

## Transformasi Asesmen Pembelajaran Lintas Negara: Studi Komparatif menuju Pendidikan yang Berkeadilan

### Cross-Country Learning Assessment Transformation: A Comparative Study Toward Equitable Education

diterima: 01-12-2025, disetujui: 26-12-2025, diterbitkan: 30 Desember 2025

DOI: <https://doi.org/10.24832/jpnk.v10i2.6870>

**Grendi Hendrastomo, Nur Endah Januarti**

Universitas Negeri Yogyakarta - Indonesia

Email: [ghendrastomo@uny.ac.id](mailto:ghendrastomo@uny.ac.id)

**Abstract:** *The global discussion regarding the role of national assessments, whether they should be graduation-oriented or learning-oriented, has become a crucial issue in education reform in various countries, including Indonesia. This study aims to analyze and compare the policies, material structure, models, methods, and consequences of national assessments across countries to provide insights for the development of contextualized and inclusive assessments in Indonesia. This study employed a comparative study approach by analyzing education policy documents, national assessment reports, and scholarly articles from nine countries through a comparative thematic analysis approach. The study findings indicate that national assessment is influenced by the social, cultural, geographical, and political context of each country, leading to diverse assessment models. These models ranging from high-stakes assessments in China and South Korea, to low-stakes assessments in countries such as Finland and Australia. Cross-country analysis shows that learning-oriented assessment has a more positive impact on students' learning quality and psychological mental health compared to assessment for graduation. This finding serves as an important reference for Indonesia, which is currently developing the Academic Ability Test (TKA) as a non-high-stakes assessment to map academic achievement. This study emphasizes that the design of TKA needs to consider regional diversity, infrastructure inequality, and the need to strengthen diagnostic functions and social justice in learning. Thus, TKA has the potential to become an inclusive and equitable national assessment.*

**Keywords:** *assessment-based learning, high-stakes assessment, low-stakes assessment, Academic Ability Test, equity in education*

**Abstrak:** *Perbedaan pendapat global mengenai peran asesmen nasional, antara berorientasi kelulusan atau pembelajaran menjadi isu penting dalam reformasi pendidikan di berbagai negara, termasuk Indonesia. Studi ini bertujuan untuk menganalisis dan membandingkan kebijakan, struktur materi, model, metode dan konsekuensi asesmen nasional lintas negara untuk memberikan wawasan bagi pengembangan asesmen yang kontekstual dan inklusif di Indonesia. Kajian ini menggunakan pendekatan studi komparatif dengan menganalisis dokumen kebijakan pendidikan, laporan asesmen nasional, dan artikel jurnal dari sembilan negara melalui pendekatan analisis tematik komparatif. Hasil kajian menunjukkan*

*bahwa asesmen nasional dipengaruhi oleh konteks sosial, budaya, geografis, dan politik setiap negara sehingga menghasilkan model asesmen yang beragam. Mulai dari asesmen high-stakes di Tiongkok dan Korea Selatan, hingga asesmen low-stakes seperti Finlandia dan Australia. Analisis lintas negara memperlihatkan bahwa asesmen yang berorientasi pembelajaran memberikan dampak lebih positif terhadap kualitas belajar dan kesehatan mental psikologis bagi siswa dibandingkan dengan asesmen untuk kelulusan. Temuan ini menjadi rujukan penting bagi Indonesia yang saat ini sedang mengembangkan Tes Kemampuan Akademik (TKA) sebagai asesmen non high-stakes untuk memetakan capaian akademik. Kajian ini menegaskan bahwa desain TKA perlu mempertimbangkan keberagaman wilayah, ketimpangan infrastruktur, dan memperkuat fungsi diagnostik dan keadilan sosial dalam pembelajaran. Dengan demikian, TKA berpotensi menjadi asesmen nasional yang inklusif dan berkeadilan.*

**Kata kunci:** *pembelajaran berbasis asesmen, high-stakes assessment, low-stakes assessment, Tes Kemampuan Akademik, keadilan pendidikan*

## **PENDAHULUAN**

Sistem pendidikan di berbagai negara saat ini menghadapi tantangan global yang semakin kompleks, terutama terkait dengan peningkatan kualitas pembelajaran dan pemerataan hasil belajar. Perubahan sosial, ekonomi, dan teknologi yang cepat menuntut sistem pendidikan untuk memastikan bahwa setiap peserta didik memiliki kesempatan yang adil untuk berkembang. Dalam konteks negara dengan keragaman sosial budaya dan geografis yang tinggi, isu keadilan pendidikan (*educational equity*) menjadi semakin krusial.

Sejumlah negara anggota OECD, seperti Australia, Kanada, Denmark, dan Amerika Serikat, menjadikan *equity* sebagai prioritas kebijakan pendidikan untuk mengurangi kesenjangan capaian belajar antarkelompok sosial dan wilayah (OECD, 2018). Sementara itu, di negara berkembang termasuk Indonesia dan Filipina, hasil asesmen internasional seperti PISA telah mendorong reformasi asesmen yang lebih berorientasi pada pembelajaran dan pemerataan mutu (Niemi, 2021). Negara-negara dengan sistem pendidikan mapan seperti Finlandia dan Estonia menegaskan bahwa asesmen nasional seharusnya mendukung pembelajaran berkelanjutan dan keadilan akses, bukan sekadar seleksi atau kelulusan (Sarv & Rõuk, 2020).

Dalam kerangka tersebut, asesmen nasional berperan sebagai instrumen strategis untuk memantau mutu pendidikan, mengidentifikasi kesenjangan, dan mendukung perumusan kebijakan berbasis bukti. Seiring meningkatnya tuntutan akuntabilitas dan kualitas pendidikan, asesmen berskala nasional dan internasional berkembang pesat dalam tiga dekade terakhir (Kamens & Benavot, 2011). Namun, berbagai studi menegaskan bahwa asesmen nasional tidak bersifat netral, karena desain, tujuan, dan konsekuensinya sangat dipengaruhi oleh konteks sosial, budaya, politik, dan geografis masing-masing negara. Oleh karena itu, kajian mengenai transformasi penilaian pembelajaran dalam perspektif lintas negara menjadi penting untuk memberikan pemahaman komprehensif tentang bagaimana berbagai negara

merancang dan menerapkan asesmen sesuai dengan karakteristik sistem pendidikan, tradisi budaya, tingkat pemerataan, dan kondisi infrastruktur yang dimilikinya.

Asesmen nasional tidak hanya dipahami sebagai tes berskala besar di akhir jenjang pendidikan, tetapi merupakan bagian integral dari kebijakan pendidikan yang mencakup fungsi monitoring, akuntabilitas, pemerataan hasil belajar, dan dukungan terhadap pembelajaran di sekolah. Dalam konteks global, banyak negara mulai menggeser paradigma asesmen dari orientasi seleksi atau kelulusan menuju fungsi diagnostik dan pembelajaran, meskipun pergeseran ini berlangsung berbeda sesuai dengan konteks nasional masing-masing (Birenbaum *et al.*, 2015; van Rijn *et al.*, 2024). Berbagai studi menunjukkan bahwa desain, pelaksanaan, dan pelaporan asesmen nasional—termasuk aspek sampling, model pengukuran, dan distribusi hasil—sangat menentukan validitas, relevansi, dan dampak sistemik asesmen, sehingga pengalaman lintas negara menjadi sumber pembelajaran kebijakan yang penting (Berman *et al.*, 2020; Fischman *et al.*, 2019).

Menurut Schellekens (2021), asesmen nasional dirancang untuk mengukur capaian belajar sebagai indikator sistem, memantau kesenjangan sosial ekonomi, memberikan umpan balik bagi perbaikan pembelajaran, serta mendukung akuntabilitas publik. OECD menegaskan bahwa meningkatnya perhatian terhadap kualitas dan hasil pembelajaran telah mendorong berkembangnya asesmen berskala nasional hingga internasional (Cresswell *et al.*, 2015). Namun, perluasan fungsi asesmen tersebut juga menimbulkan tantangan kebijakan, khususnya ketika asesmen digunakan sebagai instrumen berisiko tinggi bagi murid dan sekolah.

Perdebatan utama dalam kebijakan asesmen berkisar pada pilihan antara asesmen berisiko tinggi (*high-stakes*), yang menentukan kelulusan atau seleksi, dan asesmen berorientasi pembelajaran (*low-stakes*) yang menekankan fungsi diagnostik dan umpan balik. Sejumlah kajian menunjukkan bahwa asesmen *high-stakes* berpotensi meningkatkan tekanan psikologis, mempersempit kurikulum, dan mereproduksi ketimpangan sosial, terutama dalam sistem pendidikan yang masih timpang (Birenbaum *et al.*, 2015; Klenowski & Wyatt-Smith, 2012). Sebaliknya, asesmen yang berorientasi pembelajaran cenderung memberikan dampak lebih positif terhadap kualitas belajar dan kesejahteraan siswa (Black & William, 2018).

Perdebatan mengenai ujian nasional berpusat pada pilihan antara asesmen berisiko tinggi (*high-stakes*) yang menentukan kelulusan dan asesmen tanpa ujian akhir yang menekankan pelaporan capaian serta umpan balik pembelajaran. Perubahan sistem PSLE di Singapura menuju Achievement Levels (AL) menunjukkan upaya menurunkan tekanan ujian dan mengalihkan fokus dari skor ke pertumbuhan kompetensi (Lee & Ho, 2022). Berbagai studi menegaskan bahwa asesmen *high-stakes* mendorong pengajaran berorientasi tes dan meningkatkan tekanan psikologis bagi siswa dan guru, sehingga pendekatan asesmen diagnostik non *high-stakes*

semakin dipandang sebagai alternatif yang lebih berorientasi pembelajaran dan berkeadilan (Heissel *et al.*, 2021).

