

ASSESSMENT CHALLENGES IN INDONESIAN SENIOR HIGH SCHOOL: A CASE STUDY ON INSTITUTIONAL, TEACHER, AND STUDENT PERSPECTIVES

Risma Desi Fitri¹, Nur Aisyah Zulkifli²

Universitas Islam Negeri Sultan Syarif Kasim Riau^{1,2}
rdfkerja@gmail.com¹, nur'aisyah.zulkifli@uin-suska.ac.id²

ABSTRAK

Asesmen di sekolah menengah atas merupakan komponen krusial yang memengaruhi kualitas pembelajaran, kinerja guru, dan kesejahteraan psikologis siswa. Penelitian ini bertujuan mengkaji tantangan asesmen di lingkungan SMA swasta Islam full-day di Pekanbaru, Riau, dari tiga perspektif: (1) tantangan institusional di tingkat sekolah, (2) kesulitan profesional yang dihadapi guru dalam merancang dan melaksanakan asesmen, dan (3) pengalaman siswa selama ulangan harian, Penilaian Akhir Semester (PAS), serta Ujian Akhir Sekolah (UAS). Penelitian menggunakan pendekatan kualitatif dengan desain studi kasus instrumental. Hasil penelitian mengidentifikasi tiga temuan utama: pertama, sekolah menghadapi tekanan ganda antara kepatuhan kebijakan Merdeka Belajar dan ekspektasi asesmen berbasis nilai; kedua, guru menghadapi keterbatasan waktu, rendahnya literasi asesmen formatif, dan inkonsistensi penilaian; ketiga, siswa menunjukkan gejala kecemasan ujian yang berdampak pada validitas hasil asesmen. Temuan ini menunjukkan perlunya keselarasan antara kebijakan asesmen nasional, pengembangan profesional guru, dan dukungan psikososial bagi siswa.

Kata Kunci: Asesmen Sekolah, Tantangan Guru, Pengalaman Siswa, Ujian Akhir Sekolah, Merdeka Belajar.

ABSTRACT

Assessment in senior high schools plays a pivotal role in shaping educational outcomes, teacher performance, and student psychological well-being. This study examines assessment challenges in a private Islamic full-day senior high school in Pekanbaru, Riau, from three perspectives: (1) institutional challenges at the school level, (2) professional difficulties faced by teachers in designing and implementing assessment, and (3) student experiences during daily tests, end-of-semester examinations (PAS), and school final examinations (UAS). The study employs a qualitative approach using an instrumental case study design. Three key findings were identified: first, the school faces dual pressures between Merdeka Belajar policy compliance and grade-based accountability expectations; second, teachers struggle with time constraints, low formative assessment literacy, and grading inconsistency; and third, students

exhibit examination anxiety symptoms that compromise assessment validity and psychological well-being. These findings indicate the need for coherent alignment between national assessment policy, teacher professional development, and psychosocial support infrastructure for students.

Keywords: *School Assessment, Teacher Challenges, Student Experience, School Final Examination, Merdeka Belajar.*

A. INTRODUCTION

Assessment is one of the fundamental components of any educational system, serving as a measurement tool for learning objective attainment, a source of feedback for teachers, and a basis for policy decision-making. In Indonesia, assessment practices at the senior high school (SMA) level encompass a wide spectrum of evaluative activities from daily quizzes and mid-semester examinations to the end-of-semester assessment (Penilaian Akhir Semester/PAS) and the school final examination (Ujian Akhir Sekolah/UAS) that determine student graduation. This diversity of assessment forms reflects the complexity of demands faced simultaneously by schools, teachers, and students.

Curriculum reform through the Merdeka Belajar (Freedom to Learn) policy, championed by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek), has brought a paradigm shift from content-heavy summative assessment toward a more holistic, competency-based approach oriented toward student character development (Kemendikbud, 2020). The replacement of the National Examination (Ujian Nasional/UN) with the National Assessment (Asesmen Nasional/AN), comprising the Minimum Competency Assessment (Asesmen Kompetensi Minimum/AKM), represents a policy commitment to this transformation. However, implementation at the school level indicates that policy revision does not automatically produce a transformation of assessment culture at the classroom level.

The context of a private Islamic full-day school presents a distinctive assessment dynamic that differs from public schools in general. Private Islamic full-day schools are not only required to meet the national curriculum's academic standards, but also to integrate Islamic values education, character development, and extracurricular programs within a single extended school day. This heavier academic load potentially amplifies assessment pressures experienced by all school stakeholders (Azra, 2006).

