

Teacher-Guided Test Construction And Learners' Assessment Preferences: A Quasi-Experimental Study In Teacher Education

Kiran Thangjam, Aribam Dhaneshwari Devi
(Department Of Teacher Education, Manipur University, India)
(D.M. College Of Teacher Education, India)

Abstract:

This study investigated the effect of guided test construction on the preference for different forms of assessment tools in teacher education. 95 student-teachers participated in the quasi-experimental (pre-test and post-test non-equivalent group) design which were divided into 44 student-teachers as experimental group and 51 student-teachers as control group. The participants in the experimental group were given hands-on-activity based on blue print in making items and were given task to evaluate their peers' answer sheet according to the criteria provided for four weeks. On the other hand, participants of control group were given online lectures and reading material on the test planning, construction of blue-print and criteria for a good test. Pre-and post-Assessment preference questionnaires were used to test the significant effect of the test construction knowledge on both groups. A written test on planning and construction of test was conducted for both groups to check deep learning as a qualitative data. The result showed that test planning and construction assisted by the teachers significantly motivated the participants in preferring different forms of assessment tools with medium effect size. The findings also revealed that item making with the help of blueprints in classroom activities enhanced their knowledge of linking the different components of blue print in developing a particular question. The findings also showed that student-teachers of the treatment group gave constructive feedback while evaluating the peer response related to different forms of questions. The results of the study indicated that preferences for different assessment tools improved with the understanding and application of the test concepts in real classroom situations.

Key Word: Test Constructions, Assessment tools, Assessment preference, Constructive feedback.

Date of Submission: 21-05-2026

Date of Acceptance: 31-05-2026

I. Introduction

Assessment is the practices in which teachers intensively prepare themselves in judgement and improvement of the teaching learning process. Proper assessment tools with suitable methods improve the learning activities by diagnosing the constraints. The importance of different assessment procedure increases as different form of assessment changes (Mansory, 2020; Tosuncuoglu, 2018). Framing assessment tools like different forms of questions requires suitable hands-on activity in the preparation of student-teachers in Teacher Education Institutions. This helps to develop creativity and analytical skills while designing different forms of question aligned to the learning outcomes. Other skills like technological competencies are needed to be introduced in the curriculum along with its applications and usability (Søby & Tzu, 2015). The process of assessment in Finland is regularly conducted in the form of self-assessment under the supervision of the mentor (Voogt Joke & Kasurinen Helena, 2005).

The recommendations of NEP (National Education Policy) 2020 mention about the use of innovative assessment in the education system to enhance the teaching methods used for both offline and online mode. The NEP 2020 has also supported different innovative methods of assessment which are actively implementing in different parts of the worlds. The policy also looks in to the implementation gaps of different innovative assessment and hence support holistic transformation in the education system. The implementation of the policy depends on the process with careful planning and conducive environment for different assessment procedure (Pathak & Maity, 2024). This study focusses on the importance of activities related to the constructions of different assessment tools with suitable methods in teacher education institutions. The existing study investigates the effect of test planning, construction and evaluation on the preference of different forms of questions in a quasi-experimental setting. Furthermore, this study gives the linkage of theoretical and practical knowledge of making a test in the teaching process.

Research Questions:

1. Do assisted test constructions have significant effect on preference of different forms of question?
2. How do participants prefer different forms of questions before the intervention?
3. Is there a significant difference between participants in the control group and that of experiment group in their preference after the intervention?

Forms of Questions

Assessment in the education system is changing from traditional way of testing towards higher-order cognitive skills. Eldakhakhny & Elsamanoudy (2023) mentioned that the choice of Assessment tools (forms of questions) significantly differentiates the student study behaviours of memorization and creativity. Multiple-Choice Questions (MCQs) remain the assessment tools of high reliability, objectivity and efficiency (Desy et al., 2024). However, the cue spotted has a primary limitation where students recognize the correct answer without recalling (Janse et al., 2022). This often encourages a surface-level cognitive process but shows high reliability in scoring (Martinez, 1999). Despite these drawbacks, Palmer & Devitt (2007) argued that well-structured MCQs could assess higher-order thinking as compared to the modified essay questions. The quality of the MCQ test can be improved after post-exam review with modification of low-discrimination items resulting in more reliable test (Desy et al., 2024).

