



PENGUNAAN TEORI TES KLASIK UNTUK ANALISIS BUTIR SOAL PADA ASESMEN

Leonardus Jafar¹, Riyadi², Achmad Ridwan³

^{1,2,3}Program Studi Magister Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta

Post-el: leonardusjafar_9912822009@mhs.unj.ac.id*

Abstrak	Info Artikel
<p><i>Penelitian ini merupakan studi deskriptif yang bertujuan untuk mengidentifikasi pendekatan yang umum dan seharusnya diterapkan dalam analisis butir asesmen. Metode yang digunakan adalah Systematic Literature Review, di mana sumber-sumber berupa jurnal nasional dan internasional bereputasi serta buku-buku relevan dikumpulkan dan dianalisis untuk mengidentifikasi kesenjangan antara teori dan praktik di lapangan. Hasil penelitian menunjukkan bahwa terdapat beberapa studi yang dilakukan dalam lima tahun terakhir mengenai analisis butir soal dalam konteks pembelajaran di sekolah. Temuan ini mengindikasikan bahwa telah ada upaya untuk meningkatkan kualitas evaluasi pembelajaran dengan menggunakan pendekatan teori tes klasik. Pendekatan ini paling umum digunakan karena analisisnya relatif mudah dan sederhana. Namun, seiring dengan perkembangan, ada analisis baru yang lebih dianjurkan, yaitu Teori Respon Butir (Item Response Theory/IRT) dan Model Rasch. Teori-teori ini menawarkan pendekatan analisis butir yang menggunakan skala pengukuran yang konsisten antara data ordinal, yang mencakup informasi tentang kemampuan siswa dan kualitas butir soal yang diuji. Dengan skala yang sama, pengukuran ini dianggap mampu memberikan informasi yang lebih akurat mengenai peserta tes dan kualitas soal yang diberikan.</i></p>	<p>Diajukan: 20-07-2024 Diterima: 19-09-2024 Diterbitkan : 25-10-2024</p>
<p>Abstract</p> <p><i>This research is a descriptive study that aims to identify general approaches that should be applied in the analysis of assessment items. The method used is a Systematic Literature Review, where sources in the form of reputable national and international journals and relevant books are collected and analyzed to identify gaps between theory and practice in the field. The results of the research show that there have been several studies conducted in the last five years regarding the analysis of test items in the context of learning at school. These findings indicate that there have been efforts to improve the quality of learning evaluation using a classical test theory approach. This approach is most commonly used because the analysis is relatively easy and simple. However, along with developments, there are new analyzes that are more recommended, namely Item Response Theory (IRT) and the Rasch Model. These theories offer an item analysis approach that uses a consistent measurement scale between ordinal data, which includes information about student abilities and the quality of the items tested. With the same scale, this measurement is considered capable of providing more accurate information about test takers and the quality of the questions given.</i></p>	<p>Kata kunci: <i>Analisis butir soal asesmen, teori tes klasik, teori tes modern, IRT, rasch model</i></p> <p>Keywords: <i>Analysis of assessment items, classical test theory, modern test theory, IRT, Rasch model</i></p>

Cara mensitasi artikel:

Jafar, L., Riyadi, R., & Ridwan, A. (2023). Penggunaan Teori Tes Klasik untuk Analisis Butir Soal Pada Asesmen. *IJSH: Indonesian Journal of Social and Humanities*, 2(3), 1–10. <https://jurnal.academiacenter.org/index.php/IJSH>

