COMPUTERIZED ADAPTIVE TESTING BERBASIS IRT: PRINSIP, MEKANISME, DAN SIMULASI SEDERHANA

Maria Sumunaringtyas[1]
Edi Istiyono[2]
Lilin Rofiqotul Ilmi[3]

[123]Universitas Negeri Yogyakarta

Alamat Korespondensi

mariasumunar@gmail.com

ABSTRAK

Transformasi asesmen dari bentuk linear menuju bentuk adaptif menjadi kebutuhan penting untuk meningkatkan ketepatan pengukuran dalam bidang pendidikan. Penelitian ini bertujuan untuk mengkaji mekanisme kerja *Computerized Adaptive Testing* (CAT) melalui pendekatan simulasi menggunakan Microsoft Excel. Metode yang digunakan adalah simulasi kuantitatif yang mengintegrasikan algoritma *Maximum Likelihood Estimation* (MLE) dan bank soal yang telah dikalibrasi menggunakan model *Item Response Theory* (IRT) 1-PL. Kebaruan penelitian ini terletak pada penerapan aturan penghentian tes (*stopping rules*) yang mengombinasikan kriteria presisi (SEM ≤ 0,30) dan kriteria stabilitas (ΔSEM < 0,03) untuk mengoptimalkan durasi tes. Hasil simulasi menunjukkan bahwa mekanisme adaptif mampu mencapai konvergensi estimasi kemampuan peserta dengan tingkat akurasi yang tinggi hanya dalam rentang 12–15 butir soal. Temuan ini membuktikan efektivitas CAT dalam menghasilkan pengukuran yang efisien dan stabil, sekaligus memberikan gambaran yang transparan mengenai logika fungsional asesmen adaptif bagi para praktisi pendidikan.

Kata kunci: asesmen adaptif, *computerized adaptive testing* (CAT), IRT, bank soal, simulasi CAT.

1. Pendahuluan

Asesmen sebagai langkah yang bertujuan untuk menggali informasi penting mengenai peserta tes memerlukan instrumen atau alat ukur yang valid dan reliabel (Almanasreh et al., 2019; Azwar, 2021; Hasrul et al., 2021; Sumunaringtyas et al., 2024). Asesmen dalam pendidikan menjadi penting karena informasi yang dihasilkan akan digunakan sebagai dasar pengambilan keputusan oleh guru (Nitko & Brookhart, 2014; Ramatni et al., 2023) serta pemegang kebijakan dalam konteks yang lebih luas. Pengambilan keputusan ini tentu berkaitan dengan peserta didik di dalam lingkup kelas, sekolah, maupun secara nasional. Selain memperhatikan alat ukur, sistem pengukuran pada proses asesmen juga tidak kalah penting dalam menjamin hasil asesmen yang akurat (Popham, 2017; Tatenov et al., 2024; Thompson, 2019).

Asesmen yang selama ini digunakan telah mengalami perkembangan, yang tentunya selaras dengan pemanfaatan teknologi. Sistem asesmen dimulai secara konvensional menggunakan *paper based test* (PBT) kemudian berkembang menjadi *computer based test* (CBT) (Eshaghi, 2019; Perry et al., 2022; Shute & Rahimi, 2017). Namun, pada implementasi awal penggunaan CBT hanyalah pemindahan media dalam proses asesmen yang mulanya menggunakan kertas menjadi komputer. Sistem lain di dalam CBT pada dasarnya masih sama dengan metode konvensional PBT. Peserta tes masih mendapatkan soal yang sama sehingga kecurangan-kecurangan belum teratasi dengan adanya CBT (Berkhout et al., 2024; Prisacari & Danielson, 2017).

Seiring perkembangannya, CBT dilengkapi berbagai fitur keamanan seperti randomisasi soal, penggunaan paket tes yang berbeda, dan mekanisme proteksi akses soal untuk meminimalkan kecurangan (Berkhout et al., 2024). Meskipun demikian, sebagian besar CBT masih menggunakan pendekatan tes linear yang memberikan jumlah dan tingkat kesulitan soal yang sama kepada seluruh peserta tes. Pendekatan tersebut menyebabkan seluruh peserta tetap mengerjakan tes yang seragam tanpa mempertimbangkan profil kemampuan masing-masing. Kondisi ini berbeda dengan prinsip asesmen berdiferensiasi yang saat ini menjadi salah satu kebijakan dalam pendidikan. Asesmen berdiferensiasi menyesuaikan bentuk atau tingkat kesulitan asesmen berdasarkan karakteristik peserta didik (Koshy, 2013; Scott et al., 2014).

Asesmen merupakan proses mengukur kemampuan peserta tes layaknya mengukur ukuran badan seseorang sebelum menjahit baju. Ukuran yang dikenakan pada seseorang hendaknya pas atau dalam istilah lain tidak kebesaran dan tidak sempit. Dalam konteks ini, asesmen hendaknya memberikan butir-butir soal yang sesuai dengan kemampuan peserta tes, agar hasil pengukuran benar-benar mencerminkan kemampuan sesungguhnya (Embretson & Reise, 2000).

Butir yang sulit tidak akan efektif memberikan informasi mengenai peserta tes dengan kemampuan rendah, sedangkan butir yang mudah tidak akan efektif memberikan informasi mengenai kemampuan peserta tes yang tinggi (Schmucker & Moore, 2025; Yulia Herosian et al., 2022). Namun, dalam praktik asesmen berbasis CBT semua peserta tes biasanya diberikan jumlah soal yang sama, dengan tingkat kesulitan yang tetap. Akibatnya, terdapat soal yang terlalu mudah bagi peserta tes berkemampuan tinggi yang tidak memberikan informasi tambahan apa pun tentang kemampuannya. Sebaliknya, soal yang terlalu sulit untuk peserta tes berkemampuan rendah pun tidak banyak membantu, karena besar kemungkinan soal tersebut dijawab salah, dan tidak menambah informasi berarti tentang di mana letak kemampuan peserta tes tersebut.

Computerized Adaptive Testing (CAT) menawarkan pendekatan pengukuran yang lebih efisien dan presisi dibandingkan tes linear konvensional. Dalam konteks asesmen, CAT akan menyesuaikan tingkat kesulitan soal dengan kemampuan peserta tes secara *real-time* (Chang, 2015; Yuan et al., 2020) dan melakukannya dalam waktu yang lebih singkat dibanding sistem tes konvensional (Istiyono, 2020). Ketika peserta menjawab benar, soal berikutnya akan sedikit lebih sulit. Sebaliknya, jika peserta menjawab salah, sistem akan menyesuaikan dengan memberikan soal yang lebih mudah. Melalui mekanisme ini, setiap peserta akan dihadapkan pada soal-soal yang berada di sekitar zona kemampuannya. Karena kemampuan peserta berbeda-beda dan sistem CAT memberikan butir soal yang disesuaikan secara individual, maka jumlah soal dan durasi tes pun bisa berbeda antar peserta. Meskipun demikian, hasil asesmen tetap sah dan adil karena setiap peserta diuji secara adaptif sesuai kemampuan aktualnya.