In viewing the limitations of MCQs, Very Short Answer Questions (VSAQs) which requiring responses of few words or sentence have replaced as an alternative form of Question. They retain the benefits of automated marking with proper coding, eliminating the cueing effect (Potter & McLachlan, 2025). Janse et al. (2022) and Potter & McLachlan (2025) emphasized mainly on VSAQs for higher discriminative power, differentiating high-performing students from low-performing students. Furthermore, Eldakhakhny & Elsamanoudy (2023) and Hyo-Jin Kwon (2023) found that Short Essay Questions correlate with the academic achievement of specific learning outcomes. The lower scores obtained in VSAs are viewed as a reflection of factual knowledge (Sher et al., 2026).

Studies related to Multiple-Answer Questions (MAQs) shows more complexity that can promote more deeper analytical thinking than MCQs (Oc & Hassen, 2025). Although students faced more difficulties in MAQs, this form of questions enhanced more understanding than MCQs. In a Retrieval Practice, Smith & Karpicke (2014) found that VSAs improve long-term retention while testing all forms of questions. On the other side, van Wijk et al. (2024) found that the learning outcomes mainly depends on the difficulty of initial retrieval in VSAQs.

Test construction

Test construction is a systematic process of developing instrument for assessing the learning outcomes. Previous studies revealed that the steps of test were designed to ensure the validity and reliability of the test measurements (Tanguma, 2000). The procedure of test constructions starts with defining the objective of the test and identifying the criteria to be measured (Khanal, 2020; Tanguma, 2000). The main task in planning phase is the development of a blue print (Table of Specifications) which maintains proper balance between the weightage given to the content, the types of objectives, forms of questions and difficulty level of the test item (Ing et al., 2015; Khanal, 2020; Tanguma, 2000). Ing et al. (2015) mentioned the importance of the objective of the item which aligned with the cognitive domain of Bloom's taxonomy in blueprints leading to the understanding of higher-order thinking skills. Tanguma (2000) mention that the structure of the item and grammar is taken proper care during the development of test item for minimising the errors. The difficulty level and time consumptions also vary accordingly with forms of questions especially between MCQs and essay type questions (Tanguma, 2000).

In scoring and evaluation phase rubrics are used for subjective response to maintain the consistency in the result (Singh et al., 2026). This quantitative evaluation for good item is determined by the difficulty index, which measures the proportion of correct responses, and the discrimination index (D), which measure the ability of an item to differentiate between high and low achievers (Kissi et al., 2023).

The existing study on preference of assessment tools largely focusses on the planning and construction of test. While several studies have given emphasis on psychometric property of the test and the theoretical concepts on the development of test and other suggest on the importance for practical or hands on activity particularly in real classroom situations.

II. Material And Methods

The study was conducted at Ibotombi Institute of Education, Canchipur affiliated under Manipur University. A quasi-experimental (pre-test and post-test non-equivalent group) design was used to find the effect of planning and construction of test on the preference of the different assessment tools between the two intact groups. In real-classroom setting, quasi-experiment is suitable as randomisation is not possible (Capili & Anastasi, 2024). The balance between internal and external validity is maintained in quasi-experiment (Handley et al., 2018). In this study, the experimental group were given exposure to the construction of Test and their preference on different assessment tools will be evaluated.

Sample:95 student-teachers from second-semester of B.Ed course involved in the present study using convenient sampling method. The participants were divided into two intact groups with 44 student-teachers as experimental group and 51 student-teachers as control group without randomization. There were 20 male student-teachers and 75 female students-teachers. On the basis of teaching experiences, most of them about 52 participants have no teaching experiences. There were 8 participants with 1 year teaching experience, 11 participants with 2 to 4 years of teaching experiences and only 4 of them have more than 5 years of teaching experiences. The participants were informed about the confidentiality of their response for ethical considerations.