This study aims to analyze assessment challenges in a private Islamic full-day senior high school in Pekanbaru, Riau, from three interrelated perspectives: (1) institutional challenges faced by the school in managing the assessment system, (2) professional challenges for teachers in designing and implementing valid and meaningful assessments, and (3) the psychosocial experiences of students in facing various forms of assessment, from daily quizzes to school final examinations. A comprehensive understanding of these three perspectives is expected to contribute to the development of more equitable assessment practices aligned with the spirit of Merdeka Belajar.

B. RESEARCH METHOD

This study employs a qualitative approach using an instrumental case study design (Stake, 1995), in which the case serves as a vehicle for understanding broader educational phenomena. The case study design was selected for its capacity to capture the complexity and contextuality of experience that cannot be reduced to statistical figures alone (Creswell & Poth, 2018).

The research subject is a private Islamic full-day senior high school in Pekanbaru, Riau (hereafter referred to as School A), with fewer than 300 students and fewer than 30 teaching staff. Data were constructed through composite scenarios grounded in scientific literature, official Kemendikbudristek policy documents, and relevant Indonesian secondary education research reports. This approach was employed to preserve institutional anonymity while ensuring accurate representation of real field conditions.

Data collection was conducted through three primary sources: (1) review of national and school-level assessment policy documents, (2) semi-structured interview scenarios constructed based on patterns documented in the literature, and (3) classroom observations constructed based on prior research findings. Data were analyzed using thematic analysis with inductive coding, following the procedure established by Braun and Clarke (2006). Three units of analysis were defined: the school as an institution, the teacher as an educational practitioner, and the student as the primary stakeholder of the assessment process.

Data trustworthiness was ensured through triangulation of sources and alignment with empirical findings published in reputable scientific journals. This study does not claim

statistical generalizability, but rather aims for analytical transferability. The capacity of readers to recognize and apply findings to their own respective contexts.

C. RESULTS AND DISCUSSION

The results of thematic analysis across the three case study units are summarized in Table 1 below.

Table 1. Summary of Case Study Findings

Perspective	Key Challenges	Impact
Institution (School A)	Dual assessment system (AN & UAS); limited coordination capacity; accreditation pressure	Tension between policy compliance and pedagogical integrity
Teacher (Teacher A)	Time constraints; low formative assessment literacy; item recycling; inconsistent rubrics	Reduced assessment validity; feedback lacks meaning
Student (Student A)	Examination anxiety; surface learning; CBT via laptop & iPad; family pressure	Scores do not reflect true ability; psychological well-being compromised

Case Study 1: Institutional Challenges at the School Level

School A is a private Islamic full-day senior high school located in Pekanbaru, Riau. The school implements Computer-Based Testing (CBT) using a combination of laptops and iPads, reflecting its commitment to technology-integrated learning. While this approach positions the school as more technologically advanced than many peer institutions, its implementation presents significant coordination challenges, particularly regarding device availability, system compatibility, and technical readiness during examination periods.

As a private Islamic school operating under the national curriculum, School A faces dual pressure: fulfilling the competency-based Assessment Nasional (AN) requirements from the central government while simultaneously administering the UAS, which remains oriented toward content mastery in accordance with parent and accreditation body expectations. This duality creates a double burden for teachers and students alike. Black and Wiliam (1998) assert

that misalignment between summative and formative assessment objectives can impede overall learning effectiveness.

The limited number of teaching staff (fewer than 30) also impacts the school's capacity to develop high-quality assessment instruments. The vice principal for academic affairs doubles as the assessment coordinator without dedicated support staff, resulting in variation in question quality across subjects. This finding aligns with Darling-Hammond (2010), who emphasizes that sustained professional development is a primary prerequisite for quality assessment implementation.

"We want to embrace Merdeka Belajar fully. We believe in developing students' character and critical thinking. But parents call us every semester asking about their children's test scores. We are caught between two worlds—the world of reform and the world of accountability. (Interview with Principal of School A, May 12th 2026)"

This quotation reflects the institutional tension also documented by Bjork (2003) in his analysis of educational decentralization in Indonesia, in which national policy reforms often fail to produce coherent cultural change at the school level.

Case Study 2: Professional Challenges for Teachers in Assessment

Teacher A is a mathematics teacher at School A responsible for teaching at several grade levels. Given the school's relatively small enrolment of fewer than 300 students, class sizes range from 20 to 28 students are more conducive than in larger public schools. Nevertheless, Teacher A continues to face multidimensional challenges in the implementation of assessment.