Dengan demikian, CAT tidak hanya meningkatkan efisiensi dan akurasi asesmen, tetapi juga sejalan dengan kebijakan pembelajaran berdiferensiasi yang saat ini menjadi prioritas dalam Kurikulum Merdeka. Pemberian butir yang berada di sekitar tingkat kemampuan peserta tes memungkinkan setiap soal memberikan informasi yang optimal sehingga proses pengukuran menjadi lebih presisi (Ebenbeck et al., 2024; Way, 2006).

Meskipun memiliki berbagai keunggulan, pemahaman mengenai mekanisme kerja CAT di kalangan praktisi pendidikan di Indonesia masih relatif terbatas. Sebagian besar pembahasan CAT berfokus pada aspek matematis, algoritma, atau penggunaan perangkat lunak khusus yang kurang mudah diakses oleh guru dan pengambil kebijakan. Akibatnya, terdapat kesenjangan antara perkembangan teori CAT dan pemahaman praktis mengenai implementasinya di lapangan. Oleh karena itu, artikel ini bertujuan menjelaskan prinsip, mekanisme, dan simulasi sederhana *Computerized Adaptive Testing* (CAT) berbasis *Item Response Theory* (IRT) menggunakan *Microsoft Excel* sebagai media yang lebih mudah dipahami dan direplikasi oleh praktisi pendidikan.

Mekanisme adaptivitas pada CAT didasarkan pada penyesuaian tingkat kesulitan butir terhadap estimasi kemampuan peserta tes (Istiyono, 2020). Apabila peserta tes menjawab dengan benar sebuah soal dengan tingkat kesulitan b1, maka pada nomor selanjutnya ia akan mendapatkan soal dengan tingkat kesulitan b2 dimana b2>b1. Apabila peserta tes menghasilkan respon 0 pada butir b2 maka ia akan mendapatkan butir b3 dimana b3<b2 . Seterusnya hingga berakhirnya sistem memberikan soal sesuai dengan aturan tertentu. Mekanisme yang lebih jelas dapat dilihat pada tabel di bawah ini.

Tabel 1. Contoh Adaptasi Butir Soal Berbasis Parameter Kesulitan (b)

Tahap	Butir Soal	Tingkat Kesulitan (b)	Respon Peserta
1	B1	b1 (sedang)	1 (benar)
2	B2	b2 > b1 (lebih sulit)	1 (benar)
3	B3	b3 > b2 (lebih sulit)	0 (salah)
4	B4	b4 < b3 (lebih mudah)	1 (benar)
dst	dst	dst	dst

Aturan untuk penghentian tes atau penghentian pemberian soal dikenal sebagai *stopping rules*, yaitu kondisi tertentu di mana sistem menganggap bahwa informasi tentang kemampuan siswa sudah cukup akurat, sehingga tes dapat diakhiri. Selain aturan penghentian tes, sistem CAT juga memerlukan komponen penting lainnya seperti bank soal terkalibrasi berbasis komputer, algoritma pemilihan soal, serta sistem pelaporan hasil yang bersifat individual. Semua komponen tersebut saling berkaitan dan menjadi penentu apakah asesmen dapat berjalan secara adaptif dan adil bagi setiap peserta tes. Maka dari itu, pemahaman terhadap prinsip kerja CAT menjadi penting, bukan hanya bagi pengembang sistem asesmen, tetapi juga bagi guru dan tenaga pendidik. Dengan memahami CAT, guru dapat lebih siap menghadapi perubahan arah kebijakan asesmen di Indonesia.

2. Metode

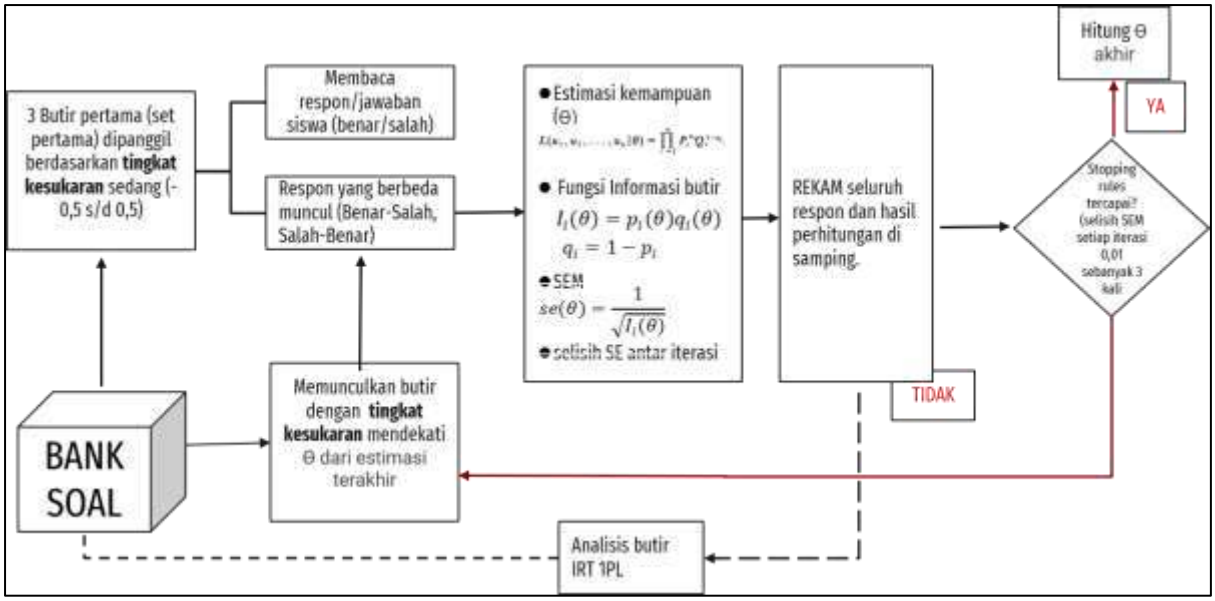
Penelitian ini menggunakan pendekatan simulasi kuantitatif untuk memodelkan mekanisme *Computerized Adaptive Testing* (CAT). Simulasi dikonstruksi menggunakan perangkat lunak *Microsoft Excel* untuk menunjukkan alur kerja algoritma adaptif secara transparan. Prosedur penelitian dibagi menjadi tiga tahap utama: pengembangan bank soal, perancangan algoritma adaptif, dan prosedur simulasi.

