Development of self-made academic achievement test on pedagogy of social sciences

Priyanka Kumari[1] and Dr. Ishrat Naaz[2*]

[1]*PhD Scholar, School of Education, Galgotias University, Greater Noida, India*
priyanka.kumari_phd2019@galgotiasuniversities.edu.in

[2*]*Associate Professor, School of Education, Galgotias University, Greater Noida, India*
ishrat.naaz@galgotiasuniversity.edu.in;

ABSTRACT

Self-made tests are a common tool used by researchers to measure various constructs and outcomes in their studies. However, the lack of standardization in developing and administering these tests can lead to issues with reliability, validity, and comparability of results across studies. The process and significance of standardizing self-made tests in research are examined in this work, along with suggestions for researchers looking to guarantee the caliber and reliability of their evaluation instruments. Standardized tests use a methodical approach. Following these procedures improves the test's validity, reliability, and usefulness. An interpretable measurement is the only test that can be considered valid and reliable. The final version of the social science achievement test consisted of 36 multiple-choice questions. Item discrimination and the difficulty index were computed to analyze the items. Experts from reputable colleges were shown the drafts, and their insightful comments were used to improve the final version.

Keywords: Academic Achievement Test, Difficulty Index, Item discrimination, Reliability, Standardized Test, Validity

INTRODUCTION

It is frequently necessary for researchers to measure particular components or outcomes that are not easily accessible through the current standardized examinations. In these situations, they might create their exams to meet the requirements of their study (Wissman et al., 2012). Although this strategy has advantages, it also poses difficulties for guaranteeing the validity and reliability of these testing. The validity and reliability of study findings may be impacted by inconsistent concept measurement resulting from a lack of uniformity in the creation and delivery of self-made tests. One essential stage in evaluating students' development and comprehension of the subject is creating a self-made academic achievement test on social science (Wiyono & Kholidya, 2018). Research and information collection, planning, product creation, field testing, and changes are all part of this methodical process (Wiyono & Kholidya, 2018). Since the test is used as a gauge of students' achievement, its quality is essential (Arsari et al., 2021). When creating the test items, the cognitive, emotional, and psychomotor domains should be carefully taken into account (Yunus & Maliki, 2021). To guarantee the test's validity and rigor, choosing the right assessment components and checklists is a crucial factor to take into account (Dreyer et al., 2014). The development of the test should also take into account the

different levels of student abilities, from basic to comprehensive and practical (Deng & Xiong, 2020). The teaching-learning process seems incomplete without a test. The test helps a teacher to identify the progress of the students. This suggests that teachers use tests (teacher-tailored tests) on almost every occasion in their academic routines. Teachers may test their students regularly, for example, monthly, termly, and/or annually. However, two fundamental questions arise in line while preparing the self-made test or teacher- tailed tests that is *"Are the tests valid?"* and *"Are the tests reliable?"* The validity and reliability of tests are fundamentally important. This is because tests without validity and reliability cannot yield results that can be understood. This means that for the test results to be used as meaningful measurements, they must possess validity and reliability. The test that is the main focus of this research has strong validity and reliability. A standardized test is the common term for this type of assessment. This is because standardized tests are accurate and valid among other things.

This paper will concentrate on developing a test that is both valid and trustworthy. Thus, this paper will start with a brief explanation of what a standardized test is, followed by a discussion of the procedures that should be followed to ensure the validity and reliability of a standardized test and a conclusion.

STANDARDIZED TEST

Contrary to popular belief, the term "test" is straightforward and slightly ambiguous. Ahmann and Glock (1981) define a test so broadly that it includes some evaluation techniques that produce simply verbal descriptions of student characteristics, and specifically, it is just a series of questions to be answered or assignments to be completed. Cronbach (1970), in contrast to Ahmann and Glock, offers a little more detailed explanation. They assert that any methodical processes for observing are considered tests and techniques to watch someone's behavior and describe it using a set of categories or a numerical scale.

