Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 4, 2025

CONSTRUCTION AND SEMI-STANDARDIZATION OF AN ACHIEVEMENT TEST FOR PROSPECTIVE TEACHERS IN METHODS OF TEACHING

Iffat Khalil[*1], Prof. Dr. Muhammad Aamir Hashmi[2]

*[*1]Phd Scholar, Institute of Education and Research, University of the Punjab*
[2]Institute of Education and Research, University of the Punjab

[*1]iffatkhalil25@gmail.com

DOI: https://doi.org/10.5281/zenodo.15314894

Keywords
Semi-Standardization, Item Difficulty, Item Discrimination, prospective teachers.

Article History
Received on 24 March 2025
Accepted on 24 April 2025
Published on 30 April 2025

Corresponding Author: *
Iffat Khalil

Abstract

The primary aim of this study was to develop a semi-standardized achievement test for prospective teachers in the Methods of Teaching course. A detailed table of specifications was prepared to guide the test development process. Initially, 60 multiple-choice questions were generated from the relevant content domain and reviewed by a panel of experts for content validity. Based on the expert feedback, necessary revisions were made, and all items were then selected for pilot testing. After administering the pilot test to a sample of 50 prospective teachers, item analysis was conducted, resulting in the selection of 35 items.The final version of the test comprised 35 items, which were administered to a larger sample of 400 prospective teachers enrolled in teacher training programs at universities. Data collected from the test were analyzed using item analysis i-e Item Difficulty and Item Discrimination was used. The study concluded that the final test was of moderate difficulty and both valid and reliable for assessing prospective teachers' academic achievement. Furthermore, the selected items may undergo additional try-outs with different samples to further standardize the test.

INTRODUCTION

Learning at its core is a transformative process that leads to lasting change in students. Effective teaching brings about this transformation by developing specific skills, shaping attitudes, and fostering a deeper understanding of scientific principles (Sequeira, 2012). In higher education, students expect to be treated as independent learners who actively engage in their learning process, ask questions, and seek clarifications (Michael & Modell, 2003). Ownership of the learning process is crucial for students, as it enhances motivation and engagement (Mitra, 2008; Pond & Rehan, 1997). It is widely recognised that the main goal of professional higher education is to help students become reflective practitioners capable of critically examining their professional roles. Consequently, teacher educators have the responsibility to develop student teachers into competent professionals who continually reflect on, develop, and refine their practices. To measure the extent of a student teacher's competency and to support their development, high-quality assessment remains crucial.

Furthermore, teaching is a dynamic and unified process that involves both instruction by teachers and learning by students. Teaching is considered one of the most noble and esteemed professions. Teachers serve as the custodians of human civilization. The preservation and transmission of humanity's greatest cultural achievements, alongside the nurturing of emerging talent, are deeply connected to the dedication and hard work of teachers. Through this

process, students acquire knowledge and skills while simultaneously developing physically, mentally, and morally. Thus, teaching plays a pivotal role in shaping students' intellectual growth and moral character.

In addition, evaluation and assessment are key components of the educational process, as they provide critical insights into students' knowledge and skill development. Assessment is a fundamental element of the instructional process, offering comprehensive information about learners' progress. Linn and Gronlund (2008) defined assessment as "any of a variety of procedures used to obtain information about the student performance." Specialized tools of assessment allow educators to measure students' performance systematically. A test, specifically, is an instrument consisting of a set of questions designed to evaluate students' achievement. Although definitions of "test" vary, Linn and Gronlund (2008) described a test as "a particular type of assessment typically consisting of a set of questions administered during a fixed period under comparable conditions for all students." Various types of tests are used to assess knowledge, skills, and behavior, with achievement tests—both informal and standardized—being the most prevalent (Bichi & Talib, 2018).

Moreover, assessment plays a central role in teacher education by evaluating not only content mastery but also pedagogical competence and readiness for real-world teaching. According to Popham (2011), assessments in teacher education must be carefully designed to provide meaningful evidence of a candidate's ability to foster student learning. However, a persistent challenge remains the inconsistency in assessment practices across institutions, resulting in variability in the preparedness of prospective teachers (Darling-Hammond, 2017).

However, while standardized testing offers a solution for comparability, it is often criticized for its rigidity and potential to constrain pedagogical creativity (Au, 2007). On the other hand, entirely localized assessments provide flexibility but often lack the reliability and generalizability necessary for large-scale comparison. Thus, a semi-standardized approach—where core components of a test are standardized while allowing contextual adaptations—emerges as a promising middle ground (Black & Wiliam, 1998).