2.1 Pengembangan Bank Soal Terkalibrasi

Simulasi dalam penelitian ini didasarkan pada bank soal simulatif yang telah dikalibrasi menggunakan model IRT Logistik satu-parameter (1-PL). Setiap butir soal dalam bank tersebut memiliki parameter tingkat kesulitan (b) yang diasumsikan terdistribusi normal pada skala logit, dengan rentang nilai antara −2,00 hingga +2,00. Pemilihan model 1-PL bertujuan untuk menyederhanakan implementasi formula dalam *Microsoft Excel* tanpa mengurangi esensi mekanisme adaptivitas pada sistem *Computerized Adaptive Testing* (CAT). Bank soal simulatif yang digunakan dalam penelitian ini terdiri atas 228 butir soal.

2.2. Perancangan Algoritma Adaptif

Mekanisme CAT dalam simulasi ini melibatkan proses pemilihan butir, estimasi kemampuan, perhitungan SEM, dan penerapan *stopping rules* secara berulang hingga diperoleh estimasi kemampuan yang stabil. Alur proses tersebut dapat dilihat pada Gambar 1.



Gambar 1. Sistem simulasi CAT

a. **Aturan Awal (*Starting Rules*):** Menentukan pemberian butir soal pertama dengan asumsi kemampuan awal peserta tes berada pada nilai 0,00 (rata-rata) sebagai titik awal yang netral. Oleh karena itu, *starting rules* menggunakan tiga butir soal dengan tingkat kesulitan sedang, yaitu berada pada rentang −0,5 hingga 0,5.
b. **Algoritma Pemilihan Butir (*Item Selection*):** Pemilihan butir soal dilakukan menggunakan metode *Maximum Fisher Information*, di mana sistem secara otomatis memilih butir yang memiliki nilai informasi maksimum pada estimasi kemampuan sementara peserta tes (θ), yang umumnya berada pada tingkat kesulitan (b) paling dekat dengan nilai $\hat{\theta}$ tersebut.
c. **Prosedur Estimasi Kemampuan:** Setelah setiap respons diberikan, estimasi kemampuan peserta diperbarui menggunakan pendekatan *Maximum Likelihood Estimation* (MLE). Dalam simulasi berbasis Microsoft Excel, proses ini dimodelkan melalui prosedur iteratif untuk memperoleh nilai $\hat{\theta}$ yang memaksimalkan likelihood, atau secara operasional meminimalkan selisih antara probabilitas respons benar yang diprediksi model dengan respons aktual peserta tes.
d. **Aturan Penghentian (*Stopping Rules*):** Tes dihentikan apabila salah satu dari kriteria berikut terpenuhi: (1) selisih nilai Standard Error Measurement (SEM) antar iterasi mencapai ≤ 0,03 yang menunjukkan kestabilan estimasi kemampuan; (2) nilai SEM pada iterasi terakhir mencapai ≤ 0,30; atau (3) peserta tes telah menyelesaikan jumlah maksimum butir soal yang telah ditetapkan.

3. Hasil dan Pembahasan

3.1. Bank Soal dan IRT

Sistem CAT hanya dapat dilakukan apabila sudah tersedia bank soal berbasis CBT. Di dalam bank soal tersebut akan disimpan soal-soal yang sudah terbukti kualitasnya dan memenuhi standar parameter butir yang baik menurut *Item Response Theory* (IRT). Dalam IRT, sebuah butir memiliki parameter tingkat kesulitan, daya beda, dan tebakan semu (Baker, 2001; Hambleton & Swaminathan, 1985; Paek & Cole, 2020; Wilson, 2023). Ketiga hal ini diperlukan

untuk mengestimasi kemampuan peserta tes. Berikut merupakan syarat butir yang baik menurut IRT.

Tabel 2. Kualitas Butir menurut IRT

Parameter	Aturan
1. Tingkat Kesulitan	-2 s/d +2
2. Daya Beda	0 s/d 1
3. Tebakan Semu	$< \frac{1}{banyaknya\ opsi}$

Aturan di atas merupakan syarat butir yang baik menurut IRT apabila dianalisis menggunakan model yang paling kompleks yaitu model 3PL. Namun, apabila butir dianalisis menggunakan model 1PL, maka hal yang perlu diperhatikan hanyalah tingkat kesulitan (b). Model 1PL merupakan model IRT yang mengasumsikan bahwa semua soal memiliki daya pembeda yang sama, yaitu sebesar 1 ($a = 1$), sehingga satu-satunya parameter yang dipertimbangkan adalah tingkat kesulitan soal (Linden & Hambleton, 1997). Apabila butir dianalisis menggunakan model 2PL, maka parameter yang perlu diperhatikan adalah tingkat kesulitan (b) dan daya beda (a). Model 2PL merupakan pengembangan dari 1PL yang mempertimbangkan bahwa tiap soal dapat memiliki daya pembeda yang berbeda-beda, yaitu seberapa tajam soal tersebut membedakan peserta tes dengan kemampuan tinggi dan rendah. Apabila butir dianalisis menggunakan model 3PL maka seluruh parameter perlu diperhatikan. Model 3PL merupakan model yang lebih kompleks karena menambahkan parameter peluang menebak (g), yang memperhitungkan kemungkinan peserta tes menjawab benar secara kebetulan pada soal pilihan ganda. Artinya, butir soal yang disimpan di dalam bank soal harus memenuhi aturan sesuai dengan model analisis yang digunakan

Pada bagian sebelumnya telah disampaikan bahwa prinsip kerja CAT salah satunya adalah algoritma pemilihan butir. Butir-butir dipilih dari bank penyimpanan soal. Maka dari itu, bank soal harus berbasis komputer agar sistem CAT dapat mengambil butir-butir secara otomatis ke dalam bank soal.

3.2. Simulasi CAT

Simulasi CAT dilakukan dengan menggunakan program yang sudah dibuat dalam Microsoft Excel. Di dalam program tersebut sudah tersedia *sheet* yang menyediakan informasi sebagai berikut:

a. *Sheet* 1: Sistem simulasi CAT

Berisi tahap pengambilan butir, nomor soal dan parameter butirnya, respon jawaban peserta tes (benar atau salah), kemampuan, SEM setiap iterasi, stabilitas SEM setiap iterasi, dan keterangan tercapainya aturan penghentian tes. *Sheet* ini merupakan *sheet* utama yang digunakan untuk menjalankan simulasi dan menampilkan hasil. Adapun perhitungan parameter butir dan fungsi peluang dilakukan pada *sheet* terpisah.

