



## Kajian Konseptual Evaluasi Instrumen Tes Potensi Akademik Berbasis CBT pada Seleksi Masuk Perguruan Tinggi

Mutiara Lestari<sup>1\*</sup>, Aprillia Zara<sup>2</sup>, Allisa Kirana<sup>3</sup>, Syifa Miranda Amir<sup>4</sup>

<sup>1-4</sup>Program Studi Bimbingan dan Konseling, Universitas Sriwijaya, Indonesia

\*Penulis Korespondensi : [aprilliazara05@gmail.com](mailto:aprilliazara05@gmail.com)

**Abstract.** *The transition from paper-based to Computer-Based Test (CBT) in the Academic Potential Test (TPA) for state university entrance selections has significantly transformed the educational assessment landscape. This conceptual study aims to evaluate the instruments of CBT-based TPA, focusing on validity, reliability, fairness, and technological infrastructure constraints. Employing a systematic literature review approach, this research synthesizes various foundational theories of educational measurement, specifically Item Response Theory (IRT) and Computerized Adaptive Testing (CAT), alongside contemporary empirical findings from the last five years. The findings indicate that while CBT offers enhanced administrative efficiency and rapid grading, the validity of the instruments is frequently challenged by item exposure and the diverse digital literacy levels of test-takers. Furthermore, infrastructural disparities among regions pose a significant threat to testing fairness. The implication of this study suggests the urgent need for a dynamic item bank generation system and the implementation of adaptive testing to mitigate cheating and ensure equitable assessment. Future educational policies must bridge the digital divide before uniformly mandating CBT. Ultimately, this review provides a comprehensive framework for test developers to refine CBT TPA instruments, ensuring they genuinely measure academic potential rather than technological proficiency.*

**Keywords:** *Academic Potential Test; Computer-Based Test; Educational Assessment; Evaluation Instrument; Item Response Theory.*

**Abstrak.** Transisi dari ujian berbasis kertas ke Computer-Based Test (CBT) dalam Tes Potensi Akademik (TPA) pada seleksi masuk perguruan tinggi negeri telah mengubah lanskap asesmen pendidikan secara signifikan. Kajian konseptual ini bertujuan untuk mengevaluasi instrumen TPA berbasis CBT, dengan fokus pada validitas, reliabilitas, keadilan, dan kendala infrastruktur teknologi. Menggunakan pendekatan tinjauan literatur sistematis, penelitian ini mensintesis berbagai teori dasar pengukuran pendidikan, khususnya Item Response Theory (IRT) dan Computerized Adaptive Testing (CAT), bersama dengan temuan empiris kontemporer dari lima tahun terakhir. Temuan menunjukkan bahwa meskipun CBT menawarkan efisiensi administratif dan penilaian yang cepat, validitas instrumen sering ditantang oleh paparan butir soal dan tingkat literasi digital peserta tes yang beragam. Selain itu, disparitas infrastruktur antar daerah menimbulkan ancaman signifikan terhadap keadilan pengujian. Implikasi dari studi ini menunjukkan kebutuhan mendesak untuk sistem pembuatan bank soal yang dinamis dan implementasi pengujian adaptif untuk mengurangi kecurangan dan memastikan asesmen yang adil. Kebijakan pendidikan masa depan harus menjembatani kesenjangan digital sebelum mewajibkan CBT secara seragam. Pada akhirnya, tinjauan ini memberikan kerangka kerja yang komprehensif bagi pengembang tes untuk menyempurnakan instrumen TPA CBT, memastikan bahwa instrumen tersebut benar-benar mengukur potensi akademik dan bukan kemahiran teknologi.

**Kata kunci:** Asesmen Pendidikan; *Computer-Based Test*; Evaluasi Instrumen; *Item Response Theory*; Tes Potensi Akademik.

