or collectively in a group by instructors. During the setting, some examination items are drawn from the examination banks with little or no regards to the performance of the items, such items may not discriminate between students who have good grasp of the subject matter and those who are apparently just sliding by (Michael, Kimberly, Mary, Veronica, Lisa & Margarita, 2019, Sept.).

This study aimed at determining the performance of EDPC 106- Educational Psychology examination by computing the difficulty and discrimination indices.

## Objective

This study was guided by the following objectives"

1. Determine the difficulty index of 'Fill in Items' and 'Matching Items' in the EDPC 106-Educational Psychology Examination.
2. Establish the discrimination index of 'Fill in Items' and 'Matching Items' in the EDPC 106-Educational Psychology Examination.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

## Review of Related Literature

Examinations aim at finding out whether the desired learning outcomes have been achieved based on a set of expected learning outcomes (Ahmad & Hamed, 2014). For the students, examinations may be used as a way for providing feedback to their learning, a mean to motivate them to learn or may be used as means for overlearning. While for the teacher assessment can be a means for to provide insight into how students are processing content taught, detecting students' misconceptions, serve as basis for grading and certification and for evaluating students' achievement of the expected learning outcome in order to make informed decisions regarding the future use and revision of instructional modalities, (Paniagua & Swygert, 2016, Hubbard, Potts & Couch, 2017).

Teacher made examinations can be effective assessment tools if planned to measure the expected learning outcomes after teaching-learning process. Quality examination items are necessary for an examination to have reliability and to draw valid conclusions from the resulting scores (Downing, 2005). The author further asserts that developing quality examination items, especially objective items (multiple choice, fill in, true or false, matching items), can be challenging. According to Downing (2006), there is existing evidence suggesting that a sizeable proportion of items within course-based examinations contain one or more flaws.

Examination Item flaws may introduce systematic errors that reduce validity and can negatively impact the performance of some learners in examinations thus hindering their ability to exhibit what has been learned (Paniagua & Swygert, 2018). There are two categories of technical flaws that the authors identified: "irrelevant difficulty" and "test-wiseness." Irrelevant difficulty occurs when there is an artificial increase in the difficulty of an item because of flaws such as options that are too long or complicated, numeric data that are not presented, continually use of "none of the above" as an option, and stems that are unnecessarily complicated or negatively phrased.

Test-wise flaws include the presence of grammatical cues, grouped options, absolute terms, correct options that are longer than others, word repetition between the stem and options, and convergence. Because of the potential danger for these and other flaws, it is strongly encouraged that examiners consider asking a colleague to review their items prior to administering the examination as a means of identifying and correcting flaws and providing some assurance of content-related validity, which aims to determine whether the test content covers a representative sample of the knowledge or behavior to be assessed (Haladyna, & Downing, 2004, Anastasi & Urbina,, 1997). Conducting an item analysis after students have completed the examination is important as this may identify flaws that may not have been clear at the time the examination was developed.

According to McDaniel (1994), item analysis is a powerful procedure for revealing the difficulty level and discriminating value of each item in a test. Some items may not discriminate between students who have good grasp of the subject matter and those who are apparently just sliding by.

The item difficulty index ($p$) is expressed as the percent of students who correctly answered the item (Thorndike & Hagen, 2008). For example, if 80% of students answered an item correctly, $p$ would be 0.80. Theoretically, $p$ can range from 0 (if all students answered the item incorrectly) to 1 (if all students answered correctly). However, Haladyna and Downing (2004) note that because of students guessing, the practical lower bound of $p$ is 0.25 rather than zero for a four-option item, 0.33 for a three-option item, and so forth.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

A principle function of examination items is to discriminate between the learners who know and those who do not know the content being examined. Items that fail to discriminate fail to make contribution to the total score of an examination. The measure of item discrimination is the item discrimination index (*d*), which measures how well an item differentiates between low- and high-performing students (Lane, Raymond & Haladyna, 2015, Kinuthia, 1999). There are several different methods that can be used to calculate *d* although it has been shown that most produce comparable results (Anastasi & Urbina, 1997). One method used to obtain a discrimination index (d) is by subtracting the number responding correctly in the low group from the number responding correctly in the high group and dividing by number in one group. Based on the results gotten, a decision is made either to throw away an item or retain it. It is recommended that only items with *d* value of 0.3 and higher are good for improvement and revision (McDaniel, 1994).