Instrument

The Questionnaire of the study is adapted from Assessment Preferences Inventory (Birenbaum, 1994). It comprises of 12 items in which 2 items are related to MCQs, 3 items to Short Answer type Questions,5 items to Essay Types Questions and 2 items to the forms of objectives (Bloom's Taxonomy) with 5 points Likert-scale. The degree of preference is coded as follows 1 for "not at all", 2 for "To a small extent", 3 for "unsure", 4 for "To a certain extent" and 5 for "to a great extent". The items were found to be reliable with Cronbach's alpha coefficient of 0.784.

Procedure methodology

Before giving treatment to student-teachers of the experimental group, all the participants were given Assessment Preferences Inventory as a pre-test Questionaries through google form. The task of the experimental group was conducted in offline real classroom mode while that of the control were conducted in online mode. Firstly, the participants of the experimental group were exposed to the building of Blue Print assisted by the teacher and were given task for making items according to blue-print. On the other side, student-teachers of the control group were given lectures on the topic of Test planning and construction for four weeks 2 hours daily. Students of the Control group were given study material and online lectures as a placebo effect while the experimental group were given hands on activities in the classroom. A written test on planning and construction of test was conducted for both groups to check deep learning as a qualitative data. The questions of the examinations were based on the importance of blue print in the construction of a test. The student-teachers of experimental group were given task to evaluate their peers' answer sheet according to the criteria provided. The post-test questionnaires were administered after the completion of the programme.

Data Analysis

The responses of the 12 items were analysed using statistics data analysis method. Homogeneity of variance was evaluated using Levene's Test with p-value 0.762. Normality test was evaluated using Shapiro-Wilk's Test with p-value 0.071. Both p-values were found to be greater than 0.05 and hence both assumptions were satisfied for parametric test. To ensure the two groups were matched before the intervention independent t-test was used as illustrated in table no.1. Pair t-test was used to check the significant improvement from pre-test score to the post-test score as shown in table no. 2. Furthermore, ANCOVA was used to check the effect of Test Construction Knowledge on the preference of different assessment tools which is illustrated in table no. 3. For checking effect size cohen's d (t-test) and partial eta square η^2 (ANCOVA) were used. The different value's cohen's d were classified as small size (d=0.2), medium size(d=0.5) and large size (d=0.8) (Cohen, 2013)while that of partial eta square were classified as small size ($\eta^2=0.0099$), medium size($\eta^2=0.0588$) and large size ($\eta^2=0.1379$) (Richardson, 2011).

III. Results

Table no. 1 Independent t test result before intervention

Groups	N	Mean	SD	df	t-value	p-value	Effect Size (Cohen's d)
Experimental	44	48.9	4.19	93.0	0.282	0.779	0.0580
Control	51	49.1	3.91				

Table no. 2 Pair sample t-test result comparing pre-test and post-test.

Groups	N	Pre-test		Post-test		df	t-value	p	Effect Size (Cohen's d)
		Mean	SD	Mean	SD				
Experimental	44	48.9	4.19	51.3	4.12	50.0	2.72	0.009	0.381
Control	51	49.1	4.12	50.2	4.17	43.0	5.25	<.001	0.792

Table no. 3 ANCOVA test result for post-test score with pre-test score as covariates

	Sum of Squares	df	Mean Square	F	p	η^2
Overall model	923.2	2	461.60	58.20	<.001	
Group	43.8	1	43.81	5.59	0.020	0.027
Pre-test total	879.4	1	879.38	112.17	<.001	0.535
Residuals	721.3	92	7.84			

Table 1 shows, the preference of assessment tool for the experimental group ($M=48.9$, $SD=4.19$) is nearly equal to that of control group ($M=49.1$, $SD=4.12$). This implies that student-teachers from both groups are at the same level before giving intervention of test constructions. The results in table 2 shows significant rise in their preference of different forms of assessment tools from pre-test score to post-test score: experimental group ($p=0.009$, $d=0.381$) and control group ($p<0.001$, 0.792). Table 3 indicates the result from one-way ANCOVA that there was significant difference between the control group and the experimental group [$F(1,92)=5.59$, $p=0.020$, $\eta^2=0.027$]. This finding indicated that the group explain 2.7% of the variance in the post-test score representing small effect size.

