

Teacher-Designed Assessment Banks and Adaptive Testing

Vivek Nair

Independent Researcher

Kerala, India

ABSTRACT

This manuscript examines the integration of teacher-designed assessment banks with adaptive testing frameworks to enhance personalized learning, assessment accuracy, and instructional efficacy. Teacher-designed assessment banks comprise repositories of carefully crafted items aligned with curricular objectives, reflecting educators' deep understanding of learner needs and pedagogical goals. Adaptive testing, powered by algorithmic item selection based on examinee performance, dynamically adjusts difficulty to optimize measurement precision and learner engagement. By combining these two approaches, educators can deliver assessments that are both contextually relevant and psychometrically robust.

Building on foundational theories of item response theory (IRT) and computerized adaptive testing (CAT), this study employs a mixed-methods investigation involving 150 secondary school teachers and 600 students across mathematics, science, and language arts. Teacher-authored item pools were developed, calibrated, and field-tested within a custom CAT platform. Quantitative analyses reveal that adaptive tests drawing from these banks achieved high reliability (marginal reliability coefficients ≥ 0.90), strong convergent validity with state achievement measures ($r = .86-.89$), and required approximately 40% fewer items to reach precision thresholds compared to fixed-form tests. Qualitative feedback indicates that students experienced greater confidence, reduced anxiety, and heightened engagement when assessment difficulty “just right” matched their ability levels. Teachers valued the curricular alignment and contextual relevance of their items, though they identified calibration procedures and time investments as implementation challenges.

This manuscript concludes by offering practical guidelines for developing high-quality teacher-designed item banks, implementing adaptive algorithms in resource-constrained settings, and designing sustained professional development to build teacher capacity in psychometric principles. It also highlights emerging opportunities—such as automated item generation, adaptive feedback loops, and equity-focused DIF analyses—to further refine the synergy between teacher expertise and adaptive testing technology.

KEYWORDS

Assessment banks; adaptive testing; personalized learning; item response theory; teacher-authored items

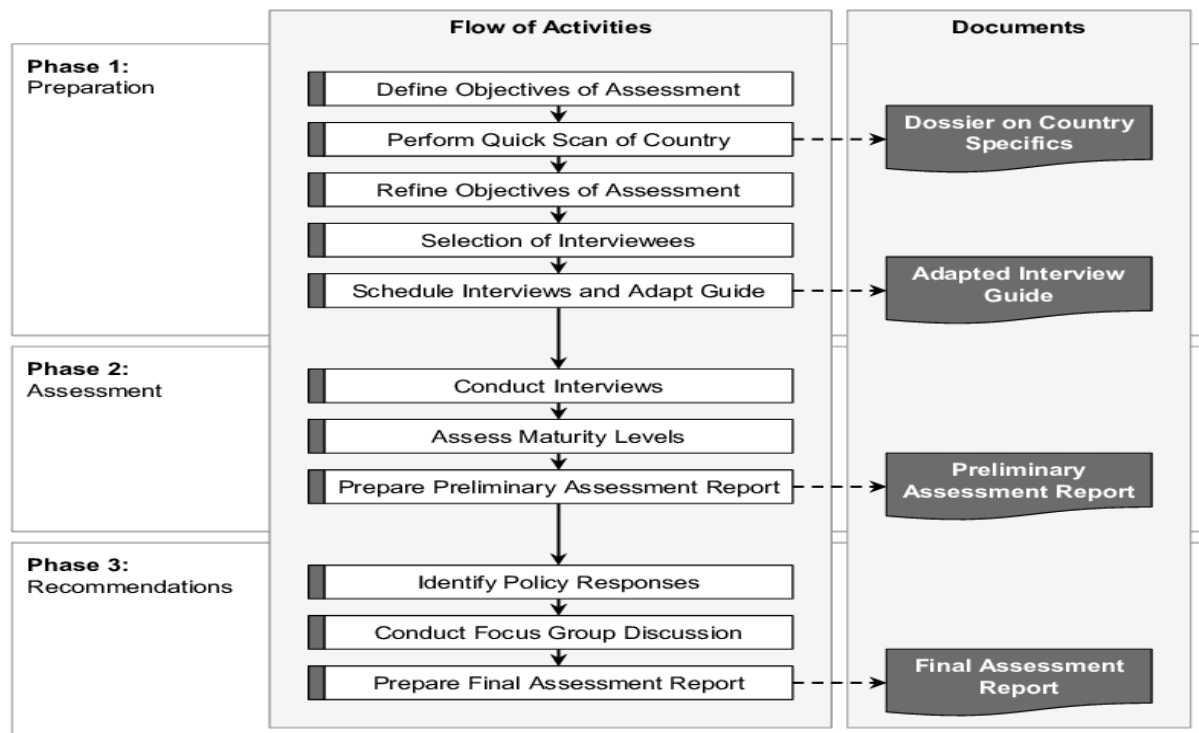


Fig.1 Assessment Process, [Source:1](#)

