



Available online at

<http://journal.pusbindiklatren.bappenas.go.id/>

Commentary

One Test Cannot Serve Two Masters

Validity Risks in Integrating Indonesia's National Assessment with the Academic Ability

Agus Saptono

*Secretariat of the Directorate General of Vocational Education, Special Education, and Special Educational Services
Ministry of Primary and Secondary Education, Jakarta, Indonesia*

*Corresponding author: agus.saptono@kemendikdasmen.go.id

Pusat Standar dan Kebijakan Pendidikan (PSKP, 2026) proposes to merge Indonesia's National Assessment (Asesmen Nasional, AN) with the Academic Ability Test (Tes Kemampuan Akademik, TKA) into a single assessment database serving both institutional quality evaluation and student selection. The administrative case for doing so is not in dispute here. What is in dispute is the measurement premise beneath it: that one instrument carrying high-stakes consequences for individual students can simultaneously yield valid low-stakes evidence about school quality. This commentary argues that it cannot, and that the indicators extracted from such an instrument will describe test-taking behaviour and access to private tutoring more faithfully than they describe teaching and learning.

The proposal begins with one central assumption, namely that a single assessment instrument is capable of producing distinct forms of evidence that support fundamentally different evaluative purposes (Newton, 2010). Before considering such integration, it is important to recognize the conceptual and operational differences between the two assessment systems. Table 1 summarizes these key distinctions, highlighting the contrasting philosophical foundations and implementation characteristics of each model.

Table 1. Comparative Overview of Assessment Paradigms

Assessment Dimension	National Assessment (AN)	Academic Ability Test (TKA)
Primary Objective	Systemic diagnostic and institutional quality mapping.	Individual certification and academic selection.

ARTICLE INFO

Received: February 02, 2026
Received in revised form: June 12, 2026
Accepted: August 30, 2026

doi: [10.46456/jisdep.v7i2.1245](https://doi.org/10.46456/jisdep.v7i2.1245)



This is an open access article under the [CC BY-SA](#) license
© Saptono (2026)

THE JOURNAL OF INDONESIA SUSTAINABLE DEVELOPMENT PLANNING

Published by Centre for Planners' Development, Education, and Training (Pusbindiklatren), Ministry of National Development Planning/National Development Planning Agency (Bappenas), Republic of Indonesia

Address: Jalan Proklamasi 70,
Central Jakarta, Indonesia 10320

Phone: +62 21 31928280/31928285

Fax: +62 21 31928281

E-mail:

journal.pusbindiklatren@bappenas.go.id

Supported by Indonesian Development Planners Association (PPPI)

Please cite this article in APA Style as:

Saptono, A. (2026). One Test Cannot Serve Two Masters: Validity Risks in Integrating Indonesia's National Assessment with the Academic Ability. *The Journal of Indonesia Sustainable Development Planning*, Vol 7(2), 429-432.
<https://doi.org/10.46456/jisdep.v7i2.1245>

Assessment Dimension	National Assessment (AN)	Academic Ability Test (TKA)
Stakes Level	Low-stakes (No direct impact on individual student progression).	High-stakes (Determines access to advanced educational tiers).
Behavioral Impact	Encourages authentic responses reflecting the true learning environment.	Triggers performance anxiety and intense, strategic test preparation.
Data Output	Aggregated institutional metrics (Education Report Card).	Individual standardized scores and percentiles.

As highlighted in Table 1, the policy would derive data for system level quality assurance from TKA subject examinations, particularly Mathematics and Indonesian Language, by applying a scoring approach designed to estimate students' literacy and numeracy competencies. This design assumes that the cognitive processes and response patterns displayed in a high stakes academic selection examination can also provide valid evidence for evaluating school performance. Such an assumption overlooks well established evidence on participant behavior in standardized testing. When examination results carry substantial consequences for students, as they do in the TKA, motivation, test taking strategies, and performance are shaped by external incentives, causing test scores to diverge from students' actual learning outcomes (Koretz, 2017).

