

## PENGGUNAAN TEORI TES KLASIK PADA ASESMEN PEMBELAJARAN DI SMK DAN PELUANG PERBAIKANNYA

Fina Luthfiyah<sup>a</sup>, Riyan Arthur<sup>b</sup>, Arris Maulana<sup>c</sup>

<sup>abc</sup>Universitas Negeri Jakarta, Indonesia

**Correspondence:** finaluthfiyah\_1503619005@mhs.unj.ac.id

### *Abstract*

*This research is a descriptive research that aims to find out the general approach and should be used for the analysis of assessment items. The research method used is a Systematic Literature Review where materials in the form of reputable national and international journals and relevant books are collected and analyzed to serve as a basis for identifying gaps that occur between a theory and reality in the field. The results obtained from this study explain that there have been several studies conducted in the past 5 years that discuss the analysis of test items carried out in learning in SMK. This explanation shows that there have been efforts to improve the quality of learning evaluation using the classical test theory approach. The classical test theory approach is the most common approach, because the analysis is easy and simple. Nonetheless, along with developments there is a new analysis that is more recommended, namely modern test theory in the form of Item Response Theory or Item Response Theory (IRT) and the Rasch Model. This theory is an item analysis approach that uses the same measurement scale between ordinal data in the form of information about the ability of students taking the test and the quality of the items being tested. The same scale, the measurement is considered to be able to provide more accurate information about the ability of person and the quality of the questions given.*

**Keyword:** *assessment, classical test theory, modern test theory, IRT, rasch model.*

### **Abstrak**

Penelitian ini merupakan penelitian deskriptif yang ditujukan untuk mengetahui pendekatan yang umum dan seharusnya digunakan untuk analisis butir asesmen. Metode penelitian yang digunakan adalah *Systematic Literature Review* dimana bahan-bahan berupa jurnal nasional maupun internasional bereputasi dan buku-buku yang relevan dikumpulkan dan dianalisis untuk dijadikan dasar dalam mengidentifikasi kesenjangan yang terjadi antara suatu teori dan kenyataan di lapangan. Hasil yang diperoleh dari penelitian ini menjelaskan bahwa terdapat beberapa penelitian yang dilakukan dalam kurun waktu 5 tahun kebelakang yang membahas mengenai analisis butir soal yang dilakukan dalam pembelajaran di SMK. Penjelasan tersebut menunjukkan bahwa sudah terdapat usaha memperbaiki kualitas evaluasi pembelajaran dengan menggunakan pendekatan teori tes klasik. Pendekatan teori tes klasik merupakan pendekatan yang paling umum dilakukan, karena analisisnya mudah dan sederhana. Meskipun demikian, seiring perkembangan terdapat analisis baru yang lebih direkomendasikan yaitu teori tes modern berupa Teori Respon Butir atau Item Response Theory (IRT) dan Rasch Model. Teori tersebut merupakan pendekatan analisis butir yang menggunakan skala pengukuran yang sama antara data ordinal berupa informasi mengenai kemampuan siswa peserta tes dan kualitas butir soal yang diujikan. Skala yang sama maka pengukuran dianggap dapat memberikan informasi yang lebih akurat mengenai peserta tes maupun kualitas soal yang diberikan.