Furthermore, Bloom's Taxonomy (1956) remains a foundational framework for educational assessment design, offering a hierarchical model for structuring learning outcomes from basic recall to higher-order thinking skills such as analysis and evaluation. Achievement tests grounded in Bloom's cognitive levels ensure that assessments extend beyond mere memorization, capturing deeper and more meaningful learning (Anderson & Krathwohl, 2001).

Moreover, Classical Test Theory (CTT) underpins much of the technical development of educational assessments. Core CTT concepts such as item difficulty, item discrimination, and test reliability are essential for ensuring that assessments are psychometrically sound (Allen & Yen, 2002). In the context of teacher education, employing CTT principles during test construction and validation enhances the credibility of assessment results and ensures fair evaluation of prospective teachers' competencies.

In addition, several researchers emphasize that assessment in teacher education must serve not only to measure knowledge but also to support student learning, inform instructional practices, and ensure teacher readiness for diverse educational contexts (Zeichner, 2010; Cochran-Smith et al., 2015). Semi-standardized tests, designed through robust theoretical and psychometric foundations, can effectively fulfill these multiple critical roles.

Thus, achievement refers to the level of skill or proficiency attained in a specific task or domain of knowledge. Tests can generally be classified into two broad categories based on their construction procedures: teacher-made tests and standardized tests. Teacher-made tests are developed by educators responsible for teaching the courses. Moreover, these tests are relatively easier to construct, requiring fewer resources, less time, and minimal expert validation. In contrast, standardized tests are developed by specialists following rigorous procedures to ensure psychometric soundness, including reliability and validity.

Moreover, standardized tests are designed following structured methodologies. Swain et al. (2005, p.39) define standardized tests as "carefully constructed assessments that ensure uniformity in administration, scoring, and interpretation of results." These assessments are typically norm-referenced, whereby

student performance is compared to a representative reference group to determine relative achievement levels. Mullis et al. (2007) emphasize that the establishment of well-defined norms is critical for ensuring the accurate interpretation of scores in standardized testing contexts.Despite their advantages, the development of standardized tests is a complex, time-consuming, and costly process (Roid, 2012). The standardization procedure includes:

- Defining the scope of the test
- Setting clear testing objectives
- Determining test item formats
- Creating a table of specifications
- Developing an item bank
- Assessing validity and reliability
- Conducting pilot testing
- Analyzing and refining test items
- Finalizing the test based on empirical data (Akhter, Akhtar & Iqbal, 2019)

A critical part of test standardization is item analysis, which includes both qualitative and quantitative evaluations. Face validity and content validity assess the overall quality of test items, while quantitative measures such as item difficulty, distractor effectiveness, and item discrimination determine statistical reliability (Rudner, 2011). Item difficulty refers to the percentage of students who answer a question correctly or incorrectly, with an optimal difficulty range of 0.20 to 0.80 (Hulin, Drasgow & Parsons, 2013). Item discrimination assesses how well a test differentiates between high-achieving and low-achieving students, with a discrimination power of at least 0.30 being considered acceptable (Kelly, 2014; Popham, 2006).

Within the broader field of educational theory and practice, teaching methodology is a critical component. Traditionally, teaching methods have been understood in a relatively narrow sense. A more precise definition describes them as structured strategies used by teachers to achieve learning objectives, deliver content effectively, and facilitate meaningful teacher-student interactions.Given the independence and significance of teaching methods, they occupy an essential place in educational activities.The present study aimed to develop semi-standardized test items for Achievement test for prospective teachers. For this purpose, the AIOU General Methods of Teaching was utilized to design an objective-type test based on multiple-choice questions (MCQs). The study followed standardization principles to create a semi-standardized test for Methods of Teaching.The goal was to construct a semi-standardized achievement test by following standard test development procedures. This initiative aims to improve the assessment process and enhance the quality of education by ensuring valid and reliable student evaluations.

Theoretical Framework

This study is grounded in two key educational theories: Bloom's Taxonomy of Educational Objectives **and** Classical Test Theory (CTT)**.**

Bloom's Taxonomy of Educational Objectives (1956):

Bloom's Taxonomy provides the foundational structure for developing the achievement test. It categorizes cognitive processes into a hierarchy: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. The current test focuses primarily on assessing the first three lower-order cognitive skills—Knowledge, Comprehension, and Application. By aligning test items with these levels, the study ensures that assessments target specific learning outcomes critical for prospective teachers. Furthermore, the study recommends expanding future assessments to cover higher-order thinking skills to foster deeper, more analytical learning.