PENDAHULUAN

Asesmen merupakan salah satu elemen krusial dalam pendidikan, terutama dalam proses pembelajaran (Anggrena et al., 2022). Tujuan dari asesmen adalah untuk menilai sejauh mana tujuan pembelajaran tercapai dan mengumpulkan informasi mengenai perkembangan belajar siswa dalam berbagai aspek, termasuk kognitif, psikomotor, dan afektif (Septiani et al., 2019). Dalam konteks pendidikan kejuruan, asesmen sangat penting untuk mendukung kompetensi keterampilan peserta didik (Sutami, 2020). Penjelasan di atas menunjukkan bahwa hasil asesmen dapat mencerminkan tingkat pencapaian pembelajaran. Hasil tersebut dapat dimanfaatkan oleh pendidik untuk membuat keputusan yang berkaitan dengan pembelajaran, sehingga berkontribusi pada peningkatan kualitas proses pembelajaran (Matondang et al., 2022). Oleh karena itu, kemampuan untuk melaksanakan asesmen menjadi salah satu kompetensi yang harus dimiliki oleh guru. Kompetensi ini dapat dikuasai jika pendidik memahami konsep pengembangan asesmen dan penerapannya dalam proses pembelajaran. Saat ini, banyak pendidik yang masih memiliki pemahaman yang kurang memadai tentang konsep asesmen meskipun telah memenuhi kualifikasi akademik dan mengikuti pelatihan (Irawati et al., 2018; Y. R. Kurniawan et al., 2016; Magdalena et al., 2020). Oleh karena itu, diperlukan pembiasaan dalam pelaksanaan asesmen oleh pendidik, mulai dari pengembangan instrumen hingga analisis item dan pengambilan keputusan berdasarkan hasil yang diperoleh (Abdul, 2015).

Asesmen dapat dilakukan melalui kegiatan tes maupun non-tes dengan menggunakan instrumen sebagai alat evaluasi (Irawati et al., 2018). Instrumen evaluasi yang umum digunakan adalah tes objektif pilihan ganda, karena bentuknya memungkinkan pengukuran pencapaian banyak pokok materi pembelajaran dalam satu instrumen sekaligus (Abdul, 2015; Suseno, 2017). Selain itu, jenis instrumen tes ini dapat mencakup berbagai tingkat berpikir, mulai dari pengetahuan dasar hingga sintesis dan analisis. Kemudahan dalam penskoran instrumen tes objektif juga menjadikannya pilihan yang lebih populer karena dapat dikoreksi oleh siapa saja (Y. R. Kurniawan et al., 2016). Proses pengembangan instrumen asesmen objektif dilakukan melalui beberapa langkah penting, termasuk konstruksi tes, analisis butir instrumen tes, dan implementasi atau penyebaran tes (Septiani et al., 2019). Dari langkah-langkah tersebut, analisis butir merupakan tahap krusial untuk menentukan kualitas instrumen dan fungsi masing-masing butir. Analisis butir umumnya dilakukan melalui dua metode: analisis kualitatif dan analisis kuantitatif (Pintastyo & Sudji, 2017). Analisis kualitatif, yang juga dikenal sebagai validitas logis, bertujuan untuk menilai kualitas butir berdasarkan penilaian dari validator atau ahli. Sementara itu, analisis kuantitatif atau validitas empiris dilakukan untuk menilai apakah sebuah instrumen tes berfungsi dengan baik setelah diuji coba pada sampel yang representatif (Hendryadi, 2017).

METODE PENELITIAN

Metode penelitian yang digunakan adalah deskriptif naratif dengan menggunakan tinjauan pustaka atau systematic literature review (Yunida & Arthur, 2023). Dimana penulis mengumpulkan bahan-bahan berupa jurnal nasional maupun internasional bereputasi dan buku- buku yang relevan untuk dijadikan dasar dalam mengidentifikasi kesenjangan yang terjadi antara suatu teori dan kenyataan di lapangan (Cahyono et al., 2019). Tahapan dalam melakukan tinjauan pustaka sistematis yaitu pengumpulan data dan informasi dari berbagai sumber, melakukan proses screening atau reduksi serta penggolongan informasi data yang diperlukan, dan menganalisa hasil ulasan untuk disajikan dalam bentuk narasi (Pérez-Escolar et al., 2023; Turmuzi et al., 2022). Kemudian berdasarkan hasil analisis yang diperoleh diharapkan dapat menjawab permasalahan yang dibahas yaitu Penggunaan Teori Tes Klasik pada Asesmen Pembelajaran dan Peluang Perbaikannya.