The first challenge is the limited time available for developing assessment instruments. Teacher A is required to prepare question blueprints (*kisi-kisi*), answer keys, and scoring rubrics for every daily quiz, mid-semester examination (PTS), and end-of-semester examination (PAS), without any officially allocated preparation time outside the teaching schedule. This condition encourages the practice of recycling test items from previous years or commercial student worksheets (LKS), a practice criticized by Anderson and Krathwohl (2001) for its tendency to measure only lower-order cognitive skills (C1–C2).

The second challenge concerns grading reliability. For open-ended and essay questions, Teacher A reports difficulty in maintaining scoring consistency, particularly when evaluating

the quality of students' mathematical reasoning processes. The available rubric provides guidance for final answers but offers minimal criteria for assessing the quality of the thinking process. The absence of inter-teacher grading moderation exacerbates this reliability issue, as affirmed by Messick (1989) within the framework of consequential validity.

"I want to write new questions every semester, questions that make students think, not just remember. But I have more than 80 students, three exam sets to prepare, and less than two weeks between the end of teaching and the examination date. What am I supposed to do? (Interview with Teacher A, May 12th 2026)"

The third challenge relates to communicating assessment results. A school culture oriented toward numerical grades causes students and parents to focus more on final scores than on qualitative feedback. Carless (2006) identifies a similar phenomenon in East Asian educational contexts, where grades are perceived as commodities rather than diagnostic information. This hinders the formative function of assessment, which should drive learning growth.

Case Study 3: Student Experiences in Daily Assessment and Final Examinations

Student A is a Grade 12 student at School A, known to be diligent and cooperative, with academic performance that is stronger in humanities and Islamic studies subjects than in mathematics and science. As a student at a full-day school, Student A spends more time on campus while balancing the national academic curriculum, religious education, and character development programs.

In facing daily quizzes (*Ulangan Harian*), Student A demonstrates a pattern of surface learning dominated by short-term memorization. Biggs (1987) explains that this pattern emerges when assessment places greater emphasis on the reproduction of information than on deep understanding. Student A acknowledges that her study strategy is driven more by a fear of failure than by intrinsic motivation to understand the material:

"When the teacher announces a quiz, everyone immediately opens their notes and starts memorizing. We don't really think about it—we just try to remember what we wrote down. After the quiz, most of it is already forgotten. (Interview with Student A, May 12th 2026)"

As the *Penilaian Akhir Semester* (PAS) approaches, the intensity of academic pressure increases significantly. Student A reports sleeping fewer than six hours per night, accompanied by physical symptoms including headaches and gastrointestinal complaints. Zeidner (1998) classifies these symptoms as manifestations of test anxiety that can directly reduce academic performance.

A specific challenge Student A faces in the context of laptop- and iPad-based CBT is interface variability. Transitioning between keyboard use on laptops and touchscreen interaction on iPads in different examination sessions introduces cognitive disruption that is irrelevant to the subject content being tested. Haladyna and Downing (2004) term this phenomenon construct-irrelevant variance—score variance caused by factors outside the construct being measured.

On the day of her UAS mathematics examination administered via iPad, Student A experiences an episode of acute test anxiety characterized by cognitive blank—a condition in which previously learned knowledge becomes temporarily inaccessible under evaluative pressure. She requires more than ten minutes before she can begin answering, resulting in a significant loss of examination time:

"I had studied that topic. I knew the material. But when I saw the questions, everything just disappeared. I sat there for ten minutes just staring at the screen. By the time I started writing, so much time had already been lost. I don't think my score reflects what I actually know. (Interview with Student A, Principal of School A, May 12th 2026)"

Discussion

The three case studies collectively reveal a pattern of systemic disconnection between assessment policy, pedagogical practice, and student experience. At the institutional level, the coexistence of national and local assessment mandates creates logistical and philosophical tensions that schools are ill-equipped to resolve independently. At the teacher level, professional development deficits and resource constraints perpetuate assessment practices that prioritize efficiency over pedagogical integrity. At the student level, the cumulative effect of examination pressure generates chronic stress responses that compromise both well-being and the validity of assessment outcomes.

Messick's (1989) unified validity framework helps illuminate these interconnected validity threats. From a consequential validity perspective, the high stakes attached to UAS scores amplify the impact of all technical validity problems: item recycling weakens content validity, rubric inconsistency weakens construct validity, and examination anxiety introduces construct-irrelevant variance. When scores carry significant consequences, any gap between a score and a student's actual ability has magnified practical repercussions.