Classical Test Theory (CTT):

CTT informs the construction, evaluation, and analysis of the test items. According to CTT, an observed score consists of a true score and an error component. To enhance the reliability and validity of the test, item difficulty, discrimination indices, and overall reliability were calculated. These psychometric properties ensure that the semi-standardized test fairly and accurately measures prospective teachers' academic achievement, distinguishing effectively between different levels of learner performance.

Together, Bloom's Taxonomy and CTT provide a strong theoretical base for designing a semi-standardized, reliable, and educationally meaningful assessment tool for methods of teaching courses.

Statement of the Problem

Teaching is a wholesome activity an activity. It is considered not just an art but the most challenging of all arts and the most profound of all sciences. Therefore, teaching is a deliberate, intricate, and complex activity. Every nation has expectation from its students to bring change in the society. Teachers are the crucial agents for bringing out the desirable changes in the system. Therefore, to accomplish the expected assignments the teachers must posses certain competences of profession. Teacher education plays a crucial role in this context, providing prospective teachers with proper guidance to enhance their knowledge and teaching competencies; it also aims to instill desirable teaching qualities. Besides all these practices and the subjects taught in this programs, methods of teaching is a core subject and the importance of this course cannot be denied as this course serve as a baseline for developing teaching skills in prospective teachers. Therefore, it is very important to assess the pedagogical knowledge of the prospective teachers. For the said purpose, there must be a valid and reliable tool. So, a semi-standardized test in methods of teaching was developed to measure the achievement of prospective teachers.

Objectives of the Study

The objectives of the study are:

To construct a semi-standardized achievement test of Methods of teaching.

To find difficulty index and item discrimination of each item in achievement

Research Questions

Following are research questions of the study:

How to construct a semi- standardized achievement test?

What is the difficulty index and item discrimination of each item of an achievement test?

SIGNIFICANCE OF THE STUDY

The semi-standardization of tests in methods of teaching for prospective teachers holds significant value for teacher education programs, educational institutions, and broader educational policy-making. Several factors influence student progress in classrooms, including instructional strategies, classroom management, and student-teacher interactions. This study highlights the importance of consistent assessment of the prospective teachers across the same core standard of the teaching profession.Educational policymakers may find the study's findings useful when developing teacher evaluation standards, offering a structured yet adaptable framework that can guide curriculum reform, faculty development, and quality assurance initiatives in teacher education. This approach could lead to more strategic teacher training practices for developing the well-rounded teacher profiles.The findings can serve as a foundation for future initiatives aimed at developing innovative intervention strategies and teacher training programs. Over time, such advancements could lead to a more resilient and effective educational system that adapts to evolving societal demands. Furthermore, it underscores the direct impact of these factors on student success and can serve as an evidence for elevating the quality and credibility of teacher training programs and recruitment practices. The long-term goal is to refine classroom practices, ensuring that both prospective teachers are better equipped to thrive in an increasingly interconnected world.

Methodology

The study was quantitative in nature, test development technique were used to develop the semi-standardized test and survey design was used to collect data. The population of the study consisted of 400 prospective teachers enrolled in teacher training programs.

To measure the academic achievement of the prospective teachers, researcher developed a semi-standardized achievement test. As the name suggest, the purpose of an achievement test is to assess a student's proficiency, abilities, and comprehension in a particular subject. Nevertheless, Callahan et al. (2010) elucidated that achievement assessments are frequently implemented in educational institutions and preparation settings. These are employed to assess the effectiveness of educational factors, teachers, and courses of study. Standardized test as defined by Good (1959) is "a test for which content has been selected and empirically verified, for which norms have been established, for which uniform methods of administration and scoring have been developed, and

for which a relatively high degree of objectivity can be scored."

Widely respected organizations such the American Psychological Association (APA) and American Educational Research Association (AERA)have proposed some directions for creating an achievement test. There are four major phase involved in the construction of a semi-standardized test such as planning, construction, evaluation and validation of a test.

Planning of the test

The first step of test construction is planning it includes the purpose of the test,to whom, what, when and how to measure. It includes designing the test and preparation of the blue print. Stanley and Hopkins (1990) observed that the planning stage of a test should include nature of the test, items of a test and administration condition. There are three major steps of planning:

(i) Defining test universe and test purpose.
(ii) Defining the content domain to be measured.
(iii) Preparing the blue print of the achievement test.