HASIL DAN PEMBAHASAN

Sistem penilaian yang digunakan guru pada umumnya masih menggunakan pola klasik dan konvensional. (Chan et al., 2014). Teori tes klasik merupakan pendekatan yang sangat sederhana dan mudah untuk dipahami dalam kegiatan analisis soal secara empiris. Analisis secara empiris dilakukan terhadap butir-butir soal untuk mengetahui informasi mengenai reliabilitas, tingkat kesukaran, daya beda serta keberfungsian distraktor/pengecoh. Pada pendidikan kejuruan khususnya, banyak diantara pendidik menggunakan teori tes klasik untuk mendeskripsikan parameter butir soal dalam mengukur kompetensi peserta didiknya. Dibuktikan dari penelitian 5 tahun terakhir yang menggunakan teori tes klasik dalam asesmen pembelajaran:

Tabel 1. Penelitian Penggunaan Teori Tes Klasik pada Asesmen

Peneliti	Tahun	Judul Penelitian	Tempat/subjek penelitian	Hasil penelitian
Muh. Syahrul Sarea	2019	Karakteristik Butir Soal: Classical Test Theory Vs Item Response Theory?	37 peserta didik yang mengikuti UAS semester gasal tema satu tingkat sekolah dasar (SD) tahun pelajaran 2018/2019 di desa paramasan bawah	Berdasarkan hasil analisis data diperoleh bahwa model yang cocok digunakan untuk menganalisis data menggunakan model 2 parameter logistik dimana tingkat kecocokan butir dengan model sebanyak 37 dari 41 butir yang di analisis.
Ismiyati	2023	Pelatihan Analisis Butir Soal Berdasarkan Teori Tes Klasik Berbantuan Iteman untuk Meningkatkan	Target utama dari pelatihan ini adalah peningkatan pengetahuan dan keterampilan para pendidik di MTs. Tarbiyatul Banin Pati dalam	Berdasarkan testimoni dari peserta, kegiatan pelatihan tentang analisis butir soal dinyatakan berhasil karena pelatihan ini bermanfaat bagi pengembangan profesi guru.

Peneliti	Tahun	Judul Penelitian	Tempat/subjek penelitian	Hasil penelitian
Anggrina Gusmizain	2022	Kualitas Instrumen Penilaian Karakteristik Butir Soal Tes Mata Kuliah Matriks & Ruang Vektor Mahasiswa Matematika	menggunakan software iteman untuk menganalisis butir soal. 75 mahasiswa matematika angkatan 2020 Universitas Sebelas Maret yang sudah mengerjakan 10 butir soal tes uraian.	Hasil analisis item soal tes subjektif menggunakan anates uraian versi 4 yang membahas validitas, reabilitas, tingkat kesukaran, dan daya pembeda. Dengan korelasi Rxy sebesar 0,67 dan butir soal didominasi kategori signifikan, butir soal dikatakan realibel tinggi dengan indeks reabilitas sebesar 0,777, tingkat kesukaran item soal masih tergolong sangat mudah dengan persentase 70%, dan untuk daya pembeda item soal termasuk soal dengan kualitas cukup baik dengan kategori baik sebesar 30%, kategori cukup sebesar 20%, dan kategori jelek sebesar 50%. Dengan demikian karakteristik butir soal tes mata kuliah matriks dan ruang vektor dikategorikan cukup baik.
Heni Purwati	2021	Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik	Data yang digunakan pada penelitian ini merupakan data sekunder yaitu data nasional pelajaran matematika tahun 2015 tingkat SMP/MTs Propinsi Papua sebanyak 8002 siswa.	Hasil analisis menunjukkan nilai rerata dari indeks kesukaran butir kelompok tinggi berada pada kategori sedang dan kelompok rendah pada kategori sukar, nilai rerata daya pembeda kelompok tinggi dan rendah menunjukkan kategori tidak baik, yang artinya soal tersebut tidak dapat membedakan kemampuan peserta didik berkemampuan tinggi dan yang berkemampuan rendah hal ini dapat disebabkan peserta tes menjawab soal dengan cara menebak