INTRODUCTION

Educational assessment serves multiple purposes: diagnosing learner strengths and weaknesses, informing instructional decisions, and certifying achievement. Traditional fixed-form assessments, while straightforward to administer, face limitations in delivering precise measurement across the full spectrum of student abilities. High-achieving learners may find items too easy, resulting in ceiling effects, whereas struggling students may encounter floor effects when items exceed their skill levels. In response, computerized adaptive testing (CAT) has emerged as a powerful methodology to match item difficulty dynamically to examinee performance, thereby maximizing measurement efficiency and examinee engagement.

However, off-the-shelf item pools used in many CAT systems often lack contextual alignment with specific classroom curricula and may fail to reflect the nuanced pedagogical intents of teachers. This gap creates a demand for teacher-designed assessment banks—collections of items authored or curated by classroom educators, mapped explicitly to learning objectives, instructional materials, and local context. When integrated

with adaptive testing algorithms, these teacher-sourced item banks hold promise for delivering assessments that are both psychometrically sound and pedagogically meaningful.

This manuscript explores the theoretical underpinnings, practical implementation, and outcomes of combining teacher-designed assessment banks with adaptive testing frameworks. We first review foundational theories of item response theory (IRT) and adaptive test assembly, then examine best practices in developing teacher-authored item repositories. A mixed-methods study involving teachers and students evaluates the feasibility, reliability, validity, and instructional utility of adaptive assessments based on teacher-designed banks. Finally, we discuss implementation challenges, professional development implications, and avenues for future research.

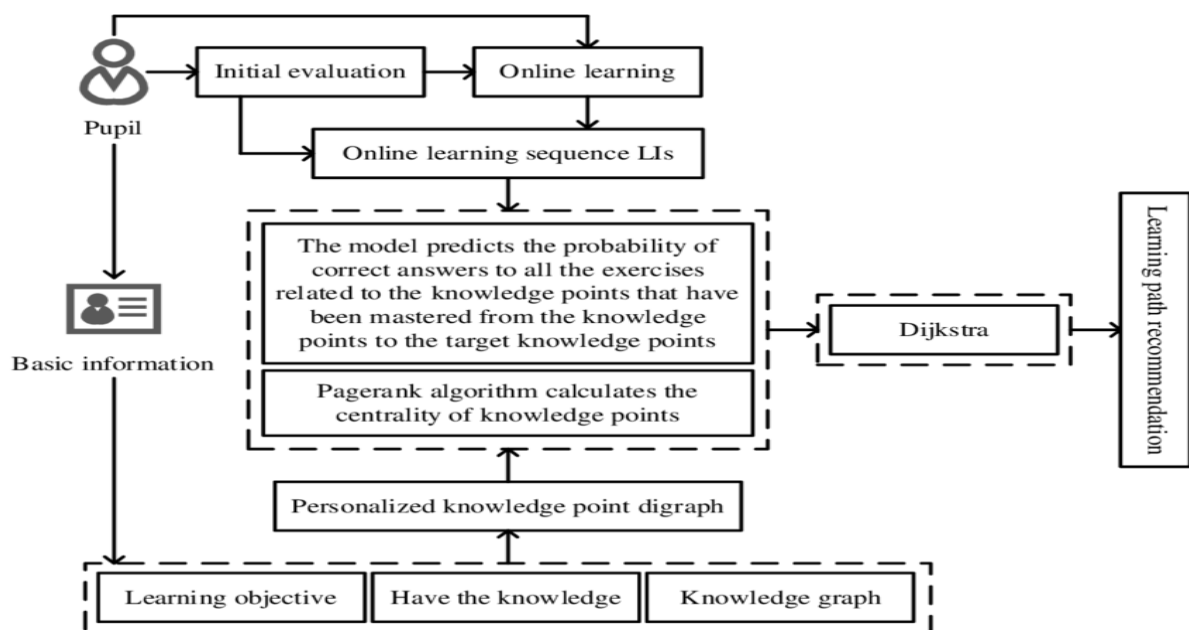


Fig.2 Personalized Learning, [Source:2](#)

LITERATURE REVIEW

Foundations of Assessment Banks

Assessment banks comprise structured item repositories categorized by content standards, cognitive demand levels, and performance objectives. Shavelson and Stanton (2010) emphasize that well-designed item banks enable flexible test assembly, repeated use, and systematic alignment with curricular goals. Teachers, drawing on intimate knowledge of learner needs and instructional sequences, are well-positioned to author items that capture nuanced understanding and context-specific applications (Black & Wiliam, 1998).

