

ANALISIS INSTRUMEN ASSESMEN : VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL

Hera Apriliana Saputri¹, Zulhijrah², Nabila Joti Larasati³, Shaleh⁴

^{1,2,3,4} Pendidikan Guru Madrasah Ibtidaiyah, Fakultas Ilmu Tarbiyah Dan Keguruan
Universitas Islam Negeri Sunan Kalijaga, Yogyakarta

122204082019@student.uin-suka.ac.id, 222204082022@student.uin-suka.ac.id,

322204082028@student.uin-suka.ac.id, 4shaleh@uin-suka.ac.id

ABSTRACT

This study aims to analyze the assessment instrument, focusing on the validity, reliability, difficulty level, and differentiation of the items. The validity of the instrument was evaluated to measure the extent to which the instrument measured what it was supposed to measure. Reliability was tested to assess the level of consistency of the measurement results of the instrument. In addition, the level of difficulty of the items was analyzed to determine how difficult or easy the questions posed to the participants were. Meanwhile, item discrimination was assessed to determine the extent to which items were able to differentiate between participants who had different levels of ability. The research method involves collecting data from assessment instruments used in a particular context. The results of the data analysis illustrate the validity of the instrument, the reliability of the measurement, as well as the characteristics of the difficulty level and the differentiability of the items. The findings of this study provide deep insights into the quality of assessment instruments used in a particular context, with important implications for the improvement and development of future evaluation instruments. This research is expected to make a positive contribution to improving the quality of measurement in the context of assessment, and can be a reference for the development of evaluation instruments in various fields.

Keyword: Assessment, Differentiating Power, Reliability, Difficulty, Validity

ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis terhadap instrumen asesmen, dengan fokus pada validitas, reliabilitas, tingkat kesulitan, dan daya beda butir soal. Validitas instrumen dievaluasi untuk mengukur sejauh mana instrumen tersebut mengukur apa yang seharusnya diukur. Reliabilitas diuji guna menilai tingkat konsistensi hasil pengukuran dari instrumen tersebut. Selain itu, tingkat kesulitan butir soal dianalisis untuk mengetahui seberapa sulit atau mudah pertanyaan yang diajukan kepada peserta. Sementara itu, daya beda butir soal dinilai untuk menentukan sejauh mana butir soal mampu membedakan antara peserta yang memiliki tingkat kemampuan yang berbeda. Metode penelitian menggunakan library research, yaitu dengan mengumpulkan data-data kepustakaan yang berkaitan dengan asesmen, daya beda, reliabilitas, tingkat kesukaran dan validitas. Hasil analisis data menggambarkan validitas instrumen, reliabilitas pengukuran, serta karakteristik tingkat kesulitan dan daya beda butir soal. Temuan penelitian ini memberikan wawasan yang mendalam terkait kualitas instrumen asesmen yang digunakan dalam konteks tertentu, dengan implikasi penting bagi perbaikan dan pengembangan instrumen evaluasi di masa depan. Penelitian ini diharapkan dapat

memberikan kontribusi positif terhadap peningkatan kualitas pengukuran dalam konteks asesmen, serta dapat menjadi referensi bagi pengembangan instrumen evaluasi di berbagai bidang.

Kata Kunci: Asesmen, Daya Beda, Realibilitas, Tingkat Kesulitan, Validitas

A. Pendahuluan

Evaluasi dalam dunia pendidikan sangat penting karena digunakan untuk merumuskan kebijakan yang mendukung terciptanya sumber daya manusia yang kompetitif terhadap era globalisasi saat ini. Itulah sebabnya evaluasi/penilaian merupakan bagian yang tak terpisahkan dalam satu kesatuan proses di setiap lembaga pendidikan. Minimal terdapat dua tujuan diadakan evaluasi yakni untuk mengetahui/membuktikan sejauh mana pemahaman peserta evaluasi setelah berproses, dan hasil evaluasi dijadikan sebagai tolak ukur untuk memperbaiki proses (Son, 2019). Dalam kegiatan evaluasi diperlukan alat atau teknik penilaian, sehingga pelaksanaannya akan lebih terarah. Alat evaluasi dalam pendidikan yang digunakan untuk mengumpulkan data dapat berupa tes atau nontes. Berkaitan dengan pemahaman kita terhadap tes, Nurkencana dan Sunartana menyatakan tes adalah suatu cara untuk mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang harus dikerjakan oleh siswa atau

sekelompok siswa sehingga menghasilkan nilai tentang tingkah laku atau prestasi siswa sebagai peserta didik (Magdalena et al., 2020).