**Katakunci:** asesmen, teori tes klasik, teori tes modern, IRT, rasch model

---

## Pendahuluan

Asesmen merupakan salah satu komponen penting dalam pendidikan, khususnya dalam proses pembelajaran (Anggrena et al., 2022). Asesmen dilakukan sebagai upaya untuk mengukur tingkat ketercapaian tujuan pembelajaran dan mengumpulkan informasi perkembangan belajar siswa pada berbagai aspek diantaranya aspek kognitif, psikomotor dan afektif (Septiani et al., 2019). Pentingnya asesmen pada pendidikan kejuruan dalam hal praktik kinerja ialah untuk menunjang kompetensi keahlian peserta didik (Sutami, 2020). Penjelasan diatas menunjukkan bahwa tingkat ketercapaian pembelajaran merupakan salah satu hal yang dapat dilihat dari hasil asesmen. Hasil dari asesmen dapat digunakan oleh pendidik untuk pengambilan keputusan pembelajaran hingga tercipta peningkatan kualitas proses pembelajaran (Matondang et al., 2022). Karena itu, kemampuan melaksanakan asesmen menjadi salah satu kompetensi yang harus dimiliki oleh guru. Kompetensi ini dapat dikuasai apabila pendidik telah mengetahui konsep pengembangan asesmen dan penerapannya pada proses pembelajaran. Saat ini, masih banyak pendidik yang memiliki pemahaman yang tergolong rendah terhadap konsep asesmen pembelajarannya meskipun telah memenuhi kualifikasi akademik dan pernah mengikuti pelatihan (Irawati et al., 2018; Y. R. Kurniawan et al., 2016; Magdalena et al., 2020). Oleh sebab itu, perlu pembiasaan dalam hal pelaksanaan asesmen yang dilakukan pendidik mulai dari pengembangan instrumen sampai pada analisis butir dan pengambilan keputusan berdasarkan hasil yang diperoleh (Abdul, 2015).

Asesmen dapat dilakukan melalui kegiatan tes maupun non-tes dengan menggunakan instrumen sebagai alat evaluasinya (Irawati et al., 2018). Instrumen evaluasi yang umum digunakan adalah tes objektif pilihan ganda, karena bentuknya yang dapat digunakan untuk mengukur ketercapaian banyak pokok materi pembelajaran dalam satu kesatuan instrumen sekaligus (Abdul, 2015; Suseno, 2017). Selain itu juga instrumen tes jenis ini dapat melingkupi banyak tingkatan berpikir mulai dari tingkat pengetahuan, sintesis hingga analisis. Kemudahan penskoran instrumen tes obyektif juga membuat jenis tes ini lebih banyak dipilih karena dapat dikoreksi oleh siapa saja (Y. R. Kurniawan et al., 2016). Proses pengembangan instrumen asesmen objektif dilakukan melalui beberapa tahap penting diantaranya, konstruksi tes, analisis butir instrumen tes, dan implementasi atau penyebarluasan tes (Septiani et al., 2019). Dari beberapa tahap tersebut, salah satu tahap penting untuk mengetahui kualitas instrumen dan berfungsi tidaknya tiap butir adalah tahap analisis. Analisis butir instrumen pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (*qualitatif control*) dan analisis kuantitatif (*quantitatif control*) (Pintastyo & Sudji, 2017). Analisis kualitatif atau disebut juga dengan validitas logis (*logical validity*) merupakan analisis validitas yang dilakukan untuk melihat kualitas butir instrumen berdasarkan penilaian validator atau ahli. Sedangkan analisis soal secara kuantitatif atau validitas empiris (*empirical validity*) dilakukan untuk melihat berfungsi atau tidaknya sebuah instrumen tes, setelah diujicobakan kepada 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 di SMK dan Peluang Perbaikannya.