Defining test universe and test purpose. To define the universe of the test according to the course outline of Methods of Teaching given by Higher Education Comission (HEC) for prospective teachers; the book Methods of Teaching for B.Ed students, published by Allama Iqbal Open University Press, Islamabad was referred. After consulting the course outline researcher decided to consider all the 9 chapters in the test universe. The purpose of test was to assess the academic achievement of prospective teachers.

Table 1: Test Universe

Unit No.	Unit Name	Page No.
1.	Introduction to Teaching	01
2.	Lesson Planning	51
3.	Student Motivation	75
4.	Inquiry Method	95
5.	Activity Methods	117
6.	Discussion Method	133
7.	Cooperative Learning	175
8.	Teaching Skills	197
9.	Teaching Tools	223

Source: Allama Iqbal Open University, Islamabad (2016)

Defining the measurable content domain. The test intended to measure the prospective teachers knowledge regarding the general methods of teaching. The researcher decided to test only three levels; knowledge, comprehension and application of cognitive domain of Bloom's taxonomy.

Preparing the table of specification. To construct items for the test a table of specification also called blue print of the test was prepared. As the achievement test was delimited to only assess the knowledge, comprehension and application levels of the cognitive domain of the Bloom's taxonomy. The table of specification which is as follows:

Table 2: Table of Specification

Unit no.	Unit Name	Knowledge (33%)	Comprehension (45%)	Application (22%)	Total (100%)
1.	Introduction to Teaching	2	2	1	5
2.	Lesson Planning	2	2	1	5
3.	Student Motivation	2	3	1	6
4.	Inquiry Method	1	3	2	6
5.	Activity Methods	2	3	2	7

6.	Discussion Method	3	4	1	8
7.	Cooperative Learning	3	4	2	9
8.	Teaching Skills	2	4	2	8
9.	Teaching Tools	3	2	1	6
Total		**20**	**27**	**13**	**60**
Percentage		33%	45%	22%	100%

CONSTRUCTION OF TEST ITEMS

This is second phase of the test construction of items of test was done. The researcher selected multiple choice questions for this achievement test, because of their objectivity and ease of scoring. Examinations with multiple-choice questions are efficient, discriminative, and can be integrated with other assessment methods to provide an inclusive evaluation (Brady, 2005, Rodriguez, 2005). The detail of each item number according to the cognitive domain levels are presented in following:

Table 3: Details of Items in each cognitive level

Unit no.	Knowledge	Comprehension	Application	Total
1.	4, 27	1, 2	3	5
2.	14, 15	13, 59	6	5
3.	7, 8	9, 10, 11	12	6
4.	23	16, 17, 18	19, 20	6
5.	21, 22	24, 25, 26	28, 29	7
6.	30, 31, 32	33, 34, 35, 36	37	8
7.	5, 39, 40	41, 42, 43, 44	45, 46	9
8.	47, 48	49, 50, 51, 52	38, 53	8
9.	54, 55, 56	57, 58	60	6
Total	**20**	**27**	**13**	**60**

At initial level 60 MCQs were constructed by the researcher. The items were having four options from which the respondent can choose the correct answer. 20 items were developed for knowledge level, 27 items were developed to measure the comprehension of students and 13 items were developed to measure the application level of the prospective teachers.

EVALUATION OF THE ITEMS

To check the content validity of the test. The first prepared draft of the test was given to five experts from the field working as professionals. They were requested to critically analyse the items that whether the items are measuring the intended variable. After approval of the constructed items from the experts that all the items are valid and are measuring the intended variable, it was piloted.

Pilot Study

A pilot study serves as an initial assessment to detect potential challenges in the main research project, determining whether it is feasible or requires modifications. This process helps evaluate the suitability, complexity, and effectiveness of the proposed methods and instruments for the study (Fraenkel, 2012). 40 prospective teachers of the last semester of teacher training program were selected for pilot testing and were provided the achievement test. The prospective teachers were selected keeping in mind that they must have knowledge of the content and have already gone through it. They were given open and it was recorded that prospective teachers complete the test in 30 minutes. After conducting the test were marked according the key. The correct item was marked 1while incorrect item was marked 0. After marking the test, item analysis was performed and difficulty index, discrimination level and distractor effect of the items were calculated.