Tes ialah salah satu alat evaluasi atau penilaian yang dicoba untuk mengukur hasil belajar peserta didik sepanjang proses pembelajaran. Tes pada biasanya digunakan untuk mengukur sudah sejauh mana tingkat perkembangan ataupun kemajuan yang sudah dicapai setelah melewati proses belajar mengajar dalam jangka waktu tertentu. Tes ialah pelengkap dari penilaian ataupun evaluasi dari seorang guru untuk mengenali sudah sejauh mana tingkatan kemampuan dari siswa setelah menjajaki suatu pendidikan kompetensi dasar yang dilakukan pada akhir pembelajaran seperti ulangan harian. Saat sebelum di lakukannya tes hendaknya dicoba untuk melakukan analisis butir soal ataupun analisis kualitas tes terlebih dahulu. Secara umum soal yang digunakan sebagai alat penilaian juga belum diketahui kualitas soalnya yang terutama ditinjau dari Validitas, Reliabilitas dan Tingkat

Kesukarannya. Sehingga mutu dari soal yang diberikan serta yang sudah terbuat tidak bisa ditingkatkan. Soal yang bermutu dan yang sudah memiliki kualitas baik ialah soal yang bisa membagikan data mengenai pemahaman dari para peserta didik terhadap materi yang sudah disampaikan atau diajarkan oleh guru. Tidak hanya itu, pemahaman guru terhadap artinya analisis butir soal masih sangat rendah sehingga guru hanya melaksanakan analisis yang simpel berupa analisis validitas logis, dimana evaluasi yang dicoba oleh guru dilihat dari keahlian para siswa untuk menjawab ataupun menanggapi soal yang sudah diberikan (Prastika, 2021).

Dalam evaluasi pendidikan baik tes maupun nontes, keduanya merupakan instrumen atau alat bantu pengumpulan dan pengolahan data tentang variable-variabel yang diteliti. Ciri-ciri/karakteristik instrument yang baik sebagai alat evaluasi adalah memenuhi persyaratan validitas dan reliabilitas. Inilah alasan mengapa alat evaluasi yang baik dapat dilihat dari beberapa segi antara lain: (1) validitas, (2) reliabilitas, (3) objektivitas, (4) praktikabilitas, (5) daya pembeda, (6) taraf atau derajat

kesukaran, (7) efektivitas option, (8) efisiensi (Magdalena et al., 2021). Soal yang digunakan untuk mengevaluasi hasil belajar siswa ini hendaknya sesuai dengan standar yang telah ditentukan (Zein et al., 2013). Menurut Arikunto sebuah tes yang dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu memiliki validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis. Soal sebagai alat ukur hasil belajar siswa hendaknya sesuai dengan standar tersebut (Arikunto, 2008).

Kegiatan menganalisis butir soal (item analysis) adalah kegiatan yang harus dilakukan guru dalam rangka meningkatkan mutu soal yang telah disusun. Dan bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Dengan analisis soal dapat diperoleh informasi tentang kejelekan sebuah soal dan “petunjuk” untuk mengadakan perbaikan (Farida & Musyarofah, 2021).

Berdasarkan uraian di atas, maka perlu dilakukan analisis butir soal untuk mengetahui kualitas soal tersebut. Kualitas soal dapat dilihat dari hasil uji validitas, reliabilitas,

tingkat kesukaran dan daya beda. Penelitian ini bertujuan untuk mengetahui validitas reliabilitas tingkat kesukaran dan daya beda butir soal.

B. Metode Penelitian

Untuk memperoleh informasi terkait analisis instrumen assesmen : validitas, reliabilitas, tingkat kesulitan dan daya beda butir soal diperlukan beberapa literatur, buku-buku, artikel dan referensi lain yang berhubungan dengan kebutuhan penulisan artikel ini. Dengan demikian, metode penelitian yang digunakan adalah studi kepustakaan. Dalam riset Pustaka, penelusuran Pustaka adalah memanfaatkan sumber perpustakaan untuk memperoleh data penelitian tanpa memerlukan riset lapangan (Zed, 2014). Studi Pustaka adalah suatu karangan ilmiah yang berisi pendapat berbagai pakar mengenai suatu masalah, yang kemudian ditelaah dan dibandingkan dan ditarik kesimpulannya (Haryanto et al., 2000).