**Methodology**

This descriptive survey was done on 1st semester 2020/2021 EDPC 106-Educational Psychology examination given in Department of Education at University of Eastern Africa, Baraton (UEAB). A retrospective record review of 125 first year education students' performance was done. Before the analysis was done the already marked papers were retrieved from registry, then the 125 papers were ranked in ascending order, from the lowest to the highest mark. Then 34 (27%) papers for either side of low scores and high scores we picked for analysis (Mahjabeen, Alan, Hassan, Zafar, Butt, Konain and Rizvi, 2017).

The examination had three items types, the filling in items, the matching items and essay items. The filling items and the matching items were picked for analysis. There were twenty items for each of the test items. The number of students who gave correct answers were tabulated and recorded in column 'h' for high score group and 'l' for low score group in Microsoft Excel 2010. Then analysis of the difficulty index and discrimination index using Microsoft Excel 2010 was done using the following formula.

Difficulty Index (p) = [(h + l)/ (H+L)] *100 where

H = Number of students' papers picked for analysis in the high score group

L = Number of students' papers picked for analysis in the low score group

h = Number of students' who gave correct answers in the high score group.

l = Number of students, who gave correct answers in the low score group.

Using criterion given by Mahjabeen, Alan, Hassan, Zafar, Butt, Konain and Rizvi (2017) for categorization of difficulty index (p), values of $p > 70\%$ were interpreted ' Too Easy, $30\% \leq p \leq 70\%$ were interpreted Average and $p < 30\%$ Too Difficulty.

Discrimination index (d) = (h – l)/ H or d = (h -l)/L where

H = Number of students papers picked for analysis in the high score group

L = Number of students papers picked for analysis in the low score group

h = Number of students who gave correct answers in the high score group.

l = Number of students who gave correct answers in the low score group.

Using criterion given by Mahjabeen, Alan, Hassan, Zafar, Butt, Konain and Rizvi (2017) for categorization of discrimination index (d), values of $d \leq 0.2$ interpreted poor, $0.21 \leq d \leq 0.24$ interpreted Acceptable, $0.25 \leq d \leq 0.35$ interpreted Good and $d \geq .36$ Excellent.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

**Presentation of Results and Discussion of Findings**

Analysis was done on 68 students grouped as high and low performers of the 125 students who took EDPC 106- Educational Psychology test in 1st Semester 2020/2021 academic in Department of Education at UEAB. The computed Difficulty index (p) and Discrimination index (d) for fill in items and matching items were presented on tables.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

Table 1*: Difficulty Index and Discrimination Index of Fill in Items*

| Items | H | L | H | l | h + l | h-l | P | D |
|---|---|---|---|---|---|---|---|---|
| 1. | 34 | 34 | 34 | 23 | 57 | 11 | **84** | **0.32** |
| 2. | 34 | 34 | 34 | 22 | 56 | 12 | **82** | **0.35** |
| 3. | 34 | 34 | 34 | 24 | 58 | 10 | **85** | **0.29** |
| 4. | 34 | 34 | 34 | 32 | 66 | 2 | **97** | **0.06** |
| 5. | 34 | 34 | 29 | 14 | 43 | 15 | **63** | **0.44** |
| 6. | 34 | 34 | 31 | 14 | 45 | 17 | **66** | **0.5** |
| 7. | 34 | 34 | 32 | 18 | 50 | 14 | **74** | **0.41** |
| 8. | 34 | 34 | 29 | 10 | 39 | 19 | **57** | **0.56** |
| 9. | 34 | 34 | 30 | 14 | 44 | 16 | **65** | **0.47** |
| 10. | 34 | 34 | 32 | 19 | 51 | 13 | **75** | **0.38** |
| 11. | 34 | 34 | 34 | 29 | 63 | 5 | **93** | **0.15** |
| 12. | 34 | 34 | 34 | 29 | 63 | 5 | **93** | **0.15** |
| 13. | 34 | 34 | 33 | 19 | 52 | 14 | **76** | **0.41** |
| 14. | 34 | 34 | 32 | 20 | 52 | 12 | **76** | **0.35** |
| 15. | 34 | 34 | 26 | 10 | 36 | 16 | **53** | **0.47** |
| 16. | 34 | 34 | 11 | 2 | 13 | 9 | **19** | **0.26** |
| 17. | 34 | 34 | 17 | 7 | 24 | 10 | **35** | **0.29** |
| 18. | 34 | 34 | 27 | 7 | 34 | 20 | **50** | **0.59** |
| 19. | 34 | 34 | 24 | 4 | 28 | 20 | **41** | **0.59** |
| 20. | 34 | 34 | 21 | 0 | 21 | 21 | **31** | **0.62** |