Item Difficulty Index

The level of difficulty of an item was determined by dividing the correct number of responses by the total responses. The formula given below was used to calculate item difficulty.

$$\text{Difficulty Level } \rho = \frac{\text{Number of respondents who response correctly}}{\text{Total number of respondents}}$$

The difficulty level for each item was determined by using criteria given by Levitov (1985)After the application of above formula on collected data, table ;

Table 4: Difficulty index distribution for test

Difficulty Index	Evaluation of Item
0.71 to 1.00	Easy
0.31 to 0.70	Middle difficult
0.21 to 0.30	Difficult
0 to 0.20	Very difficult

Item Discrimination

The discrimination level of test items was calculated by determining the difference in correct responses between the upper 27% and lower 27% of respondents, then dividing this difference by the total number of responses. Bahoo (2015) reported that this criterion was carefully selected to assess the test's ability to distinguish between high and low achievers. The formula used to calculate the discrimination level of a test item is as follows:

$$\text{Discrimination level } D = \frac{HA — LA}{n}$$

Where
n = Total number of students
HA = number of High Achievers
LA = number of Low Achievers
The items were distributed according to the discrimination level proposed by Wiersma (1990) after the application of above formula on collected data.

Table 5: Discrimination level distribution for test

Item Discrimination Index	Quality
D≥ 0.50	Very Good Item
0.40 - 0.49	Good Item
0.30 - 0.39	Reasonably Fair Quality
0.20 - 0.29	Potentially Poor Item
D < 0.20	Potentially Very Poor

Distractor Effect

The distractor effect indicates the percentage of respondents who selected a particular distractor. To analyze this, the percentage of responses for each option provided in the test item was calculated. According to Akhter (2014) the established criteria, at least 2% of respondents must select each option to ensure the effectiveness of the distractors. The effectiveness of a distractor of a test item can be determined using the following formula:

$$\text{Distractor \%} = \frac{\text{The number of students who choose distracters}}{\text{Total number of test taker}} \text{ X } 100$$

Table 6: Item Analysis

Item no.	Correct Response of Upper Group	Correct Response of Lower Group	Discrimination Index	Difficulty Index
1.	11	10	0.07	0.80
2.	6	4	0.15	0.38
3.	10	6	0.3	0.6
4.	9	1	0.61	0.38
5.	7	2	0.38	0.34
6.	6	3	0.23	0.34
7.	6	1	0.38	0.26
8.	2	1	0.07	0.11
9.	3	3	0	0.23
10.	4	2	0.15	0.23
11.	6	5	0.07	0.42
12.	9	4	0.38	0.5
13.	8	3	0.38	0.42
14.	9	7	0.15	0.61
15.	1	2	-0.07	0.11
16.	8	1	0.53	0.34
17.	9	5	0.30	0.87
18.	10	4	0.46	0.53
19.	9	3	0.46	0.46
20.	4	3	0.07	0.26
21.	8	4	0.30	0.46
22.	8	2	0.46	0.38
23.	12	3	0.69	0.57
24.	8	0	0.61	0.30
25.	10	4	0.46	0.53
26.	9	5	0.30	0.53
27.	10	4	0.46	0.53
28.	9	4	0.38	0.5
29.	8	0	0.61	0.30
30.	10	4	0.46	0.53
31.	7	3	0.30	0.38
32.	10	2	0.61	0.46
33.	10	1	0.69	0.42
34.	9	3	0.46	0.46
35.	11	3	0.61	0.53
36.	10	2	0.61	0.46
37.	9	4	0.38	0.5
38.	7	5	0.15	0.46
39.	7	1	0.46	0.30
40.	8	2	0.46	0.38
41.	7	1	0.46	0.30
42.	9	5	0.30	0.53
43.	7	5	0.15	0.46

44.	7	2	0.38	0.34
45.	9	4	0.38	0.5
46.	10	7	0.23	0.65
47.	5	3	0.15	0.30
48.	6	1	0.38	0.26
49.	9	4	0.38	0.5
50.	9	4	0.38	0.5
51.	9	1	0.61	0.38
52.	5	2	0.23	0.26
53.	11	3	0.61	0.53
54.	9	1	0.61	0.38
55.	11	3	0.61	0.53
56.	3	1	0.15	0.15
57.	7	3	0.30	0.38
58.	9	3	0.46	0.46
59.	6	3	0.23	0.34
60.	11	2	0.69	0.5

The results of item analysis indicated that 44 items were at middle difficult level, 11 items were difficult, 3 items were very difficult and only 2 items were easy. On the other hand, 18 items were at reasonably fair discrimination level, 14 items were very good at discriminating, 11 items were ranked good and the rest had poor discriminating power. Those items having difficulty level between 0.20-0.80 and discrimination level between 0.30 and above was selected. Only 35 items were selected. So, the achievement test contain 35 analysis after piloting.