Tahap	Nomor Soal	Tingkat Kesulitan	Respon (Jawaban)	*Kemampuan (MLE)*	*SE < 0.3*	*STABILITAS SEM*	*ATURAN PENGHENTIAN TES*
1	5	-0,50	1	4,00	26,973	-	TIDAK TERCAPAI
1	113	-0,13	1	4,00	15,910	11,063	TIDAK TERCAPAI
1	57	0,40	0	0,35	1,174	14,736	TIDAK TERCAPAI
2	165	0,13	1	0,65	0,869	0,305	TIDAK TERCAPAI
3	205	0,78	0	0,40	0,712	0,158	TIDAK TERCAPAI
4	173	0,35	1	0,65	0,614	0,098	TIDAK TERCAPAI
5	105	0,66	1	0,85	0,546	0,068	TIDAK TERCAPAI
6	145	0,72	0	0,65	0,495	0,051	TIDAK TERCAPAI
7	163	0,64	0	0,50	0,457	0,038	TIDAK TERCAPAI
8	39	0,51	1	0,65	0,426	0,031	TIDAK TERCAPAI
9	175	0,71	1	0,75	0,401	0,025	TERCAPAI

Gambar 2. Sistem simulasi CAT

b. *Sheet* 2: Peluang

Berisi peluang menjawab benar dan salah pada rentang *theta* $-4\ s/d\ 4$. Pada setiap butir selalu ada nilai peluang menjawab butir dengan benar ataupun salah pada peserta tes dengan kemampuan terendah hingga tertinggi ($-4\ s/d\ 4$).

NOMOR SOAL	PARAMETER BUTIR			PELUANG MENJAWAB BENAR				
	DAYA BEDA	TINGKAT KESULITAN	TEBAKAN SEMU	-4	-3,95	-3,9	-3,85	-3,8
1	1,00	-1,46	0,00	0,01315125	0,014301	0,01555	0,016907	0,018379
2	1,00	1,04	0,00	0,000190056	0,000207	0,000225	0,000245	0,000267
3	1,00	-1,18	0,00	0,00821129	0,008933	0,009718	0,010571	0,011498
4	1,00	1,77	0,00	5,49518E-05	5,98E-05	6,51E-05	7,09E-05	7,72E-05
5	1,00	-0,50	0,00	0,002599068	0,002829	0,003079	0,003351	0,003648
6	1,00	1,48	0,00	8,9965E-05	9,79E-05	0,000107	0,000116	0,000126
7	1,00	-0,22	0,00	0,001616297	0,001759	0,001915	0,002085	0,002269
8	1,00	1,60	0,00	7,33643E-05	7,99E-05	8,7E-05	9,47E-05	0,000103
9	1,00	1,03	0,00	0,000193314	0,00021	0,000229	0,000249	0,000272
10	1,00	0,16	0,00	0,000847815	0,000923	0,001005	0,001094	0,001191

Gambar 3. Peluang Menjawab

Peluang menjawab benar disimbolkan dengan P, dan peluang menjawab salah disimbolkan dengan Q. Karena dalam model ini hanya terdapat dua kemungkinan (benar atau salah), maka jumlah total peluangnya adalah 1. Oleh karena itu, secara matematis, Q dihitung sebagai $1 - P$, yakni komplemen dari peluang menjawab benar. Peluang menjawab benar tersedia dari kemampuan $-4\ s/d\ 4$. Peluang menjawab benar dan salah ini tentu selaras dengan kemampuan. Apabila kemampuan tinggi, maka peluang menjawab benar pasti akan lebih besar daripada peluang menjawab salah. Peluang peserta tes menjawab benar pada model IRT satu-parameter (1PL) (Hambleton et al., 1991; Linden & Hambleton, 1997) dinyatakan sebagai berikut:

$$P(\theta) = \frac{e^{(\theta - b)}}{1 + e^{(\theta - b)}}$$

Keterangan:
P(θ) : Peluang peserta dengan kemampuan θ menjawab benar.
θ : Estimasi kemampuan (*ability*) peserta tes pada skala logit.
b: Tingkat kesulitan butir soal.
e: Bilangan Euler (≈ 2,718).

c. *Sheet* 3: Respon *lifetime*

Respon langsung dari sistem simulasi (benar/salah) berisi estimasi peluang berdasarkan nilai pada *sheet* peluang yang diambil berdasarkan respon jawaban peserta tes (benar atau salah). Apabila peserta tes menjawab salah, maka peluang yang dimunculkan adalah peluang menjawab salah, sebaliknya apabila peserta tes menjawab benar maka yang dimunculkan adalah peluang menjawab benar. Nilai-nilai ini dimunculkan dari *sheet* peluang.

Nomor Soal	Respon	-4	-3,95	-3,9	-3,85	-3,8
5	1	0,0026	0,00283	0,00308	0,00335	0,00365
113	1	0,00139	0,00151	0,00164	0,00179	0,00195
57	0	0,99944	0,99939	0,99933	0,99927	0,99921
165	1	0,00089	0,00097	0,00106	0,00115	0,00125
205	0	0,9997	0,99968	0,99965	0,99962	0,99958
173	1	0,00061	0,00067	0,00073	0,00079	0,00086
105	1	0,00036	0,00039	0,00043	0,00047	0,00051
145	0	0,99967	0,99964	0,99961	0,99958	0,99954
163	0	0,99962	0,99959	0,99956	0,99952	0,99947
39	1	0,00047	0,00051	0,00055	0,0006	0,00066

Gambar 4. Peluang berdasarkan Respon

Nilai peluang dipilih sesuai dengan respon jawaban yang diberikan oleh peserta tes. Apabila respon 1 (benar), maka nilai-nilai yang diambil adalah P. Sebaliknya, apabila respon 0

(salah), maka nilai-nilai yang dimunculkan adalah Q. Seperti yang sudah dijelaskan sebelumnya, apabila kemampuan rendah, peluang menjawab salah lebih besar daripada peluang menjawab benar (*lihat tanda merah pada theta -4*).

d. *Sheet* 4: Fungsi informasi dan SEM

Sheet ini memuat informasi mengenai fungsi informasi butir dan SEM pada setiap butir dari rentang *theta* $-4\ s/d\ 4$, serta *theta* dimana fungsi informasi maksimal terjadi.

Dalam simulasi CAT ini, perhitungan SEM untuk keperluan evaluasi kestabilan estimasi kemampuan dan penerapan *stopping rules* didasarkan pada SEM kumulatif, yaitu SEM yang diturunkan dari akumulasi fungsi informasi tes (TIF) hingga tahap tertentu. Pendekatan ini digunakan untuk menilai apakah penambahan butir selanjutnya masih memberikan peningkatan presisi pengukuran yang bermakna.