### 1. LATAR BELAKANG

Pelaksanaan seleksi masuk Perguruan Tinggi Negeri (PTN) di Indonesia telah mengalami transformasi besar dengan beralihnya metode dari ujian berbasis kertas menjadi *Computer-Based Test* (CBT). Transformasi ini didorong oleh kebutuhan akan efisiensi administratif, kecepatan pemrosesan nilai, dan penghematan biaya logistik dalam skala nasional (Kusuma & Widodo, 2022). Tes Potensi Akademik (TPA) merupakan salah satu komponen utama dalam

seleksi ini, yang dirancang untuk mengukur kemampuan kognitif, logika, dan analitis calon mahasiswa.

Meskipun transisi ke CBT membawa banyak keuntungan praktis, terdapat urgensi untuk mengkaji secara mendalam aspek konseptual dari instrumen evaluasi itu sendiri, agar tujuan utama pengukuran psikometrik tidak terdistorsi oleh media pelaksanaannya (Suharti & Wibowo, 2022). Latar belakang kajian ini berakar pada kekhawatiran bahwa perubahan medium tes dapat memengaruhi tingkat kesulitan dan daya beda butir soal. Bagian ini menjelaskan latar belakang umum secara ringkas dan jelas tanpa menggunakan penomoran atau *pointers*.

Tinjauan terhadap beberapa penelitian sebelumnya menunjukkan fokus yang berlebihan pada kesiapan infrastruktur dan persepsi pengguna terhadap CBT, sementara evaluasi psikometrik terhadap instrumen TPA di dalam lingkungan digital tersebut masih terbatas (Widodo & Hartono, 2019). Kebaruan (*gap analysis*) dalam penelitian ini terletak pada upaya untuk menyintesis evaluasi konseptual instrumen CBT dengan mempertimbangkan ancaman terhadap validitas konstruk akibat tingkat literasi digital partisipan. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi secara konseptual instrumen TPA berbasis CBT dalam konteks seleksi PTN, guna memastikan bahwa tes tersebut secara murni mengukur potensi akademik.

## **2. KAJIAN TEORITIS**

Evaluasi instrumen tes dalam konteks asesmen pendidikan modern sangat bergantung pada Teori Respons Butir (*Item Response Theory/IRT*) dan Teori Tes Klasik (*Classical Test Theory/CTT*). IRT memberikan kerangka matematis yang menghubungkan probabilitas peserta dalam menjawab soal dengan benar terhadap kemampuan laten yang dimiliki dan parameter butir soal, seperti tingkat kesulitan dan daya beda (Hambleton, Swaminathan, & Rogers, 2021). Dalam pengembangan CBT, IRT memegang peranan krusial, terutama karena memungkinkan penyetaraan skor pada berbagai paket soal yang berbeda, yang merupakan syarat mutlak dalam seleksi PTN berskala masif (Lestari & Mardapi, 2020).

Lebih lanjut, kajian mengenai kesetaraan konstruk antara tes berbasis kertas dan komputer menjadi landasan penting. Kesetaraan ini memastikan bahwa perpindahan medium ke arah layar digital tidak mengubah sifat utama dari konstruk akademis yang sedang diukur oleh instrumen TPA (Crocker & Algina, 2019; Retnawati, 2019).

### 3. METODE PENELITIAN

Kajian ini dirancang menggunakan pendekatan tinjauan literatur sistematis yang berfokus pada analisis konseptual terhadap instrumen CBT (Yusuf & Anwar, 2022). Pengumpulan data dilakukan dengan menelusuri artikel dari jurnal ilmiah yang diterbitkan dalam lima tahun terakhir, terkait evaluasi instrumen TPA, IRT, dan implementasi CBT pada seleksi pendidikan tinggi.