VALIDATION OF THE TEST

The last phase of development of a semi-standardized test is to measure its reliability and validity. The reliability of an instrument is its degree of consistency which is usually expressed as internal consistency. It is measured by Cronbach Alpha and the reliability coefficient for the test was 0.75 which is acceptable. It means the test is reliable. The validity of the test was already measured. The content and face validity of the test was measured while expert review.

Achievement Test

A semi-standardized achievement test of general methods of teaching was constructed by the researcher to analyze the academic achievement of the prospective teachers. Item analysis of each item of the test is presented below to analyze the quality of the items and the scores of prospective teachers and comparison between the scores of male prospective teachers and female prospective teachers was made to analyze their achievement.

Table 7: Discrimination power of all items

Item No.	High Achievers	Low Achievers	Discrimination Index= D	Quality	Recommendation
1	0.63	0.22	0.41	Good	Retain
2	0.49	0.18	0.31	Fair	Retain
3	0.56	0.18	0.39	Fair	Retain
4	0.51	0.12	0.39	Fair	Retain
5	0.62	0.18	0.44	Good	Retain
6	0.62	0.20	0.42	Good	Retain
7	0.28	0.09	0.19	Poor	Need to review
8	0.46	0.11	0.35	Fair	Retain

9	0.59	0.27	0.32	Fair	Retain
10	0.38	0.08	0.3	Fair	Retain
11	0.77	0.15	0.62	Excellent	Retain
12	0.53	0.18	0.35	Fair	Retain
13	0.74	0.09	0.65	Excellent	Retain
14	0.45	0.16	0.30	Fair	Retain
15	0.53	0.21	0.32	Fair	Retain
16	0.63	0.11	0.52	Excellent	Retain
17	0.47	0.12	0.35	Fair	Retain
18	0.50	0.20	0.30	Fair	Retain
19	0.49	0.15	0.34	Fair	Retain
20	0.54	0.14	0.40	Good	Retain
21	0.61	0.09	0.52	Excellent	Retain
22	0.50	0.19	0.31	Fair	Retain
23	0.52	0.14	0.38	Fair	Retain
24	0.19	0.14	0.06	Poor	Discard
25	0.71	0.23	0.48	Good	Retain
26	0.59	0.28	0.31	Fair	Retain
27	0.59	0.09	0.50	Excellent	Retain
28	0.36	0.06	0.30	Fair	Retain
29	0.64	0.31	0.33	Fair	Retain
30	0.51	0.20	0.31	Fair	Retain
31	0.65	0.11	0.54	Excellent	Retain
32	0.56	0.15	0.41	Good	Retain
33	0.50	0.15	0.35	Fair	Retain
34	0.49	0.16	0.33	Fair	Retain
35	0.55	0.20	0.34	Fair	Retain

Table 7 presents the discrimination power of all the items. It is shown that all the items were meeting the criteria of discrimination except item no. 7, and 24, these two items were not discriminating and needs to be reviewed.

Table 8: Difficulty Index of all the items

Item No.	No. of Students answered Correctly	P value	Quality	Recommendation
1	162	0.41	Good	Retain
2	154	0.39	Good	Retain
3	139	0.35	Good	Retain
4	134	0.34	Good	Retain
5	151	0.38	Good	Retain
6	141	0.35	Good	Retain
7	57	0.14	Very Difficult	Need to Revise
8	106	0.27	Good	Retain
9	166	0.42	Good	Retain
10	103	0.26	Good	Retain
11	148	0.37	Good	Retain
12	130	0.33	Good	Retain

13	140	0.35	Good	Retain
14	117	0.29	Good	Retain
15	136	0.34	Good	Retain
16	134	0.34	Good	Retain
17	139	0.35	Good	Retain
18	127	0.32	Good	Retain
19	111	0.28	Good	Retain
20	138	0.35	Good	Retain
21	132	0.33	Good	Retain
22	121	0.39	Good	Retain
23	127	0.32	Good	Retain
24	78	0.20	Difficult	Retain
25	179	0.45	Good	Retain
26	176	0.44	Good	Retain
27	115	0.29	Good	Retain
28	96	0.24	Good	Retain
29	179	0.45	Good	Retain
30	158	0.40	Good	Retain
31	119	0.30	Good	Retain
32	126	0.32	Good	Retain
33	113	0.28	Good	Retain
34	109	0.27	Good	Retain
35	136	0.34	Good	Retain

Table 8 presents the difficulty index of all the items of achievement test. It is evident from the table that all items that majority of items were good meaning they are neither too difficult nor too easy to attempt. Only item no 24 is difficult. The item no 7 was found very difficult and it must be revised.