### Hasil Penelitian 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 di SMK:

**Tabel 1. Penelitian Penggunaan Teori Tes Klasik pada Asesmen di SMK**

Peneliti	Tahun	Judul Penelitian	Tempat/Subjek Penelitian	Hasil Penelitian
Agus Prasetyo	2016	Analisis Butir Soal Sertifikasi Kompetensi P1 Bidang Teknik Mesin di LSP SMK N 1 Magelang	109 peserta sertifikasi kompetensi teknik pemesinan pada bidang pekerjaan Operasi Mesin dan Proses LSP SMK N 1 Magelang	Hasil penelitian menunjukkan bahwa 5.71% soal tidak memenuhi aspek materi, 45.71% soal tidak memenuhi aspek konstruksi, dan memenuhi aspek bahasa. Skor yang tidak valid sebagai ukuran kompetensi peserta tes LSP SMK N 1 Magelang tahun 2015. Soal masuk dalam kualitas kurang baik, dengan 62.86% soal mudah, 54.29% soal yang dapat diterima, pengecoh berfungsi sempurna sebanyak 4.29%, dan reliabilitas soal sebesar 0.613 pada uji mandiri. Dalam uji unjuk kerja bubut 92.85% butir soal dengan kategori peka dan 72.97% butir soal dengan kategori peka dalam uji unjuk kerja frais.
Dian Normalitasari, Fitri Alfarisa	2020	Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi SMK Berdasarkan Teori Tes Klasik dan Teori Respons Butir	Siswa Kelas XII SMK Jurusan Akuntansi di Sleman	Hasil penelitian menunjukkan: (1) karakteristik soal kualitatif secara umum tergolong baik. Pada aspek materi terdapat 92,5% soal yang baik, aspek konstruksi dan bahasa 100% soal baik. (2) karakteristik soal kuantitatif berdasarkan pendekatan teori tes klasik memiliki tingkat kesukaran kategori sangat mudah = 7,5%, mudah = 25%, sedang = 67,5% sukar dan sangat sukar = 0%. (3) karakteristik soal kuantitatif berdasarkan pendekatan teori respons butir memiliki tingkat kesukaran kategori sangat mudah = 7,5%, mudah = 15%, sedang = 60%, sukar = 17,5% dan sangat sukar = 0%.

Sutami, Yusak Hudiyo, M. Ilyas	2020	Pengembangan Instrumen Asesmen Higher Order Thinking Skills (Hots) pada Mata Pelajaran Bahasa Indonesia SMK Dan SMK	Siswa SMA Negeri 1 Samarinda, SMA Negeri 3 Samarinda, SMK Negeri 15 Samarinda, dan SMKS TI Airlangga Samarinda.	Hasil analisis menunjukkan perangkat tes mata diklat akuntansi Paket A dan Paket B sebagian besar butir soal yang disiapkan belum memenuhi kriteria tingkat kesukaran “sedang”. Daya beda butir soal dominan “jelek”, dan pengecoh tidak berfungsi secara efektif. Berdasarkan analisis kemampuan, dapat disimpulkan kompetensi peserta ujian mata diklat akuntansi SMK di kabupaten Ende belum mencapai kompetensi yang diharapkan.
Fransiskus Korosando, Manggu Ngguna Raji	2020	Kemampuan Akademik Peserta Ujian Kompetensi Sekolah Menengah Kejuruan	Siswa SMK di Kabupaten Ende	instrumen asesmen HOTS berupa soal tes HOTS yang terdiri dari 40 butir soal pilihan ganda dan 5 butir soal uraian dari aspek materi, konstruksi, dan bahasa dinyatakan valid dan layak digunakan. Instrumen tersebut mempunyai koefisien reliabilitas sebesar 0,88 kategori sangat tinggi untuk soal pilihan ganda dan sebesar 0,79 kategori tinggi untuk soal uraian. Soal pilihan ganda memiliki rata-rata tingkat kesukaran 0,57 (sedang), rata-rata daya pembeda 0,44 (baik), dan semua pengecoh berfungsi baik. Soal uraian memiliki rata-rata tingkat kesukaran 0,60 (sedang) dengan rata-rata daya pembeda 0,45 (baik).
Ninda Ayu Narassati, Rosmawita Saleh, Riyan Arthur	2021	Pengembangan Alat Evaluasi Berbasis Hots Menggunakan aplikasi Quizizz Pada Mata Pelajaran Mekanika Teknik dalam Pembelajaran Jarak Jauh	X DPIB dan KGSP di beberapa SMK se-DKI Jakarta	Dari 20 Butir Soal yang dikembangkan, diperoleh Validitas alat evaluasi menggunakan aplikasi Quizizz dengan jumlah sampel 68 siswa sebanyak 3 soal tidak valid dan 17 soal valid. Reliabilitas alat evaluasi menggunakan Aplikasi Quizizz dinyatakan reliabel sebesar 0,871. Kualitas butir soal dari segi tingkat kesukaran menunjukkan 19 soal sedang dan 1 soal sukar. Indeks daya pembeda butir soal menunjukkan tingkat yang semakin baik pada pemberian tes untuk jumlah sampel yang semakin besar. Kualitas butir soal dipengaruhi oleh karakteristik dan jumlah sampel. Dan berdasarkan analisis faktor konfirmatori terdapat 16 butir soal yang signifikan dan 4 butir soal tidak signifikan pada penggunaan Aplikasi Quizizz.