Selain perdebatan antara pendekatan tes dan non-tes, asesmen nasional juga melibatkan berbagai aspek teknis, seperti penentuan peserta, cakupan materi, moda pelaksanaan, skala pengukuran, serta pelaporan dan pemanfaatan hasil oleh pemangku kepentingan (van Rijn *et al.*, 2024). Aspek-aspek teknis tersebut berimplikasi langsung pada keadilan dan inklusivitas, khususnya terkait aksesibilitas bagi siswa di daerah terpencil, pendidikan nonformal, serta kesiapan infrastruktur, terutama teknologi digital. Studi Anghel, Khorramdel, dan von Davier (2024) menegaskan bahwa pengelolaan data dalam asesmen skala besar berperan penting dalam meningkatkan validitas dan inklusivitas tes melalui pemahaman yang lebih mendalam terhadap respons siswa.

Berbagai studi menunjukkan bahwa karakteristik sosial dan geografis suatu negara memengaruhi desain asesmen nasional, khususnya dalam penentuan model sampling dan moda pelaksanaan. Negara dengan wilayah terpencil dan populasi kecil cenderung menerapkan desain berbasis wilayah atau model digital hibrida untuk mengendalikan biaya dan logistik, sebagaimana terlihat pada regional assessment seperti ERCE di Amerika Latin yang lebih sensitif terhadap konteks geografis dan berpotensi menghasilkan pengukuran yang lebih adil (Carrasco, 2023).

Berbagai kajian menegaskan pentingnya studi lintas negara untuk memahami praktik baik dan tantangan dalam pemilihan model asesmen nasional, mengingat setiap negara memiliki konteks sosial, budaya, ekonomi, dan geografis yang berbeda. Fischman (2019) menunjukkan bahwa International Large-scale Assessment (ILSA) kerap digunakan sebagai instrumen legitimasi kebijakan, meskipun dampak positifnya terhadap pengembangan pendidikan nasional belum terbukti kuat. Hal ini menegaskan bahwa adopsi model asesmen tanpa adaptasi konteks lokal berisiko kurang efektif. Selain itu, Bearman (2020) menekankan pentingnya mempertimbangkan aspek teknis, seperti sampling, skala, pelaporan, dan karakteristik siswa, agar hasil asesmen dapat dimaknai secara tepat dan mendukung perumusan kebijakan asesmen nasional yang lebih kontekstual, proporsional, dan berorientasi pada pembelajaran yang inklusif dan berkeadilan.

Kajian lintas negara bagi Indonesia yang memiliki keberagaman sosial budaya yang tinggi, kondisi geografis kepulauan yang luas, dan ketimpangan kualitas pendidikan antardaerah menjadi relevan sebagai dasar pembelajaran kebijakan, bukan hanya meniru model (*policy borrowing*). Melalui pemahaman terhadap pengalaman negara lain dalam menyeimbangkan fungsi penilaian sebagai alat akuntabilitas dan sebagai instrumen pembelajaran, Indonesia dapat mengevaluasi secara kritis pilihan model penilaian nasionalnya. Hal ini menjadi semakin penting di tengah perdebatan mengenai penggunaan model ujian nasional yang bersifat *high-stakes* (menentukan kelulusan) atau model tanpa ujian akhir yang berorientasi pada pembelajaran, diagnosa dan laporan capaian secara individual.

Dalam konteks Indonesia, keberagaman kondisi geografis dan kualitas sekolah menuntut desain asesmen nasional yang mempertimbangkan akses, konektivitas, serta kesiapan guru dan siswa agar hasil asesmen mencerminkan kondisi pendidikan secara inklusif. Sejalan dengan itu, penghapusan Ujian Nasional diikuti dengan pengembangan Asesmen Kompetensi Minimum (AKM) dan Tes Kemampuan Akademik (TKA) sebagai asesmen non *high-stakes* yang berfungsi memetakan capaian akademik siswa secara obyektif, mendukung pemerataan hasil belajar, meningkatkan literasi asesmen guru, dan menyediakan umpan balik individual (Kemendikdasmen, 2025a; 2025b).

Meskipun demikian, kajian komprehensif yang membandingkan transformasi asesmen nasional lintas negara dan implikasinya bagi konteks Indonesia masih terbatas. Padahal, perbandingan lintas negara penting untuk memahami bahwa tidak ada satu model asesmen yang bersifat universal. Setiap negara mengembangkan asesmen nasional sebagai respons terhadap karakteristik sosial, budaya, geografis, dan tingkat ketimpangan sistem pendidikannya.

Artikel ini bertujuan mengisi celah penelitian melalui kajian komparatif sistematis tentang perancangan dan implementasi asesmen nasional di berbagai negara sesuai dengan karakteristik sosial, budaya, dan geografisnya. Dengan menganalisis dasar kebijakan, tujuan, pelaksanaan, dan implikasinya bagi siswa, kajian ini diharapkan dapat memberikan pijakan empiris dan konseptual bagi pengembangan asesmen nasional Indonesia, khususnya Tes Kemampuan Akademik (TKA), agar transformasi asesmen tidak sekadar bersifat teknis, tetapi juga mendukung peningkatan kualitas pembelajaran.

## **METODE**

Penelitian ini menggunakan metode studi pustaka komparatif untuk menganalisis perumusan, pelaksanaan, dan pemanfaatan asesmen nasional di berbagai negara. Pendekatan ini dipahami sebagai metode sistematis untuk membandingkan kebijakan, praktik, dan kerangka konseptual lintas konteks melalui analisis dokumen kebijakan, laporan resmi, dan publikasi ilmiah (Brislin, 1976; van Rijn *et al.*, 2024). Studi ini tidak bertujuan menghasilkan generalisasi, melainkan mengidentifikasi pola kebijakan, variasi pendekatan, serta elemen kontekstual dan lintas konteks yang relevan sebagai dasar penilaian kritis terhadap kemungkinan penerapannya dalam konteks Indonesia.

Pendekatan komparatif dipilih karena kebijakan pendidikan tidak dapat dilepaskan dari konteks sosial, politik, budaya, dan geografis masing-masing negara, sehingga pemahaman lintas konteks menjadi prasyarat pembelajaran kebijakan yang bermakna (Bray *et al.*, 2007). Studi kebijakan pendidikan juga menegaskan bahwa perbandingan asesmen lintas negara memerlukan analisis sistematis terhadap desain tes, metodologi analitik, dan kerangka pelaporan untuk memahami bagaimana negara membangun sistem asesmen yang kredibel dan relevan secara kebijakan (van Rijn, 2024). Secara operasional, pendekatan ini bertujuan menyusun

kerangka analitis yang memetakan pola dan variasi desain asesmen nasional lintas negara sebagai dasar refleksi dan pengembangan kebijakan asesmen nasional Indonesia.

Pemilihan negara dalam penelitian ini dilakukan secara purposif untuk merepresentasikan variasi kebijakan asesmen nasional dan tingkat risikonya dalam konteks sosial, budaya, dan geografis yang berbeda. Australia, Amerika Serikat, dan Finlandia dipilih sebagai contoh asesmen *low-stakes* yang berorientasi pada monitoring dan pembelajaran, sementara Singapura, Korea Selatan, dan Tiongkok merepresentasikan tradisi asesmen *high-stakes* dengan dinamika reformasi yang beragam. Jepang dan Vietnam dipilih untuk menggambarkan pendekatan pemantauan mutu yang relatif stabil.

Pengumpulan data dilakukan melalui penelusuran literatur sistematis terhadap sumber akademik bereputasi dan dokumen resmi kebijakan pendidikan. Penelusuran literatur akademik dilakukan menggunakan basis data Scopus, Web of Science, ERIC, dan Google Scholar dengan kata kunci "national assessments", "educational measurement systems", "comparative educational policy", dan "large-scale assessments". Proses ini mengikuti prinsip *systematic search strategy* yang mencakup tahap identifikasi, penyaringan, dan evaluasi kualitas referensi untuk menjamin validitas temuan penelitian (Xiao & Watson, 2019).

Data penelitian mencakup dua kategori utama, yaitu dokumen kebijakan resmi asesmen nasional dari berbagai negara dan artikel jurnal ilmiah terkait asesmen pendidikan. Dokumen kebijakan meliputi laporan dan panduan asesmen nasional, seperti NAPLAN Australia, NAEP Amerika Serikat, PSLE Singapura, NAAA Jepang, NAEA Korea Selatan, sistem evaluasi pendidikan Finlandia (FINEEC), NLA Vietnam, serta Gaokao dan Zhongkao di Tiongkok. Selain itu, penelitian ini juga memanfaatkan literatur ilmiah dari kajian empiris dan teoretis mengenai asesmen pendidikan, termasuk studi tentang dampak asesmen berisiko tinggi, sains pembelajaran dan perkembangan, serta komparabilitas asesmen skala nasional (Klenowski & Wyatt-Smith, 2012; Darling-Hammond *et al.*, 2020; Berman *et al.*, 2020).

Seleksi literatur dilakukan secara bertahap melalui penyaringan judul dan abstrak untuk memastikan relevansi dengan asesmen nasional pada jenjang pendidikan dasar dan menengah, kemudian dilanjutkan dengan seleksi berdasarkan kriteria inklusi yang mencakup relevansi substantif, kredibilitas sumber, rentang tahun publikasi, dan kejelasan landasan metodologis. Evaluasi kualitas literatur mengacu pada kriteria Snyder (2019), meliputi reputasi sumber, relevansi topik, kebaruan, dan ketepatan metodologi. Data yang terpilih dianalisis menggunakan analisis tematik komparatif untuk mengidentifikasi pola, perbedaan, dan hubungan antarkategori kebijakan asesmen lintas negara. Analisis difokuskan pada pemetaan elemen utama asesmen nasional, meliputi dasar kebijakan dan tujuan, struktur domain atau materi, model pelaksanaan, serta konsekuensinya bagi siswa.

Tahap selanjutnya dilakukan pengkodean tematik terhadap isu diagnostik dan akuntabilitas serta spektrum *high-stakes* dan *low-stakes* dengan merujuk pada

pendekatan analisis tematik Braun dan Clarke (2006). Proses ini meliputi pembacaan mendalam sumber data, pengkodean awal unit makna, pengelompokan kode ke dalam tema utama, serta peninjauan dan pemaknaan tema secara komparatif lintas negara. Pendekatan ini memungkinkan identifikasi pola konvergensi kebijakan menuju asesmen diagnostik dan pola divergensi kebijakan yang tetap mempertahankan asesmen *high-stakes*.