Table 9: Distractor Analysis of all the items

Item No.	No. of respondents select A	% of respondents select A	No. of respondents select B	% of respondents select B	No. of respondents select C	% of respondents select C	No. of respondents select D	% of respondents select D
1	162	41%	86	22%	128	32%	24	6%
2	54	14%	154	39%	92	23%	100	25%
3	133	33%	139	35%	85	21%	43	11%
4	93	23%	134	34%	53	13%	120	30%
5	103	26%	151	38%	87	22%	59	15%
6	94	24%	123	31%	141	35%	42	11%
7	84	21%	122	31%	137	34%	57	15%
8	115	29%	65	16%	106	27%	114	29%
9	166	42%	91	23%	76	19%	67	17%
10	132	33%	103	26%	84	21%	81	20%
11	77	19%	148	37%	112	28%	63	16%
12	130	33%	82	21%	133	33%	55	14%
13	145	36%	140	35%	51	13%	64	16%
14	101	25%	110	28%	117	30%	72	18%
15	83	21%	101	25%	136	34%	80	20%
16	140	35%	134	34%	67	17%	59	15%
17	78	20%	139	35%	85	21%	98	25%
18	111	28%	122	31%	127	32%	40	10%
19	93	23%	111	28%	64	16%	132	33%
20	88	22%	138	35%	78	20%	96	24%

21	132	33%	92	23%	104	26%	72	18%
22	64	16%	121	31%	148	37%	67	17%
23	110	28%	127	32%	80	20%	83	21%
24	149	37%	78	20%	82	20%	88	22%
25	57	14%	179	45%	56	14%	108	27%
26	176	44%	88	22%	81	20%	55	14%
27	90	23%	115	29%	76	19%	119	30%
28	79	20%	88	22%	137	34%	96	24%
29	92	23%	179	45%	84	21%	45	11%
30	107	27%	158	40%	77	19%	58	15%
31	91	23%	106	27%	119	30%	83	21%
32	87	22%	126	32%	90	23%	97	24%
33	119	30%	106	27%	113	28%	62	16%
34	100	25%	111	28%	109	28%	80	20%
35	108	27%	136	34%	89	22%	67	17%

Table 9 presents the distractor analysis of all the items. All the distractors were functional having percentage above 5%.

Discussion

The findings of this study highlight the practicality and need for semi-standardized assessment models in methods of teaching courses for prospective teachers. The item analysis revealed that a majority of the test items demonstrated appropriate levels of difficulty and acceptable discrimination power, suggesting that the semi-standardized test reliably measured the intended cognitive domains.

The reliance on Bloom's Taxonomy in the test construction ensured that assessment items aligned with essential learning outcomes at the knowledge, comprehension, and application levels. However, the limitation to the lower three levels suggests a need for future assessments to encompass the higher-order skills of analysis, synthesis, and evaluation. Such expansion would provide a more comprehensive measure of critical thinking and instructional decision-making abilities, both of which are vital competencies for effective teachers.

Applying Classical Test Theory principles throughout the test development and validation process strengthened the reliability and validity of the instrument. Calculations of item difficulty and discrimination indices allowed the identification and improvement of weak test items, ensuring a fair evaluation process. The observed reliability coefficient (e.g., KR-20) indicated satisfactory internal consistency, affirming the technical soundness of the semi-standardized test.

Importantly, the academic achievement data showed that a significant proportion of prospective teachers scored below 30%, revealing potential gaps in knowledge and understanding that must be addressed through curriculum and instructional improvements. This finding emphasizes the role of assessment not merely as a tool for evaluation but as a diagnostic measure that informs teaching practices and supports student development.

Overall, the semi-standardization of tests offers an effective strategy for balancing the need for uniformity and flexibility in teacher education assessments. By establishing core content and psychometric quality standards while allowing contextual adaptation, semi-standardized tests can enhance both the equity and relevance of evaluation practices in teacher education programs.