Sementara itu, *sheet* IIF & SEM tidak hanya digunakan untuk memantau nilai SEM pada butir yang sedang disajikan, tetapi juga berperan penting dalam pengambilan keputusan pemilihan butir berikutnya. Pada setiap iterasi, sistem menghitung nilai IIF seluruh kandidat butir pada estimasi kemampuan terakhir ($\hat{\theta}$), kemudian memilih butir dengan nilai informasi maksimum sebagai soal berikutnya. Dengan demikian, IIF berfungsi sebagai dasar seleksi butir secara lokal pada $\hat{\theta}$ terkini, sedangkan SEM kumulatif berfungsi sebagai indikator global untuk menentukan penghentian tes.

NOMOR SOAL	Tingkat Kesulitan	Daya Beda	Tebakan Semu	-4	-3,95	3,95	4	IIF MAX	THETA (IIF MAX)	SEM MIN
1	-1,46	1,00	0,00	0,038	0,041	0,000	0,000	0,722	-1,450	1,177
2	1,04	1,00	0,00	0,001	0,001	0,020	0,019	0,722	1,050	1,177
3	-1,18	1,00	0,00	0,024	0,026	0,000	0,000	0,722	-1,200	1,177
4	1,77	1,00	0,00	0,000	0,000	0,068	0,062	0,722	1,750	1,177
5	-0,50	1,00	0,00	0,007	0,008	0,001	0,001	0,723	-0,500	1,176
6	1,48	1,00	0,00	0,000	0,000	0,042	0,039	0,722	1,500	1,177
7	-0,22	1,00	0,00	0,005	0,005	0,002	0,002	0,722	-0,200	1,177
8	1,6	1,00	0,00	0,000	0,000	0,051	0,047	0,723	1,600	1,176
9	1,03	1,00	0,00	0,001	0,001	0,020	0,018	0,722	1,050	1,177
10	0,16	1,00	0,00	0,002	0,003	0,005	0,004	0,722	0,150	1,177

Gambar 5. IIF dan SEM

e. *Sheet* 5: *Theta*

Memuat informasi estimasi kemampuan (θ) menggunakan MLE (*Maximum Likelihood*). Perhitungan θ bergantung pada pola respon atau jawaban peserta tes pada setiap tahap. Estimasi kemampuan menggunakan MLE hanya dapat dilakukan apabila terdapat respon yang berbeda pada setiap tahap. Sebagai contoh, pada Gambar 1 ketika peserta tes menjawab butir nomor 5 dan 113 dengan benar secara berturut-turut kemampuan yang muncul mencapai batas atas estimasi yang ditetapkan dalam simulasi (θ = 4). Sebaliknya apabila peserta tes menjawab butir tersebut salah terus-menerus maka kemampuan yang muncul adalah -4. MLE baru akan terestimasi setelah terdapat respon berbeda, contohnya pada butir 57. Pada metode MLE, estimasi kemampuan diperoleh dengan mencari nilai θ yang memaksimalkan fungsi likelihood berdasarkan pola respons peserta tes. Fungsi likelihood menunjukkan seberapa besar kemungkinan respons yang diamati muncul pada suatu tingkat kemampuan tertentu. Nilai θ yang menghasilkan likelihood terbesar kemudian digunakan sebagai estimasi kemampuan peserta tes. Formula MLE adalah sebagai berikut.

$$L(\theta) = \Pi_{i=1}^{n} P_i(\theta)^{ui} . (1 - P_i(\theta))^{1-ui}$$

Keterangan:

$L(\theta)$: Nilai likelihood yang menunjukkan seberapa besar kemungkinan pola respons peserta terjadi pada kemampuan θ tertentu.
$Pi(\theta)$: Peluang menjawab benar pada butir ke-i
ui: Respons peserta (1 = benar, 0 = salah)

Sebagai ilustrasi, apabila peserta tes lebih banyak menjawab benar pada butir dengan tingkat kesulitan tinggi, maka nilai θ yang memaksimalkan likelihood cenderung lebih tinggi dibanding peserta yang banyak menjawab salah.

Kelima informasi di atas sangat berguna bagi berjalannya sistem adaptif karena semua butir yang digunakan untuk simulasi sudah terjamin kualitasnya melalui analisis butir secara IRT.

3. 2.1 Aturan Awal Pemberian Soal

Starting rules yang selanjutnya menggunakan istilah aturan awal pemberian soal adalah aturan yang digunakan untuk memilih butir pertama atau seperangkat butir pertama yang akan disajikan kepada peserta tes (Chang, 2015; Meijer & Nering, 1999; Weiss & Gage Kingsbury, 1984; Yuan et al., 2020). Terdapat berbagai aturan awal pemberian soal yang bisa digunakan, salah satunya dengan memberikan soal yang terdiri atas 3 butir dengan tingkat kesulitan sedang (-0,5 s/d 0,5 atau -1 s/d 1).

Butir pertama yang dimunculkan ke dalam sistem sebanyak 3 butir dengan tingkat kesulitan sedang. Rentang tingkat kesulitan sedang yang biasa digunakan adalah -0,5 s/d 0,5 atau -1 s/d 1. Ketiga butir yang ditampilkan ini bisa dibuat sama untuk semua peserta tes. Namun ada juga yang menampilkan 3 butir yang berbeda pada masing-masing peserta tes.

Tabel 3. Contoh 1 Pemilihan Butir Tahap 1

Tahap	Nomor Soal	Tingkat Kesulitan	Respon	Kemampuan
1	5	-0,50	1	4
1	113	-0,13	1	4
1	57	0,40	0	0,35

Pada tahap 1, butir yang diberikan kepada peserta tes berada pada rentang -0,5 s/d 0,5. Pada tahap 1 ini diharapkan terdapat respon yang berbeda (tidak benar semua ataupun salah semua). Hal ini dikarenakan kemampuan yang diestimasi menggunakan MLE hanya dapat terestimasi apabila sudah terdapat respon yang berbeda, contohnya pada baris ketiga kemampuannya adalah 0,35 sedangkan pada baris 2 dan 1 kemampuannya adalah 4 (maksimum) karena peserta tes menjawab 2 butir berturut-turut dengan benar. Pemilihan MLE untuk mengestimasi kemampuan dikarenakan peneliti belum memiliki informasi mengenai kemampuan awal peserta tes. Apabila informasi mengenai kemampuan awal peserta tes sudah tersedia, maka estimasi kemampuan pada CAT dapat menggunakan metode Bayesian (Bock & Mislevy, 1982). Contoh lain pada pemilihan butir tahap 1 akan disajikan pada tabel di bawah ini.