Computerized Adaptive Testing (CAT)

CAT relies on algorithms—typically rooted in item response theory (IRT)—to select items that maximize information about an examinee’s latent ability. Lord’s (1980) three-parameter logistic model remains a foundational IRT framework, accounting for item difficulty, discrimination, and guessing. Subsequent research has refined selection criteria (e.g., maximum information, Bayesian methods) and assembly constraints to ensure content balance (Weiss, 2011).

Integration of Teacher-Designed Banks with CAT

Researchers have begun exploring hybrid models that leverage teacher-generated items within CAT environments. Studies demonstrated that teacher-designed items, once calibrated, achieve psychometric properties comparable to commercially developed items, while offering greater curricular relevance. Nevertheless, challenges include item calibration processes, technological infrastructure in classrooms, and teacher training in psychometric principles (Wang & Chinnappan, 2015).

Professional Development and Teacher Engagement

Effective implementation of teacher-designed item banks requires sustained professional development. Popham (2006) argues that teachers need grounding in item-writing guidelines, bias review procedures, and basic IRT concepts to author high-quality items. Collaborative item-review workshops and calibrated scoring panels have proven effective in enhancing item quality and teacher buy-in.

Gaps and Emerging Trends

Despite growing interest, gaps remain in understanding how teacher-designed adaptive assessments impact student motivation, equity, and long-term learning trajectories. Emerging research directions include automated item generation using natural language processing (NLP), adaptive feedback mechanisms that integrate explanatory hints, and multi-stage testing designs that blend formative and summative functions.

METHODOLOGY

Research Design

A sequential explanatory mixed-methods design was employed. Phase one involved quantitative calibration and field testing of teacher-authored item banks within a CAT platform. Phase two gathered qualitative insights through teacher focus groups and student interviews to contextualize quantitative findings.

Participants

One hundred fifty secondary school teachers across five public schools volunteered to develop and review test items in mathematics, science, and language arts. Six hundred students (grades 7–10) participated in the

adaptive test administration. Demographics reflected urban and suburban populations with diverse socioeconomic backgrounds.

Item Development and Calibration

Teachers authored an initial pool of 1,200 multiple-choice and constructed-response items aligned to curriculum standards. Items underwent expert review for clarity, bias, and alignment. A pilot field test involving 300 students provided response data for IRT calibration using a two-parameter logistic (2PL) model. Items with poor fit statistics (infit or outfit mean square outside 0.7–1.3) were revised or discarded, resulting in a final pool of 900 calibrated items.

Adaptive Test Administration

Students undertook four parallel forms of adaptive tests—one each in mathematics, science, and language arts—each with a stopping rule of 20 items or measurement precision threshold of standard error ≤ 0.30 . Item exposure controls (Simpson–Hetter method) ensured item security and content balancing across strands (e.g., algebra vs. geometry).

Data Collection and Analysis

Primary quantitative measures included test reliability (marginal reliability coefficients), validity evidence (convergent validity with standardized achievement tests), measurement precision (test information functions), and test length. Student motivation scores were collected via a validated engagement questionnaire. Qualitative data from teacher focus groups (six groups, eight teachers each) and student interviews (n=40) were thematically analyzed to identify perceived benefits and challenges.

RESULTS

Psychometric Properties

Adaptive tests drawing from teacher-designed banks achieved high reliability: marginal reliability coefficients of 0.92 for mathematics, 0.90 for science, and 0.91 for language arts. Test information functions peaked around ability levels corresponding to grade-level expectations, indicating effective targeting of contents. Convergent validity correlations with state achievement tests were strong ($r = .86-.89$), supporting validity evidence.

Test Efficiency and Precision

Adaptive administrations required an average of 18 items to reach the precision threshold, reducing student testing time by approximately 40% compared to fixed-form tests of 30 items. Item exposure rates remained within acceptable limits ($<20\%$ for any single item) due to implemented security controls.