Source: (Field Data, 2020)

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

Table 2: Analysis of Difficulty Index (p) of Fill in Items

| Categorization | Number of Items | Percentage | Interpretation |
|---|---|---|---|
| $p > 70\%$ | 10 | 50% | Too Easy |
| $30\% \leq p \leq 70\%$ | 9 | 45% | Average |
| $p < 30\%$ | 1 | 5% | Too Difficulty |

Source: (Field Data, 2020)

Analysis of Difficulty Index (p) of Fill in Items shows that 10 (50%) items had p-values of greater than 70% implying that more than 70% of the students could answer the items correctly hence items were too easy, 9 (45%) items had p- values of 30% to 70% implying between 30% to 70 % of the students could answer the items correctly indicating they were of average difficulty and 1(5%) item had p-values of less than 30% meaning less than 30% of the students could answer the item correctly signifying that it was too difficult.

The essence of an examination is to differentiate those who know from those who do not know. The better prepared students should answer correctly items more often than less prepared students. "Items that are moderately difficult better differentiate between well-prepared and less-prepared students than items that are either very difficult or items that are very easy" (Blerkom, 2009, p. 127). Therefore, one "should strive to develop test items that are moderately difficult" (Blerkom, 2009, p. 127). Going by Blerkom's assertion, the analyzed Fill in Items of EDPC 106- Educational Psychology examination, at greater extent meet the criterion of a good test items since 45% of the items are averagely difficulty. However, effort should be made to improve the too easy items.

Table 3: *Analysis of Discrimination Index (d) of Fill in Items*

| Categorization | Number of Items | Percentage | Interpretation |
|---|---|---|---|
| $d \leq 0.2$ | 0 | 0% | poor |
| $0.21 \leq d \leq 0.24$ | 3 | 15% | Acceptable |
| $0.25 \leq d \leq 0.35$ | 6 | 30% | Good |
| $d \geq 0.36$ | 11 | 55% | Excellent |

Source: (Field Data: 2020)

Analysis of Discrimination Index (d) of the Fill in Items reveals that 11 (55%) items had d-values of 0.36 and above implying that these items were excellent in discriminating well prepared students from less prepared ones, 6 (30%) items had d-values of between 0.25 to 0.35 interpreted that the items were good in differentiating well prepared from less prepared students and 3(15%) items had d-values of between 0.21 to 0.24 signifying that the items were of acceptable discrimination. These findings were in agreement with Blerkom's assertion:

item discrimination index of .40 or higher indicates that it is an exceptionally good item in terms of its ability to discriminate between well-prepared and less-prepared students. Item discrimination indices in the range of .20 to .39 are very desirable and help the reliability of the test. When item discrimination indices are near zero (.00), the item is neither helping nor harming the reliability of the test. (Blerkom, 2009, p.129)

From the above direct quote, we can conclude that all Fill in Items analyzed were within the acceptable ranges.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