Validitas analitis penelitian ini dijaga melalui triangulasi sumber dengan membandingkan temuan dari jurnal ilmiah, dokumen kebijakan, dan laporan resmi (Denzin, 2009). Studi komparatif ini tidak bertujuan menentukan model asesmen terbaik, melainkan memahami strategi negara dalam merancang asesmen sesuai dengan karakteristik sosial dan geografisnya. Pendekatan ini sejalan dengan kritik Fischman (2019) yang menegaskan bahwa praktik asesmen lintas negara tidak dapat diadopsi secara langsung tanpa adaptasi kebijakan. Oleh karena itu, penelitian ini menempatkan perbandingan asesmen nasional dalam perspektif global untuk memperoleh pemahaman kontekstual tentang bagaimana kebijakan asesmen berkembang dan diterapkan di berbagai negara.

## **HASIL DAN PEMBAHASAN**

Kajian komparatif ini membandingkan asesmen nasional di sembilan negara, yaitu Australia, Amerika Serikat, Inggris, Finlandia, Singapura, Jepang, Korea Selatan, Vietnam, dan Tiongkok. Meskipun sama-sama bertujuan memotret perkembangan siswa sebagai acuan pencapaian kompetensi global, setiap negara mengembangkan asesmen dengan perbedaan pada landasan kebijakan, tujuan, struktur materi, serta konsekuensinya bagi siswa. Berdasarkan perbandingan tersebut, temuan penelitian disajikan dalam empat subbagian, yaitu kebijakan dan tujuan asesmen nasional, materi serta model pelaksanaan asesmen lintas negara, konsekuensi asesmen bagi siswa, dan implikasi pengalaman lintas negara bagi Indonesia.

### **Kebijakan dan Tujuan Asesmen Lintas Negara**

Studi lintas negara menunjukkan bahwa kebijakan dan tujuan asesmen nasional sangat dipengaruhi oleh konteks sosial, geografis, politik, dan historis masing-masing negara. Australia, Amerika Serikat, Finlandia, Jepang, dan Vietnam cenderung memanfaatkan asesmen nasional sebagai instrumen monitoring dan diagnostik sistem pendidikan, sementara Tiongkok, Korea Selatan, dan Inggris menggunakannya sebagai bagian dari mekanisme seleksi dan akuntabilitas. Perbedaan kebijakan dan tujuan asesmen tersebut disajikan secara komparatif pada Tabel 1.

Perbandingan di atas menunjukkan adanya pola konvergensi kebijakan asesmen yang cukup signifikan. Hampir di semua negara mulai bergerak dari pola *high-stakes evaluation* (yang menitikberatkan pada kelulusan) menuju *learning-oriented diagnostic assessment* (Darling-Hammond *et al.*, 2020; OECD, 2023). Australia, Finlandia, Jepang dan Vietnam menekankan asesmen sebagai instrumen pembelajaran, bukan seleksi. Sementara Amerika Serikat menggunakan NAEP untuk

Tabel 1 Komparasi Kebijakan dan Tujuan Asesmen Lintas Negara

Negara	Konsep & Arah Kebijakan	Orientasi Kebijakan
Australia (NAPLAN)	Asesmen nasional berbasis kurikulum untuk memetakan kemahiran lintas jenjang & wilayah; diarahkan pada pelaporan standar profisiensi dan diagnostik pembelajaran sambil tetap menjaga akuntabilitas sistem.	Hibrid (diagnostik; akuntabilitas moderat). Non-high-stakes bagi siswa (tidak menentukan kelulusan digunakan untuk pemantauan mutu nasional, analisis kesenjangan, dan umpan balik instruksional ke sekolah.
Amerika Serikat (NAEP)	Sebagai indikator nasional berbasis sampling untuk memonitor tren capaian & kesenjangan; kuat pada akuntabilitas sistem dan bukti kebijakan; bukan merupakan tes sekolah harian.	Akuntabilitas sistem murni. Non-high-stakes untuk siswa/ sekolah (hasil setingkat negara/negara bagian). Dipakai untuk tren nasional; disparitas kelompok, dan evaluasi kebijakan pendidikan.
Inggris – England (SATs/GCSE)	Kombinasi tes standar dan ujian akhir nasional untuk menentukan akuntabilitas dan seleksi jenjang pendidikan. Inggris merupakan negara yang menganut sistem sentralistik	KS2: akuntabilitas sekolah (dampak pada <i>accountability framework</i> ). NRT: indikator non-high-stakes untuk memantau pergeseran performa, membantu penetapan standar ujian menengah (GCSE).
Jepang (NAEA – MEXT)	Asesmen nasional kurikulum-based untuk mengevaluasi implementasi kurikulum & memperbaiki pengajaran; kuat pada perbaikan instruksional dengan transparansi publik.	Akuntabilitas sistem dan perbaikan pembelajaran. Non-high-stakes bagi siswa; hasil dipakai untuk identifikasi kelemahan sistemik dan penyesuaian kurikulum & pengajaran.
Korea Selatan (NAEA – KICE)	Dari akuntabilitas kuat bergeser menuju reform pembelajaran (kurangi <i>over-testing</i> ); National Assessment dipakai sebagai monitoring capaian kurikulum dan penopang kebijakan pemerataan, termasuk digunakan untuk seleksi perguruan tinggi	Akuntabilitas secara hibrid. <i>Non-high-stakes</i> bagi siswa; hasil dimanfaatkan untuk <i>school improvement</i> , pemetaan kesenjangan, dan revisi kurikulum.
Finlandia (FINEEC/KARVI – sample-based)	Formative system evaluation: evaluasi hasil belajar berbasis sampel yang menekankan trust & improvement; tanpa ujian nasional kelulusan di pendidikan dasar; fokus equity & kapasitas guru.	Formatif–diagnostik sistemik. Non-high-stakes secara penuh; data dipakai untuk rekomendasi kebijakan dan dukungan profesional guru, bukan ranking.
Vietnam (ANLAS / National Assessment)	Agenda penguatan mutu dengan mengaitkan asesmen & kurikulum; gunakan temuan PISA sebagai <i>leverage</i> reform; National Assessment difungsikan sebagai diagnostik sistem & koherensi kebijakan.	Monitoring & reformasi kurikulum. <i>Non-high-stakes</i> pada komponen sistemik; ujian kelulusan tetap ada (terpisah). Data dipakai untuk penyesuaian kurikulum–pengajaran

Singapura (PSLE – AL System)	<i>High-capability but growth-oriented</i> : PSLE direformasi ke Achievement Levels (AL 1–8) untuk mengurangi <i>fine ranking</i> ; menekankan kesiapan & penempatan berbasis kompetensi.	<i>High-stakes</i> moderat (untuk penjurusan, bukan kelulusan sisten akhir). Tujuan: keadilan penempatan, fokus pembelajaran, kurangi kompetisi skor tipis.
Tiongkok (NAEQ/Quality Monitoring; APT/Huikao; Gaokao)	<i>Dual track</i> : (a) Monitoring mutu wajib belajar (NAEQ) untuk akuntabilitas sistem; (b) <i>Academic Proficiency Test</i> (APT/Huikao) di SMA untuk kelulusan; (c) Gaokao sebagai seleksi Perguruan Tinggi. Ada arah reformasi ke kualitas holistik.	Akuntabilitas sistem dan <i>high-stake</i> (APT/Gaokao). Monitoring nasional <i>non-high-stakes</i> ; Tes memengaruhi kelulusan SMA; Gaokao high-stake: untuk seleksi masuk PT.

Sumber: Olahan Penulis, 2025

melakukan evaluasi kesetaraan hasil belajar antarnegara bagian tanpa memberi dampak langsung bagi siswa (NCES, 2019). Fenomena ini mencerminkan *global paradigm shift dari testing for accountability* menjadi *assessment for learning and equity* (Black & Wiliam, 2018; Moss *et al.*, 2006).

Australia melalui NAPLAN menerapkan asesmen nasional untuk mengukur literasi dan numerasi sebagai sarana pemetaan mutu pendidikan, khususnya di wilayah rural dan terpencil yang luas, dengan fungsi utama sebagai alat monitoring dan diagnostik sekaligus pemantauan kesenjangan (*equity monitoring*) (Cranley *et al.*, 2022). Pendekatan serupa diterapkan di Amerika Serikat, di mana asesmen nasional tidak menentukan kelulusan, tetapi digunakan untuk memetakan kesenjangan capaian antarras, antarnegara bagian, dan antarkelompok sosial ekonomi. Kebijakan asesmen di AS mencerminkan karakter masyarakat yang plural dan terfragmentasi, sehingga asesmen dirancang untuk menampilkan analisis tren dan *achievement gaps* sebagai dasar perumusan kebijakan pendidikan (Darling-Hammond *et al.*, 2020).

Berbeda dengan Amerika Serikat dan Australia, Inggris memanfaatkan asesmen nasional sebagai instrumen akuntabilitas sekolah dan seleksi jenjang melalui GCSE, sejalan dengan tradisi pendidikan yang lebih tersentralisasi dan penerapan standar asesmen yang seragam (Johnson & Johnson, 2010). Sebaliknya, Finlandia menolak pendekatan high-stakes dengan mengandalkan asesmen berbasis sampel untuk memantau sistem, didukung oleh tingkat kepercayaan tinggi terhadap profesionalisme guru dan kualitas pendidikan yang merata (Sahlberg, 2021). Sementara itu, Singapura yang dikenal memiliki kontrol sistem pendidikan yang kuat, mereformasi fungsi PSLE dari mekanisme seleksi ketat menjadi sistem Achievement Levels (AL) guna mengurangi kompetisi ekstrem dan mengalihkan fokus dari perankingan menuju pertumbuhan kompetensi siswa (Lee & Ho, 2022).