Tabel 4. Contoh 2 Pemilihan Butir Tahap 1

Tahap	Nomor Soal	Tingkat Kesulitan	Respon	Kemampuan
1	5	-0,50	0	-4
1	113	-0,13	0	-4
1	215	-1,00	1	1

Pada baris pertama dan kedua, peserta tes menjawab dengan salah berturut-turut sehingga kemampuannya adalah -4 (minimum), sedangkan pada baris ketiga setelah ia berhasil menjawab dengan benar kemampuannya meningkat menjadi -1. Pada baris ketiga inilah MLE dapat terestimasi.

3.2.2 Algoritma Pemilihan Butir

Algoritma pemilihan butir adalah aturan dan ketentuan yang digunakan untuk memilih butir yang akan disajikan kepada peserta tes berdasarkan jawaban pada butir sebelumnya (Cao et al., 2024; Min & Aryadoust, 2021; Pan et al., 2023; Robin, 2005). Berbagai algoritma pemilihan butir telah dikembangkan dalam sistem CAT, antara lain algoritma pohon, algoritma linear, pendekatan fuzzy, dan algoritma berbasis fungsi informasi maksimum. Pembahasan mengenai algoritma-algoritma tersebut berada di luar cakupan artikel ini.

Simulasi pada artikel ini menggunakan algoritma informasi maksimum dikombinasikan dengan algoritma pohon. Pada algoritma pohon, pemilihan butir didasarkan pada dua hal yaitu estimasi kemampuan pada tahap sebelumnya dan respon jawaban (Cao et al., 2024; Istiyono, 2020; Pan et al., 2023; Robin, 2005). Misalnya, pada tahap kedua peserta tes menjawab benar dan dihasilkan kemampuan 0,4, maka pada tahap ketiga peserta tes akan diberikan butir dengan tingkat kesulitan lebih besar dari kemampuan tahap 2 (>0,4) karena ia menjawab benar. Namun apabila pada tahap 2 ia menjawab salah, maka butir yang diberikan pada tahap 3 harus lebih rendah dari estimasi kemampuan tahap 2 (<0,4).

Tabel 5. Contoh Algoritma Pemilihan Butir

Tahap	Nomor Soal	Tingkat Kesulitan	Respon	Kemampuan
1	5	-0,50	1	4
1	113	-0,13	1	4
1	57	0,40	0	0,35
2	165	0,13	1	0,65
3	205	0,78	0	0,4

Lihat tahap kedua, butir yang dipilih adalah butir nomor 165 yang memiliki tingkat kesulitan 0,13. Hal ini didasarkan pada kemampuan pada tahap 1 baris ketiga MLE yang dihasilkan adalah 0,35 sedangkan respon terakhir pada baris ketiga adalah 0 atau salah, maka butir yang dipilih harus mendekati 0,35 namun karena salah maka pilih yang tingkat kesulitannya di bawah 0,35. Pada tahap ketiga, dikarenakan tahap kedua peserta tes mampu menjawab butir dengan benar. Maka, butir yang dipilih pada tahap 3 harus di atas estimasi kemampuan pada tahap kedua yaitu $b \geq 0{,}65$.

3.2.3 Aturan Penghentian Tes

Stopping rules yang selanjutnya menggunakan istilah aturan penghentian tes adalah seperangkat aturan untuk menghentikan pengambilan butir dan estimasi akhir kemampuan peserta tes. Disebut sebagai seperangkat aturan karena umumnya terdapat lebih dari satu ketentuan dalam penghentian tes. Penggunaan satu aturan penghentian saja cenderung sulit diterapkan secara efektif (Ayanwale & Ndlovu, 2024; Ebenbeck et al., 2024; Istiyono, 2020; Stafford et al., 2019). Beberapa pilihan aturan penghentian tes yang umum digunakan antara lain stabilitas standar eror (<= 0,03), panjang tes tetap, durasi tes tetap, stabilitas estimasi kemampuan, dan standar eror tetap (< 0,3). Untuk menyederhanakan simulasi, aturan penghentian tes yang akan digunakan pada penelitian ini hanya merujuk pada stabilitas standar error pengukuran (Hambleton et al., 1991) yang diperoleh dari selisih standar error pada iterasi setiap tahap dan SEM tetap yaitu SEM < 0,3.

Standard Error Measurement (SEM) merupakan ukuran ketepatan estimasi kemampuan peserta tes. Nilai SEM menunjukkan sejauh mana estimasi kemampuan tersebut dapat dipercaya; semakin kecil SEM, semakin tinggi presisinya. Dalam konteks simulasi ini, SEM dihitung pada setiap iterasi dan dijadikan acuan untuk menghentikan tes apabila $SEM \leq 0{,}3$ atau perubahannya telah cukup stabil ($\Delta SEM \leq 0{,}03$). Stabilitas error menggambarkan bahwa error pengukuran mencapai angka yang sama. Error pengukuran berkaitan langsung dengan fungsi informasi. Semakin stabil error pengukuran, maka semakin dapat dipercaya hasil pengukuran tersebut. Hambleton menyatakan salah satu *stopping rules* yang dapat digunakan secara spesifik adalah apabila stabilitas error mencapai 0,03. Stabilitas error didapatkan dari dikurangi $SEM_n - SEM_{(n-1)}$.

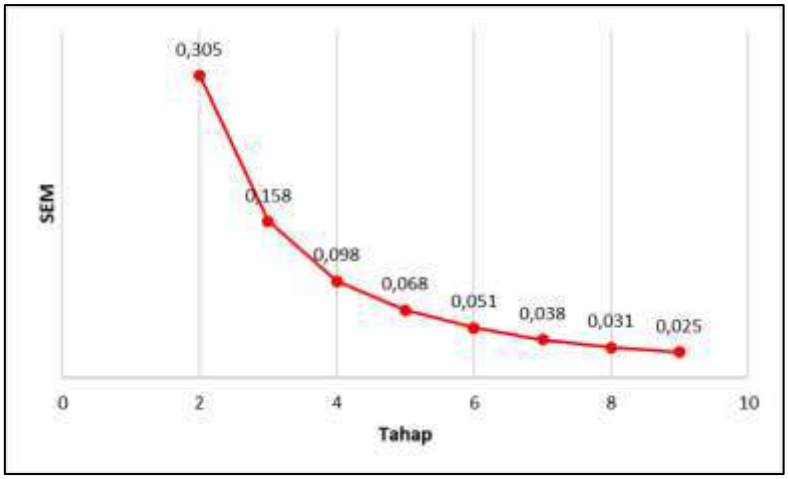
Tabel 6. Contoh Stabilitas SEM

Tahap	Nomor Soal	Tingkat Kesulitan	Respon	Kemampuan	SEM	Stabilitas SEM
1	5	-0,50	1	4	26,973	-
1	113	-0,13	1	4	15,910	11,063
1	57	0,40	0	0,35	1,174	14,736
2	165	0,13	1	0,65	0,869	0,305
3	205	0,78	0	0,4	0,712	0,158
4	173	0,35	1	0,65	0,614	0,098
5	105	0,66	1	0,85	0,546	0,068
6	145	0,72	0	0,65	0,495	0,051
7	163	0,64	0	0,5	0,457	0,038
8	39	0,51	1	0,65	0,426	0,031
9	175	0,71	1	0,75	0,401	0,025