Peneliti	Tahun	Judul Penelitian	Tempat/subjek penelitian	Hasil penelitian
Ali Ridho	2021	Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik Dan Teori Respon Aitem	Subjek penelitian ini adalah siswa SMA yang mengikuti UAN Matematika SMA tahun pelajaran 2003/2004 di Daerah Istimewa Yogyakarta. Jumlah subjek adalah 7000 orang (3500 laki-laki dan 3500 perempuan).	sehingga tidak memberikan hasil yang sesuai. Berdasarkan pendekatan TTK yang diterapkan, mean skor yang diperoleh adalah 18.628 dengan standar deviasi 6.910 , range: 3-39. Reliabilitas berdasarkan Alpha adalah 0.844 dengan standard error of measurement SEM = 2.733. Tingkat kesukaran aitem p berkisar dari 0.152 (aitem 37) sampai dengan 0.928 (aitem 23). Sementara itu, korelasi point biserial rpbis berkisar dari 0.011 (aitem 37) sampai dengan 0.543 (aitem 9).
Heridanti	2023	Karakteristik Tes Kemampuan Berpikir Kritis Siswa SMA Pada Materi Momentum dan Impuls: Perbandindinag Classical Theory Test (CTT) dan Model Rasch	Partisipan penelitian ini adalah 75 orang siswa SMA dari kelas XI dan XII dengan jumlah laki-laki sebanyak 23 orang dan jumlah perempuan sebanyak 52 orang yang berasal dari 21 sekolah menengah atas (SMA) dengan mempertimbangkan bahwa partisipan telah mempelajari materi momentum dan impuls.	Berdasarkan data keragaman (raw variance) yang diperoleh dari software Winstep adalah 60,9% tidak jauh beda dengan nilai ekspektasinya yaitu 60,7%. Hal ini menunjukkan persyaratan minimum batas unidimensi rasch terpenuhi yaitu diatas 40% (Linacre, 2011). Hal tersebut menunjukkan bahwa instrumen tes tersebut mampu mengukur keragaman kemampuan berpikir kritis siswa.
Suwanda Priyadi	2024	Analisis Tes dan Butir Soal pada Moodle	Dirancang untuk memfasilitas penilaian hasil belajar siswa.	Beberapa ahli merekomendasikan agar menggunakan formula (4) atau KR21 untuk estimasi reliabilitas tes yang datanya dikotomi (jawaban benar diberi skor 1 dan salah diberi skor 0). Studi yang dilakukan Wibowo (2016) menunjukkan bahwa penggunaan koefisien alfa (α) pada data dikotomi menghasilkan nilai reliabilitas yang underestimate. Meskipun demikian, perbedaan hasil estimasi kedua formula ini dirasa tidak terpaut jauh karena formula estimasinya relatif identik.
Sri Nurfadillah Ningsih	2023	An Application of Classical Test Theory for Item	Subjek penelitian sebanyak 80 siswa kelas XI MIPA SMA	Berdasarkan hasil analisis diperoleh koefisien reliabilitas sebesar 0,834, tingkat kesukaan

Peneliti	Tahun	Judul Penelitian	Tempat/subjek penelitian	Hasil penelitian
		Characteristic Analysis of Chemical Literacy Instruments	Negeri 4 Takalar	87% butir soal berada pada kategori sedang, dan 13% berada pada kategori sukar.

Penelitian pertama yang dibahas yaitu Muh. Syahrul Sarea (2019), memiliki tujuan untuk mengetahui karakteristik soal UAS Kelas IV Tema 1 Sekolah Dasar tahun pelajaran 2018/2019 di Desa Paramasan Bawah. Respon peserta didik dianalisis secara empiris berdasarkan pendekatan teori tes klasik dan teori respon butir untuk menguji kelayakan butir soal. Penelitian dilaksanakan pada bulan november sampai dengan Januari 2019 di Desa Paramasan Bawah Kabupaten Banjar Provinsi Kalimantan Selatan. Berdasarkan pendekatan teori tes klasik, untuk menganalisis karakteristik butir soal baik tingkat kesukaran maupun daya pembeda secara empiris digunakan bantuan program Iteman. Berdasarkan tingkat kesukaran soal, yang dikategorikan butir soal yang baik adalah rentang antara 0,3 sampai dengan 0,8 dan berdasarkan daya pembeda butir soal yang dapat ditekategorikan butir soal baik adalah 0,2 sampai dengan +2.