Conclusion

The item analysis of the achievement test indicated that a majority of the items exhibited a moderate level of difficulty. Specifically, 60.5% of the items (24 out of 40) fell within the middle difficulty range, with difficulty indices between 0.31 and 0.70. Additionally, 28.5% of the items (10 items) were classified as difficult, showing difficulty indices between 0.20 and 0.30. Notably, only one item (3%) was identified as very difficult, with a difficulty index below 0.20.

In terms of the items' discrimination power; their ability to differentiate between high and low achievers; the analysis revealed that 57% of the items (20 items) demonstrated a reasonably fair discrimination level, with discrimination indices ranging between 0.30 and 0.39. Moreover, 17% of the items (6 items) showed very good discrimination, scoring 0.50 or higher. Another 17% (6 items) were categorized as good discriminators, falling within the range of 0.41 to 0.49. However, 9% of the items (3 items) were considered poor discriminators, having

discrimination indices below the acceptable threshold.

The analysis of prospective teachers' academic achievement scores further highlighted notable trends. Nearly half of the participants, 194 individuals (49%), scored between 30% and 70% on the test, reflecting moderate achievement. Conversely, 193 participants (48%) scored below 30%, indicating significant challenges in mastering the tested content. Only a small fraction, 13 participants (3%), achieved high performance by scoring above 75%. These findings suggest a generally moderate level of difficulty across the test items, with a reasonable capacity for discriminating between varying levels of student achievement, but also point to a need for further instructional and assessment improvements to better support prospective teachers' learning outcomes.

Recommendations for future researchers

- It is recommended that an additional achievement test be developed specifically for the Methods of Teaching course, with a focus on evaluating prospective teachers' learning outcomes at the higher cognitive levels of Bloom's Taxonomy.
- Furthermore, it is recommended that semi-standardized tests be constructed for other core subjects within teacher education programs.

REFERENCES

A., Akhtar, S., & Iqbal, Z. (2019). Standardization of educational tests: Procedures and best practices. Journal of Educational Research, 12(3), 245–259.

Allen, M. J., & Yen, W. M. (2002). Introduction to measurement theory (2nd ed.). Waveland Press.

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. Educational Researcher, 36(5), 258–267.

Bichi, M. H., & Talib, S. (2018). Types of achievement tests in educational settings. International Journal of Education and Research, 6(2), 77–88.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.

Cochran-Smith, M., Villegas, A. M., & Adams, K. L. (2015). The political economy of teacher education policy. Journal of Teacher Education, 66(2), 113–129.

Darling-Hammond, L. (2017). Teacher education and the quest for equitable schools. Educational Researcher, 46(4), 186–196.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (2013). Item response theory and its applications. Sage Publications.

Kelly, M. (2014). Measuring item discrimination in educational tests. Journal of Educational Measurement, 51(4), 437–452.

Linn, R. L., & Gronlund, N. E. (2008). Measurement and assessment in teaching (10th ed.). Pearson.

Michael, J., & Modell, H. (2003). Active learning in higher education. Journal of College Science Teaching, 32(5), 298–301.

Mitra, D. (2008). Student voice in school reform: Building youth-adult partnerships that strengthen schools and communities. Pedagogy and Education, 13(1), 1–21.

Mullis, I. V., Martin, M. O., & Foy, P. (2007). TIMSS 2007: International trends in mathematics and science achievement. TIMSS & PIRLS International Study Center.

Pond, M., & Rehan, M. (1997). The importance of student engagement in learning. Journal of Educational Psychology, 89(3), 395–402.

Popham, W. J. (2006). The pitfalls of assessment-driven instruction. Educational Leadership, 64(3), 25–29.

Popham, W. J. (2011). The value of assessment in teacher education. Journal of Teacher Education, 62(4), 399–411.

Roid, G. H. (2012). The development and application of standardized testing in educational settings. Psychological Testing and Assessment, 11(2), 142–151.

Rudner, L. M. (2011). Statistical analysis of test items. Educational Measurement, 14(1), 22–33.

Sequeira, E. (2012). Effective teaching strategies in higher education. Journal of Educational Research, 15(1), 33–45.

Swain, J., Smith, D., & Roberts, J. (2005). Standardized tests in education: A critical overview. Educational Testing, 39(1), 38–56.

Zeichner, K. (2010). Teacher education and the role of assessment in preparing teachers for diverse classrooms. Journal of Teacher Education, 61(1-2), 7–22