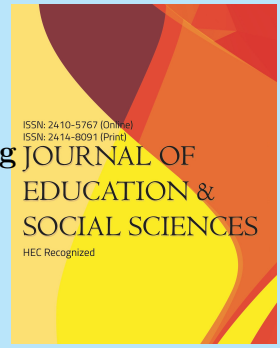
Journal of Education & Social Sciences

ISSN: 2410-5767 (Online)
ISSN: 2414-8091 (Print)



Teacher- Made Test Design and Uses: Exploring Teachers' Practices in Pakistan

Affiliation:

Bashir Uddin
Student, Master of Education Programme, The Aga Khan University Institute for Educational Development, Karachi, Pakistan.
Email: bashir.uddin@akesp.org

Razia Fakir Mohammad
Assistant Professor, Aga Khan University Institute for Educational Development, Karachi, Pakistan. Email: razia.fakir.m@gmail.com

Manuscript Information
Submission Date: July 14, 2021
Reviews Completed: October 19, 2021
Acceptance Date: October 29, 2021
Publication Date: November 10, 2021

Citation in APA Style:
Uddin, B., & Mohammad, R. F. (2021). Teacher- Made Test Design and Uses: Exploring Teachers' Practices in Pakistan, *Journal of Education & Social Sciences, 9*(2), 88-98.

DOI: https://doi.org/ 10.20547/jess0922109202

.

Journal of Education & Social Sciences
Vol. 9(2): 88-98, 2021
DOI: 10.20547/jess0922109202

Teacher- Made Test Design and Uses: Exploring Teachers' Practices in Pakistan

Bashir Uddin * Razia Fakir Mohammad †

Abstract: *Teachers' practices of achievement tests design have not been systematically explored in context. The current qualitative case study investigated the teachers' perceptions and practices of designing and using achievement tests at elementary level (Grade VII and VIII). The participants were four science teachers selected from two schools. Data was collected mainly by conducting semi-structured telephonic interviews and document collection. The findings revealed that teachers' practices of test design were not linked with their existing perceptions. There were no explicit criteria of achievement test design and using the assessment evidence. The findings also showed limited knowledge of achievement test design among the teachers. Moreover, it was noticed that teachers do not have exposure to reflect on their practices of test design. That is why teachers lack the essential assessment literacy and skills of developing items for tests. The findings suggest for especial considerations towards enhancing the teachers' assessment literacy.*

Keywords: Achievement test, alignment, summative assessment.

Theoretical Framework

The emphasis of existing literature is mostly on formative purpose of assessment which aims to improve instructional practices (Black & Wiliam, 2018; Wiliam, 2011; McCallum & Milner, 2021). Although formative assessment is crucial for enhancing teaching-learning as an ongoing process however, the role of summative assessment is also crucial to draw valid and fair inferences about students' progress towards the end of an instructional period either month, term, or year (Brookhart, 1999). Teachers are involved in designing and implementing the summative assessment practices at school level hence, summative assessment is based on teachers' conceptions of the assessment principles. Literature indicates that the tests, whether teacher–made or large–scale standardised achievement tests, have visible effects on decisions taken to promote students to next grades, their motivation towards continuity of education as well as their decisions about selection of specialised subjects for further learning (Rind & Malik, 2019). Usually, students of elementary grades require to take summative tests which provide evidence of their achieve-

*Student, Master of Education Programme, The Aga Khan University Institute for Educational Development, Karachi, Pakistan. Email: bashir.uddin@akesp.org
†Assistant Professor, Aga Khan University Institute for Educational Development, Karachi, Pakistan. Email: razia.fakir.m@gmail.com

Received July 14, 2021; Received in revised form October 19, 2021; Accepted October 29, 2021; Available online November 10, 2021

ment to their parents, teachers, and school. Since these tests primarily intend to measure achieved students' learning outcomes for the purpose of promotions, placement and/or certification (Fautley & Savage, 2008) they are termed as achievement tests. The achieved evidence of students' learning also guide to drive the educational changes in terms of decision making about reforms in curricular contents, teachers' trainings and measuring the accountability of stakeholders involved in the formal education system (Leighton & Gierl, 2011). Tests, to measure students' progress and achievement are considered helpful in reinforcing instruction to "provide a better foundation for self-assessment, both for students as well as for their teachers and schools" (Geiser, 2009). However, the current literature on test design reports issues about fairness, authenticity and validity of inferences drawn about students' achievements (Harlen & James, 1997). Moreover, limited assessment literacy among the teachers to design quality tests has been reported as the main issue with the achievement tests (Zolfaghari & Ahmadi, 2016).