IV. Discussion

Given the importance of every component of test construction, it is required to be given exposure in the real world of assessment and evaluation for making an effective learning. The findings shows that test construction with the help of the teacher significantly increases the preference of assessment tool for experimental group as compared to preference of the control group. The increase in the preference is also consistent with many previous studies. For instance, Hsu et al. (2023) found that attitude of student towards changes when they are exposed to different scenarios. In another study, training in writing MCQ examination motivates students to take part in future examinations with MCQs. (Kurtz et al., 2019). Although many students did not like the exposure to different forms of questions, they still learn more higher order skills than that of students with one form of question (Stanger-Hall, 2012). In the present study it is also found that students in the treatment group give constructive feedback they assess their answer sheet. But conducting training programme for students to give constructive feedback was not much useful in the improvement of the programme (Hajhamid & Somogyi-Ganss, 2021).

V. Conclusions

The study was conducted to fill the gap of practical implication of Test Construction activities over the theoretical perspectives of different forms of Questions. The results of the study indicate that preferences for different assessment tools improved with the understanding and application of the test concepts in real classroom situations. One of the limitations in this study is the exposure of traditional forms of assessment tools which is one of the main factors in which preference of assessment tools depends. Exposure to different forms of assessment tools is necessary to build the ability of the students in selecting the different forms of Questions (Van De Watering et al., 2008). Another limitation of the study is the use of only one treatment i.e construction and planning of a test. Further study is recommended on other treatment like item analysis or use of innovative assessment tools which will help the participants to understand more deeply about the different forms of assessment tools.

References

- [1]. M. Mansory, 'The Significance Of Non-Traditional And Alternative Assessment In English Language Teaching: Evidence From Literature', *International Journal Of Linguistics*, Vol. 12, No. 5, Pp. 210–225, Oct. 2020, Doi: 10.5296/IJL.V12I5.17782.
- [2]. I. Tosuncuoglu, 'Importance Of Assessment In ELT', *J. Educ. Train. Stud.*, Vol. 6, No. 9, Pp. 163–167, Aug. 2018, Doi: 10.11114/JETS.V6I9.3443.
- [3]. M. Soby And L. Tzu, 'Finnish Education System', *Nordic Journal Of Digital Literacy*, Vol. 2015, No. 2, Pp. 64–68, 2015, Doi: 10.18261/ISSN1891-943X-2015-02-01;WGROU:STRING: PUBLICATION.
- [4]. Voogt Joke And Kasurinen Helena, *Formative Assessment IMPROVING LEARNING IN SECONDARY CLASSROOMS*. 2005. Accessed: Aug. 07, 2025. [Online]. Available: [www.Sourceoecd.org](http://www.sourceoecd.org).
- [5]. J. S. Pathak And S. Maity, 'Innovative Assessment Practices Aligned With India's National Educational Policy-2020', *Asian Journal Of Education And Social Studies*, Vol. 50, No. 9, Pp. 241–253, Sep. 2024, Doi: 10.9734/AJESS/2024/V50I91584.
- [6]. B. Eldakhkhny And A. Z. Elsamanoody, 'Discrimination Power Of Short Essay Questions Versus Multiple Choice Questions As An Assessment Tool In Clinical Biochemistry', *Cureus*, Vol. 15, No. 2, Feb. 2023, Doi: 10.7759/CUREUS.35427.
- [7]. J. Desy Et Al., 'The Impact Of Systematically Repairing Multiple Choice Questions With Low Discrimination On Assessment Reliability: An Interrupted Time Series Analysis', *Can. Med. Educ. J.*, Vol. 15, No. 3, Mar. 2024, Doi: 10.36834/CMEJ.77596.
- [8]. R. J. Janse Et Al., 'Comparison Of Very Short Answer Questions And Multiple Choice Questions In Medical Students: Reliability, Discrimination, Acceptability And Effect On Knowledge Retention', *Medrxiv*, P. 2022.07.13.22277583, Jul. 2022, Doi: 10.1101/2022.07.13.22277583.
- [9]. M. E. Martinez, 'Cognition And The Question Of Test Item Format', *Educ. Psychol.*, Vol. 34, No. 4, Pp. 207–218, 1999, Doi: 10.1207/S15326985EP3404_2.
- [10]. E. J. Palmer And P. G. Devitt, 'Assessment Of Higher Order Cognitive Skills In Undergraduate Education: Modified Essay Or Multiple Choice Questions? Research Paper', *BMC Medical Education* 2007 7:1, Vol. 7, No. 1, Pp. 49–, Nov. 2007, Doi: 10.1186/1472-6920-7-49.
- [11]. H. G. Potter And J. C. Mclachlan, 'Assessing Medical Knowledge: A 3-Year Comparative Study Of Very Short Answer Vs. Multiple Choice Questions', *Med. Teach.*, Vol. 47, No. 10, Pp. 1669–1677, 2025, Doi: 10.1080/0142159X.2025.2496382.
- [12]. A. Sher Et Al., 'Very Short Answer Questions Vs. Multiple Choice Questions In Health Professions Education: A Systematic Review And Meta-Analysis', *BMC Medical Education* 2026, May 2026, Doi: 10.1186/S12909-026-09359-5.
- [13]. Y. Oc And H. Hassen, 'Comparing The Effectiveness Of Multiple-Answer And Single-Answer Multiple-Choice Questions In Assessing Student Learning', *Marketing Education Review*, Vol. 35, No. 1, Pp. 44–57, Jan. 2025, Doi: 10.1080/10528008.2024.2417106.