Table 4*: Difficulty Index and Discrimination Index of Matching Item*

| Items | H | L | H | l | h + l | h-l | P | d |
|---|---|---|---|---|---|---|---|---|
| 1. | 34 | 34 | 31 | 15 | 46 | 16 | **68** | **0.47** |
| 2. | 34 | 34 | 34 | 23 | 57 | 11 | **84** | **0.32** |
| 3. | 34 | 34 | 25 | 5 | 30 | 20 | **44** | **0.59** |
| 4. | 34 | 34 | 16 | 10 | 26 | 6 | **38** | **0.18** |
| 5. | 34 | 34 | 28 | 13 | 41 | 15 | **60** | **0.44** |
| 6. | 34 | 34 | 24 | 4 | 28 | 20 | **41** | **0.59** |
| 7. | 34 | 34 | 24 | 6 | 30 | 18 | **44** | **0.53** |
| 8. | 34 | 34 | 20 | 13 | 33 | 7 | **49** | **0.21** |
| 9. | 34 | 34 | 33 | 26 | 59 | 7 | **87** | **0.21** |
| 10. | 34 | 34 | 18 | 3 | 21 | 15 | **31** | **0.44** |
| 11. | 34 | 34 | 27 | 7 | 34 | 20 | **50** | **0.59** |
| 12. | 34 | 34 | 23 | 11 | 34 | 12 | **50** | **0.35** |
| 13. | 34 | 34 | 18 | 4 | 22 | 14 | **32** | **0.41** |
| 14. | 34 | 34 | 10 | 8 | 18 | 2 | **26** | **0.06** |
| 15. | 34 | 34 | 27 | 22 | 49 | 5 | **72** | **0.15** |
| 16. | 34 | 34 | 28 | 17 | 45 | 11 | **66** | **0.32** |
| 17. | 34 | 34 | 12 | 9 | 21 | 3 | **31** | **0.09** |
| 18. | 34 | 34 | 20 | 16 | 36 | 4 | **53** | **0.12** |
| 19. | 34 | 34 | 29 | 18 | 47 | 11 | **69** | **0.32** |
| 20. | 34 | 34 | 31 | 15 | 46 | 16 | **68** | **0.47** |

Source: (Field Data, 2020)

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

Table 5: *Analysis of Difficulty Index (p) of Matching Items*

| Categorization | Number of Items | Percentage | Interpretation |
|---|---|---|---|
| $p > 70\%$ | 3 | 15% | Too Easy |
| $30\% \leq p \leq 70\%$ | 16 | 80% | Average |
| $p < 30\%$ | 1 | 5% | Too Difficulty |

Source: (Field Data, 2020)

Analysis of Difficulty Index (p) of Matching in Items shows that 3 (15%) items had p-values of greater than 70% implying that more than 70% of the students could answer the items correctly hence items were too easy, 16 (80%) items had p- values of 30% to 70% implying between 30% to 70 % of the students could answer the items correctly indicating they were of average difficulty and 1(5%) item had p-values of less than 30% meaning less than 30% of the students could answer the item correctly signifying that it was too difficult. .

Blerkom (2009) counsels that one "should strive to develop test items that are moderately difficult" (p. 127). Considering Blerkom's counsel, the Matching Items analyzed are within the acceptable range since 80% of items had average difficulty.

Table 6: *Analysis of Discrimination Index (d) of Matching Items*

| Categorization | Number of Items | Percentage | Interpretation |
|---|---|---|---|
| $d \leq 0.2$ | 5 | 25% | poor |
| $0.21 \leq d \leq 0.24$ | 2 | 10% | Acceptable |
| $0.25 \leq d \leq 0.35$ | 4 | 20% | Good |
| $d \geq 0.36$ | 9 | 45% | Excellent |

Source: (Field Data, 2020)

Analysis of Discrimination Index (d) of the Matching Items reveals that 9 (45%) items had d-values of 0.36 and above implying that these items were excellent in discriminating well prepared students from less prepared ones, 4 (20%) items had d-values of between 0.25 to 0.35 interpreted that the items were good in differentiating well prepared from less prepared students, 2(10%) items had d-values of between 0.21 to 0.24 signifying that the items were of acceptable discrimination and 5 (25%) items had d-value of 0.2 and below signifying that they had poor discriminating ability.

According to Blerkom (2009), test items with discrimination indexes of above 0.2 are within acceptable ranges. Looking at the results of analyzed Matching items 75% are within acceptable discrimination index. However, the 25% with poor discrimination need to be deleted or improved to fall within acceptable ranges.