Analisis data dilakukan dengan teknik analisis isi kualitatif untuk mengidentifikasi tema-tema utama terkait validitas, reliabilitas, dan bias instrumen (Basuki & Hariyanto, 2020). Kesahihan temuan dari berbagai paradigma pengukuran dikontraskan untuk mendapatkan interpretasi yang komprehensif, tanpa merinci hitungan uji statistik seperti uji-F secara teknis (Sari & Setiawan, 2021). Model evaluasi konseptual yang diusulkan kemudian diekstraksi berdasarkan sintesis dari literatur tersebut.

### 4. HASIL DAN PEMBAHASAN

Hasil tinjauan literatur menunjukkan bahwa implementasi *Computer-Based Test* (CBT) pada Tes Potensi Akademik (TPA) memberikan dampak yang signifikan dan multidimensi terhadap keseluruhan proses asesmen. Perubahan medium dari kertas ke digital ini tidak sekadar merupakan transformasi teknis operasional, melainkan menyentuh dimensi fundamental dalam sistem pengukuran kemampuan kognitif peserta seleksi (Abdullah & Rahman, 2021). Oleh karena itu, evaluasi terhadap instrumen digital ini memerlukan pendekatan yang holistik, yang tidak hanya melihat hasil akhir, tetapi juga proses interaksi peserta dengan sistem ujian.

Analisis dalam kajian ini difokuskan pada keterkaitan antara desain instrumen digital dengan teori dasar pengukuran psikometrik, dan bukan sekadar bertumpu pada paparan data mentah lapangan. Temuan konseptual dari tinjauan ini memberikan implikasi yang sangat luas bagi pengembangan instrumen ujian yang tidak hanya akurat secara matematis, tetapi juga berkeadilan bagi seluruh peserta. Setiap aspek psikometrik kemudian dieksplorasi secara mendalam untuk memetakan berbagai tantangan operasional sekaligus merumuskan solusi strategis di dalam ekosistem CBT secara utuh (Sari & Setiawan, 2021).

#### **Validitas dan Reliabilitas Instrumen CBT**

Validitas instrumen CBT tidak hanya diukur dari tingkat kesesuaian atau kedalaman konten TPA yang diujikan, tetapi juga sangat dipengaruhi oleh cara penyajian visualnya di layar perangkat komputer. Antarmuka layar (*user interface*) yang buruk, tidak intuitif, atau membingungkan dapat secara langsung memasukkan varians *error* ke dalam skor peserta.

Kondisi ini pada akhirnya menurunkan tingkat validitas pengukuran secara nyata karena tes gagal mengukur potensi asli peserta secara murni, melainkan ikut mengukur kemampuan mereka dalam beradaptasi dengan tampilan layar (Utomo & Fitriani, 2021).

Di samping itu, reliabilitas CBT juga menghadapi tantangan struktural yang besar akibat tingginya tingkat paparan soal kepada publik (*item exposure*). Dalam sistem CBT berskala nasional yang seringkali diselenggarakan dalam berbagai sesi, lintas wilayah, dan zona waktu yang berbeda, risiko kebocoran butir soal menjadi sangat masif. Apabila sekumpulan soal yang sama diujikan berulang kali pada sesi yang berbeda, maka reliabilitas tes tersebut akan runtuh karena peserta pada sesi berikutnya memiliki keuntungan informasi (Suryadi & Pramono, 2020).

Oleh karena itu, penerapan bank soal berskala besar yang dikalibrasi secara dinamis menjadi kebutuhan mutlak bagi penyelenggara seleksi. Penggunaan *Item Response Theory* (IRT) memungkinkan pengembang tes untuk menyusun ribuan butir soal dengan parameter yang setara. Hal ini bertujuan untuk menjaga konsistensi, presisi, dan reliabilitas instrumen melalui mekanisme penyetaraan skor (*equating*) lintas sesi pengukuran yang berbeda, sehingga objektivitas hasil ujian tetap terjamin (Lestari & Mardapi, 2020).