STEPS OF DEVELOPING SELF-MADE ACADEMIC ACHIEVEMENT TEST ON SOCIAL SCIENCES

Many educational researchers have emphasized the importance of using standardized tests to measure academic achievement (Gur et al., 2010). One approach to developing a standardized test is through the use of an instrument blueprint, which can guide the creation of items and the collection of valid evidence (Menold et al., 2015). There are usually a few essential elements involved in creating a standardized academic self-assessment test (Wiyono & Kholidya, 2018). The researcher must first collect data and investigate pertinent constructs and current evaluation instruments (Wiyono & Kholidya, 2018; Menold et al., 2015). The researcher can then plan the creation of the new test based on this preliminary study (Wiyono & Kholidya, 2018). A small sample of students is then used for field testing once the researcher creates an initial version of the test (Wiyono & Kholidya, 2018). As a result, the exam can be revised to address any problems or areas that need improvement (Wiyono & Kholidya, 2018). Because it enables the researcher to assess the test's validity and reliability, this primary field test is an essential part of the development process (Wiyono & Kholidya, 2018). To make sure the test satisfies the requirements for usage in research and practice, the researcher can perform additional operational field testing and make additional adjustments to the test based on the findings of the primary field test (Wiyono & Kholidya, 2018).

ACHIEVEMENT TEST ON THE PEDAGOGY OF SOCIAL SCIENCE

Standardized exams and teacher-made tests are the two types of achievement testing. The former are commonly employed in educational environments to assess how well pupils do in relation to the instruction they get there. Experts in test construction develop standardized assessments, usually with feedback from educators, school administrators, and curriculum specialists. Comparing a student's performance to that of other students in the same grade and age group is the aim of the creation of standardized tests. The development of standardized examinations can be time-consuming. These tests are referred to as standardized because they are administered and graded according to exacting and uniform standards. In other words, a standardized test administered and graded in a school would be administered and graded in the same manner as it would be anywhere else in the country.

Development of a Self-made Academic Achievement Test on Pedagogy of Social Sciences One essential stage in evaluating students' development and comprehension of the subject is creating a self-made academic achievement test on social science (Wiyono & Kholidya, 2018). Research and information collection, planning, product creation, field testing, and changes are all part of this methodical process (Wiyono & Kholidya, 2018). Since the test is used as a gauge of students' achievement, its quality is essential (Arsari et al., 2021). When creating the test items, the cognitive, emotional, and psychomotor domains should be carefully taken into account (Yunus & Maliki, 2021). To guarantee the test's validity and rigor, choosing the right assessment components and checklists is a crucial factor to take into account (Dreyer et al., 2014). The test's design should also consider the range of student skill levels, from fundamental to thorough and useful (Deng & Xiong, 2020). Students' performance can be accessed via standardized exams, but it's crucial to make sure the test questions are customized to the unique requirements and aptitudes of the target audience (Deng & Xiong, 2020). Teachers can create a self-made academic achievement test on social science that accurately gauges student learning and guides instructional decisions by using a methodical approach and best practices in educational assessment.

VALIDITY AND RELIABILITY OF A TEST

Creating a standardized test is the same as creating a good test. A good test should be created in a way that is both widely accepted and scientifically reasonable. It should also have unique characteristics that set it apart from a bad test. Finally, it should be valid and reliable. Expert opinions can be consulted to determine the validity of test items. It is possible to get expert opinions prior to testing the test. Nonetheless, the test creator should first create rankings or classifications of suitability and degree of difficulty of the test item so that the specialists can provide logical conclusions about the exam items. The test items' validity may differ from one kind of test to another. In the words of Rust and Golombok (1989), five types of validity can be obtained from one test.
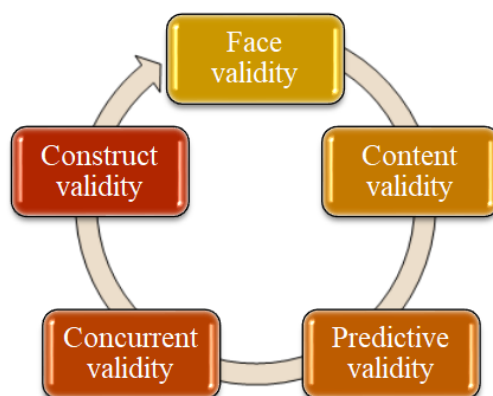
Figure 1.1 Types of Validity