Penelitian pertama yang dibahas yaitu Prasetyo (2016) memiliki tujuan untuk melihat karakteristik beserta kualitas soal ujian mandiri yang di tinjau dari aspek analisis kuantitatif dan

kualitatif serta persentase butir dengan kategori peka terhadap soal tes uji unjuk kerjadalam sertifikasi kompetensi bidang teknik mesin di LSP SMKN 1 Magelang. Penelitian ini berupa penelitian analisis terhadap dokumen dengan memanfaatkan desain penelitian survei. Pengumpulan data menggunakan sarana dokumentasi. Populasi dalam penelitian ini ialah soal beserta semua jawaban tes uji mandiri dan uji unjuk kerja tes sertifikasi kompetensi P1 di LSP SMKN 1 Magelang pada tahun 2015. Hasil penelitian ini menunjukkan bahwa 5.71% soal tidak memenuhi aspek materi, serta 45.71% soal tidak memenuhi aspek konstruksi dan memenuhi aspek bahasa. Skor yang tidak valid digunakan sebagai ukuran kompetensi peserta tes LSP SMKN 1 Magelang tahun 2015. Butir soal dalam tes masuk dalam kategori soal dengan kualitas kurang baik dengan rincian 62.86% soal mudah, 54.29% soal dapat diterima. Terdapat sebanyak 4.29% pengecoh dalam butir soal yang dapat berfungsi secara sempurna serta nilai reliabilitas soal sebesar 0.613 pada uji mandiri. Dalam uji unjuk kerja bubut 92.85% butir soal dengan kategori peka dan 72.97% butir soal dikategorikan peka dalam uji unjuk kerja frais.

Selanjutnya penelitian Purnama & Alfarisa, (2020) yang dilakukan dengan tujuan mengetahui karakteristik butir soal tes *Try Out* Teori Kejuruan Akuntansi. Analisis yang dilakukan untuk mengetahui karakteristik butir soal yaitu analisis kualitatif berupa aspek materi, konstruksi, bahasa dan secara karakteristik butir kuantitatif yaitu tingkat kesukaran. Subjek penelitian yang digunakan dalam penelitian ini adalah siswa Kelas XII SMK Jurusan Akuntansi di Sleman., dengan objek penelitian berupa jawaban tes sejumlah 300. Teknik pengumpulan data menggunakan tes dengan instrumen sejumlah 40 butir. Teknik analisis data kualitatif dilakukan dengan telaah butir oleh ahli sedangkan secara kuantitatif menggunakan pendekatan Teori Tes Klasik dan Teori Respons Butir Model 1 Parameter Logistik dengan program QUEST. Hasil penelitian menunjukkan bahwa pada analisis kualitatif secara umum butir soal tergolong dalam kategori baik. Pada aspek materi terdapat 92,5% soal tergolong baik, sedangkan aspek konstruksi dan bahasa 100% soal tergolong baik. Kemudian pada analisis kuantitatif berdasarkan pendekatan teori tes klasik memiliki tingkat kesukaran kategori sangat mudah = 7,5%, mudah = 25%, sedang = 67,5% sukar dan sangat sukar = 0%. (3) karakteristik soal kuantitatif berdasarkan pendekatan teori respons butir memiliki tingkat kesukaran kategori sangat mudah = 7,5%, mudah = 15%, sedang = 60%, sukar = 17,5% dan sangat sukar = 0%.