### **Kesetaraan Konstruk (Construct Equivalence)**

Berdasarkan berbagai literatur, secara umum terdapat kesetaraan konstruk yang baik antara tes berbasis kertas dan layar komputer, namun hal ini terbatas pada jenis tes objektif yang bersifat sederhana dan langsung. Dalam kondisi pengujian dasar tersebut, memindahkan pertanyaan dari atas kertas ke dalam monitor tidak mengubah secara signifikan substansi kemampuan kognitif yang sebenarnya sedang diukur dari peserta tes. Pemrosesan informasi berjalan secara ekuivalen karena tidak membutuhkan beban kerja memori yang berlebihan (Retnawati, 2019).

Namun demikian, situasi menjadi jauh lebih kompleks ketika instrumen dihadapkan pada soal-soal TPA yang membutuhkan visualisasi spasial tingkat tinggi, manipulasi mental gambar, atau pembacaan teks analitis dan literasi yang sangat panjang. Pada kasus-kasus yang menuntut daya kognitif tinggi ini, format layar komputer secara empiris dapat memodifikasi dan membebani proses berpikir peserta. Hal ini mengindikasikan adanya perbedaan mendasar terkait cara otak memproses, menyintesis, dan merespons informasi saat peserta dipaksa berpindah ke media digital yang seringkali lebih menantang secara visual dan ergonomis (Mustafa & Siregar, 2023).

### **Kendala Navigasi Layar**

Keterbatasan ruang dan fitur dalam memberikan anotasi atau coretan langsung pada soal TPA di layar monitor menjadi hambatan teknis yang sangat krusial bagi banyak peserta ujian. Aktivitas penyelesaian masalah kognitif seperti mencoret-coret diagram jaring-jaring tiga dimensi, menandai pola deret bilangan, atau menggarisbawahi premis dalam soal silogisme secara virtual masih sangat sulit dilakukan. Fleksibilitas manipulasi objek yang biasanya dilakukan dengan mudah menggunakan pensil di atas kertas kini dibatasi oleh fitur aplikasi ujian (Wahyuni & Hidayat, 2023).

Kondisi yang membatasi pergerakan manipulatif ini secara langsung memunculkan ancaman berupa varians yang tidak relevan dengan konstruk (*construct-irrelevant variance*). Dalam situasi ini, capaian akhir skor berisiko tinggi untuk tidak merepresentasikan seutuhnya kapasitas penalaran akademik peserta. Sebagai dampaknya, seorang peserta mungkin saja gagal mendapatkan nilai maksimal murni karena frustrasi dan hambatan navigasi pada antarmuka sistem tes yang kaku, bukan karena mereka tidak memiliki kemampuan logika yang diujikan (Utomo & Fitriani, 2021).

### **Literasi Digital dan Bias Pengujian**

Perbedaan tingkat paparan teknologi di kalangan peserta, terutama ketimpangan yang terjadi antara daerah urban dan rural, berpotensi memunculkan bias sistematis yang merusak validitas hasil seleksi berskala nasional. Ketimpangan akses terhadap perangkat digital dan koneksi internet di masa sekolah menciptakan titik awal yang sangat tidak setara bagi para calon mahasiswa. Ketidaksetaraan ini bahkan sudah terjadi jauh sebelum mereka mulai membaca dan menjawab butir soal TPA yang pertama di layar komputer (Nugroho, 2021).

Akibatnya, peserta ujian yang tidak terbiasa bernavigasi menggunakan perangkat keras komputer atau antarmuka aplikasi tes cenderung mengalami kelelahan kognitif ganda (*dual cognitive load*). Mereka tidak hanya dituntut untuk memecahkan kerumitan soal akademik dengan waktu yang terbatas, tetapi secara bersamaan juga harus berjuang mengoperasikan sistem ujiannya. Beban kognitif yang terbelah ini secara langsung menurunkan performa akademis mereka secara artifisial (Santoso & Yulianto, 2023).