Selanjutnya penelitian Ismiyati (2023), yang dilakukan dengan tujuan pengabdian kepada masyarakat ini dilaksanakan untuk memecahkan permasalahan yang ada, yaitu memberikan pelatihan dan pendampingan analisis butir soal dengan pendekatan teori tes klasik berbantuan iteman guna meningkatkan mutu instrumen penilaian di MTs. Tarbiyatul Banin Winong. Apabila butir-butir soal dianalisis secara manual, maka pendidik membutuhkan waktu yang panjang. Dengan demikian, pemanfaatan software iteman dapat membantu para pendidik dalam menganalisis butir soal secara komputerisasi. Berdasarkan hasil pelaksanaan kegiatan, pelatihan ini dapat meningkatkan kemampuan para pendidik dalam menganalisis butir soal dan meningkatkan kemauan para pendidik dalam menganalisis butir soal. Kegiatan ini juga bermanfaat untuk menambah pengetahuan dan keterampilan para guru dalam meningkatkan kualitas proses pembelajaran dan pendidikan, dimana guru dapat melakukan kegiatan pengolahan analisis butir soal pada perangkat tes dengan berbantuan software iteman secara mandiri.

Kemudian penelitian yang dilakukan oleh Anggrina Gusmizain (2022). Penelitian ini memiliki tujuan menganalisis butir soal adalah untuk memperoleh informasi mengenai karakteristik setiap butir soal, baik melalui telaah butir soal maupun analisis empiris. Analisis empiris bertujuan untuk memberikan informasi mengenai karakteristik setiap butir soal tes (Purwati, 2021). Hasilnya dapat digunakan untuk mengetahui mutu soal dan mutu belajar mahasiswa dari analisis hasil tes ujian. Karakteristik tes yang berkualitas maka butir soal harus dianalisis secara tepat. Karakteristik tes yang baik menurut Nana Sudjana suatu alat penilaian tes soal dikatakan berkualitas apabila tes tersebut memenuhi dua hal yaitu validitas (ketepatan) dan reabilitas (ketetapan) (Sudjana, 2014). Selain itu, butir soal dikatakan baik juga dapat dilihat dari tingkat kesukaran, daya beda, dan keberfungsian pengecoh (Wibawa, 2019). Karakteristik butir soal uraian pada mata kuliah matriks dan ruang vektor memiliki kualitas butir soal yang baik ditinjau dari validitas, reabilitas, tingkat kesukaran, dan daya pembeda dengan korelasi Rxy sebesar 0,67 dan butir soal didominasi kategori signifikan, butir soal dikatakan realibel tinggi dengan indeks reabilitas sebesar 0,777, tingkat kesukaran item soal masih tergolong sangat

mudah dengan persentase 70%, dan untuk daya pembeda item soal termasuk soal dengan kualitas cukup baik dengan kategori baik sebesar 30%, kategori cukup sebesar 20%, dan kategori jelek sebesar 50%. Dengan demikian karakteristik butir soal tes mata kuliah matriks dan ruang vektor dikategorikan cukup baik.

