

The Impact of the Test Length On the Accuracy of Estimates of the Ability Indicators for Individuals and The Information Function of the Test

Geoffrey Anthony Bao Condor

Teodoro Edgardo Leonardo Aguilar

Abstract

This study aimed at verifying the impact of the test length on the accuracy of estimates of the ability indicators for individuals and the information function of the test according to the item response theory, to achieve the objective of the study, a multiple-choice test in chemistry was constructed for students of the 10th grade in Jordan. The test is made up of three models of the same content, but vary in the number of items, the extended test, the intermediate test, and the short test. The(Bilog-Mg) program was used to analyze the 915 students' responses for all three test models according to the two-parameter logistic model. The study results revealed no statistically significant differences between the means of the standard error for estimating individual ability indicators attributed to the length of the test. The results of the study showed that the information function of the test differed with the variation of the test models, as well as the absence of statistically significant differences attributed to the values of the empirical validity coefficients.



Test Length On the Accuracy of Estimates ...

Keywords Test Length, Item Response Theory, Information Function,

1. Introduction Since the foundation of the psychological movement, psychologists have been interested in achieving the highest degree of objectivity in tests while using tests and psychological scales. The field has witnessed increasing developments designing, Constructing, and analyzing test items to achieve this objective (Abu Hatab, 1992). With criticism of classical measurement theory (CCT), contemporary measurement scientists have made innovative research efforts since the 1970s to develop a modern sociometric theory that can overcome many of the traditional measurement problems called the Item Response Theory (IRT) or Latent Traits Theory (LTT). This theory was evolved, and multiple models of significant development emanated since that time until the present time, these developments were mainly based on the advanced technology and computer software, which became available to researchers in several areas (Allam, 2005). In the item response theory, psychological and educational tests generally assume that certain traits or characteristics are shared among all individuals, but vary in size. Although these traits are unobservable, it can be inferred from the observed behavior of an individual whose responses to the test items justifies its designation as latent traits (Allam, 1987). The item response theory consists of a set of mathematical models showing how examinees respond from different levels of ability to test items. These models are also used to estimate the item's and individuals' characteristics in these traits through several

indicators. Using the estimated values of these indicators can explain the performance of each individual in the test, Using the estimated values of these indicators, we can interpret the performance of each individual in the test, and because these features are difficult to observe and measure directly, they are inferred by using these estimated values, they are called latent traits (Salah al-Din Allam, 2005, p. 47; Crocker & 1986, p450). Psychologists have been interested in achieving the highest degree of objectivity when using psychological tests and measurements. To achieve this goal, this field has witnessed increasing developments in designing, constructing, and analyzing test items (Abu Hatab, 1992). The testing is one of the essential tools that provide us with the data used to make many important decisions related to the individual and society. The test is a systematic procedure for measuring a sample of the behavior of individuals. The use of tests has been widespread in many fields. Such as choosing a person for a job from among a group of applicants, or for classification purposes such as determining the students' path in proportion to their abilities and skills, as well as evaluating the achievement of students through the grades obtained in classroom tests, etc. (Allen & Yen, 1979). The items of the multiple-choice test are considered the most flexible among tests. It can be used effectively in the appropriate coverage of the educational material, and the ease and objectivity of the correction because it requires the identification of the correct answer, unlike the essay questions that require recalling answer, they were known for their 177

Test Length On the Accuracy of Estimates ...

178 accuracy, reliability, their correctness and is not affected by the subjective factors of the corrector (Samarah et al., 1989). The multiple-choice test items are considered the most flexible among all tests; this type of tests can be used effectively in the appropriate coverage of the educational material. Moreover, it is the ease and objectivity of the correction because it requires the identification of the correct answer, as correctors never dispute the answers, unlike the essay questions that require recalling answers. They were known for their accuracy, reliability, validity, and is not affected by the subjective factors of the corrector (Samarah et al., 1989). Because of the widespread of these tests and their importance in several areas, one of the essential aspects of the tests is the length of the test, as some consider the best test is the extended test, others prefer the medium length test. In contrast, some consider the short test is the best, hence the importance of the study to examine the impact of the length of the test on the accuracy of the individual ability estimates and the information function of the test according to the binary parameter model, one of the models of the item response theory. Despite the interest of psychometric researchers and educational scientists in several aspects of the tests in general, but the subject of this study did not receive excessive attention by the scientists of psychological and educational measurement, there is a paucity in the studies that examined the subject of this study. The study of (Al -Omari, 1985) is one of the studies related to the subject of this study, which aimed to know the impact of the length of the test time on the

performance of the examinees, and its psychometric characteristics such as, validity and reliability among examinees related to the general concern and risk levels when taking the test using multiple-choice items. The test consisted of (25) question; each question has (4) alternative; only one is correct. The results of the statistical analysis of the binary comparisons of the differences between the reliability coefficients of the test showed a differences between the reliability coefficients of the test attributed to the length of the test. Among the studies related to the subject of the study is the study of (AL -Diyabat, 2007), which aimed to identify the effect of the length of the test on the reliability of the test. The study results showed the differences between the test attributes' reliability estimates and the change in the length of the test. The results indicate that the reliability estimates increase with the length of the test, as the total test had the highest reliability coefficient estimate (0.9585). The least length had the lowest reliability coefficient estimate (0.832).