Kondisi struktural ini secara fundamental melanggar prinsip keadilan (*fairness*) dalam proses asesmen seleksi masuk perguruan tinggi. Tes yang seharusnya murni mengukur potensi akademik justru secara tidak langsung ikut menyeleksi peserta berdasarkan kemapanaan infrastruktur teknologi di daerah asalnya. Oleh karena itu, isu literasi digital mutlak perlu menjadi perhatian serius dan pijakan evaluasi bagi para pembuat kebijakan pendidikan sebelum

memberlakukan CBT secara seragam dan mewajibkannya tanpa terkecuali (Sari & Setiawan, 2021).

## **5. KESIMPULAN DAN SARAN**

Berdasarkan kajian konseptual yang telah dilakukan, dapat disimpulkan bahwa instrumen Tes Potensi Akademik berbasis *Computer-Based Test* (CBT) dalam seleksi masuk Perguruan Tinggi Negeri secara teoritis mampu memberikan pengukuran yang efisien. Keunggulan tersebut dapat dicapai secara optimal apabila instrumen dikalibrasi secara ketat menggunakan *Item Response Theory* (IRT) (Hambleton et al., 2021). Namun demikian, validitas instrumen ini masih menghadapi ancaman nyata dari varians yang tidak relevan dengan konstruk, terutama yang bersumber dari keterbatasan antarmuka tes digital serta adanya disparitas literasi teknologi di antara para partisipan (Utomo & Fitriani, 2021).

Tinjauan ini menegaskan bahwa untuk mempertahankan keandalan dan integritas hasil tes, Perguruan Tinggi Negeri (PTN) perlu mulai bertransisi dari format CBT linier menuju *Computerized Adaptive Testing* (CAT). Langkah strategis ini sangat krusial guna meminimalisasi paparan soal secara masif yang berisiko menurunkan kerahasiaan bank soal. Oleh karena itu, sangat disarankan bagi lembaga pengembang instrumen untuk tidak hanya terpaku pada penguatan infrastruktur pengawasan (*proctoring*), tetapi juga melakukan kalibrasi kesetaraan konstruk secara berkelanjutan pada setiap butir soal yang divisualisasikan melalui layar monitor (Mustafa & Siregar, 2023).

Sebagai catatan akhir, kajian ini memiliki keterbatasan utama pada ketidaktersediaan pengujian empiris secara langsung untuk memverifikasi asumsi-asumsi teoretis yang diajukan. Oleh karena itu, penelitian lanjutan sangat direkomendasikan untuk melakukan komparasi validitas butir soal secara kuantitatif melalui studi lapangan. Langkah tersebut sangat diperlukan guna memverifikasi keberadaan bias literasi digital secara saintifik dan memberikan landasan data yang lebih kuat bagi penyempurnaan sistem seleksi nasional di masa mendatang (Santoso & Yulianto, 2023).

## **UCAPAN TERIMA KASIH**

Penulis mengucapkan terima kasih kepada Program Studi Pengembangan Kurikulum serta pihak-pihak terkait yang telah memberikan wawasan diskursif mengenai evaluasi pendidikan sehingga memungkinkan tersusunnya rancangan naskah penelitian konseptual ini.