- [14]. M. A. Smith And J. D. Karpicke, 'Retrieval Practice With Short-Answer, Multiple-Choice, And Hybrid Tests', *Memory*, Vol. 22, No. 7, Pp. 784–802, 2014, Doi: 10.1080/09658211.2013.831454.
- [15]. E. V. Van Wijk, M. De Jonge, F. M. Van Blankenstein, R. J. Janse, And A. M. J. Langers, 'The Battle Of Question Formats: A Comparative Study Of Retrieval Practice Using Very Short Answer Questions And Multiple Choice Questions', *BMC Medical Education* 2024 24:1, Vol. 24, No. 1, Pp. 1547–, Dec. 2024, Doi: 10.1186/S12909-024-06538-0.
- [16]. E. G. B. Ph.D., 'Essential Steps In Constructing Commercial Tests', *The Journal Of Business Education*, Vol. 7, No. 5, Pp. 9–10, Apr. 1932, Doi: 10.1080/08832323.1932.10116891.
- [17]. J. Tanguma, 'Steps In Test Construction.', Apr. 2000.
- [18]. P. Khanal, 'Key Considerations In Test Construction, Scoring And Analysis: A Guide To Pre-Service And In-Service Teachers', *International Journal Of Research Studies In Education*, Vol. 9, No. 5, Jul. 2020, Doi: 10.5861/IJRSE.2020.5027.
- [19]. L. M. Ing, M. B. Musah, S. H. V. Al-Hudawi, L. M. Tahir, And N. M. Kamil, 'Validity Of Teacher-Made Assessment: A Table Of Specification Approach', *Asian Soc. Sci.*, Vol. 11, No. 5, P. P193, Feb. 2015, Doi: 10.5539/ASS.V11N5P193.
- [20]. G. Singh, R. Singh, And N. Ananthkrishnan, 'Assessing The Assessments: Quality Metrics For Short Answer Questions And Essay Type Questions', *Journal Of Marine Medical Society*, Feb. 2026, Doi: 10.4103/JMMS.JMMS_113_25.
- [21]. P. Kissi, D. Baidoo-Anu, E. Anane, And R. K. Annan-Brew, 'Teachers' Test Construction Competencies In Examination-Oriented Educational System: Exploring Teachers' Multiple-Choice Test Construction Competence', *Front. Educ. (Lausanne)*, Vol. 8, P. 1154592, Apr. 2023, Doi: 10.3389/FEDUC.2023.1154592/TEXT.
- [22]. B. Capili And J. K. Anastasi, 'An Introduction To The Quasi-Experimental Design (Nonrandomized Design)', *Am. J. Nurs.*, Vol. 124, No. 11, P. 50, Nov. 2024, Doi: 10.1097/01.NAJ.0001081740.74815.20.