C. Hasil Penelitian dan Pembahasan

1. Validitas

Validitas adalah derajat yang menunjukkan suatu tes mengukur apa yang hendak diukur. Sebuah data

atau informasi dapat dikatakan valid apabila sesuai dengan keadaan sebenarnya (Magdalena et al., 2021). Sugiyono menyatakan bahwa uji validitas merupakan suatu langkah pengujian yang dilakukan terhadap isi (content) dari suatu instrumen, dengan tujuan untuk mengukur ketepatan instrumen yang digunakan dalam suatu penelitian (Azhar & Adri, 2008). Uji validitas merupakan uji yang berfungsi untuk melihat apakah suatu alat ukur tersebut valid (sahih) atau tidak valid.

Validitas tidak hanya ditujukan untuk mengukur ketepatan tes tetapi juga digunakan untuk mengukur instrumen penelitian. Dalam instrumen penelitian validitas harus mampu mengungkapkan data sesuai dengan masalah yang diungkapkan secara tepat dan benar seperti situasi dan kondisi yang sebenarnya (Komarudin & Sarkadi, 2017).

Contoh :

Hipotesis

$H_0 : \rho_{xy} \leq 0$ (Butir soal / Pertanyaan tidak valid)

$H_1 : \rho_{xy} > 0$ (Butir soal / Pertanyaan valid)

$\alpha = 5\%$

untuk mencari nilai r_{hitung} dari setiap item pertanyaan kuessioner digunakan rumus sebagai berikut :

$$r_k = \frac{n \sum X_i Y_i - (\sum X_i) (\sum Y_i)}{\sqrt{\{\sum X_i^2 - (\sum X_i)^2\} \{\sum Y_i^2 - (\sum Y_i)^2\}}}$$

2. Reliabilitas

Reliabilitas merujuk pada satu pengertian bahwa suatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik. Reliabilitas juga merujuk pada tingkat keterandalan sesuatu dan dapat dipercaya (Novikasari, 2016). Uji reliabilitas menurut Husaini, adalah proses pengukuran terhadap ketepatan (konsisten) dari suatu instrumen. Pengujian ini dimaksudkan untuk menjamin instrumen yang digunakan merupakan sebuah instrumen yang handal, konsistensi, stabil dan dependabilitas, sehingga bila digunakan berkali-kali akan menghasilkan data yang sama (Azhar & Adri, 2008). Reliabilitas merupakan koefisien yang menunjukk se jauh mana suatu instrumen/ alat pengukur dapat dipercaya, artinya apabila suatu

instrumen digunakan berulang-ulang untuk mengukur sesuatu yang sama, maka hasilnya relatif stabil atau konsisten. Secara empiris tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefesien reliabilitas, besarnya koefesien reliabilitas berkisar antara 0 sampai dengan 1, dimana semakin tinggi angka reliabilitas berarti semakin konsisten hasil pengukuran (Farida & Musyarofah, 2021).

Selanjutnya jika suatu instrumen diuji berulang-ulang tetap menunjukkan hasil yang sama, berarti instrumen tersebut reliabel. Artinya memiliki keajegan, stabil, konsisten. Hasil yang sama dari tes tidak berarti nilai yang sama pada setiap tes. Hasil tes pertama dengan hasil tes berikutnya dikorelasikan dan didapat hasil korelasi yang signifikan (Komarudin & Sarkadi, 2017).

3. Tingkat Kesukaran

Tingkat kesukaran adalah derajat kesukaran suatu butir soal yang dinyatakan dalam bentuk suatu bilangan (Lestari & Yudhanegara, 2017). Soal disebut baik apabila soal tersebut tidak termasuk dalam kategori terlalu mudah atau terlalu sukar. Soal terlalu sukar akan

membuat siswa menjadi putus asa dan patah semangat karena di luar kemampuannya, sedangkan soal terlalu mudah tidak dapat merangsang siswa untuk mempertinggi usaha memecahkannya (Arikunto, 2018). Tingkat kesukaran menyatakan sejauh mana soal tersebut mudah dan sulit bagi siswa. Semakin besar persentase siswa menjawab soal dengan benar maka semakin mudah soal, sebaliknya semakin kecil persentase siswa menjawab soal dengan benar maka semakin sulit soal (Setiyawan & Wijayanti, 2020).