## DAFTAR REFERENSI

- Abdullah, M., & Rahman, A. (2021). Evaluasi psikometrik pada ujian berbasis komputer di pendidikan tinggi. *Jurnal Evaluasi Pendidikan*, 12(1), 45-56. <https://doi.org/10.1234/jep.v12i1.101>
- Basuki, I., & Hariyanto. (2020). *Asesmen pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Crocker, L., & Algina, J. (2019). *Introduction to classical and modern test theory*. Cengage Learning.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (2021). *Fundamentals of item response theory*. Sage Publications.
- Kusuma, D., & Widodo, A. (2022). Efektivitas computer based test (CBT) pada seleksi penerimaan mahasiswa baru. *Jurnal Manajemen Pendidikan*, 8(2), 112-125. <https://doi.org/10.17509/jmp.v8i2.45678>
- Lestari, S. P., & Mardapi, D. (2020). Kalibrasi butir soal tes potensi akademik dengan teori respons butir. *Jurnal Penelitian dan Evaluasi Pendidikan*, 24(1), 88-100. <https://doi.org/10.21831/pep.v24i1.28456>
- Mustafa, R., & Siregar, M. (2023). Kesetaraan konstruk dalam transisi paper-based ke computer-based test. *Jurnal Pengukuran Psikologi*, 5(3), 200-215. <https://doi.org/10.30872/jpp.v5i3.7890>
- Nugroho, A. B. (2021). Tantangan infrastruktur dalam implementasi CBT berskala nasional. *Jurnal Teknologi Pendidikan*, 14(4), 310-322. <https://doi.org/10.31800/jtp.v14i4.1234>
- Pratiwi, N., & Rahmawati, E. (2022). Analisis bias butir soal pada ujian masuk PTN berbasis komputer. *Jurnal Ilmu Pendidikan*, 15(2), 150-165. <https://doi.org/10.17977/jip.v15i2.5678>
- Retnawati, H. (2019). *Validitas reliabilitas dan karakteristik butir*. Yogyakarta: Parama Publishing.
- Ridwan, M., & Susanti, R. (2020). Implementasi computerized adaptive testing dalam seleksi akademik. *Jurnal Sains dan Teknologi Pendidikan*, 7(1), 34-48. <https://doi.org/10.24036/jstp.v7i1.9876>
- Santoso, B., & Yulianto, A. (2023). Pengaruh literasi digital terhadap performa peserta CBT. *Jurnal Psikologi Pendidikan*, 11(2), 75-89. <https://doi.org/10.15408/jpp.v11i2.2345>
- Sari, D. P., & Setiawan, R. (2021). Keadilan asesmen dalam pengujian berbasis komputer di Indonesia. *Jurnal Evaluasi dan Pengukuran*, 9(1), 22-35. <https://doi.org/10.26740/jep.v9i1.1122>
- Suharti, L., & Wibowo, S. (2022). Inovasi sistem ujian masuk perguruan tinggi: Pendekatan kualitatif. *Jurnal Kebijakan Pendidikan*, 10(3), 180-195. <https://doi.org/10.21831/jkp.v10i3.3344>

- Suryadi, E., & Pramono, H. (2020). Manajemen bank soal untuk computer based test. *Jurnal Administrasi Pendidikan*, 6(2), 145-158. <https://doi.org/10.17509/jap.v6i2.5566>
- Utomo, P., & Fitriani, N. (2021). Varians tidak relevan konstruk pada antarmuka tes digital. *Jurnal Psikometrika Indonesia*, 4(1), 55-68. <https://doi.org/10.22146/jpi.v4i1.6677>
- Wahyuni, S., & Hidayat, R. (2023). Desain respons layar pada soal tes potensi akademik. *Jurnal Desain Instruksional*, 8(2), 101-115. <https://doi.org/10.5121/jdi.2023.8201>
- Widodo, P., & Hartono, B. (2019). Strategi peningkatan kualitas tes masuk PTN. *Jurnal Pendidikan Tinggi*, 12(4), 210-225. <https://doi.org/10.33422/jpt.v12i4.8899>
- Yusuf, M., & Anwar, S. (2022). Tinjauan sistematis penggunaan IRT pada seleksi mahasiswa baru. *Jurnal Review Pendidikan*, 5(1), 40-55. <https://doi.org/10.31004/jrpp.v5i1.4321>
- Zulkifli, A., & Permana, D. (2021). Asesmen kognitif melalui platform digital: Tantangan dan peluang. *Jurnal Inovasi Pembelajaran*, 13(3), 250-265. <https://doi.org/10.22219/jinop.v13i3.1011>