It is imperative for the teachers to be skilful in assessment tasks and they should know the process involved in test design because the test designer's vocabulary selection for test items and technical terms involved affect the quality of the test (Crisp, Johnson, & Constantinou, 2019). Because choosing item format for designing quality test is a complex process, and the broad principles must construct the main considerations in assessment programmes. Therefore, teachers need to develop deeper understanding of subject content, pedagogical knowledge and assessment information for effective test construction otherwise the test will not ensure the validity, reliability and fairness of judgments made based on the students' responses in tests. One of the fundamental aspects to test design is selecting and writing the test items. Nevertheless, the teacher-made tests are unreliable because of the teachers' lack of skills in developing proper items for the test. One of the comebacks to this issue is the knowledge to draw Table of Specifications (TOS) or test blueprint. The study of Notar, Zuelke, Wilson, and Yunker (2004) illustrate that TOS helps the teachers to align objective, instruction, and assessment. Moreover, Fives and DiDonato-Barnes (2013) consider the use of TOS by teachers to be helpful for developing summative assessments that are "well aligned to the subject matter studied and the cognitive processes used during instruction" (p. 1). Hence this paper discusses, how teachers, at elementary level in Pakistan, design the test.

Methodological Approach

A qualitative case study method (Creswell & Poth, 2016; Merriam & Tisdell, 2015) was adapted to entail the insights of teachers' practices about designing school–based achievement tests in science classrooms. Four science teachers of elementary grades participated in this research. After seeking consent, online interviews were conducted with each four participants to know about their practices of designing the achievement tests. Additionally, documents such as test papers, sample of teachers 'marking, and science curriculum were gathered through emails and WhatsApp attachments.

Teachers' Practices of Designing Test

Overall, all the teachers perceived assessment as a method to gather information about their students' academic progress. The findings suggested that the teachers monitor and examine students' learning mainly through paper-pencil tests.

Figure 1
Sample of teacher-made test

Q 1. Choose the correct options.

1. During which phase of mitosis the chromosomes attached with the spindle fibre and arrange themselves at the equator of the cell.
   a) prophase. b) metaphase. C) anaphase. D) telophase
2. the process of crossing over occur during.
   a) Prophase. B) metaphase. C) anaphase. D) telophase
3. _____ is considered as powerhouse of the cell
   A) Mitochondria. B) Golgi bodies. C) nucleus. D) cell membrane.
4. Replication of DNA occurs during.
   A) G1. B) G2. C) S. d) prophase.
5. Which of the following is the largest part of brain.
   A). cerebrum. B) cerebellum. C) medulla.
6. Insulin is secreted by.
   A) Adrenal gland. B) pancreas. C) thyroid gland. C) liver
7. The process of photosynthesis takes place in which part of the cell.
   A) nucleus. B) mitochondria. C) chloroplasts. D) cell wall
8. the transfer of characters to offspring is called.
   A) heredity. B) variation. C) gamete and zygote
9. If AA is crossed with aa then all the offsprings will be.
   A) Aa. B) AA. C) aa. D) both A and B
10. If 2n = 26 for a particular cell, then the chromosome numbers in the egg cell after meiosis would be.
    A) 26. B) 8. C) 16. D) 13.

Q 2. Define the following.

1. Stimulus. 2. Sensitivity. 3. Reflex action. 4. Heredity. 5. Sensory neuron

In the field of science, assessment is to judge or evaluate what you have taught in science. . . how the students are progressing . . . this you [as a teacher] can find out either

your goals in science are achieved or not. (Interview B, July 5th, 2020) The cognitive learning outcomes of students, both overall, as well as from individual units are presented in the science curriculum however, the teachers did not mention explicitly about the learning outcomes. Since their teaching was textbook content oriented, their assessment mainly focuses on measuring the extent of the knowledge that students remembered. Hence the test items were aligned with the teacher's teaching approach but not holistically with the learning outcomes stated in the curriculum.

Together with developing scientific literacy among students, the goal of teaching science up to Grade VIII mentioned in the National Curriculum of Science (2006), is to "enable students to use science and technology to acquire new knowledge and solve problems" (p. 2). The teachers found test as a tool to force students' learning. For example, the teacher in his interview revealed "if we do not give tests then students will not learn". However, it has negative consequences as well, since testing all the time could promote a culture of teaching and learning to test. The literature indicates that when a test is in place regularity, it becomes so deeply rooted in the culture of classrooms that it becomes part of that culture.