While constructing the standardized test researchers must have a particular syllabus, in this test researchers followed the syllabus of the Guru Gobind Singh Indraprastha University of Delhi. The researcher had prepared four modules of social science each module was prepared by prescribed four quadrants of MOOCs. A multiple-choice question was prepared and this test contained 46 test items from four different modules before the finalization of the draft. The final draft contains only 36 items of the test. The Finalization of the test has gone through many steps. To ensure the development of a robust and reliable academic achievement test, it is essential to involve a team of content experts who can provide valuable insights into the subject matter and the desired learning outcomes (Menold et al., 2015). Additionally, a thorough review of existing instruments can help inform the development process and identify best practices (Menold et al., 2015). The test was presented to the experts of the different universities and the final test was prepared by selecting 36 questions.

ITEM ANALYSIS

According to Downie and Health (1974), an item analysis must be performed on test results to assess the validity of any test. The analysis of test items, as Downie and Health further claim, leads to three kinds of information:
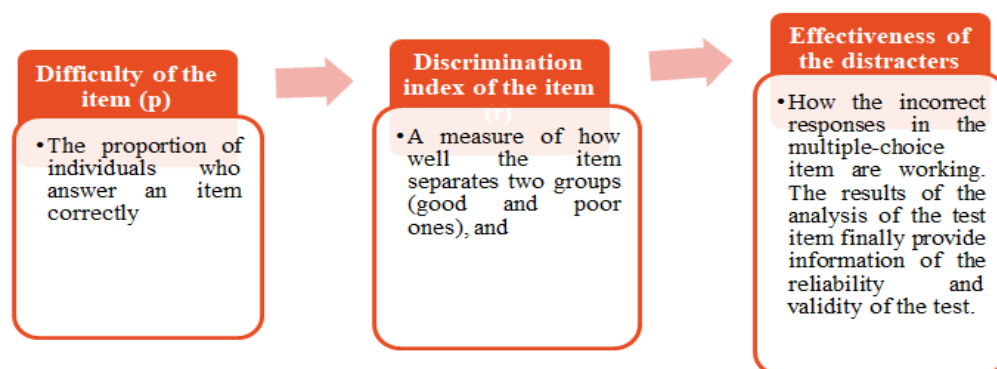
FIGURE: 1.2 STEPS OF ITEM ANALYSIS

Item analysis is the process by which the quality of test items can be assessed. It is the process of collecting, and summarizing the collected data and using student responses. Two parameters are used for determining the Difficulty Index (P) and Discrimination Index (D) of the items in a test paper for evaluating the standards of MCQ questions used in the test paper. In this study, 46 test items from four modules were taken to assess the student's achievement test and retention. The difficulty Index and Discrimination Index of each question were analyzed by using Microsoft Excel.

The formula for calculating the Difficulty Index (P) $\frac{RU+RL}{TOTAL\ VALUE} * 100$

RU= Response of Upper group

RL= Response of Lower Group

Total Value=RL+RU

The index of difficulty and discrimination for all 46 test items administered for the pilot study were thus computed. Any item whose difficulty index is lower than 10% or higher than 90% should be considered worthless for measurement (Remmers et al., 1967) any item whose index of discrimination is above 0.30 should be regarded as a reasonably good item (Ebel and Frisbie, 1986). In the present study, only such items with difficulty indices ranging from 55% to 81% and whose indices of discrimination ranged from 0.31 to 0.48 were selected. Thus, 8 items were deleted on the above principle and only 36 items were retained for the final test. The final form of the test had 36 multiple-choice items for 36.

The relationship between the item difficulty index and discrimination index for each test item was determined by Pearson correlation. Thirty-six out of forty-five test items showed an excellent discrimination index.

COEFFICIENT CORRELATION

The coefficient correlation of the test ranges from **0.76** which is highly reliable.