Penelitian yang dilakukan oleh Heni Purwati (2021), penelitian ini bertujuan untuk menganalisis karakteristik butir soal ujian nasional matematika SMP/MTs berdasarkan teori klasik. Penelitian ini merupakan penelitian eksploratif dengan pendekatan kuantitatif. Data yang digunakan pada penelitian adalah data sekunder dengan mengambil data ujian nasional mata pelajaran matematika tahun 2015 untuk Propinsi Papua. Pada penelitian ini, responden dikelompokkan menjadi dua kelompok yaitu kelompok tinggi dan kelompok rendah berdasarkan jumlah skor yang diperoleh setiap peserta tes. Berdasarkan hasil analisis, faktor penyebab letak kesulitan siswa dalam menyelesaikan soal di atas yaitu; pertama belum bisa menerapkan prinsip gradien garis lurus, kedua belum bisa mengubah persamaan yang ada kedalam bentuk persamaan $y = mx \pm c$, ketiga kurang penguasaan dasar-dasar aljabar, dan keempat belum memahami unsur-unsur pada persamaan linier satu variabel. Indeks daya pembeda pada kelompok rendah dari 40 butir soal tidak ada butir dalam kategori baik. Rerata indeks daya pembeda sebesar $0,088 < 0,3$, artinya semua butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Reliabilitas butir tes untuk kelompok rendah sebesar 0,98 menunjukkan reliabilitas tinggi artinya hasil tes siswa mampu menggambarkan sejauh mana konsistensi hasil ukur yang dilakukan berulang-ulang terhadap responden yang memiliki gejala sama dengan alat ukur yang sama.

Penelitian yang dilakukan oleh Ali Ridho (2022), tujuan penelitian ini adalah untuk mengungkap secara empirik karakteristik Tes UAN Matematika SMA tahun pelajaran 2003/2004 berdasarkan pendekatan TRA, yaitu: (1) invariansi traits level peserta θ berdasar model 1PL, 2PL dan 3PL, (2) invariansi parameter aitem pada model 1PL, 2PL dan 3PL, serta (3) membandingkan hasil pendekatan metode TTK dan TRA. Evaluasi menggunakan TTK menunjukkan bahwa tes UAN Matematika SMA memiliki reliabilitas internal sebesar 0.850. Tingkat kesukaran p terentang dari 0.234 sampai dengan 0.928 dan daya beda $rpbis$ terentang dari 0.210 sampai dengan 0.545. Satu kesimpulan penting yang dapat ditarik adalah bahwa masing-masing aitem memiliki daya beda yang berbeda-beda.

Penelitian yang dilakukan oleh Suwanda Priyadi (2024), bertujuan untuk mengulas analisis tes dan butir soal pada Moodle. Ulasan tersebut meliputi formula estimasi analisis tes dan butir soal yang digunakan Moodle beserta interpretasinya. Secara umum, analisis tes dan butir soal pada Moodle dapat dimanfaatkan untuk mengevaluasi tes yang akan digunakan menilai hasil belajar siswa. Baik itu tes yang sifatnya sumatif, maupun tes yang sifatnya formatif. Interpretasi yang dilakukan dapat menggunakan pendekatan teori tes klasik. Selain itu, analisis tes dan butir soal pada Moodle memiliki nilai praktis dan ekonomis bagi para guru. Mengingat analisis tes dan butir soal membutuhkan komputasi yang rumit. Selain itu, perangkat lunak komersil untuk analisis butir soal dan tes dijual dengan harga yang relatif mahal.

Terakhir penelitian yang dilakukan oleh Sri Nurfadillah Ningsih (2024), tujuan penelitian ini adalah mendeskripsikan karakteristik butir instrument penilaian berbasis literasi kimia materi laju reaksi dengan pendekatan teori tes klasik. Hasil analisis diperoleh

koefisien reliabilitas sebesar 0,834, tingkat kesukaan 87% butir soal berada pada kategori sedang, dan 13% berada pada kategori sukar. Uji daya beda, sebanyak 73% butir soal berada pada kategori sangat baik, dan untuk uji distraktor, 93% dari keseluruhan butir soal telah menunjukkan distraktor yang berfungsi dengan baik,