Penelitian Sutami, (2020) menjadi penelitian ketiga yang dibahas pada artikel ini. Penelitian tersebut bertujuan untuk menghasilkan instrumen asesmen *HOTS* bahasa Indonesia yang valid dan reliabel serta mendeskripsikan kualitas instrumen *HOTS* tersebut untuk mengukur keterampilan berpikir tingkat tinggi pada peserta didik SMA dan SMK kelas X semester 1. Penelitian ini merupakan penelitian pengembangan Borg & Gall, dengan tahapan penelitian pendahuluan, perencanaan, pengembangan produk awal, revisi produk awal, uji coba terbatas, uji coba lapangan, dan revisi produk akhir. Hasil penelitian menunjukkan bahwa instrumen asesmen *HOTS* berupa soal tes *HOTS* yang terdiri dari 40 butir soal pilihan ganda dan 5 butir soal uraian dari aspek materi, konstruksi, dan bahasa dinyatakan valid dan layak digunakan. Instrumen tersebut mempunyai koefisien reliabilitas sebesar 0,88 kategori sangat tinggi untuk soal pilihan ganda dan sebesar 0,79 kategori tinggi untuk soal uraian. Dengan menggunakan pendekatan teori tes klasik, didapatkan hasil analisis soal pilihan ganda memperoleh rata-rata tingkat kesukaran sejumlah 0,57 (sedang), rata-rata daya pembeda sejumlah 0,44 (baik), dan semua pengecoh berfungsi baik. Sedangkan soal uraian memiliki rata-rata tingkat kesukaran sejumlah 0,60 (sedang) dengan rata-rata daya pembeda 0,45 (baik).

Kemudian penelitian yang dilakukan oleh Korosando & Raji (2020). Penelitian ini memiliki tujuan mendeskripsikan parameter butir soal pada instrument ujian kompetensi mata diklat paket A dan paket B, serta mendeskripsikan ability peserta tes kompetensi peserta didik SMK di Kabupaten Ende. Riset ini menggunakan jenis penelitian kuantitatif. Dengan pengumpulan data menggunakan teknik dokumentasi jawaban peserta tes pada lembar jawaban. Data kemudian dianalisis menggunakan pendekatan teori tes klasik. Hasil analisis menunjukkan perangkat tes mata diklat akuntansi Paket A dan Paket B sebagian besar butir soal belum memenuhi kriteria tingkat kesukaran butir yang tergolong “sedang”. Kemudian daya beda butir soal cenderung “jelek”, dan

pengecoh tidak berfungsi dengan baik. Sedangkan berdasarkan analisis kemampuan, dapat disimpulkan kompetensi peserta ujian mata diklat akuntansi SMK di Kabupaten Ende belum mencapai kompetensi sebagaimana yang diharapkan.

Terakhir penelitian yang dilakukan oleh Narasati (2021), yang termasuk dalam model penelitian dan pengembangan. Penelitian tersebut bertujuan untuk mengembangkan alat evaluasi menggunakan aplikasi *Quizizz* pada mata pelajaran Mekanika Teknik dalam pembelajaran jarak jauh pada kelas X DPIB dan KGSP di beberapa SMK se-DKI Jakarta. Model penelitian pengembangan yang digunakan ialah ADDIE (*Analysis, Design, Development, Evaluation*). Analisa data yang digunakan adalah analisa kualitatif dengan menghitung validitas isi menggunakan formula *aiken's V* dan kuantitatif dengan teori tes klasik dan *confirmatory factor analysis*. Hasil penelitian menunjukkan bahwa pengembangan alat evaluasi Mekanika Teknik mendapatkan kelayakan ahli materi sebesar 85,88% dengan kategori sangat layak, ahli instrumen sebesar 80,62% dengan kategori layak, dan ahli bahasa sebesar 85,14% dengan kategori sangat layak. Hasil uji validitas alat evaluasi menggunakan aplikasi *Quizizz* dengan jumlah sampel 68 siswa sebanyak 3 soal tidak valid dan 17 soal valid. Reliabilitas alat evaluasi menggunakan Aplikasi *Quizizz* dinyatakan reliabel sebesar 0,871. Kualitas butir soal dari segi tingkat kesukaran menunjukkan 19 soal sedang dan 1 soal sukar. Indeks daya pembeda butir soal menunjukkan tingkat yang semakin baik pada pemberian tes untuk jumlah sampel yang semakin besar. Kualitas butir soal dipengaruhi oleh karakteristik dan jumlah sampel. Dan berdasarkan analisis faktor konfirmatori terdapat 16 butir soal yang signifikan dan 4 butir soal tidak signifikan pada penggunaan Aplikasi *Quizizz*.