- [23]. M. A. Handley, C. R. Lyles, C. Mcculloch, And A. Cattamanchi, 'Selecting And Improving Quasi-Experimental Designs In Effectiveness And Implementation Research', *Annu. Rev. Public Health*, Vol. 39, No. Volume 39, 2018, Pp. 5–25, Apr. 2018, Doi: 10.1146/ANNUREV-PUBLHEALTH-040617-014128/CITE/REFWORKS.
- [24]. M. Birenbaum, 'Toward Adaptive Assessment — The Student's Angle', *Studies In Educational Evaluation*, Vol. 20, No. 2, Pp. 239–255, Jan. 1994, Doi: 10.1016/0191-491X(94)90011-6.
- [25]. J. Cohen, 'Statistical Power Analysis For The Behavioral Sciences', *Statistical Power Analysis For The Behavioral Sciences*, Pp. 1–567, Jan. 2013, Doi: 10.4324/9780203771587/STATISTICAL-POWER-ANALYSIS-BEHAVIORAL-SCIENCES-JACOB-COHEN/RIGHTS-AND-PERMISSIONS.
- [26]. J. T. E. Richardson, 'Eta Squared And Partial Eta Squared As Measures Of Effect Size In Educational Research', *Educ. Res. Rev.*, Vol. 6, No. 2, Pp. 135–147, Jan. 2011, Doi: 10.1016/J.EDUREV.2010.12.001.
- [27]. J. L. Hsu, N. Clark, K. Hill, And M. Rowland-Goldsmith, 'Investigating The Influence Of Assessment Question Framing On Undergraduate Biology Student Preference And Affect', *CBE Life Sci. Educ.*, Vol. 22, No. 4, Dec. 2023, Doi: 10.1187/CBE.22-12-0249;REQUESTEDJOURNAL:JOURNAL:LSE;PAGE:STRING:ARTICLE/CHAPTER.
- [28]. J. B. Kurtz, M. A. Lourie, E. E. Holman, K. L. Grob, And S. U. Monrad, 'Creating Assessments As An Active Learning Strategy: What Are Students' Perceptions? A Mixed Methods Study', *Med. Educ. Online*, Vol. 24, No. 1, P. 1630239, Jan. 2019, Doi: 10.1080/10872981.2019.1630239.
- [29]. K. F. Stanger-Hall, 'Multiple-Choice Exams: An Obstacle For Higher-Level Thinking In Introductory Science Classes', *CBE—Life Sciences Education*, Vol. 11, No. 3, Pp. 294–306, Sep. 2012, Doi: 10.1187/CBE.11-11-0100.
- [30]. B. Hajhamid And E. Somogyi-Ganss, 'Improving Effectiveness Of Dental Students' Feedback And Course Evaluation', *J. Dent. Educ.*, Vol. 85, No. 6, Pp. 794–801, Jun. 2021, Doi: 10.1002/JDD.12548;REQUESTEDJOURNAL:JOURNAL:19307837;WGROUP:STRING:PUBLICATION.
- [31]. G. Van De Watering Et Al., 'Students' Assessment Preferences, Perceptions Of Assessment And Their Relationships To Study Results', *Higher Education* 2008 56:6, Vol. 56, No. 6, Pp. 645–658, Feb. 2008, Doi: 10.1007/S10734-008-9116-6.