Jepang menerapkan asesmen nasional berbasis prefektur yang berfungsi untuk memetakan dan meningkatkan kualitas pembelajaran. Korea Selatan

mengombinasikan asesmen diagnostik melalui National Assessment of Educational Achievement (NAEA) dan asesmen selektif melalui College Scholastic Ability Test (CSAT), mencerminkan budaya kompetitif yang tetap mempertahankan ujian seleksi meskipun mengakui pentingnya fungsi diagnostik (Kwon *et al.*, 2017). Vietnam mengadaptasi kerangka PISA untuk memetakan kualitas dan kesenjangan pendidikan nasional (World Bank, 2020). Sementara itu, Tiongkok tetap mempertahankan asesmen high-stakes karena fungsi sosialnya sebagai sarana mobilitas sosial, menjadikan ujian sebagai penentu masa depan siswa dan fenomena sosial-budaya yang kuat (Cheng & Hamid, 2025; Liu, 2013, 2016).

Kajian lintas negara menunjukkan adanya pergeseran paradigma asesmen dari penentuan kelulusan menuju asesmen untuk pembelajaran. Asesmen dipahami bukan sekadar alat pengukuran hasil akhir, tetapi bagian integral dari proses pembelajaran yang dapat meningkatkan interaksi kelas dan capaian siswa (Black & Wiliam, 2018). Sejumlah studi menegaskan bahwa asesmen formatif yang sistematis dan berbasis umpan balik memiliki reliabilitas lebih tinggi dalam memprediksi hasil belajar serta berdampak positif terhadap capaian siswa dan pengurangan disparitas sosial ekonomi (Bennett, 2011; Kingston & Nash, 2011; Brookhart, 2019). Pergeseran ini menandai perubahan kebijakan asesmen global dari sekadar *judging learning* menuju *guiding learning*.

Kebijakan asesmen nasional merupakan hasil interaksi kompleks antara pengaruh global, tuntutan akuntabilitas domestik, dan upaya mewujudkan pembelajaran yang berkeadilan. Dalam konteks ini, asesmen nasional berfungsi sebagai instrumen politik pendidikan yang dirancang untuk merespons prioritas nasional masing-masing negara (Kamens & Benavot, 2011). Asesmen skala besar juga menjadi bagian dari governing by numbers dalam pendidikan, di mana data kuantitatif digunakan untuk memengaruhi kebijakan dan memetakan kinerja sistem pendidikan (Steiner-Khamsi *et al.*, 2024).

Negara dengan keragaman sosial tinggi seperti Amerika Serikat memanfaatkan asesmen nasional (NAEP) sebagai instrumen monitoring untuk memetakan kesenjangan dan mendukung kebijakan pendidikan yang lebih berkeadilan dibandingkan pendekatan berisiko tinggi (Darling-Hammond *et al.*, 2020). Sebaliknya, negara dengan populasi besar dan tingkat kompetisi tinggi, seperti Tiongkok dan Korea Selatan, masih mempertahankan asesmen *high-stakes* karena fungsi selektifnya yang signifikan secara sosiologis. Studi Liu (2013, 2016) menunjukkan bahwa Gaokao berperan sebagai simbol nasional dalam mempertahankan legitimasi birokrasi, menegaskan bahwa asesmen nasional merupakan keputusan kebijakan yang dipengaruhi oleh struktur sosial, bukan semata-mata pertimbangan teknis.

Negara dengan tingkat profesionalitas guru yang tinggi dan sistem berbasis kepercayaan, seperti Finlandia, cenderung menerapkan asesmen berskala kecil dan non high-stakes yang berfungsi untuk memahami sistem pendidikan secara makro, bukan untuk mengontrol kinerja guru (Sahlberg, 2021). Sementara itu,

negara dengan wilayah luas dan kondisi geografis menantang, seperti Australia dan Vietnam, memilih asesmen berbasis sampel dan fleksibel agar realistis secara operasional dan relevan dengan tujuan pembelajaran (Carrasco, 2023). Dalam konteks ini, asesmen nasional tidak hanya menuntut keterbandingan lintas konteks, tetapi juga keadilan interpretatif (*interpretive fairness*) (Pellegrino, 2014). Dari perspektif pedagogis, asesmen yang diarahkan untuk mendukung pembelajaran terbukti memberikan dampak yang lebih besar terhadap kualitas belajar siswa dibandingkan asesmen selektif semata (Black & Wiliam, 2018).

### Struktur Materi, Model dan Metode Asesmen Lintas Negara

Asesmen nasional di berbagai negara dirancang dengan struktur materi, model, dan metode yang beragam, mencerminkan filosofi kebijakan pendidikan, kurikulum, serta konteks sosial budaya masing-masing negara. Struktur materi yang diukur memiliki konsekuensi langsung terhadap beban tes, orientasi belajar, dan implikasi bagi siswa. Meskipun banyak negara memfokuskan asesmen pada literasi dan numerasi, kajian lintas negara menunjukkan adanya variasi dalam penambahan materi, kedalaman pengukuran, dan konsekuensi asesmen bagi siswa (Tabel 2).

Tabel 2 Perbandingan Struktur Materi, Model dan Metode Asesmen Lintas Negara

Negara	Domain/Materi Asesmen Utama	Model Psikometrik	Metode Pelaksanaan
Australia (NAPLAN)	Reading, writing, numeracy, language conventions	Rasch-based IRT <sup>1</sup>	CBT <sup>2</sup> & paper (transisi bertahap)
AS (NAEP)	Reading, math, science, civics, arts	Multidimensional IRT	Sampling-based, CBT
Inggris (SATs/ GCSE)	English, math, science, humanities	CTT <sup>3</sup> dan partial IRT	Paper-based exam
Finlandia (FINEEC)	Literacy, numeracy, scientific literacy	IRT (sample)	Sampling, school-based test
Singapura (PSLE)	English, math, science	CTT with scaling <sup>4</sup>	National exam
Jepang (NAAA)	Math & literacy	IRT	National sampling
Korea (NAEA/ CSAT)	Literacy, math, social studies, science, English	IRT & CTT (CSAT <sup>5</sup> )	Paper/computer
Vietnam (NLA)	Reading, math, science (PISA-linked)	IRT	Sample-based
Tiongkok (Gaokao/ Zhongkao)	Mandarin, English, math, sciences	CTT	Pen-and-paper

Sumber: Olahan Penulis, 2025

Keterangan:

<sup>1</sup> Item Response Theory (IRT) adalah pendekatan psikometrik modern yang memodelkan hubungan antara kemampuan peserta dan probabilitas menjawab benar setiap butir soal,

sehingga memungkinkan perbandingan hasil yang lebih adil lintas wilayah, waktu, dan kelompok murid

- 2 Computer-Based Testing (CBT) mengacu pada pelaksanaan asesmen berbasis komputer.
- 3 Classical Test Theory (CTT) merupakan pendekatan pengukuran tradisional yang memandang skor tes sebagai gabungan antara kemampuan sejati peserta dan kesalahan pengukuran. Pendekatan ini relatif sederhana namun sangat bergantung pada karakteristik kelompok peserta dan paket soal tertentu.
- 4 CTT with scaling merujuk pada pengembangan CTT dengan teknik penskalaan skor untuk meningkatkan perbandingan antar bentuk tes, meskipun masih memiliki keterbatasan dalam komparabilitas lintas populasi.
- 5 CSAT (College Scholastic Ability Test) merupakan contoh asesmen berisiko tinggi (high-stakes) yang berfungsi sebagai ujian seleksi masuk perguruan tinggi dan memiliki konsekuensi langsung terhadap masa depan akademik murid.

Perbandingan lintas negara menunjukkan bahwa materi asesmen nasional umumnya berfokus pada literasi dan numerasi sebagai kompetensi dasar bagi keberhasilan akademik dan sosial, sejalan dengan kerangka OECD dan PISA. Meskipun demikian, setiap negara mengembangkan struktur materi dengan karakteristik yang berbeda (Tabel 2). Di Australia, NAPLAN mengukur literasi membaca, menulis, komunikasi berbahasa, dan numerasi, mencerminkan penekanan pada kemampuan komunikasi dan penerapan numerasi dalam kehidupan nyata. Pelaksanaannya telah beralih dari *paper-based* ke *computer-based* dengan penerapan bertahap Computerized Adaptive Testing (CAT) untuk memungkinkan diferensiasi tingkat kesulitan antarsiswa.

Di Amerika Serikat, NAEP menilai cakupan materi yang lebih luas, tidak hanya literasi membaca dan matematika, tetapi juga literasi kewarganegaraan, seni, dan geografi. Asesmen ini menggunakan matrix sampling sehingga setiap siswa mengerjakan sebagian butir soal, memungkinkan cakupan materi yang luas tanpa membebani siswa secara berlebihan. Sementara itu, Inggris mempertahankan materi inti seperti bahasa Inggris, matematika, dan sains dalam asesmen nasional, serta menambahkan mata pelajaran humaniora, kejuruan, dan kreatif pada GCSE. Model pengukuran di Inggris masih didominasi *Classical Test Theory* (CTT), meskipun pada beberapa mata pelajaran telah digunakan *Item Response Theory* (IRT) untuk keperluan penskalaan.

Finlandia sebagai negara dengan asesmen non high-stakes menerapkan asesmen berbasis sampel yang mengukur literasi, numerasi, dan sains, dengan peran guru yang sentral sehingga asesmen terintegrasi dengan penilaian kelas. Meskipun cakupan materi tampak terbatas, pendekatan berbasis praktik pembelajaran memungkinkan pengukuran kompetensi yang luas. Sementara itu, Singapura melalui PSLE menilai bahasa Inggris, matematika, dan sains dengan penekanan pada pemecahan masalah tingkat tinggi, serta menerapkan sistem Achievement Levels (AL) yang berorientasi pada pertumbuhan kemampuan siswa dibandingkan peringkat.

Di Jepang, National Assessment of Academic Ability (NAAA) menilai literasi dan matematika dengan fokus pada pemecahan masalah sebagai sarana memantau efektivitas kurikulum dan perkembangan pendidikan nasional. Korea Selatan menerapkan dua model asesmen, yaitu National Assessment of Educational Achievement (NAEA) yang bersifat diagnostik dan College Scholastic Ability Test (CSAT) sebagai asesmen selektif untuk masuk perguruan tinggi. Materi CSAT mencakup bahasa Korea, matematika, bahasa Inggris, serta mata pelajaran ilmu sosial dan sains yang dapat dipilih oleh siswa.