The document analysis also indicated the test items were weakly designed mostly focusing on students' lower level thinking ability with most of the items of memorization level, only a few items of understanding/comprehension level, while no single item was noted that could assess the students' analytical skills and ability of creating new knowledge from their classroom learning. The test items were randomly selected from the content in the textbook that would ask the students to define, describe or explain, rather than analyse, interpret, evaluate or create knowledge from learning of scientific opinions and phenomena.

Marking the Test and Use of Students' Responses

The findings revealed that the teachers would not use any explicit criteria, marking scheme and/or rubrics for marking the students' responses to constructed responses items given in the test. The teachers would assess the students' responses as per the textbook knowledge, and/or the content provided to the students in classrooms. The interviews revealed that they knew the answers because of their long teaching experiences and using of science textbooks "there is no particular criteria for this [marking], I have taught the book for many years, I know the answer and if there is difficulty in assigning marks then I refer to the textbook" (Interview C, 8th July 2020).

Additionally, the criteria used for marking and assessing students' responses were implicit and could not have been reflected in the test question and/or discussed with the students. For example, the teacher revealed in his interview "Sir, if any student has elaborated by example then I give more marks to him, and if explained by diagram then I provide more marks so in this way I mark the test and there is no special criteria" (Interview C, 8th July 2020). The teachers' responses indicated implicit criteria for marking the students' work however, it was not clear whether or not the students were aware of the teachers' expectations. This data also showed that one teacher had implicit criteria in mind, while the other teacher used the best response as a norm to judge other students'

work. It is important to recognize that teachers' practices of marking the students' responses could not be analysed. Hence, the analysis showed that since the teachers would not consider any specification or criteria to design the test, they appeared ignorant of using the rubric and marking scheme. Mainly the questions were taken from the textbooks and the responses were assessed through the texts in the book. However, a set of expected elements should be considered while marking students' responses and this is guided by the marking scheme. But the teachers were unaware of the use of marking schemes. Thus, it can be concluded that assigning marks to students' responses was not guided by uniform criteria but based on teachers' own will and self-controlled estimation.

This raises an issue of drawing valid inferences about students' performance. Because these tests are of high stake since their results are used to make decision about students' promotion or failure. If the students receive low marks or fail the examination, students are not aware of the criteria marked by the relevant subject teachers that what exactly they need to do, while the results are used for the promotion and/or failure of students. Without using a marking scheme, marking a test referring to the content from the textbook may not ensure fair and accurate marking to better performance and poorer responses. Because the valid judgment of students' responses demands the effective use of normative/criterion judgment. But this practice was completely ignored in assessment practice of teachers.

The data informed that students' responses are mostly used for documentation, recording and reporting purposes as well as to make judgment about the promotion of students to next grades. They would check the students' responses in mid-term and annual examinations and share their judgments with them in numerical forms/marks or grades. The teachers find it difficult to reach to every student with feedback and suggestions based on their performance in the tests as illustrated:

Sir, first we check the papers, then we make a list of these results in a paper with the name of the child, then we prepare an award list of the whole subject in a large size paper . . . we keep the papers with us, and we share the results with the students . . . if there is some mistake in marking then we review the marking. (Interview D, July 10, 2020). Further, the teacher was also of the view that "the mid-term and annual exams are meant to promote children to next levels by assigning grades and marks to them" (Interview D, 10th July 2020).

The teachers did not see further uses of the test responses such as examining students' learning gaps, providing feedback and instructional design to fill/minimize the learning gap. The findings revealed that they would store the test papers for their reference or demonstration of accountability to show their efforts to monitor the students' progress. No proper feedback is provided to students for further improvement based on their performance in achievement tests. The mid-term and annual assessments are meant mainly to promote students to next grades or retain them in the same grade. These tests are not meant to provide any kind of feedback to students except the numerical grades/marks. Grades alone would not disclose the study needs of students because students do not look beyond the numbers/grades (Black & Wiliam, 2010). Thus, the main purpose of assessment that is to bring improvement in teaching and learning is ignored here. The teachers take assessment/tests as a means of judgment rather than making remedial decisions to

bring improvement in further learning.