Berdasarkan pembahasan yang diperoleh dari beberapa penelitian diatas dapat diketahui bahwa teori tes klasik merupakan pendekatan yang digunakan dalam menganalisis hasil ujian, dimana asumsi dasar yang digunakan adalah skor yang didapat atau skor mentah terdiri dari skor murni hasil ujian dan error pengukuran (Prieto et al., 2003; Sumintono, 2018). Skor mentah adalah satu-satunya yang bersifat nyata sedangkan skor murni dan error pengukuran bersifat laten atau tidak dapat diprediksi dan diamati secara langsung yang dapat muncul menjadi nyata setelah melalui proses estimasi (Dwiliesanti & Yudiarto, 2022; Prieto et al., 2003; Putra et al., 2021; Sumintono, 2018). Pada dasarnya, penggunaan skor mentah untuk menganalisis hasil ujian memiliki beberapa kelemahan. Kelemahan yang pertama adalah skor mentah merupakan jumlah jawaban benar dari tes yang dikerjakan siswa sehingga tidak dapat dikatakan sebagai hasil pengukuran (U. Kurniawan & Andriyani, 2018). Skor mentah hanya sebagai informasi awal yang biasanya dinyatakan dalam bentuk presentase (%), sehingga skor mentah tidak bisa mencerminkan kemampuan seseorang terhadap suatu tes juga tidak dapat menjelaskan tingkat kesulitan tiap butirnya (Soleha & Miftahus, 2020). Kelemahan selanjutnya adalah skor mentah memiliki definisi kuantitatif yang rendah, data yang didapatkan akan selalu berubah bergantung pada banyaknya soal, sedangkan presentase jumlah jawaban betul selalu tergantung pada tingkat kesulitan suatu butir instrumen (Sumintono, 2016). Kemudian juga antara skor mentah dengan presentase jawaban benar tidak selalu bersifat linier, artinya untuk data hasil pengukuran dengan CTT, siswa yang memiliki skor 80 (dalam skala 0-100) selalu berarti lebih tinggi jika dibandingkan dengan siswa yang memiliki skor 75. Padahal secara empirik, antara kedua siswa tersebut terdapat kemungkinan memiliki kemampuan yang setara, atau lebih jelas jika kedua siswa tersebut memiliki skor yang sama, maka tidak dapat diketahui secara pasti perbedaan kemampuan kedua siswa tersebut. Hal ini menandakan ketidakpastian makna dari skor yang didapat (Molenaar, 2020; Ramdani et al., 2019). Oleh karena itu diperlukan pendekatan yang berbeda dari pada menggunakan skor mentah.

Pada hakikatnya pendekatan yang digunakan dalam analisis instrumen pengukuran terbagi menjadi dua kelompok yaitu Classical Test Theory dan Latent Trait Model (Adebule, 2013; Cursio et al., 2019). Dalam praktek penilaian terdapat beberapa hal yang memerlukan perhatian eksplisit diantaranya tiga elemen berupa aspek kognitif, ujian, interpretasi hasil ujian serta korelasi antara ketiga elemen tersebut (Sumintono & Widhiarso, 2015). Ketiga elemen tersebut harus di dasarkan pada pengetahuan modern tentang bagaimana siswa belajar dan bagaimana sebaiknya proses pembelajaran tersebut diukur (Warsono et al., 2020; Yunida & Arthur, 2023). Pengembangan instrumen saat ini sudah mengarah pada konsep modern test salah satu pendekatan yang muncul adalah Teori Respon Butir atau Item Response Theory (IRT) dan Rasch Model (Erfan et al., 2020). Tujuan utama pendekatan modern adalah menghasilkan suatu skala pengukuran yang sama antara data ordinal berupa informasi mengenai kemampuan siswa peserta tes dan kualitas butir soal yang diujikan (Long et al., 2020; Sumintono, 2018; Sumintono & Widhiarso, 2015). Skala yang dihasilkan dalam pendekatan ini kualitasnya sama seperti

pengukuran dalam dimensi-dimensi fisika berupa data interval. Proses penyamaan skala dengan standar yang digunakan dalam ilmu eksakta tersebut biasa disebut dengan proses kalibrasi instrumen (Sumintono & Widhiarso, 2013, 2015; Widhiarso, 2011). Dengan skala yang sama maka pengukuran dapat memberikan informasi yang lebih akurat mengenai peserta tes maupun kualitas soal yang diberikan.