2. Previous Studies

Alqudah& Al-Shraifin (2020) study aimed at identifying the effect of the length of the questionnaire on the Accuracy of Ability Estimation of the parameters of examinees and the parameters of Items and scale in the light of IRT. The results showed statistically significant differences at the level of ($\alpha = 0.05$) between the standard error average in the estimation of the abilities of individual related to the length of the questionnaire; the results showed differences between the standard

Test Length On the Accuracy of Estimates ...

error averages in favor Length of questionnaire Al-Qaisi (2016) study aimed at investigating the effect of the sample size and length of test on the accuracy estimation of the item-parameter by using the Non- Parametric item- response theory. The Findings showed statistically significant differences at ($\alpha = 0.05$) in the means of Bias in the estimation of the ability parameter θ attributed to the (Sample size and test length). Allen (2016) stated that many factors influence the response rate of a survey or questionnaire. The BYU alumni questionnaire was initially a lengthy survey with over 200 questions. After a short version of the questionnaire was created and administered, response rates appear to have increased substantially. Male respondents appear particularly more inclined to respond to the shortened version compared to the long version. Alhawari (2015) study investigated the Effect of Test Length and Ability distribution form on The Estimation of a person's Ability, item difficulty, and the information function of test and its items, According to Rasch Model in Item Response Theory (IRT). The results showed that there were statistically significant differences at ($\alpha=0.05$) among the standard error means of item in the estimation of difficulty parameters, such estimations in a person's ability were more accurate in the positive and negative skewed. The test consisted from 30 items.

Conclusion

Through the previous sections, it is clear that the test's study achieved its goals by answering the questions and determining the effect of the length of the tests. The

main results included the following: The results of the study showed that the information function of the test differed with the variation models, as well as the absence of statistically significant differences attributed to the values of the empirical validity coefficient. Further studies along the tests are recommended to determine other effects.

References

Abu Allam, R.(2005). LearningEvaluation, Amman: Dar Al Masirah Printing & Publishing house

Abu Hatab, F. (1992). Teacher's Guide in Student's Assessment National Center for Examinations and Educational Assessment in cooperation with the Ministry of Education, Cairo: Dar Ghraib for Printing.

Adas, A. (2002). Teacher's Guide for Building Achievement Tests: Dar Al Fikr for Printing, Publishing and Distribution.

Al-Diabat, L. (2007). The effect of the length of the test on the characteristics of the distribution of real marks according to the logistic parameter model. An unpublished master thesis. Yarmouk University 185

Alhawari, A. (2015). The Effect of Test Length and Ability Distribution Form on the Estimation of A person's

Test Length On the Accuracy of Estimates ...

Ability According to Rasch Model in Item, An-Najah University Journal for Research: Humanities 29 (8), 1463 – 1488.

Alqudah, A.& Al-Shraifin, N. (2020). The effect of the length of a questionnaire on the accuracy estimations of the ability and the psychometric properties of the item and scale in the light of item response theory,

An-Najah University Journal for Research: Humanities 34 (6), 953 – 982.

Allam, S.(1987). A critical balance study of the latent trait s' models, and classical models of psychological and educational measurement. Kuwait University, The Arab Journal of Human Sciences, Issue (27), 18-44

Allam, S. (2005). Response models for Uni- dimensional and multidimensional test items and its applications in psychological and educational measurement.

Cairo: Arab Thought House Allen, D. (2016). The Impact of Shortening a Long Survey on Response Rate and Response Quality. Brigham Young University

Allen, M., & Yen, W. (1979). Introduction to Measurement theory. California: Cole Publishing company. AlOmri, H. (1985).

Effect of the length of the test time on the performance of the examinees and on its psychometric characteristics among other subjects on risk degree and their anxiety level. A master degree unpublished thesis. Yarmouk University

Al-Qaisi, A. (2016). The Effects of the Sample Size and the Length of the Test on the Accuracy Estimation of the Item Parameters by Using Non-Parametric Item Response Theory, *Mu'tah Research and Studies, Humanities and Social Sciences Series*, (31)5, 203-246

Crocker, L., & Algina, J. (1986). *Introductions to classical and modern Test Theory*. Orlando, FL: Hacourt Brace Jovanovich.

Gronlund, N. (1985). *Measurement and evaluation in teaching*. New York: Macmillan.

Hakstain, A. & Whalen, T. (1976). A k-sample significance test for independent alpha coefficients. *Psychometric*, 41, 219-231.

Hambleton, R. & Swaminathan, H. (1985) *Item Response Theory, Principles and Applications*. Boston: Kluwer.

Nijhoff Publishing. A member of the Kluwer Academic Publishers Group

Samara et al. (1989). *Principles of Measurement and Evaluation in Education: Second Edition*, Dar Al-Fikr: Amman