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 & Yudianto, 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.

### Simpulan dan Rekomendasi

Asesmen merupakan komponen penting pada pendidikan kejuruan terutama SMK dalam menunjang kompetensi peserta didik. Oleh sebab itu, sudah seharusnya pendidik memberikan seperangkat instrumen asesmen yang berkualitas. Salah satu usaha untuk mendeskripsikan kualitas instrumen tes adalah dengan melakukan analisis butir soal. Penelitian yang dilakukan dalam kurun waktu 5 tahun kebelakang tentang analisis butir soal yang dilakukan dalam pembelajaran di SMK, menunjukkan bahwa sudah terdapat usaha memperbaiki kualitas evaluasi pembelajaran dengan menggunakan pendekatan teori tes klasik. Pendekatan teori tes klasik merupakan pendekatan yang paling umum dilakukan, karena analisisnya mudah dan sederhana. Meskipun demikian, seiring perkembangan terdapat analisis baru yang lebih direkomendasikan yaitu teori tes modern berupa Teori Respon Butir atau *Item Response Theory* (IRT) dan Rasch Model. Teori tersebut merupakan pendekatan analisis butir yang menggunakan skala pengukuran yang sama antara data ordinal berupa informasi mengenai kemampuan siswa peserta tes dan kualitas butir soal yang diujikan. Skala yang sama maka pengukuran dianggap dapat memberikan informasi yang lebih akurat mengenai peserta tes maupun kualitas soal yang diberikan.

### Daftar Pustaka

- Abdul, K. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Al-Ta'dib*, 8(2), 70–81.
- Adebule, S. O. (2013). Item Response Theory as a Basis for Measuring Latent Trait of Interest. *Greener Journal of Social Sciences*, 3(7), 378–382. <https://doi.org/10.15580/gjss.2013.7.062513691>
- Anggrena, Y., Ginanti, D., Felicia, N., Herutmi, I., Leli, A., Iswoyo, S., Hartini, Y., & Mahardika, R. L. (2022). *Pembelajaran dan Asesmen* (S. Pratiwi & M. F. Jubaedi (eds.); 1st ed.).
- Cahyono, E. A., Sutomo, & Harsono, A. (2019). Literatur Review: Panduan Penulisan dan Penyusunan. *Jurnal Keperawatan*, 12.
- Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *Procedia - Social and Behavioral Sciences*, 129(May), 133–139. <https://doi.org/10.1016/j.sbspro.2014.03.658>
- Cursio, J. F., Mermelstein, R. J., & Hedeker, D. (2019). Latent trait shared-parameter mixed models for missing ecological momentary assessment data. *Statistics in Medicine*, 38(4), 660–673. <https://doi.org/10.1002/sim.7989>
- Dwiliesanti, W. G., & Yudiarso, A. (2022). Rasch Analysis of The Indonesian Version of Individual Work Performance Questionnaire (IW PQ). *JP3I (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, 11(2), 152–166. <https://doi.org/10.15408/jp3i.v11i2.24157>
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19.
- Hendryadi. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen Dan Bisnis (JRMB) Fakultas Ekonomi UNIAT*, 2(2), 169–178. <https://doi.org/10.36226/jrmb.v2i2.47>
- Irawati, H., Saifuddin, M. F., & Ma'rifah, D. R. (2018). Pengembangan Instrumen Tes Dan Non Tes Dalam Rangka Menyiapkan Penilaian Autentik Pada Kurikulum 2013 Di Smp/Mts Muhammadiyah Se-Kabupaten Bantul. *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian Kepada Masyarakat*, 1(2), 503. <https://doi.org/10.12928/jp.v1i2.362>
- Korosando, F., & Raji, M. N. (2020). Kemampuan Akademik Peserta Ujian Kompetensi Sekolah