Vietnam menerapkan National Learning Assessment (NLA) yang mengacu pada kerangka PISA dengan fokus pada literasi dan numerasi berbasis kompetensi. Sebaliknya, Tiongkok tetap mempertahankan sistem asesmen high-stakes melalui Gaokao yang menguji bahasa Mandarin, matematika, dan bahasa Inggris sebagai mata pelajaran wajib, serta pilihan ilmu alam atau ilmu sosial. Karakter kompetitif sistem tersebut menjadikan materi dan tuntutan asesmen di Tiongkok sangat tinggi bagi siswa.

Analisis lintas negara tentang struktur materi, model dan metode pelaksanaan memperlihatkan bahwa materi asesmen di tingkat nasional tidak netral. Pilihan materi mencerminkan urgensi dalam memahami kompetensi inti bagi warga/siswanya untuk meningkatkan daya saing. Ada pilihan untuk mempertimbangkan kompetensi dasar dan kompetensi untuk masa depan. Penelitian yang dilakukan Mullis (2023) tentang TIMSS *framework* menegaskan bahwa materi numerasi dan literasi memprediksi keberhasilan pembelajaran tingkat lanjut dan karir kerja, sehingga hampir semua negara tetap mempertahankan domain tersebut sebagai materi inti asesmen. Desain materi asesmen berkaitan dengan tujuan pendidikan nasional. Darling-Hammond & Adamson (2015) menunjukkan bahwa asesmen yang menekankan kompetensi tinggi memiliki kontribusi pada kualitas pembelajaran jangka panjang. Fokus pada materi tertentu cenderung akan mengabaikan materi lainnya. Asesmen yang fokus pada literasi dan numerasi saja menyebabkan pengabaian materi lain (Ronksley-Pavia, 2023). Konsekuensinya, terjadi peningkatan tekanan terhadap literasi dan numerasi saja dan pada akhirnya berdampak pada pengajaran yang berfokus pada tes (Cranley *et al.*, 2022). Berbeda halnya dengan model TIMSS, studi Raeder (2022) menunjukkan bahwa pengukuran numerasi menerapkan *vertical scaling* antar jenjang kelas sehingga materi numerasi tidak hanya diukur sebagai kompetensi dasar, tetapi kompetensi yang terus berkembang. Hal ini menunjukkan bahwa struktur materi yang lebih kompleks dapat memberikan gambaran tren perkembangan siswa. Konsekuensi bagi siswa adalah bahwa mereka diposisikan untuk mengembangkan terus menerus kemampuan sehingga orientasi belajar menjadi longitudinal.

Kajian lintas negara menunjukkan bahwa meskipun struktur materi asesmen nasional bervariasi, hampir semua negara menempatkan literasi dan numerasi sebagai domain inti, sehingga pilihan tersebut dalam Tes Kemampuan Akademik (TKA) Indonesia memiliki legitimasi yang kuat. Secara internasional, desain TKA

paling relevan mengadopsi model asesmen low-stakes dan diagnostik seperti di Vietnam dan Jepang, yang berfokus pada pemetaan mutu dan kesenjangan belajar tanpa konsekuensi kelulusan. Kesamaan tantangan ketimpangan wilayah dan infrastruktur menjadikan pendekatan Vietnam—dengan penggunaan model psikometrik berbasis Item Response Theory (IRT)—sangat kontekstual bagi Indonesia karena mendukung komparabilitas dan keadilan interpretasi hasil. Pengalaman Australia dan Vietnam juga menegaskan pentingnya fleksibilitas pelaksanaan melalui pendekatan hibrid digital dan non-digital agar asesmen tidak memperkuat bias struktural. Dengan demikian, TKA sebaiknya diposisikan sebagai asesmen nasional yang berorientasi pada pembelajaran dan keadilan sosial, dengan fokus materi yang terbatas namun mendalam pada literasi dan numerasi kontekstual, serta berfungsi sebagai alat pemetaan dan diagnosis sistem pendidikan, bukan sebagai instrumen seleksi atau penentu kelulusan.

### **Konsekuensi Asesmen bagi Siswa**

Pilihan asesmen—baik dari sisi kebijakan, struktur materi, model, maupun metode pelaksanaan—selalu membawa konsekuensi langsung bagi siswa. Temuan lintas negara menunjukkan adanya perbedaan yang signifikan dalam konsekuensi asesmen nasional terhadap pengalaman dan posisi siswa dalam sistem pendidikan, sebagaimana dirangkum dalam Tabel 3.

Di negara-negara dengan asesmen berisiko tinggi (*high-stakes*) seperti Tiongkok, Korea Selatan, Inggris, dan Singapura, asesmen nasional memiliki konsekuensi langsung terhadap masa depan akademik siswa. Dalam sistem seperti Gaokao dan CSAT, hasil asesmen menjadi penentu utama akses ke perguruan tinggi unggulan sehingga dipersepsikan sebagai penentu masa depan individu. Konsekuensinya, berkembang budaya belajar yang sangat intens dan kompetitif. Meskipun sistem ini dapat meningkatkan motivasi akademik, dampak negatif berupa tekanan psikologis yang tinggi juga tidak terhindarkan. Studi Liu (2013, 2016) menunjukkan bahwa siswa di Tiongkok mengalami stres akademik yang signifikan menjelang Gaokao, dengan beban belajar yang melampaui standar internasional. Di Korea Selatan, CSAT bahkan menjadi fenomena sosial tahunan yang memengaruhi aktivitas publik, mencerminkan kuatnya tekanan struktural terhadap siswa dan keluarga. Dinamika ini membentuk identitas sosial dan budaya berbasis kompetisi, namun sekaligus meningkatkan risiko kelelahan mental, depresi, dan tekanan psikososial.

Pada sistem asesmen moderate-stakes seperti Singapura, sebelum penerapan *Achievement Levels* (AL), perbedaan skor yang sangat kecil dapat mengubah jalur pendidikan siswa, sehingga meningkatkan kecemasan siswa dan orang tua. Setelah sistem AL diterapkan, tingkat kompetisi relatif menurun karena orientasi bergeser dari perankingan ke rentang capaian, meskipun ekspektasi budaya terhadap prestasi tetap tinggi.

Tabel 3 Perbandingan Konsekuensi Asesmen bagi Siswa

Negara	Dampak Akademik	Dampak Psikologis	Dampak Sosial
Australia (NAPLAN) – Low-moderate stakes	Meningkatkan <i>awareness</i> terhadap literasi-numerasi; umpan balik individu	Stress moderat; beberapa studi mencatat kecemasan siswa	Perbedaan antar sekolah direproduksi
AS (NAEP) – Low stakes	Tidak berdampak pada siswa individual ( <i>no stakes</i> )	Minimal karena sampling	Membuka diskursus ketimpangan
Inggris (GCSE/SATs) – High stakes	Menentukan jenjang pendidikan; mempengaruhi kurikulum	Tinggi: anxiety, test pressure	Stratifikasi akademik
Finlandia (FINEEC) – No/Low stakes	Tidak berdampak individual; guru terbantu dalam perbaikan pembelajaran	Sangat rendah	Mendorong kultur kepercayaan
Singapura (PSLE) – Moderate stakes	Menentukan penempatan; memotivasi belajar	Tinggi: stress komunitas ( <i>parental pressure</i> )	Mobilitas sosial melalui sekolah
Jepang (NAEA) – Low stakes	Informasi perbaikan pembelajaran; tidak memengaruhi siswa langsung	Rendah	Diskursus mutu antar prefektur
Korea Selatan (CSAT/NAEA) – Very high stakes	CSAT menentukan masa depan pendidikan; NAEA diagnostik	Sangat tinggi pada CSAT	Budaya kompetisi ekstrem
Vietnam (NLA) – Low stakes	Perbaikan mutu dan identifikasi kesenjangan	Rendah	Representasi ketimpangan wilayah
Tiongkok (Gaokao/Zhongkao) – Very high stakes	Menentukan akses pendidikan tinggi; <i>strongly predictive</i>	Sangat tinggi; <i>kelelahan mental</i>	Mobilitas sosial berbasis nilai

Sumber: Olahan Penulis, 2025

Negara yang menerapkan asesmen low-stakes seperti Australia dan Jepang menunjukkan konsekuensi yang lebih seimbang bagi siswa. Di Australia, NAPLAN tidak menentukan kelulusan dan berfungsi sebagai asesmen diagnostik, namun tekanan tetap muncul akibat praktik perbandingan dan perankingan antarsekolah, terutama di sekolah unggulan. Sebaliknya, di Jepang asesmen nasional digunakan murni sebagai alat pemantauan mutu pendidikan di tingkat prefektur, sehingga tekanan hampir tidak dirasakan oleh siswa dan lebih dialihkan kepada sekolah serta pemerintah daerah.

Vietnam sebagai negara dengan pendekatan low-stakes memanfaatkan asesmen nasional untuk memetakan kualitas dan kesenjangan pendidikan antarwilayah. Konsekuensi langsung bagi siswa relatif rendah, namun hasil asesmen memberikan dampak signifikan bagi perumusan kebijakan dan perbaikan sistem

pendidikan secara nasional.

Berbeda dengan itu, negara yang menerapkan no-stakes testing seperti Finlandia hampir tidak memberikan konsekuensi apa pun bagi siswa. Asesmen yang dikelola oleh Finnish Education Evaluation Centre (FINEEC) tidak ditujukan untuk menilai individu, melainkan untuk memahami kinerja sistem pendidikan secara makro. Guru menjadi aktor utama dalam asesmen formatif di kelas, sehingga asesmen nasional tidak menimbulkan beban psikologis maupun akademik bagi siswa.

Temuan lintas negara ini menunjukkan bahwa tingkat stakes, cakupan asesmen, serta konteks sosial dan budaya merupakan faktor kunci yang menentukan dampak asesmen nasional terhadap pengalaman belajar dan kesejahteraan siswa.