KESIMPULAN DAN SARAN

Asesmen adalah elemen penting dalam pendidikan, karena berperan dalam mendukung kompetensi peserta didik. Oleh karena itu, pendidik seharusnya menyediakan instrumen asesmen yang berkualitas. Salah satu cara untuk menggambarkan kualitas instrumen tes adalah melalui analisis butir soal. Penelitian yang dilakukan dalam lima tahun terakhir mengenai analisis butir soal di SMK menunjukkan bahwa telah ada upaya untuk meningkatkan kualitas evaluasi pembelajaran dengan menerapkan pendekatan teori tes klasik. Pendekatan ini adalah yang paling umum digunakan karena analisisnya yang sederhana dan mudah. Namun, seiring dengan perkembangan, kini ada analisis baru yang lebih dianjurkan, yaitu teori tes modern seperti Teori Respon Butir (Item Response Theory/IRT) dan Model Rasch. Teori-teori ini menawarkan pendekatan analisis butir yang menggunakan skala pengukuran yang konsisten antara data ordinal yang mencerminkan kemampuan siswa dan kualitas butir soal yang diuji. Dengan skala yang sama, pengukuran ini dianggap dapat memberikan informasi yang lebih akurat mengenai peserta tes dan kualitas soal yang diberikan.

DAFTAR RUJUKAN

- Ali Ridho (2022). Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik Dan Teori Respon Aitem. <https://journal.unair.ac.id>
- Anggrina G. (2022). Karakteristik Butir Soal Tes Mata Kuliah Matriks & Ruang Vektor Mahasiswa Matematika. *Jurnal Evaluasi Pendidikan*, Volume 13, Nomor 2
- Haridanti, Winny L., Yuyu R.T. (2023). Karakteristik Tes Kemampuan Berpikir Kritis Siswa SMA Pada Materi Momentum dan Impuls: Perbandingan Classical Theory Test (CTT) dan Model Rasch. <https://doi.org/10.17509/wapfi.v8i1.30958>
- Heni P., Heri R., Jailani, Trie H. R. (2021). Analisis Karakteristik Butir Soal Ujian Nasional Matematika SMP/MTs Berdasarkan Pendekatan Teori Tes Klasik. *Jurnal Sains dan Edukasi Sains* Vol.4, No.2, Agustus 2021: 46-51 <https://doi.org/10.24246/juses.v4i2p46-51>
- Ismiyati, Teguh H. R., Tusyanah, Mar'atus S. (2023). Pelatihan Analisis Butir Soal Berdasarkan Teori Tes Klasik Berbantuan Iteaman untuk Meningkatkan Kualitas Instrumen Penilaian. *JAPI (Jurnal Akses Pengabdian Indonesia)*, Volume 8, Nomor 2
- Muh. S. S., Rosnia R. (2019). Karakteristik Butir Soal: Classical Test Theory Vs Item Response Theory? *Didaktika: Jurnal Kependidikan Fakultas Tarbiyah IAIN Bone*, Vol. 13, No. 1
- Prasetyo, A. (2016). Analisis Butir Soal Sertifikasi Kompetensi P1 Bidang Teknik Mesin Di LSP SMK N 1 Magelang *Item Analysis of P1 Certification of Competence in Mechanical Engineering At Lsp Smk N 1 Magelang*. 385–394.
- Sri N. N., Edi I., Widihastuti (2024). An Application of Classical Test Theory for Item Characteristic Analysis of Chemical Literacy Instruments. *Jurnal Pendidikan Kimia*

Indonesia. Volume 7, Issue 2, pp. 58-68. Open Access:
<https://doi.org/10.23887/jpki.v7i2.68983>

Suwanda P. (2024). Analisis Tes dan Butir Soal pada Moodle. *Jurnal Evaluasi Pendidikan*
Volume 15, Nomor 1