- Menengah Kejuruan. *Journal of Education and Instruction (JOEAI)*, 3(2), 142–150. <https://doi.org/10.31539/joeai.v3i2.335>
- Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, 6(1), 34–39.
- Kurniawan, Y. R., Marpaung, R. R. T., & Yolida, B. (2016). *Identifikasi Butir Soal Biologi pada SMK Kesehatan Nurul Islam Husada Lampung Selatan*. 1–23.
- Long, S. K., Sato, T., Millner, N., Loranger, R., Mirabelli, J., Xu, V., & Yamani, Y. (2020). Empirically And Theoretically Driven Scales On Automation Trust: A Multi-Level Confirmatory Factor Analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 1829–1832. <https://doi.org/10.1177/1071181320641440>
- Magdalena, I., Fauzi, H. N., Putri, R., & Tangerang, U. M. (2020). Pentingnya Evaluasi Dalam Pembelajaran Dan Akibat Memanipulasinya. *Pendidikan Dan Sains - Stitpn*, 2, 244–257. <https://ejournal.stitpn.ac.id/index.php/bintang/article/download/986/680/>
- Matondang, P., Azizah, B., Masdalipa, R., & ... (2022). Workshop Analisis Butir Soal Pada Dosen Institut Teknologi Pagar Alam. *Jurnal Pengabdian Kepada Masyarakat*, 5(01), 12–18. <http://ejournal.lppmsttpagaralam.ac.id/index.php/ngabdimas/article/view/491%0Ahttp://ejournal.lppmsttpagaralam.ac.id/index.php/ngabdimas/article/download/491/350>
- Molenaar, D. (2020). Review of Handbook of Item Response Theory, Volume II: Statistical Tools . *Journal of Educational and Behavioral Statistics*, 45(4), 507–511. <https://doi.org/10.3102/1076998620911932>
- Narasati, N. A., Saleh, R., & Arthur, R. (2021). Pengembangan Alat Evaluasi Berbasis HOTS Menggunakan Aplikasi Quizizz pada Mata Pelajaran Mekanika Teknik dalam Pembelajaran Jarak Jauh. *Jurnal Pendidikan Teknik Sipil*, 3(December), 169–180. <https://doi.org/10.21831/jpts.v3i2.43919>
- Pérez-Escolar, M., Lilleker, D., & Tapia-Frade, A. (2023). A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation. *Media and Communication*, 11(2), 1–12. <https://doi.org/10.17645/mac.v11i2.6453>
- Pintastyo, L., & Sudji. (2017). Analisis Soal Ujian Akhir Semester Genap Pelajaran Gambar Teknik Mesin SMK Muhammadiyah 1 Bantul. *Jurnal Pendidikan Vokasi Teknik Mesin*, 5(6), 411–418. <https://journal.student.uny.ac.id/index.php/mesin/article/view/9441>
- 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.
- Prieto, L., Alonso, J., & Lamarca, R. (2003). *Classical test theory versus Rasch analysis for quality of life questionnaire reduction*. 13, 1–13.
- Purnama, D. N., & Alfarisa, F. (2020). Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi Smk Berdasarkan Teori Tes Klasik Dan Teori Respons Butir. *Jurnal Pendidikan Akuntansi Indonesia*, 18(1), 36–46. <https://doi.org/10.21831/jpai.v18i1.31457>
- Putra, Z. H., Hermita, N., & Alim, J. A. (2021). Analisis Pengetahuan Matematika, Didaktika, dan Teknologi Calon Guru Sekolah Dasar Menggunakan Rasch Model. *Mosharafa: Jurnal Pendidikan Matematika*, 10(3), 345–356. <https://doi.org/10.31980/mosharafa.v10i3.1042>
- Ramdani, Z., Widyastuti, T., & Ferdian, F. R. (2019). Penerapan Analisis Teori Klasik Model Rasch dan Computer Based Test Moodle: Sebuah Pilot Studi. *Indonesian Journal of Educational Assesment*, 1(2), 21. <https://doi.org/10.26499/ijea.v1i2.9>
- Septiani, D., Widiyawati, Y., & Nurwahidah, I. (2019). Pengembangan Instrumen Tes Literasi Sains PISA Aspek Penilaian literasi sains berskala Internasional adalah Programme of International Student Assesment ( PISA ) dan Trends in International Mathematics and Science Study ( TIMSS ) yang. *Science Education and Application Journal (SEAJ)*, 1(2), 46–55.