Sistem CAT akan berhenti dan menghasilkan estimasi kemampuan yang merupakan output utama yang dibutuhkan apabila *stopping rules* telah tercapai. Berdasarkan aturan yang dibuat Hambleton (1985), aturan penghentian tes pada simulasi ini tercapai pada tahap kesembilan dengan total butir yang dikerjakan oleh peserta tes sebanyak 11 butir setelah stabilitas SEM mencapai 0,025.

Stabilitas hasil pengukuran menggunakan CAT berdasarkan simulasi menunjukkan bahwa error pengukuran sangat diperhitungkan untuk menjamin tingkat kepercayaan hasil pengukuran. Selain itu, pemberian butir selalu pada rentang hasil estimasi kemampuan peserta tes, tidak dalam kategori sangat sulit maupun sangat mudah. Apabila butir yang diberikan kepada peserta tes terlalu mudah untuk tingkat kemampuannya, sama halnya dengan menjahit baju yang terlalu sempit yaitu tidak dapat digunakan karena ukuran tubuhnya (dalam konteks ini adalah ukuran kemampuannya) jauh lebih besar daripada ukuran baju (tes) yang diberikan. Selaras dengan hal itu, apabila tes yang diberikan cenderung sangat sulit sama seperti baju yang terlalu besar, ukuran asli (kemampuan) yang terlalu kecil tidak dapat terukur dengan baik.

Sistem pengukuran menggunakan CAT memberikan pengukuran yang lebih tepat sesuai kemampuan peserta tes. Pemberian butir soal yang didasarkan pada estimasi kemampuan sebelumnya lebih tepat sasaran untuk mengukur kemampuan peserta tes. Selain itu, stabilitas pengukuran juga menjadi nilai tambah pengukuran menggunakan CAT.



Gambar 6. Stabilitas Error Pengukuran CAT
Sumber: Simulasi

Estimasi kemampuan dalam CAT dilakukan secara iteratif, di mana setiap respons peserta digunakan untuk memperbarui estimasi kemampuan dan menghitung SEM. Jika nilai SEM telah mencapai tingkat kestabilan yang ditentukan, hal ini menunjukkan bahwa proses estimasi telah konvergen. Artinya, penambahan butir baru tidak lagi memberikan peningkatan signifikan terhadap akurasi estimasi, sehingga tes dapat dihentikan dengan aman tanpa mengorbankan ketepatan hasil. Hal ini dibuktikan pada tabel di bawah ini.

Tabel 7. Stabilitas SEM sebagai *Stopping Rules*

Tahap	Nomor Soal	Tingkat Kesulitan	Respon	Kemampuan	SEM	Stabilitas SEM
1	5	-0,50	1	4	26,973	-
1	113	-0,13	1	4	15,910	11,063
1	57	0,40	0	0,35	1,174	14,736
2	165	0,13	1	0,65	0,869	0,305
3	205	0,78	0	0,4	0,712	0,158
4	173	0,35	1	0,65	0,614	0,098
5	105	0,66	1	0,85	0,546	0,068
6	145	0,72	0	0,65	0,495	0,051
7	163	0,64	0	0,5	0,457	0,038
8	39	0,51	1	0,65	0,426	0,031
9	175	0,71	1	0,75	0,401	0,025

Pada tahap ke-8 hingga ke-9, perubahan nilai *Standard Error Measurement* (SEM) menunjukkan kecenderungan yang semakin kecil dan mencapai kondisi stabil (<0,03). Hal ini mengindikasikan bahwa penambahan butir soal pada tahap tersebut tidak lagi memberikan peningkatan yang signifikan terhadap ketepatan estimasi kemampuan, sehingga proses pengukuran dapat dianggap telah konvergen.

4. Kesimpulan

Berdasarkan hasil analisis dan simulasi mekanisme *Computerized Adaptive Testing* (CAT) berbasis Microsoft Excel, dapat disimpulkan bahwa CAT mampu meningkatkan efisiensi dan presisi pengukuran melalui pemilihan butir yang adaptif terhadap kemampuan peserta tes. Berbeda dengan tes konvensional yang menggunakan seperangkat soal tetap, CAT menyajikan

butir yang tingkat kesulitannya menyesuaikan estimasi kemampuan peserta sehingga pengukuran dapat dilakukan dengan jumlah soal yang lebih sedikit tanpa mengurangi akurasi.
Keberhasilan implementasi CAT sangat bergantung pada ketersediaan bank soal yang telah terkalibrasi menggunakan Item Response Theory (IRT). Parameter butir, khususnya tingkat kesulitan, menjadi dasar dalam proses pemilihan soal dan estimasi kemampuan peserta tes. Selain itu, penggunaan *stopping rules* berbasis *Standard Error of Measurement* (SEM) memungkinkan sistem menghentikan tes secara objektif ketika tingkat presisi pengukuran telah mencapai batas yang ditentukan.
Secara praktis, penelitian ini menunjukkan bahwa mekanisme CAT dapat dipahami dan divisualisasikan melalui simulasi sederhana berbasis Microsoft Excel tanpa memerlukan perangkat lunak khusus yang kompleks. Simulasi ini berpotensi menjadi media pembelajaran bagi guru, peneliti, dan pengembang asesmen untuk memahami prinsip dasar CAT sebelum mengimplementasikannya dalam sistem asesmen yang lebih komprehensif.
Sebagai tindak lanjut, guru, peneliti, dan pengembang asesmen disarankan menggunakan simulasi berbasis Microsoft Excel sebagai media pembelajaran awal untuk memahami mekanisme CAT sebelum mengimplementasikannya menggunakan perangkat lunak yang lebih kompleks. Penelitian selanjutnya juga dapat mengembangkan simulasi ini dengan model IRT yang lebih kompleks, seperti 2PL atau 3PL, serta mengujinya menggunakan data empirik peserta tes.