High-stakes testing cenderung meningkatkan performa tes, tetapi tidak selalu berdampak pada peningkatan kualitas pembelajaran. Tekanan asesmen berisiko tinggi sering memicu inflasi nilai tanpa mencerminkan kemampuan aktual siswa (Koretz, 2017). Sebaliknya, asesmen low-stakes menghasilkan gambaran capaian yang lebih akurat karena meminimalkan perilaku manipulatif siswa, sehingga lebih kuat dari sisi validitas konsekuensial (Messick, 1995).

Dari sisi psikologis, asesmen berisiko tinggi berkorelasi dengan kecemasan akademik yang lebih tinggi, sedangkan asesmen yang tidak dikaitkan dengan kelulusan cenderung mendorong motivasi intrinsik siswa (Putwain & Daly, 2014). Selain itu, asesmen *high-stakes* berpotensi memperkuat ketimpangan sosial karena skor tes sering merefleksikan latar belakang sosial ekonomi dibandingkan kemampuan murni siswa (Reardon, 2011).

Meskipun sistem seperti Gaokao di Tiongkok berfungsi sebagai sarana mobilitas sosial, manfaat tersebut lebih banyak dinikmati oleh siswa dengan akses ke pendidikan berkualitas. Sebaliknya, asesmen low-stakes lebih berorientasi pada keadilan, karena memungkinkan identifikasi kesenjangan belajar tanpa memberikan konsekuensi individual yang merugikan. Pola perbedaan dampak tersebut dirangkum dalam Tabel 4.

Indonesia berada pada posisi antara, tidak memiliki budaya kompetisi ekstrem seperti Korea Selatan atau Tiongkok, namun juga belum didukung sistem kepercayaan profesional yang kuat seperti Finlandia. Dalam kondisi tersebut, asesmen berisiko tinggi berpotensi menimbulkan tekanan psikologis bagi siswa tanpa dukungan struktural yang memadai. Di tengah ketimpangan antarwilayah,

Tabel 4 Dampak Pemilihan Tingkat Asesmen

<b>Level <i>High-Stakes</i></b>	<b>Dampak Terhadap Siswa</b>
Tinggi	Stres tinggi, kompetisi ekstrem, kesenjangan sosial melebar
Sedang	Tekanan moderat, masih berpotensi mengarah ke <i>teaching-to-test</i>
Rendah	Fokus pembelajaran meningkat, dampak psikologis rendah
Nol	Motivasi intrinsik tinggi, asesmen untuk perbaikan sistem

Sumber: Olahan Penulis, 2025

latar belakang sosial ekonomi, dan kualitas sekolah yang beragam, asesmen dengan konsekuensi individual langsung cenderung mereproduksi ketidakadilan pendidikan. Oleh karena itu, penempatan Tes Kemampuan Akademik (TKA) sebagai asesmen non high-stakes memiliki implikasi strategis untuk menurunkan kecemasan siswa dan menggeser fungsi asesmen menuju pemetaan capaian serta diagnosis kebutuhan belajar. Namun demikian, pengalaman lintas negara menunjukkan bahwa tekanan tetap dapat muncul secara tidak langsung melalui pelabelan sekolah atau penggunaan hasil asesmen yang sempit oleh publik.

Konsekuensi TKA bagi siswa Indonesia sangat bergantung pada kebijakan pelaporan dan tindak lanjut. Jika hasil TKA digunakan untuk intervensi pembelajaran, dukungan remedial, dan penguatan sekolah, asesmen dapat meningkatkan kesejahteraan belajar dan pemerataan mutu. Namun, jika dijadikan alat kompetisi tersembunyi, TKA berpotensi menimbulkan dampak negatif seperti asesmen *high-stakes*. Oleh karena itu, kebijakan TKA harus melindungi siswa dari tekanan sosial dan psikologis serta memastikan asesmen menjadi alat untuk meningkatkan kualitas dan keadilan pendidikan. Siswa perlu dipandang sebagai subjek pembelajaran utuh, dan asesmen yang adil serta berorientasi pada pembelajaran mencerminkan komitmen negara dalam melindungi martabat, kesehatan mental, dan hak setiap anak untuk belajar secara bermakna.

### **Implikasi Perbandingan Asesmen Lintas Negara bagi Indonesia**

Perbandingan asesmen lintas negara menunjukkan bahwa efektivitas asesmen nasional bergantung pada kesesuaian fungsi asesmen dengan konteks ketimpangan pendidikan. Di Indonesia, TKA sebaiknya ditempatkan sebagai asesmen non high-stakes karena adanya ketimpangan pendidikan antarwilayah. Data BPS (2025) menunjukkan rata-rata lama sekolah penduduk usia 15 tahun ke atas di perkotaan 10,08 tahun, sementara di perdesaan hanya 7,97 tahun. Selain itu, sekitar 53% anak akhir SD belum mampu memahami bacaan (World Bank, 2024), menunjukkan banyak siswa memasuki pendidikan lanjutan dengan fondasi belajar yang rapuh dan tidak merata.

Dalam pendekatan *Assessment for Learning* (AfL), asesmen menjadi bagian integral dari proses pembelajaran yang berfungsi memberikan umpan balik, bukan sebagai mekanisme seleksi akhir (Bennett, 2011; Black & Wiliam, 2018). Negara-negara seperti Finlandia, Jepang, dan Vietnam yang menempatkan asesmen nasional dalam kerangka AfL cenderung menggunakan asesmen berskala nasional untuk diagnosis sistem dan perbaikan pembelajaran, bukan untuk menentukan kelulusan atau masa depan individual siswa. Temuan lintas negara dalam kajian ini menunjukkan bahwa pendekatan tersebut menghasilkan konsekuensi akademik dan psikologis yang lebih konstruktif dibandingkan sistem *high-stakes*, terutama dalam konteks masyarakat yang heterogen dan berketimpangan.

Kualitas asesmen tidak hanya bergantung pada ketepatan pengukuran, tetapi juga pada dampak sosial dan kebijakan yang ditimbulkannya (Messick, 1995).

Asesmen berisiko tinggi dapat menimbulkan kecemasan akademik, penyempitan kurikulum, dan reproduksi ketimpangan sosial karena dipengaruhi latar belakang sosial ekonomi dan akses terhadap sumber belajar (Koretz, 2017; Putwain & Daly, 2014). Di Indonesia, dengan ketimpangan kualitas sekolah dan dukungan keluarga yang tinggi, asesmen dengan konsekuensi individual langsung berisiko mengubah ketidaksetaraan struktural menjadi kegagalan personal siswa. Oleh karena itu, dari perspektif validitas konsekuensial, asesmen nasional harus dirancang untuk meminimalkan kerugian sosial dan psikologis, terutama bagi siswa di daerah tertinggal dan kelompok rentan.

Dalam kebijakan nasional, Permendikdasmen Nomor 9 Tahun 2025 menegaskan TKA sebagai pengukuran capaian akademik siswa dengan fungsi ganda: seleksi akademik pada konteks tertentu dan penjaminan mutu. Kajian lintas negara menekankan pentingnya melengkapi ekosistem asesmen, di mana TKA digunakan untuk pelaporan capaian terstandar dan pemetaan mutu serta kesenjangan antar sekolah/daerah. Ketegasan fungsi ini penting agar tekanan pada siswa tidak meningkat dan hasil asesmen tidak disalahgunakan untuk memperdalam ketimpangan.

Meskipun secara konseptual TKA sejalan dengan paradigma asesmen berorientasi pembelajaran dan keadilan sosial, implementasinya di Indonesia berpotensi menghadapi sejumlah hambatan struktural dan kultural yang perlu diantisipasi. Ketimpangan akses dan kesiapan digital masih signifikan, terutama antara wilayah perkotaan dan perdesaan serta barat dan timur Indonesia sehingga pelaksanaan asesmen berbasis digital berisiko memperdalam eksklusi jika tidak disertai skema adaptif dan dukungan infrastruktur (BPS, 2024; OECD, 2001). Selain itu, pengalaman historis dengan ujian nasional berisiko tinggi membentuk budaya asesmen yang selektif dan kompetitif, sehingga perubahan menuju asesmen *low-stakes* dapat menghadapi resistensi dari orang tua, sekolah, dan pemerintah daerah yang masih memaknai skor sebagai indikator keberhasilan utama (Steiner-Khamsi et al., 2024). Hambatan lain terletak pada kapasitas penyelenggara dan literasi asesmen guru, karena asesmen berorientasi pembelajaran hanya efektif apabila hasilnya dapat ditafsirkan dan dimanfaatkan secara pedagogis; tanpa penguatan kapasitas institusional dan profesional berkelanjutan, TKA berisiko menjadi aktivitas pengukuran administratif semata (Bennett, 2011; Darling Hammond & Adamson, 2015).

Pelajaran dari negara lain menunjukkan variasi tata kelola asesmen yang dapat dipetakan ke dua kutub yaitu (a) asesmen yang berorientasi diagnostik dan *low-stakes*, dan (b) asesmen yang melekat pada mekanisme akuntabilitas atau seleksi yang lebih kuat. Australia, misalnya, memposisikan NAPLAN sebagai asesmen berskala nasional dengan pelaporan dan analisis teknis yang kuat, termasuk perhatian pada perbedaan hasil menurut wilayah/karakteristik dan penguatan kualitas analisis tes (ACARA, 2025). Amerika Serikat melalui NAEP (*Nation's Report Card*) menempatkan asesmen nasional untuk memantau capaian nasional dan negara

bagian (NCES, 2019). Inggris mengatur asesmen *Key Stage* sebagai bagian dari tata kelola kurikulum nasional dan pelaporan, dengan pedoman administratif yang ketat (Standards and Testing Agency, 2024). Singapura memperkenalkan pembaruan sistem skor PSLE melalui *Achievement Levels* untuk menekankan capaian absolut dan mengurangi kompetisi peringkat yang sangat tajam, walau tetap berada dalam ekosistem seleksi transisi sekolah (MOE Singapore, 2022). Di sisi lain, Finlandia menonjol sebagai contoh evaluasi pendidikan yang kuat melalui asesmen berbasis sampel dan evaluasi nasional yang mendukung pengambilan keputusan kebijakan tanpa menempatkan ujian nasional sebagai penentu nasib individu pada pendidikan dasar (FINEEC, 2022; OECD, 2024).