Historically, the National Assessment was originally developed as a low stakes assessment to produce an objective picture of educational quality. Because its results are not linked to graduation or student selection, it was intended to capture instructional conditions with minimal influence from strategic test preparation or performance pressure. Integrating system level indicators into a high stakes examination such as the TKA creates a substantial risk that the resulting data will reflect test anxiety, coaching practices, and score optimization strategies rather than genuine improvements in teaching and learning. Although PSKP (2026) recognize the possibility of teaching to the test, their proposed safeguards, including public communication campaigns and the continued use of higher order thinking items, are unlikely to offset the strong incentives generated by a high stakes assessment environment. Consequently, if the extracted Minimum Competency Assessment (AKM) indicators primarily capture strategic test behavior instead of authentic educational progress, the evidence used to inform policy decisions and regional resource allocation may become systematically biased.

Curricular Distortion and The Shadow Education Threat

The implications of integrating these assessments at the classroom level deserve greater attention than is provided in the current analysis. Combining system level evaluation with an examination that determines individual academic outcomes is likely to reshape teaching practices. Research on high stakes assessment consistently shows that when examination results influence students' educational opportunities, classroom instruction tends to become closely aligned with the content and format of the test (Au, 2007). The proposed framework assumes that embedding literacy and numeracy competencies within subject based assessments will encourage teachers to implement the broader goals of the national curriculum more consistently. However, this expectation is unlikely to be realized without substantial investment in professional development that equips teachers to deliver inquiry based and higher order learning. In the absence of such support, instructional practice is more likely to shift toward repetitive test preparation rather than meaningful learning.

This issue also raises important concerns regarding educational equity, a principle emphasized in Indonesia's National Long Term Development Plan 2025 to 2045. The focal study acknowledges the risk of epistemic injustice, whereby standardized assessments may reflect differences in students' socioeconomic backgrounds rather than their actual academic potential. Nevertheless, assigning greater consequences to integrated assessment results may unintentionally expand the role of private supplementary education. Students from economically advantaged families are more likely to gain access to commercial tutoring that improves examination performance, independent of the quality of instruction provided in their schools (Sari et al., 2025). As a result, the integrated assessment system could generate data that capture disparities in access to private educational support instead of differences in school effectiveness. If such data become the basis for policy decisions, existing educational inequalities may be reinforced despite the appearance of objective and comprehensive measurement.

Governance, Statistical Noise, And Regional Planning

From a macro-governance perspective, reducing assessment fatigue and eliminating administrative duplication offer clear advantages for the education system. Conducting two separate national assessments requires substantial financial resources at the local level and interrupts the regular flow of school activities and the academic calendar. Aligning the implementation of these assessments through a single operational framework is expected to improve logistical efficiency and support local governments in integrating Education Report Card data into their Regional Government Work Plans (RKPD). In addition, expanding the assessment sample to include all students participating in the Academic Ability Test (TKA) is expected to strengthen the representativeness of institutional performance data, thereby providing a more comprehensive and equitable basis for evaluating educational quality.

While the proposed framework offers important administrative advantages, it also introduces significant operational challenges, particularly in relation to data extraction mechanisms. Figure 1 illustrates the proposed implementation pathway. More importantly, it highlights where the policy may become vulnerable. Once the assessment carries consequences for individual students, behavioural changes become difficult to avoid. Those changes may eventually affect the quality of the institutional indicators derived from the same assessment.

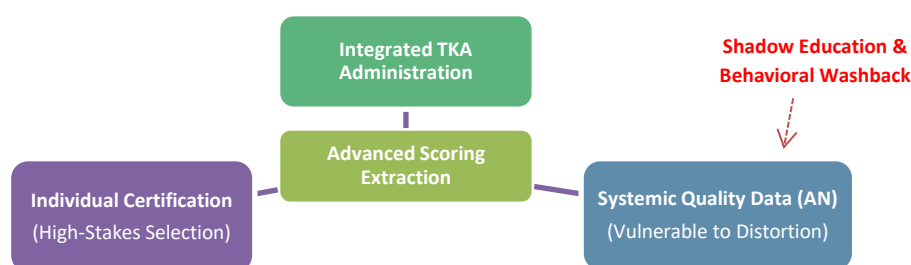


Figure 1. The Dual-Extraction Data Architecture and Washback Vulnerabilities

The policy proposes deriving low-stakes diagnostic information from an assessment primarily designed for high-stakes purposes. This approach creates a hybrid dataset that combines two fundamentally different measurement objectives. As a result, the data are likely to contain greater statistical variability, making interpretation more difficult. Expecting local education authorities, many of which continue to face limitations in data literacy and analytical capacity, to distinguish meaningful patterns from this complexity and translate them into effective instructional strategies may be unrealistic (Schildkamp, 2019). Table 2 outlines the principal operational risks associated with this policy and identifies the strategic measures needed to support its successful implementation.