RELIABILITY OF THE TEST

Split-half method was used to calculate the reliability of the test. In this method, the tool is first divided into two equivalent 'halves'. The final test contains 36, test was divided into two equivalent halves 18 items each. It was done by having alternate items. The measure of the first half of the I tool is correlated with the measure of the other half. The measures are correlated to find the reliability of tests. The main limitation of this method is that a tool can be divided into two halves in several ways and, thus, the estimate of the reliability may not have a unique value. The test was

0.86 reliable which was highly reliable. Test items No. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 34, 37, 38, 40, 41, 43, 44, 45, Rejected Items

were 16, 25, 30, 32, 33, 35, 36, 39, 42, 46. Through these processes the test was standardized and it was shown to the expert of Jamia Millia for face validity and content validity. Experts' suggestions were taken and improvement was made.
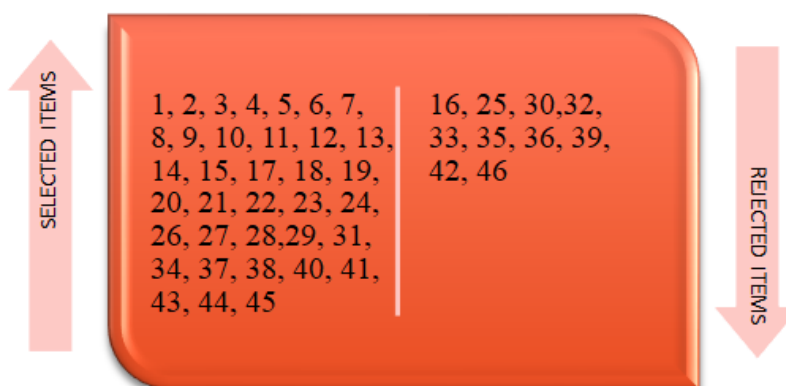
FIGURE 1.3 SELECTED/REJECTED ITEMS

CONCLUSION

In conclusion, the development of a standardized academic achievement test requires a systematic and rigorous approach that involves several key steps, including research and planning, item development, field testing, and revision (Wiyono & Kholidya, 2018) (Menold et al., 2015) (Gur et al., 2010). By following this process, researchers can ensure that the resulting test is a valid and reliable measure of the constructs of interest. This is critical for ensuring that the test can be used effectively in research and practice to assess student learning and inform educational decision- making.

REFERENCES

1. Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement*, *22*(3), 37–51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

2. Bashar, U. & Naaz, I. (2024). Digital Literacy: The Importance, Initiatives and Challenges. *International Research Journal of Modernization in Engineering Technology and Science,* 06(05), 6218-6223. https://www.doi.org/10.56726/IRJMETS56658

3. Bashar, U. & Naaz, I. (2024). Digital Literacy among B.Ed. Student Teachers. *International Journal of Scientific Research and Engineering Development*, 7 (3), 712-717.

4. Bhadana, S. & Naaz, I. (2023). Techno-Pedagogical Skills in the 21st Century and Their Relevance to NEP 2020 For Achieving the Vision of G20. *Bharat Manthan Multidisciplinary Research Journal*, 01 (02), pp. 73-81.

5. Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure? *Proceedings of the National Academy of Sciences of the United States of America,* 113(47). 13354–13359. https://doi.org/10.1073/pnas.1601135113

6. Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, *78*(2), 141–146. https://doi.org/10.1037/0022-0663.78.2.141

7. Crawford, L., Tindal, G., & Stieber, S. (2001). Using Oral Reading Rate to Predict Student Performance on Statewide Achievement Tests. *Educational Assessment/Educational Assessment*, *7*(4), 303–323. https://doi.org/10.1207/s15326977ea0704_04

8. Ercan, O. (2014). The Effects of Multimedia Learning Material on Students' Academic Achievement and Attitudes Towards Science Courses. *Journal of Baltic Science Education*, *13*(5), 608–621. https://doi.org/10.33225/jbse/14.13.608

9. Ercan, O. (2014). The Effects of Multimedia Learning Material on Students' Academic Achievement and Attitudes towards Science Courses. *Journal of Baltic Science Education*, *13*(5), 608–621. https://doi.org/10.33225/jbse/14.13.608

10. Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R., Sheridan, M. A., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2014). Cognitive Skills, Student Achievement Tests, and Schools. *Psychological Science*, *25*(3), 736–744. https://doi.org/10.1177/0956797613516008

11. Gresham, F., Evans, S., & Elliott, S. (1988). Academic and Social Self-Efficacy Scale: Development and Initial Validation. *Journal of Psychoeducational Assessment*, 6, 125 - 138. https://doi.org/10.1177/073428298800600204

12. Guay, F., Marsh, H., & Boivin, M. (2003). Academic Self-Concept and Academic Achievement: Developmental Perspectives on Their Causal Ordering. *Journal of Educational Psychology*, 95, 124-136. https://doi.org/10.1037/0022-0663.95.1.124.