Discussion

The analysis shows that the teachers lack the essential knowledge about standard procedures of constructing, marking, scoring and grading of tests as also discussed in Tshabalala, Mapolisa, Gazimbe, and Ncube (2016). Assessment is centrally important in an education system, and universally serves to approve students' learning. Classroom assessment plays a vital role in enhancing student's learning and achievement. Extensive research exists which critique on the summative role of existing assessment practices. Out learning is that teachers' practices of traditional assessment, such as paper pencil tests, promote students' expectations to acquire certain knowledge and skills around the course of their studies and mainly determines who is granted a privilege such as learned or not learned but is not congruent with or does not address the mandate to support meaningful learning as per standard goals of science education. Construction and reconstruction of the nature and use of assessment from monitoring/controlling learning to facilitating/expanding learning simultaneously are the keys to reform learning practices (Black & Wiliam, 2010). However, the reflection on current data indicates that there is lack of clarity about how to design a good test. The teachers need to be equipped with processes and theoretical underpinning of test design, marking and interpretation of test scores. Attention should be given to reform the way toward summative assessment, since the results are used about students' learning which has implications for their continuity of learning, admission and certification. It is need of the hour to rethink and modify summative assessment in a creative and logical way to make it more useful. Efforts are required to highlight the purpose and make due connections between formative and summative assessment that are necessary to preserve the true essence of assessment in education system. Additionally, government and education department must pay attention towards teachers' professional development to include training in different types of assessment skills to make it practically applicable in education system.

The overall findings argue that the teachers' assessment practices (in this case test design) are limited and weak. The analysis of the findings indicate that teachers possess limited knowledge about assessment. The findings also illustrated that teachers perceive achievement tests as a means of collected indications about students' learning that either the students have attempted the questions correctly or vice versa. In the field of science, assessment is to judge or evaluate what you have taught in science. . . how the students are progressing . . . this you [as a teacher] can find out either student has learned the topic taught or not. (Interview B, 5th July 2020)

These findings are aligned with the research which indicates issues of the real use of the assessment information. It has been a global dilemma about teachers' deeper understanding of summative assessment too, where teachers get involved in the mechanics and administration of tests rather than its design and analysis (Popham, 2003). The findings also exemplify that teachers' knowledge about assessment is limited to only collecting evidence about students' learning. The achievement test results are not used to know

what students know (Pellegrino, Chudowsky, & Glaser, 2001). Teachers do not see further uses of the assessment (tests) such as providing constructive feedback to the students, analysing the evidence to identify the learning gaps, and using the collected evidence for guiding the students on how to close the learning gaps. The collected evidence were not used for setting goals to students for future learning, identifying what they have done right and what they need to work on, so that students may develop insights into themselves to consider themselves responsible as a learners. This might be due to lack of teachers' essential knowledge about the purpose and significance of assessment.

Due to lack of fundamental awareness about assessment practices, test design, its purposes and uses, the teachers' practices of achievement tests not only indicate the issue of alignment between curriculum standards and assessment but also marking of the students' responses on tests do not appear to go parallel with the criteria guided by literature. There is no evidence if the teachers know the explicit purpose of tests and/or get engaged in systematic processes of test design and its validation. The analysis depicted that factual questions from the textbooks and/or previous papers are selected randomly.

Moreover, the assessment practices did not perform to meet the standards of the National Curriculum for Science (2006) which focus on the balance of lower and higher cognitive levels of students' learning outcomes to be assessed. The teachers' practices remain to assess mainly the memorization of knowledge and factual information about students' learning in science. It can be analysed that assessment is practiced at surface level and the means of assessment is more for reporting rather than bringing improvement in the learning standards. According to the benchmarks mentioned in the Education Policy (2017) Pakistan, students up to grade VIII level are expected to develop their analytical and creative skills on scientific concepts, phenomena, and opinions. But the findings did not demonstrate any evidence in assessment (tests) that could judge the analytical and creative skills of students. The focus of assessment was to measure the students' lower cognitive abilities mostly referred to the textbook. These findings are in line with the study of Rind and Malik (2019), who analysed the English examination papers of grade 10 and 12 under BISE (Board of Intermediate and Secondary Education) Sukkur and their study findings showed that majority of the Multiple-Choice Items (MCQs), Constructed Response Items (CRQs) and Extended Response Items (ERQs) were constructed in such a way that focus the lower cognitive order of Bloom's taxonomy. This indicates an overall issue of achievement test designing in Pakistan. Moreover, these findings do not meet the benchmarks of the National Curriculum for General Science IV–VIII (2006), which expects the assessment of both the lower and higher cognitive abilities of students' erudition.