## Conclusion

In conclusion, item analysis of the EDPC 106- Educational Psychology examination, of the Fill in Items based on Difficulty Indexes (p) reported; 50% , 45 % and 5% items were too easy, average and too difficult respectively and Discrimination Indexes (d), 55%, 30% and 15% items were excellent, good and acceptable respectively. Generally, fill in items were within satisfactory levels of difficulty and discrimination though there were some few items which need to be improved in level of difficulty.

**Citation:** Kinuthia, B. N. (2023). Improving Quality of Fill-In and Matching Items in an Examination Through Test Items Analysis. *Journal of African Interdisciplinary Studies*, 7(2), 18 – 28.

Also item analysis of Matching Items supported by Difficulty Indexes (p) reported; 15%, 80% and 5% items were too easy, average and too difficult respectively and Discrimination Indexes (d), 45%, 20% ,10% and 25% items were excellent, good, acceptable and poor respectively. Based on the findings one can conclude that matching items were within satisfactory level of difficulty and discrimination however some few items need to be improved in levels of difficulty and/or discrimination.

**Recommendations**

The study recommended that both fill in and matching items with satisfactory levels of difficulty and discrimination should be stored in a test item bank for later use, improve test items with near satisfactory level while continually ascertain levels of difficulty and discrimination of new items.

## References

Ahmad RG, Hamed OAE.(2014) Impact of adopting a newly developed blueprinting method and relating it to item analysis on students' performance. *Med Teach*. 2014;36(SUPPL.1):55 62.

Anastasi A, Urbina S. (1997). *Psychological Testing*. 7th ed. New York, NY: Pearson.

Blerkkom, M. L. (2009). Measurement and Statistics for Teachers. New York: Taylor & Francis, Routledge 270 Madison Ave, NY 10016.

Downing SM. (2005). The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. *Adv Health Sci Educ*., 10:133–143. doi: 10.1007/s10459-004-4019-5

Downing SM. (2006).Selected-response item formats in test development. In: Downing SM, Haladyna T, editors. *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum; pp. 287–302.

Haladyna TM, Downing SM. (2004). Construct-irrelevant variance in high-stakes testing. *Educ Meas Issues Prac*. 2004;23(1):17–27.

Haladyna, T.M., Downing, S.M. & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(2), 309-334.

Hubbard JK, Potts MA, Couch BA.(2017) How question types reveal student thinking: an experimental comparison of multiple-true-false and free-response formats. *CBE Life Sci Educ*. 16(2):1–13.

Kinuthia, B. N. (1999). Achievement Testing: General considerations in achievement test construction. A paper presented in partial fulfillment of the course EDRM 618-Educational Measurement and Evaluation. Department of Education at University of Eastern Africa, Baraton. (Unpublished.)

Kocdar, S., Karadag. N. & Sahin, M.D., (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. TOJET: The Turkish Online Journal of Educational Technology, 15(4), 16-24.

Lane, S, Raymond MR, Haladyna TM. (2015).*Handbook of Test Development (Educational Psychology Handbook)* 2nd ed. New York, NY: Routledge.

Mahjabeen, W., Alan, S., Hassan, U., Zafar, T., Butt, R., Konain, S.& Rizvi, M., (2017). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions.

McDaniel, E. (1994). Understanding Educational Measurement. Madison: Brown and Bench Mark Publishers.

Michael, J. R., Kimberly, K.D., Mary, E.R., Veronica, P.S., Lisa, L., & Margarita, V.D. (2019, Sept.). Best Practices Related to Examination Item Construction and Post- hoc Review. *American Journal of Pharmaceutical Education*, 83(7): 7204.doi: 10.5688/ajpe7204

Paniagua MA, Swygert KA. (2016). National Board of Medical Examiners: Constructing Written Test Items for the Basic and Clinical Sciences .http://www.nbme.org/publications/item- writing-manual.html.

Thorndike RL, Hagen EP. (2008) *Measurement and Evaluation in Psychology and Education*. 8th ed. New York, NY: Pearson.