13. Huck, S. W., & Bowers, N. D. (1972). Item Difficulty Level and Sequence Effects in Multiple-Choice Achievement Tests. *Journal of Educational Measurement*, *9*(2), 105–111. https://doi.org/10.1111/j.1745-3984.1972.tb00765.x

14. Juttner, M., & Neuhaus, B. J. (2012). Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *International Journal of Science Education*, *34*(7), 1125–1143. https://doi.org/10.1080/09500693.2011.606511

15. Kumari, P. & Naaz, I. (2022). Web-Based Learning in Inclusive Education. In Gupta Tania & Karun K. N. (Ed.) *Teaching-Learning Models and Strategies for Inclusive Education.* (pp. 233-245). Ankur Book Distributors Delhi-110053.

16. Kumari, P. & Naaz, I. (2020), Digital learning through MOOCs: Advantages, Outcomes & Challenges. *Sambodhi, UGC CARE Journal,* 43.4, (VII), 18-22

17. Kumari, P. & Naaz, I. (2023). Online Pedagogy Adopted for Effective Teaching Learning. In Aerum Khan, Naheed Zahoor and Jasim Ahmad (Ed.) *ICT Integration in Education - Issues, Challenges, and Prospects* (188-195). Shipra Publications, New Delhi.

18. Kumari, P. & Naaz, I. (2022), The Role of MOOCs in Vocational Education with Special Reference to National Education Policy-2020. In proceeding *'Umteb International Scientific Research Congress-XII*, pp. 606-613, organized by Mata Sundri College for Women, University of Delhi, Delhi held on 13-15th August 2022 by IKSAD Global 2022.

19. Marsh, H., Pekrun, R., Murayama, K., Arens, A., Parker, P., Guo, J., & Dicke, T. (2017). An Integrated Model of Academic Self-Concept Development: Academic Self-Concept, Grades, Test Scores, and Tracking Over 6 Years. *Developmental Psychology*, 54, 263–280. https://doi.org/10.1037/dev0000393.

20. Naaz, I. (2015). ICT as a catalyst for learning for differently-abled learners. *In proceedings of International Education Conference 2015 Learning Technologies in Education (IEC2015,)* (Pp. 224-234). organized by Faculty of Education, Jamia Millia Islamia.

21. Naaz, I. (2020). Innovations in Open and Distance Learning (ODL). In Husain, A, & Bahuguna R.P. (Ed*.) Distance Education, and Educational Technology* (pp. 338-344). Maktaba Jamia Ltd. New Delhi.

22. Nuthall, G., & Alton-Lee, A. (1995). Assessing Classroom Learning: How Students Use Their Knowledge and Experience to Answer Classroom Achievement Test Questions in Science and Social Studies. *American Educational Research Journal*, *32*(1), 185–223. https://doi.org/10.3102/00028312032001185

23. Nengsi, Ayu Rahma. The Development of Multiple-Choice Achievement Tests for Evaluation of Geography Learning Outcomes for Students of Second Grade at Senior High School. *UNES Journal of Education Scientists*, 2(2), 118–118, doi:10.31933/ujes.2.2.118-126.2018.

24. Shoemaker, D. M. (1975). Toward a Framework for Achievement Testing. *Review of Educational Research*, *45*(1), 127–147. https://doi.org/10.3102/00346543045001127

25. Thorndike, R.S, & Hagen, E. P. (1961). Measurement and Evaluation in Psychology and Education. *Journal of the American Statistical Association*, *56*(296), 1029–1029. https://doi.org/10.2307/2282039

26. Verma, A. & Naaz, I. (2021). The Divides of Online Learning during Covid-19. *Edu link: Journal of Education*, 7, 32-43.

27. Verma, A. & Naaz, I. (2022). Computer-aided learning: Influence of e-learning on the education system. In Gupta Tania & Karun K. N. (Ed.) *Teaching-Learning Models and Strategies for Inclusive Education*, (pp.258-272). Ankur Book Distributors Delhi-110053.