Assessment is viewed more as an administrative and mechanical task since teachers' focus on assessment is to produce good annual results and to satisfy the parents by producing good results. Due to this reason, teaching to test has become the habit of teachers. Previous studies have also noticed the teaching to test approach being followed in most of the schools in Pakistan (Rehmani, 2012). The teachers were unaware of the use of marking schemes while assigning marks to students' responses on tests. Thus, it can be concluded that marking the students' responses was based on implicit criteria that would not ensure fairness in marking.

The MCQs mentioned in the collected documents did not show proper relationship

between stem, key and distractors. The items mentioned in the MCQs section from the collected documents showed issues of setting plausible distractors and their consistency with the stem. Similarly, the question items mentioned in the CRQs and ERQs sections indicated that students were not clearly informed of the teachers' expectations. Hence, assessment of students' learning was mainly brought about superficially. Whereas education in the era of 21st century skills demand well designed tests, that help the students to practice their analytical skills and problem-solving capabilities to ensure the appropriate level of cognitive processing (DiDonato-Barnes, Fives, & Krause, 2014). Moreover, Fives and DiDonato-Barnes (2013), suggest the use of TOS as a planning tool for teacher-made tests. Because TOS not only guides the teachers to align objectives, instruction, and assessment, but also helps them to identify items for the tests bearing in mind the different cognitive abilities of students learning like analysing information, evaluating a specific opinion, or creating something new (Kitiashvili, 2014). The findings showed that this process is not being practiced effectively in teacher-made tests. This has also been reported by Khattak (2012), that the Pakistani state schools are poorly designing test papers that discourage many good students. If the focus of assessment (in this case tests) is to assess the factual knowledge ignoring the higher cognitive levels of students learning, then in such case, not only the root cause of assessment is mislaid, but the assessment also fails in directing improvements in students' learning and instructional reforms.

These findings raise serious issues related to assessment practices (achievement tests) that are mostly used to make decisions about students' performance, progression, and promotions. The teachers did not use any explicit criteria to mark and/or evaluate students' work. This could raise an issue of misinterpretation and misjudged valuation of students' learning. Perhaps this may not have implications for teachers but have a long-term impact on students' motivation in learning and their future education. It can be analysed that assessment is a highly responsible phenomenon however, the teachers remain unaware of its significance for students' learning. Unless the significance, validity and reliability of the tests are not considered, the fair judgement of students' performance may not be ensured and thus the assessment practices analysed in the present study have issues of fairness and validity.

Thus, it can be analysed that the authenticity of students' learning becomes questionable. The decision guided by the evidence gathered through these tests becomes questionable either they inform students' learning or not. Teachers neither have enough time to carefully design these tests nor they are sensible to consider their responsibility of bringing improvement in learning through the collected evidence. Therefore, assessment for the sake of assessment has a more negative impact on both students' learning and teachers' own potential to identify the learning needs of students and overcome the challenges. The findings revealed that achievement tests employed by teachers for judgment of students' learning are not in the line of literature. Teachers' perceptions and their practices reflected the assessment (achievement tests) to be weakly constructed without focusing the purpose of assessing students' higher cognitive skills, scientific literacy, and technological advancement (National Curriculum 2006; National Education Policy, 2009; 2017).

Moreover, there is no evidence of engaging teachers in effective assessment practices therefore, teachers mainly go through the textbook to design and mark the tests. Referring

to the textbook from cover to cover to judge students' progress may not fulfill the curriculum standards and the real-world expectations of teaching and learning science. Teachers' use of the textbook frequently for designing and marking the science tests shows the lack of assessment literacy among the teachers. Thus, the findings also revealed that teachers are not assessment literate but assessment illiterate. Popham (2003), considers the assessment illiteracy of teachers a "prescription for professional suicide" (p. 82). The data showed poor assessment literacy among teachers based on their perceptions and practices of the achievement tests since none of them had any learning exposure in the area of classroom assessment and/or test design. To become assessment literate, teachers are required to develop diverse assessment practices and skills (Stiggins & Chappuis, 2005), because sound assessment and rigorous grading help promote the students' motivation towards learning and increasing their achievement levels (Brookhart, 1999; White, 2009). Earlier studies in Pakistani context have also reported the lack of training to enhance the assessment literacy among teachers. Khattak (2012), reports on assessment in Pakistani schools that due to lack of proper training, teachers test only the information, and the features of the whole personality are ignored. That is why the focus of assessment techniques employed by schools is limited to judge the memory of students and the effective psychomotor domains are not assessed. Moreover, Hussain, Shaheen, Ahmad, and Islam (2019), have reported that majority of the teachers in Pakistani context are "untrained in classroom assessment practices which is considered the most devastating element of teacher's professional life" (p. 94). Hence, teachers' assessment skills are based on trial and error methods which have developed falsified beliefs about assessment that result in negative wash back of students' learning. A similar report has also been presented by Kitiashvili (2014), that in Georgia's general educational institutions there is lack of Professional Development (PD) trainings on teachers' assessment literacy that is one of the main barriers in teachers' assessment practices.