Temuan lintas negara ini menegaskan bahwa praktik asesmen tidak dapat ditransfer secara langsung tanpa adaptasi konteks. Pendekatan *policy borrowing and lending* menekankan bahwa keberhasilan kebijakan ditentukan oleh kesesuaian fungsi dengan struktur sosial dan kapasitas institusional, bukan oleh peniruan bentuk kebijakan negara lain (Bray et al., 2007; Steiner-Khamsi et al., 2024). Oleh karena itu, adopsi model asesmen kompetitif ala Korea Selatan atau Tiongkok, yang mempertahankan ujian berisiko tinggi sebagai mekanisme mobilitas sosial tidak sepenuhnya kompatibel dengan konteks Indonesia yang bercirikan heterogenitas budaya, ketimpangan wilayah, dan dukungan belajar yang belum merata.

Bagi Indonesia, relevansi *low-stakes assessment* menjadi semakin kuat karena tantangan pemerataan lebih dominan dibanding sekadar peningkatan rerata capaian. Sistem asesmen yang terlalu berorientasi seleksi akan cenderung mendorong respons sosial seperti intensifikasi bimbingan belajar, *teaching to the test*, dan kompetisi simbolik antarsekolah yang pada banyak konteks justru memperlebar jarak antara keluarga berdaya dan tidak berdaya. Literatur akademik menunjukkan bahwa kecemasan tes berkorelasi negatif dengan capaian dan dapat dipicu oleh persepsi risiko tinggi terhadap masa depan siswa (Jerrim, 2023; Mostafa, 2017). Dalam masyarakat dengan ketimpangan tinggi, konsekuensi psikologis dan sosial ini tidak netral. Kelompok rentan menanggung beban lebih besar karena keterbatasan dukungan belajar di rumah maupun akses sumber daya pendidikan. Hal ini berarti, pilihan kebijakan asesmen bukan hanya soal akurasi pengukuran, melainkan soal keadilan konsekuensial, bagaimana hasil asesmen memengaruhi peluang hidup siswa secara nyata.

Kajian lintas negara juga menegaskan bahwa untuk negara dengan disparitas wilayah, fleksibilitas moda pelaksanaan, isu kebijakanlah yang menentukan. Kebijakan yang mengizinkan transisi bertahap (hibrid digital dan non-digital) dan desain pelaporan yang menekankan diagnosis, bukan peringkat, terbukti lebih kompatibel bagi konteks yang infrastrukturnya tidak homogen. Hal ini relevan untuk Indonesia mengingat tantangan akses dan kualitas layanan pendidikan yang beragam antarwilayah. Apabila pelaksanaan TKA terlalu seragam tanpa mitigasi, risiko bias struktural meningkat. Hasil tes dapat lebih banyak merefleksikan tentang ketersediaan peluang belajar dibanding kemampuan akademik.

Berdasarkan analisis lintas negara, sistem asesmen yang efektif pada negara berketimpangan tinggi cenderung menempatkan asesmen nasional sebagai instrumen perbaikan sistem (pemetaan, diagnosis, intervensi), sementara fungsi seleksi dibatasi dan diatur ketat agar tidak mendominasi seluruh praktik pembelajaran. Dengan memasukkan data ketimpangan (desa–kota) dan indikator krisis belajar (*learning poverty*), argumentasi *low-stakes assessment* tidak lagi bersifat normatif semata, melainkan berbasis kebutuhan kebijakan yang konkret untuk melindungi siswa dari konsekuensi sosial-psikologis yang tidak adil sekaligus mengarahkan data asesmen untuk mempercepat pemerataan mutu (BPS, 2025; World Bank, 2024).

Agar TKA relevan dan adil, kebijakan pelaporan perlu mengedepankan *diagnostic reporting* (profil kompetensi dan rekomendasi tindak lanjut). Kebijakan pemanfaatan hasil perlu dipadukan dengan intervensi pembelajaran dan penguatan sekolah, serta desain implementasi perlu adaptif terhadap kesiapan daerah. Pada akhirnya, asesmen nasional yang manusiawi dan berkeadilan bukan sekadar pilihan teknis, tetapi strategi negara untuk memastikan data pengukuran tidak mengabadikan ketimpangan, melainkan menjadi perangkat utama untuk mempercepat pemerataan kualitas pendidikan.

## **SIMPULAN DAN SARAN**

### **Simpulan**

Kajian ini menyimpulkan bahwa desain asesmen nasional tidak dapat dilepaskan dari konteks sosial, budaya, geografis dan ketimpangan sistem pendidikan di suatu negara. Berikut ini beberapa simpulan dari hasil kajian yang telah dilakukan:

1. Fungsi asesmen nasional harus ditentukan oleh konteks sistem pendidikan, bukan mengadopsi model global secara seragam. Bukti lintas negara menunjukkan bahwa asesmen berisiko tinggi (*high-stakes*) cenderung bermasalah ketika diterapkan dalam sistem pendidikan yang masih timpang karena memperlebar disparitas capaian dan meningkatkan beban psikologis murid.
2. Asesmen berorientasi diagnostik dan pembelajaran (*low-stakes*) lebih konsisten dengan tujuan pemerataan mutu, terutama pada negara yang memiliki wilayah luas dan beragam serta kapasitas sekolah yang tidak merata. Negara-negara seperti Australia, Vietnam, dan Amerika Serikat secara sadar membatasi fungsi selektif asesmen nasional dan memaksimalkan perannya sebagai instrumen pemetaan dan perbaikan sistem.
3. Dalam konteks Indonesia, posisi TKA secara konseptual lebih tepat sebagai asesmen pemetaan dan pembelajaran. Model ini lebih memberikan keadilan dan perlindungan kepada kelompok siswa yang rentan.
4. Pilihan kebijakan TKA merupakan keputusan normatif dan pedagogis bukan semata teknokratis. Asesmen yang adil adalah asesmen yang meminimalkan risiko kerugian sosial dan psikologis, serta mampu memaksimalkan fungsi perbaikan pembelajaran dan penguatan mutu sekolah.

Hasil kajian ini menguatkan bahwa pendekatan studi komparatif lintas negara dapat memberikan pijakan empiris untuk merancang asesmen nasional di Indonesia yang lebih kontekstual dan berkeadilan. Dengan demikian, transformasi asesmen di Indonesia tidak hanya menjadi perubahan teknis, tetapi perubahan paradigma menuju pendidikan yang inklusif, kontekstual dan berpihak pada pembelajaran.

## **Saran**

Berdasarkan temuan dan analisis komparatif lintas negara, beberapa saran strategis diajukan agar implementasi Tes Kemampuan Akademik (TKA) di Indonesia dapat efektif, adil, dan berorientasi pada pembelajaran.

### *1. Kebijakan Asesmen Nasional*

Pemerintah perlu menegaskan secara konsisten bahwa TKA merupakan asesmen diagnostik nasional yang mendukung pembelajaran, bukan alat seleksi atau penentu kelulusan. Penegasan ini harus tercermin dalam regulasi turunan, pedoman teknis, dan komunikasi publik. Sistem pelaporan TKA harus dirancang untuk menampilkan profil capaian kompetensi dan kesenjangan belajar, bukan skor tunggal atau ranking, serta dikaitkan langsung dengan mekanisme tindak lanjut kebijakan seperti penguatan sekolah, program remedial, dan intervensi berbasis wilayah.

### *2. Penguatan Kapasitas Guru dan Sekolah*

Keberhasilan TKA sangat bergantung pada kemampuan guru dan sekolah dalam menafsirkan serta memanfaatkan data asesmen. Oleh karena itu, pemerintah perlu memberikan prioritas pada penguatan literasi asesmen guru melalui program pengembangan profesional berkelanjutan yang berfokus pada pemanfaatan hasil TKA untuk perbaikan pembelajaran. Pelatihan tidak hanya mencakup pemahaman teknis skor, tetapi juga menerjemahkan data asesmen menjadi strategi pembelajaran yang berdiferensiasi, menuju perbaikan, dan pengayaan yang kontekstual dengan kebutuhan murid.

### *3. Infrastruktur dan Model Implementasi Asesmen*

Pemerintah wajib memastikan kesiapan infrastruktur asesmen, khususnya di daerah terpencil, tertinggal, dan di kepulauan untuk menjamin keadilan pada saat pelaksanaan. TKA perlu diimplementasikan melalui pendekatan hibrid (digital dan non digital) yang adaptif terhadap kondisi daerah, disertai dengan kebijakan kompensasi logistik dan dukungan teknis bagi satuan pendidikan dengan keterbatasan sarana. Pendekatan ini penting agar asesmen tidak menjadi sumber bias struktural yang merugikan siswa akibat ketimpangan akses teknologi.

### *4. Prinsip Keadilan Sosial dan Inklusivitas*

TKA perlu secara eksplisit mengadopsi prinsip inklusivitas lintas budaya, bahasa, dan kondisi sosial. Pemerintah disarankan untuk mengembangkan adaptasi konten

asesmen yang peka terhadap keberagaman konteks lokal, termasuk penggunaan bahasa yang ramah bagi murid dengan latar belakang budaya dan bahasa daerah tertentu. Selain itu, hasil TKA harus dapat dimanfaatkan sebagai dasar kebijakan afirmatif bagi kelompok rentan sehingga asesmen nasional berfungsi sebagai alat untuk mengurangi ketimpangan pendidikan.

Dengan struktur kebijakan yang jelas, penguatan kapasitas pendidikan, dukungan infrastruktur yang adil dan komitmen pada keadilan sosial, TKA berpotensi menjadi instrumen strategis dalam transformasi asesmen nasional Indonesia yang berorientasi pembelajaran inklusif dan berkelanjutan.