Table 2. Potential Integration Risks and Strategic Mitigations

Identified Risk Domain	Operational Manifestation	Required Strategic Mitigation
Curricular Narrowing	Instructional focus shifts exclusively to testable mechanics, stalling deep learning.	Mandatory, intensive localized teacher training on formative assessments and higher-order pedagogy.
Epistemic Injustice	Extracted scores reflect shadow education access rather than institutional efficacy.	Government provision of free, highly accessible preparatory modules to equalize testing familiarity.
Data Misinterpretation	District leaders misuse hybrid data for punitive measures against schools.	Development of strict interpretive guidelines and capacity building for municipal education planners.

The gap between generating statistically robust extracted scores and ensuring that local policymakers interpret those results appropriately remains a critical challenge for the proposed policy. Without clear guidance, district education authorities may view extracted AKM scores as direct measures of teacher performance rather than indicators shaped by multiple contextual factors, including students' socioeconomic conditions and school environments. Such misinterpretations could encourage punitive

policy responses, undermining collaboration among education stakeholders and weakening the systemic improvements that the integration is intended to achieve.

Conclusion

The proposal to integrate the National Assessment with the Academic Ability Test represents a substantial restructuring of Indonesia's educational assessment system. Although the initiative is primarily justified by the goals of improving administrative efficiency and consolidating educational data into a single platform, it is built on the assumption that information generated from high-stakes individual assessments can also serve as a reliable indicator of system-level educational quality. This commentary contends that this assumption overlooks the well-documented influence of high-stakes testing on student behavior. As a result, the integration risks compromising the validity of the evidence used to evaluate educational performance and to inform long-term planning at the regional level.

While integrating the two assessments may appear administratively efficient, its broader educational consequences warrant closer scrutiny. Administrative streamlining alone cannot guarantee more equitable educational outcomes. The current integration strategy pays insufficient attention to the ways in which combining two high-stakes assessment purposes within a single testing framework may intensify curriculum narrowing and encourage greater dependence on private tutoring and other forms of shadow education. If the policy is intended to contribute meaningfully to Indonesia's long-term human capital development, it should be accompanied by targeted capacity-building measures that respond to local educational contexts and safeguard classroom learning from becoming dominated by test preparation. The take-home message is therefore narrow but firm: the two assessments may reasonably be aligned in scheduling and logistics, but system-level quality indicators should not be extracted from a high-stakes selection test.

Future research should focus on establishing the long-term scientific and practical validity of this policy. Longitudinal equating studies are needed to examine whether AKM scores extracted from the integrated assessment remain comparable to results produced by independent system-level diagnostic instruments. In parallel, empirical research should investigate changes in household spending on education after the policy is implemented to determine whether it contributes to the growth of the shadow education sector. Continuous evaluation of both the psychometric consistency of the extracted scores and the policy's socio-economic effects at the local level will be essential. Such evidence will help ensure that the integrated assessment system advances educational equity and improves policy effectiveness, rather than serving solely as an administrative restructuring initiative.

Reference

- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Koretz, D. (2017). *The testing mirage: An illusion of educational quality*. University of Chicago Press.
- Newton, P. E. (2010). Evaluating assessment systems. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (pp. 269-274). Elsevier.
- Pusat Standar dan Kebijakan Pendidikan. (2026). *Kajian akademik integrasi pelaksanaan asesmen nasional dan tes kemampuan akademik*. Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan Dasar dan Menengah
- Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational research*, 61(3), 257-273.
- Sari, K. A., Fajriani, S. W., Gunawan, G., Zulqoifah, A., & Hartani, M. (2025). The commodification of education and inequality in Indonesia: A sociological perspective. *Priviet Social Sciences Journal*, 5(8), 349-362.