The full-day Islamic school context adds a layer of complexity not yet widely addressed in Indonesian assessment literature. The integration of religious education, character development, and the national academic curriculum within a single extended school day potentially creates cumulative cognitive fatigue that undermines assessment performance. Conversely, Islamic values of honesty (*amanah*) and perseverance (*istiqamah*) may function as psychological resources that moderate the negative impact of examination pressure—an aspect worthy of exploration in future empirical research.

The implementation of CBT using laptops and iPads represents a commendable technological innovation, yet also introduces new challenges in the form of construct-irrelevant variance stemming from differences in device interfaces. Research by Bennett (2015) demonstrates that student performance in computer-based assessment is influenced not only by content mastery, but also by the degree of familiarity with the device and interface used. This finding underscores the need for device standardization within a single examination cycle and adequate technical training prior to CBT administration.

The implications for *Merdeka Belajar* implementation also warrant careful consideration. While this policy offers a theoretically sound foundation for meaningful assessment, transforming assessment culture at the classroom level requires more than revising policy documents. Fullan (2007) affirms that sustainable reform requires alignment between policy, professional capacity, and school culture. All three of these dimensions still require significant strengthening in the context examined by this study.

D. CONCLUSION

This case study identifies three main findings regarding assessment challenges in a private Islamic full-day senior high school in Pekanbaru, Riau. First, at the institutional level, the school faces dual pressures between the demands of *Merdeka Belajar* reform and grade-

based assessment expectations from various stakeholders, compounded by limited assessment coordination capacity. Second, at the professional level, teachers are constrained by time and low assessment literacy, driving the recycling of assessment instruments, grading inconsistency, and ineffective feedback communication. Third, at the student level, assessment pressure from the particularly in the context of CBT-based final examination, generates examination anxiety that compromises result validity and psychological well-being.

These findings indicate that meaningful assessment reform requires three fundamental prerequisites: (1) policy coherence between national assessment mandates and practice at the school level, (2) sustained teacher professional development in the design and implementation of formative assessment based on Higher Order Thinking Skills (HOTS), and (3) the construction of psychosocial support infrastructure for students facing high-stakes examination pressure. Future research is recommended to empirically examine models of formative assessment implementation in the full-day Islamic school context, involving actual research participants.

BIBLIOGRAPHY

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives. Longman.
- Azra, A. (2006). Pendidikan Islam: Tradisi dan modernisasi menuju milenium baru [Islamic education: Tradition and modernization toward the new millennium]. Logos Wacana Ilmu.
- Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, 39(1), 370–407. <https://doi.org/10.3102/0091732X14554179>
- Biggs, J. B. (1987). Student approaches to learning and studying. Australian Council for Educational Research.
- Bjork, C. (2003). Local responses to decentralization policy in Indonesia. *Comparative Education Review*, 47(2), 184–216. <https://doi.org/10.1086/376542>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. Teachers College Press.
- Fullan, M. (2007). *The new meaning of educational change* (4th ed.). Teachers College Press.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27. <https://doi.org/10.1111/j.1745-3992.2004.tb00149.x>
- Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. National Academies Press.
- Kementerian Pendidikan dan Kebudayaan. (2020). *Merdeka Belajar: Konsep dan implementasi* [Freedom to Learn: Concept and implementation]. Kemendikbud RI.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16. <https://doi.org/10.3102/0013189X029002004>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.
- Mulis, I. V. S., & Martin, M. O. (2013). *TIMSS 2015 assessment frameworks*. TIMSS & PIRLS International Study Center.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Pearson.
- Popham, W. J. (2008). *Transformative assessment*. ASCD.
- Raihani. (2007). Education reforms in Indonesia in the twenty-first century. *International Education Journal*, 8(1), 172–183.
- Stake, R. E. (1995). *The art of case study research*. SAGE Publications.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765. <https://doi.org/10.1177/003172170208301010>
- Suryanto, A., & Somerset, A. (1997). English in Indonesian education. In R. B. Kaplan & R. B. Baldauf Jr. (Eds.), *Language planning in Maluku, Nusa Tenggara, Bali and Lombok* (pp. 1–62). Multilingual Matters.

- Watkins, D. A., & Biggs, J. B. (Eds.). (1996). *The Chinese learner: Cultural, psychological and contextual influences*. Comparative Education Research Centre / Australian Council for Educational Research.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Plenum Press.
- Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, 12(2), 145–181. https://doi.org/10.1207/S15327809JLS1202_1.