## PUSTAKA ACUAN

- ACARA. (2025). *NAPLAN Technical Report 2024*.
- Anghel, E., Khorramdel, L., & von Davier, M. (2024). The use of process data in large-scale assessments: a literature review. *Large-Scale Assessments in Education*, 12(1), 13. <https://doi.org/10.1186/s40536-024-00202-1>
- Bennett, R.E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Berman, A., Haertel, E., & Pellegrino, J. (Eds.). (2020). *Comparability of Large-Scale Educational Assessments: Issues and Recommendations*. National Academy of Education. <https://doi.org/10.31094/2020/1>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- BPS. (2024). *Statistik Pendidikan*. Badan Pusat Statistik.
- BPS. (2025). *Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Klasifikasi Desa*. Badan Pusat Statistik.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bray, M., Adamson, B., & Mason, M. (2007). *Comparative Education Research: Approaches and Methods*. Springer Science & Business Media.
- Brislin, R. W. (1976). Comparative research methodology: Cross cultural studies. *International Journal of Psychology*, 11(3), 215–229. <https://doi.org/10.1080/00207597608247359>

- Brookhart, S. M., & McMillan, J. H. (2019). *Classroom Assessment and Educational Measurement*. Routledge. <https://doi.org/10.4324/9780429507533>
- Carrasco, D., Rutkowski, D., & Rutkowski, L. (2023). The advantages of regional large-scale assessments: Evidence from the ERCE learning survey. *International Journal of Educational Development*, *102*, 102867. <https://doi.org/10.1016/j.ijedudev.2023.102867>
- Cheng, Y., & Hamid, M. O. (2025). Social impact of Gaokao in China: a critical review of research. *Language Testing in Asia*, *15*(1), 22. <https://doi.org/10.1186/s40468-025-00355-y>
- Cranley, L., Robinson, C., Hine, G., & O'Connor, D. (2022). The desks have changed; it must be NAPLAN time: How NAPLAN affects teaching and learning of mathematics. *Issues in Educational Research*, *32*(4), 1306–1320.
- Cresswell, J., Schwantner, U., & Waters, C. (2015). *A Review of International Large-Scale Assessments in Education*. OECD. <https://doi.org/10.1787/9789264248373-en>
- Darling Hammond, L., & Adamson, F. (2015). *Beyond the Bubble Test*. Wiley. <https://doi.org/10.1002/9781119210863>
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. <https://doi.org/10.1080/10888691.2018.1537791>
- Denzin, N.K. dan Y.S.L. (2009). *Handbook of Qualitative Research*. Pustaka Pelajar.
- FINEEC. (2022). *National Education Evaluation Plan*. [https://www.karvi.fi/sites/default/files/sites/default/files/documents/National-Education-Plan\\_2022-2023\\_updated-S2022\\_web.pdf](https://www.karvi.fi/sites/default/files/sites/default/files/documents/National-Education-Plan_2022-2023_updated-S2022_web.pdf)
- Fischman, G.E., Topper, A.M., Silova, I., Goebel, J., & Holloway, J.L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, *34*(4), 470–499. <https://doi.org/10.1080/02680939.2018.1460493>
- Guo, L., Huang, J., & Zhang, Y. (2019). Education Development in China: Education Return, Quality, and Equity. *Sustainability*, *11*(13), 3750. <https://doi.org/10.3390/su11133750>
- Heissel, J.A., Adam, E.K., Doleac, J.L., Figlio, D.N., & Meer, J. (2021). Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing. *Education Finance and Policy*, *16*(2), 183–208. [https://doi.org/10.1162/edfp\\_a\\_00306](https://doi.org/10.1162/edfp_a_00306)
- Jerrim, J. (2023). Test anxiety: Is it associated with performance in high-stakes examinations? *Oxford Review of Education*, *49*(3), 321–341. <https://doi.org/10.1080/03054985.2022.2079616>

- Johnson, S., & Johnson, R. (2010). *Component reliability in GCSE and GCE*.
- Kamens, D.H., & Benavot, A. (2011). National, regional and international learning assessments: trends among developing countries, 1960–2009. *Globalisation, Societies and Education*, 9(2), 285–300. <https://doi.org/10.1080/14767724.2011.577337>
- Kemendikdasmen. (2025a). *Tes Kemampuan Akademik (TKA)*. Pusat Asesmen Pendidikan, Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Kemendikdasmen. (2025b, 22 Desember). *Kemendikdasmen tekankan TKA jadi instrumen pemetaan capaian akademik nasional*. Kemendikdasmen. <https://www.kemendikdasmen.go.id/siaran-pers/14445-kemendikdasmen-tekankan-tka-jadi-instrumen-pemetaan-capaian->
- KICE. (2022). *National Assessment of Educational Achievement Overview*.
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65–79. <https://doi.org/10.1080/0969594X.2011.592972>
- Koretz, D. (2017). *The Testing Charade: Pretending to Make Schools Better*. University of Chicago Press.
- Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, 24(1), 60–77. <https://doi.org/10.1080/0969594X.2015.1074540>
- Lee, Y.-J., & Ho, J. (2022). *Basic Education in Singapore* (pp. 1–25). [https://doi.org/10.1007/978-981-16-8136-3\\_6-1](https://doi.org/10.1007/978-981-16-8136-3_6-1)
- Liu, Y. (2013). Meritocracy and the Gaokao/ : a survey study of higher education selection and socio-economic participation in East China. *British Journal of Sociology of Education*, 34(5–6), 868–887. <https://doi.org/10.1080/01425692.2013.816237>
- Liu, Y. (2016). *Higher Education, Meritocracy and Inequality in China*. Springer Singapore. <https://doi.org/10.1007/978-981-10-1588-5>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- MEXT. (2023). *National Assessment of Academic Ability Report 2023*.

- MOE Singapore. (2022). *PSLE Scoring (Achievement Levels)*. Ministry of Education Singapore.
- MOET. (2019). *ANLAS Vietnam: Country Report - Analysis of National Learning Assessment Systems*.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Chapter 4: Validity in Educational Assessment. *Review of Research in Education*, 30(1), 109–162. <https://doi.org/10.3102/0091732X030001109>
- Mostafa, T. (2017). *Is too much testing bad for student performance and well-being?* <https://doi.org/10.1787/2109a667-en>
- Mullis, I. V. S., Martin, M. O., & Davier, M. von. (2023). *TIMSS 2023 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement (IEA).
- NCES. (2019). *The Nation's Report Card: Reading and Mathematics 2019*.
- Niemi, H. (2021). Education Reforms for Equity and Quality: An Analysis from an Educational Ecosystem Perspective with Reference to Finnish Educational Transformations. *Center for Educational Policy Studies Journal*, 11(2), 13–35. <https://doi.org/10.26529/cepsj.1100>
- OECD. (2001). *Understanding the Digital Divide*. <https://doi.org/10.1787/236405667766>
- OECD. (2018). *Equity in Education*. OECD. <https://doi.org/10.1787/9789264073234-en>
- OECD. (2023). *PISA 2022 Results (Volume I)*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2024). *Finnish Education Evaluation Centre (FINEEC)*. OECD Publishing. <https://doi.org/10.1787/b1c0b194-en>
- OECD. (2025). *Education Policy Outlook 2025*. OECD Publishing. <https://doi.org/10.1787/c3f402ba-en>
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 20(2), 65–77. <https://doi.org/10.1016/j.pse.2014.11.002>
- Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies*, 40(5), 554–570. <https://doi.org/10.1080/03055698.2014.953914>
- Ræder, H. G., Andersson, B., & Olsen, R. V. (2022). Numeracy across grades – vertically scaling the Norwegian national numeracy tests. *Assessment in Education: Principles, Policy & Practice*, 29(6), 653–673. <https://doi.org/10.1080/0969594X.2022.2147483>
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G.

- Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. Russell Sage Foundation Press.
- Ronksley-Pavia, M. (2023). The Fallacy of Using the National Assessment Program–Literacy and Numeracy (NAPLAN) Data to Identify Australian High-Potential Gifted Students. *Education Sciences*, 13(4), 421. <https://doi.org/10.3390/educsci13040421>
- Sahlberg, P. (2021). *Finnish Lessons 3.0: What Can the World Learn from Educational Change in Finland?* Teachers College Press.
- Sarv, E.-S., & Rõuk, V. (2020). Estonian Curriculum: Becoming Independent. In *Pedagogy and Educational Sciences in the Post-Soviet Baltic States, 1990–2004: Changes and Challenges* (pp. 84–101). University of Latvia Press. <https://doi.org/10.22364/bahp-pes.1990-2004.05>
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Standards and Testing Agency. (2024). *Key stage 2: assessment and reporting arrangements (ARA)*. <https://www.gov.uk/government/publications/key-stage-2-assessment-and-reporting-arrangements-ara>
- Steiner-Khamsi, G., Martens, K., & Ydesen, C. (2024). Governance by numbers 2.0: policy brokerage as an instrument of global governance in the era of information overload. *Comparative Education*, 60(4), 537–554. <https://doi.org/10.1080/03050068.2024.2308348>
- van Rijn, P., Por, H.-H., McCaffrey, D. F., Bhaduri, I., & Bertling, J. (2024). A framework for comparing large-scale survey assessments: contrasting India’s NAS, United States’ NAEP, and OECD’s PISA. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1422030>
- World Bank. (2020). *Vietnam - High quality education for all by 2020 (Vol. 1 of 2) / : Overview/policy report (English)*.
- World Bank. (2024). *Indonesia Learning Poverty Brief*. World Bank. <https://documents1.worldbank.org/curated/en/099082924151529593/pdf/P179209-8b1c2fc9-3312-46f6-abb4-e1a5d85e5dd2.pdf>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>

- Yu, J. (2023). Exam Culture and Formative Assessment in China: The Gaokao Reform and Its Sociocultural Hindrance. *Journal of Education, Humanities and Social Sciences*, 23, 291–301. <https://doi.org/10.54097/ehss.v23i.12900>
- Yuan, A. (2024). *The Impact of New Gaokao Reform on the Implementation of High School Teaching in the Context of Educational Objectives* (pp. 500–508). [https://doi.org/10.2991/978-2-38476-291-0\\_62](https://doi.org/10.2991/978-2-38476-291-0_62)