- 
- Soleha, S., & Miftahus, S. (2020). Pengujian Validitas dan Reliabilitas Instrumen Manajemen Diri Remaja: Rasch Model Analysis. *Jurnal Bimbingan Dan Konseling Terapan*, 4(1), 103–118.
- Sumintono, B. (2016). *Aplikasi Pemodelan Rasch pada asesmen pendidikan: Implementasi penilaian formatif (assessment for learning)*. March, 1–19.  
[http://eprints.um.edu.my/15876/1/ITS\\_rasch\\_model\\_asesment\\_for\\_learning.pdf](http://eprints.um.edu.my/15876/1/ITS_rasch_model_asesment_for_learning.pdf)
- Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assessment for Learning*. October 2017. <https://doi.org/10.2991/icei-17.2018.11>
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial* (B. Trim (ed.); 1st ed.). Trim Komunikata Publishing House.
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch pada Assessment Pendidikan. In *Penerbit Trim Komunikata*.
- Suseno, I. (2017). Komparasi Karakteristik Butir Tes Pilihan Ganda Ditinjau Dari Teori Tes Klasik. *Jurnal Ilmiah Kependidikan*, 4(1), 1–8.
- Sutami. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) pada Mata Pelajaran Bahasa Indonesia SMA dan SMK. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 3(1), 102–113. <https://doi.org/10.30872/diglosia.v3i1.24>
- Turmuzi, M., Ratnaya, I. G., Al Idrus, S. W., Paraniti, A. A. I., & Nugraha, I. N. B. S. (2022). Literature Review: Evaluasi Keterlaksanaan Kurikulum 2013 Menggunakan Model Evaluasi CIPP (Context, Input, Process, dan Product). *Jurnal Basicedu*, 6(4), 7220–7232. <https://doi.org/10.31004/basicedu.v6i4.3428>
- Warsono, Nursuhud, P. I., Darma, R. S., & Supahar. (2020). Empirical analysis of diagrammatic representation test instruments using partial credit model in realizing learning outcomes. *International Journal of Instruction*, 13(3), 617–632. <https://doi.org/10.29333/iji.2020.13342a>
- Widhiarso, W. (2011). Aplikasi Teori Respon Butir Untuk Menguji Invariansi Pengukuran Psikologi Guna Keperluan Survei dan Seleksi Pekerjaan. *Psikobuana*, 3(2), 104–117.
- Yunida, H., & Arthur, R. (2023). Bloom ' s Taxonomy Approach to Cognitive Space Using Classic Test Theory and Modern Theory. *East Asian Journal of Multidisciplinary Research (EAJMR)*, 2(1), 95–108.