Conclusion

Assessment is increasingly imperative for teaching-learning since the evidence gathered are used for drawing valid inferences about students' learning and making instructional decisions. Summative assessment involves collection, analysis and uses of the students' responses to report their performance and progress usually over a certain instructional period. In Pakistani context, school-based achievement tests are the main source of collecting evidence to report students' learning through summative practices. Although limited empirical studies are present on summative assessment which have reported the summative practices to judge students' lower cognitive learning, memorization and focusing the grading/marking. It was crucial to find out either the inferences drawn, and judgment made about the students' performance based on the achievement tests are valid or not. Therefore, this study aimed to explore teachers' practices of designing and using achievement tests at elementary level in the rural context of Pakistan. This study provided the contextual realities regarding assessment practices (test design) and related issues in context. The findings have significance for teachers and the school management regarding

summative assessment practices (achievement tests) and suggest for thoughtful considerations to be paid towards teachers' assessment literacy and skills development in tests design and marking.

References

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, Policy & Practice*, *25*(6), 551–575.

Brookhart, S. M. (1999). *The art and science of classroom assessment. the missing part of pedagogy.* ERIC.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Crisp, V., Johnson, M., & Constantinou, F. (2019). A question of quality: Conceptualisations of quality in the context of educational test questions. *Research in Education*, *105*(1), 18–41.

DiDonato-Barnes, N., Fives, H., & Krause, E. S. (2014). Using a table of specifications to improve teacher-constructed traditional tests: an experimental design. *Assessment in Education: Principles, Policy & Practice*, *21*(1), 90–108.

Fautley, M., & Savage, J. (2008). *Assessment for learning and teaching in secondary schools*. Learning Matters.

Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research, and Evaluation*, *18*(1), 1-7.

Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change: The Magazine of Higher Learning*, *41*(1), 16–23.

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365–379.

Hussain, S., Shaheen, N., Ahmad, N., & Islam, S. U. (2019). Teachers' classroom assessment practices: Challenges and opportunities to classroom teachers in Pakistan. *Dialogue*, *14*(1), 88.

Khattak, S. G. (2012). Assessment in schools in pakistan. *SA-eDUC Journal*, *9*(2).

Kitiashvili, A. (2014). Teachers' attitudes toward assessment of student learning and teacher assessment practices in general educational institutions: The case of georgia. *Improving Schools*, *17*(2), 163–175.

Leighton, J. P., & Gierl, M. J. (2011). *The learning sciences in educational assessment: The role of cognitive models*. Cambridge University Press.

McCallum, S., & Milner, M. M. (2021). The effectiveness of formative assessment: student views and staff reflections. *Assessment & Evaluation in Higher Education*, *46*(1), 1–16.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). The table of specifications: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, *31*(2).

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* ERIC.

Popham, W. J. (2003). Seeking redemption for our psychometric sins. *Educational Measurement: Issues and Practice*, *22*(1), 45–48.

Rehmani, A. (2012). Changing assessment practices in Pakistani schools: A case of AKU-EB middle school assessment framework. Retrieved from `https://ecommons.aku.edu/cgi/viewcontent.cgi?article=1050&context=books`

Rind, I. A., & Malik, A. (2019). The examination trends at the secondary and higher secondary level in pakistan. *Social Sciences & Humanities Open*, *1*(1), 1-7.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, *44*(1), 11–18.

Tshabalala, T., Mapolisa, T., Gazimbe, P. G., & Ncube, A. C. (2016). Establishing the effectiveness of teacher-made tests in nkayi district primary schools. *Nova Journal of Humanities and Social Sciences*, *4*(1).

White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal*, *3*(1), 3–25.

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

Zolfaghari, F., & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, *3*(1), 1-16.