5. Referensi

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Ayanwale, M. A., & Ndlovu, M. (2024). The feasibility of computerized adaptive testing of the national benchmark test: A simulation study. *Journal of Pedagogical Research*, *8*(2), 95–112. https://doi.org/10.33902/JPR.202425210

Azwar, S. (2021). *Reliabilitas dan Validitas: XI*. Pustaka Pelajar.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Berkhout, E., Pradhan, M., Rahmawati, Suryadarma, D., & Swarnata, A. (2024). Using technology to prevent fraud in high stakes national school examinations: Evidence from Indonesia. *Journal of Development Economics*, *170*. https://doi.org/10.1016/j.jdeveco.2024.103307

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431–444. https://doi.org/10.1177/014662168200600405

Cao, X., Lin, Y., Liu, D., Zheng, F., & Duh, H. B. L. (2024). Novel item selection strategies for cognitive diagnostic computerized adaptive testing: A heuristic search framework. *Behavior Research Methods*, *56*(4), 2859–2885. https://doi.org/10.3758/s13428-023-02228-9

Chang, H. H. (2015). Psychometrics Behind Computerized Adaptive Testing. *Psychometrika*, *80*(1), 1–20. https://doi.org/10.1007/s11336-014-9401-5

Ebenbeck, N., Bastian, M., Mühling, A., & Gebhardt, M. (2024). Duration versus accuracy—what matters for computerised adaptive testing in schools? *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.13074

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologist*. Lawrence Erlbaum Associates, Inc.

Eshaghi, M. (2019). An Introduction to Computer-Based Assessment. *Strides in Development of Medical Education*, *In Press*(In Press). https://doi.org/10.5812/sdme.86326

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer.

Hambleton, R. K., Swaminathan, Hariharan., & Rogers, H. Jane. (1991). *Fundamentals of item response theory*. Sage Publications.

Hasrul, H., Mukhtar, M. I., & Roddin, R. (2021). View of Validity and Reliability of Practical Teaching Practice Instruments among Construction Technology Lecturers in Vocational Colleges. *JTET*, *13*(3), 162–171. https://doi.org/https://doi.org/10.30880/jtet.2021.13.03.016

Istiyono, E. (2020). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern* (2nd ed.). UNY Press.

Koshy, S. (2013). *Differentiated assessment activities: customising to support learning*. https://ro.uow.edu.au/dubaipapers/549

Linden, W. J. van der, & Hambleton, R. K. (1997). Handbook of modern item response theory. In *Handbook of Modern Item Response Theory*. Springer. https://doi.org/10.1007/978-1-4757-2691-6

Meijer, R. R., & Nering, M. L. (1999). *Computerized Adaptive Testing: Overview and Introduction*.

Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. In *Studies in Educational Evaluation* (Vol. 68). Elsevier Ltd. https://doi.org/10.1016/j.stueduc.2020.100963

Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (6th ed.). Pearson Education Limited.

Paek, I., & Cole, K. (2020). *Using r for item response theory model applications*. Routledge Taylor & Francis Group.

Pan, Y., Livne, O., Wollack, J. A., & Sinharay, S. (2023). Item Selection Algorithm Based on Collaborative Filtering for Item Exposure Control. *Educational Measurement: Issues and Practice*, *42*(4), 6–18. https://doi.org/10.1111/EMIP.12578;PAGEGROUP:STRING:PUBLICATION

Perry, K., Meissel, K., & Hill, M. F. (2022). Rebooting assessment. Exploring the challenges and benefits of shifting from pen-and-paper to computer in summative assessment. In *Educational Research Review* (Vol. 36). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2022.100451

Popham, W. James. (2017). *Classroom assessment : what teachers need to know*. Pearson Education.

Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior*, *77*, 1–10. https://doi.org/10.1016/j.chb.2017.07.044

Ramatni, A., Anjely, F., Cahyono, D., Rambe, S., & Shobri, M. (2023). Proses Pembelajaran dan Asesmen yang Efektif. *Journal on Education*, *05*(04). https://doi.org/https://doi.org/10.31004/joe.v5i4.2687

Robin, F. (2005). Item Exposure. *Encyclopedia of Statistics in Behavioral Science*. https://doi.org/10.1002/0470013192.BSA320

Schmucker, R., & Moore, S. (2025). *The Impact of Item-Writing Flaws on Difficulty and Discrimination in Item Response Theory*. http://arxiv.org/abs/2503.10533

Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and equitable assessment practices for all students. *Assessment in Education: Principles, Policy and Practice*, *21*(1), 52–70. https://doi.org/10.1080/0969594X.2013.776943

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. In *Journal of Computer Assisted Learning* (Vol. 33, Number 1, pp. 1–19). Blackwell Publishing Ltd. https://doi.org/10.1111/jcal.12172

Stafford, R. E., Runyon, C. R., Casabianca, J. M., & Dodd, B. G. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods*, *51*(3), 1305–1320. https://doi.org/10.3758/s13428-018-1068-x

Sumunaringtyas, M., D, T. F., Istiyono, E., & Setiawati, F. A. (2024). Character Scale for Assessing Discipline and Learning Responsibilities Character in Elementary School. *Jurnal Ilmiah Sekolah Dasar*, *8*(2), 328–337. https://doi.org/10.23887/JISD.V8I2.69078

Tatenov, A., Toktaugaliyeva, S., Sandibaeva, N., Karymbai, A., & Ismailova, G. (2024). Development of measures, system of assessment of learning process parameters in the secondary education system. *Scientific Herald of Uzhhorod University Series Physics*, (56), 2444–2452. https://doi.org/10.54919/physics/56.2024.244ak4

Thompson, N. (2019). *Psychometric resources: Assessment systems*. https://assess.com/psychometric-resources/

Way, W. D. (2006). *Practical Questions in Introducing Computerized Adaptive Testing for K-12 Assessments*. http://www.pearsonassessments.com/research

Weiss, D. J., & Gage Kingsbury, G. (1984). APPLICATION OF COMPUTERIZED ADAPTIVE TESTING TO EDUCATIONAL PROBLEMS. In *JOURNAL OF EDUCATIONAL MEASUREMENT* (Vol. 21, Number 4).

Wilson, M. (2023). *Constructing measures: An item response modeling approach*. *2*.

Yuan, Y., Xia, H., Han, Y., & Hu, M. (2020). Advances in computerized adaptive testing. *Proceedings - 2020 International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2020*, 202–205. https://doi.org/10.1109/ICHCI51889.2020.00051

Yulia Herosian, M., Yeni, ;, Sihombing, R., Delyanti, ;, & Pulungan, A. (2022). *ITEM RESPONSE THEORY ANALYSIS ON STUDENT STATISTICAL LITERACY TESTS* (Vol. 9, Number 2). https://ejournal.unuja.ac.id/index.php/pedagogik