Student Motivation and Engagement

Mean engagement scores were significantly higher for adaptive tests ($M = 4.1$ on a 5-point Likert scale) compared to traditional fixed-form assessments ($M = 3.6$; $p < .01$). Students reported that adaptive tests felt more “just right” in difficulty, reducing anxiety and increasing confidence.

Teacher Perspectives

Qualitative analysis revealed that teachers valued curricular alignment and contextual relevance of their authored items. Challenges included initial unfamiliarity with IRT calibration processes and the time investment required for item development. Teachers recommended collaborative item-writing sessions and dedicated release time for assessment design.

CONCLUSION

The integration of teacher-designed assessment banks with adaptive testing frameworks represents a significant advancement in the pursuit of personalized, precise, and pedagogically meaningful assessment practices. Empirical findings from this study demonstrate that assessments leveraging teacher-authored item pools not only meet rigorous psychometric standards—evidenced by high reliability, validity, and measurement efficiency—but also substantially enhance student motivation and confidence. Students reported feeling more at ease when item difficulty was tailored to their individual ability, thereby reducing test anxiety and promoting sustained engagement. Teachers, empowered to craft items aligned closely with instructional content and learner needs, experienced a renewed sense of ownership over the assessment process and deeper insights into student learning patterns.

Nevertheless, successful implementation hinges on addressing several critical factors. First, robust calibration workflows and user-friendly CAT platforms are essential to streamline the transition from item creation to adaptive delivery. Second, targeted professional development initiatives must equip teachers with foundational knowledge of IRT principles, bias review procedures, and item-writing best practices. Collaborative item-writing workshops, peer review panels, and access to calibration software can mitigate the initial time investment and enhance item quality. Third, technological infrastructure—reliable hardware, stable internet connectivity, and technical support—must be secured, particularly in under-resourced or rural settings.

Looking ahead, future research should explore automated item generation techniques using natural language processing to reduce teacher workload, adaptive feedback systems that integrate explanatory hints directly within the test interface, and longitudinal studies examining the impact of adaptive assessments on long-term learning outcomes. Equity considerations remain paramount; ongoing DIF analyses and calibration checks

will help ensure that items function fairly across diverse demographic groups. By harnessing the complementary strengths of teacher expertise and adaptive algorithms, educators and assessment designers can create dynamic assessment ecosystems that drive personalized learning, inform targeted instruction, and ultimately foster deeper, more equitable educational experiences for all students.

REFERENCES

- <https://www.researchgate.net/publication/320298990/figure/fig3/AS:547896317939712@1507640069561/Flowchart-of-Assessment-Process.png>
- <https://www.researchgate.net/publication/383402198/figure/fig3/AS:11431281273507837@1724591730093/Personalized-learning-path-recommendation-flowchart.png>
- Black, P., & Wiliam, D. (1998). *Assessment and classroom learning*. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. SAGE Publications.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge & Kegan Paul.
- Mislevy, R. J., & Stocking, M. L. (1999). *A primer on model-based approaches to test equating, scaling, and scoring*. ETS Research Report Series, 1999(1), i–55.
- Popham, W. J. (2006). *Educational assessment: Principles, policies, and practices* (2nd ed.). Allyn & Bacon.
- Reckase, M. D. (2009). *Multidimensional item response theory*. Springer.
- Shavelson, R. J., & Stanton, G. C. (2010). *Assessing student learning on the way: Some models and methods*. *Educational Researcher*, 39(3), 205–222.
- Van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. Springer.
- Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum Associates.
- Wang, L., & Chinnappan, M. (2015). *Teacher-generated item banks in mathematics: Effects on student achievement and engagement*. *Mathematics Education Research Journal*, 27(4), 505–523.
- Weiss, D. J. (2011). *Better data from better measurements using computerized adaptive testing*. ETS Research Report Series.
- Wu, M. L., Adams, R. J., & Wilson, M. (1997). *CONQUEST: Software for Rasch analysis*. MESA Press.
- York, L., & Pendleton, L. (2001). *Collaborative item-writing workshops: Building capacity for classroom assessment*. *Phi Delta Kappan*, 82(10), 766–770.
- Zenisky, A. L., & Hambleton, R. K. (2012). *Psychometric challenges in implementing computerized adaptive testing in K–12 contexts*. *Journal of Technology, Learning, and Assessment*, 10(3), 1–30.
- Zieky, M. J., & Perie, M. (2006). *Item-response theory*. In *Educational measurement* (4th ed., pp. 111–153). American Council on Education/Macmillan.