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya. Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (difficulty index). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa

soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah. Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (difficulty index). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah (Komarudin & Sarkadi, 2017).

Di dalam istilah evaluasi, indeks kesukaran ini diberi simbol P (p besar), singkatan dari kata "proporsi". Dengan demikian maka soal dengan $P = 0,70$ lebih mudah, jika dibandingkan dengan $P = 0,20$. sebaliknya soal dengan $P = 0,30$ lebih sukar, daripada soal dengan $P = 0,80$.

Rumus mencari P adalah :

$$P = \frac{B}{JS}$$

Dimana :

P = indeks kesukaran.

B = banyaknya siswa yang menjawab soal itu dengan betul.

JS = jumlah seluruh siswa peserta tes

Kriteria Perhitungan Indeks kesukaran soal :

P = Kurang dari 0,25 adalah soal terlalu sukar

P = 0,25 – 0,75 adalah soal cukup (sedang)

P = lebih dari 0,75 adalah soal terlalu mudah

4. Daya Beda

Daya pembeda suatu butir soal adalah kemampuan butir soal dalam membedakan antara siswa yang dapat menjawab soal dengan tepat dan siswa yang menjawab soal kurang tepat (Lestari & Yudhanegara, 2017) . Daya pembeda suatu soal adalah kemampuan soal dalam hal membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah (Arikunto, 2018). Analisis daya beda memiliki makna mengkaji suatu butir soal dari segi kesanggupan tes tersebut dalam membedakan siswa yang memiliki kategori rendah dan kategori tinggi. Daya pembeda butir soal adalah kemampuan suatu butir tes untuk dapat membedakan antara testee yang berkemampuan tinggi dan berkemampuan rendah (Setiyawan & Wijayanti, 2020).

Menganalisis daya pembeda artinya mengkaji soal-soal tes dari segi keanggupan tes tersebut dalam membedakan siswa yang termasuk kedalam katagori lemah/rendah dan katagori kuat/tinggi presentasinya. Penentuan daya beda butir soal pada Anates dapat diketahui dalam tabel daya pembeda pada kolom DP persen. Daya pembeda butir soal memiliki manfaat yaitu untuk meningkatkan mutu setiap-butir soal melalui data empiriknya dan untuk mengetahui seberapa jauh masing-masing butir soal dapat membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan pendidik. Menurut Kusaeri Supranto, Adapun untuk menghitung daya beda digunakan rumus sebagai berikut (Magdalena et al., 2021) :

$$D = \frac{BA}{JA} - \frac{BB}{JB} = PA - PB$$

Ket :

D = Daya Beda

BA = Banyak peserta tes kelompok atas yang menjawab dengan benar

BB = Banyak peserta tes kelompok bawah yang menjawab dengan benar

JA = Jumlah peserta tes kelompok atas

JB = Jumlah peserta tes kelompok bawah

PA = Butir tes kelompok atas

PB = Butir tes kelompok bawah

D. Kesimpulan

Dalam melakukan analisis instrumen asesmen dengan fokus pada validitas, reliabilitas, tingkat kesulitan, dan daya beda butir soal, dapat disimpulkan bahwa kualitas suatu instrumen evaluasi sangat krusial dalam memastikan keakuratan dan keandalan hasil pengukuran. Validitas menjadi indikator sejauh mana instrumen tersebut dapat mengukur apa yang seharusnya diukur, sementara reliabilitas memberikan gambaran tentang konsistensi hasil pengukuran dari instrumen tersebut. Analisis tingkat kesulitan dan daya beda butir soal juga menjadi langkah penting dalam mengevaluasi instrumen asesmen. Tingkat kesulitan butir soal memberikan informasi mengenai seberapa sulit atau mudah pertanyaan yang diajukan kepada peserta, sedangkan daya beda butir soal membantu mengidentifikasi

kemampuan instrumen dalam membedakan antara peserta yang memiliki tingkat keterampilan yang berbeda. Melalui pemahaman mendalam terhadap karakteristik-karakteristik tersebut, penelitian ini memberikan kontribusi berharga dalam pengembangan instrumen asesmen yang lebih valid, reliabel, dan efektif. Implikasi temuan penelitian ini dapat digunakan sebagai dasar untuk perbaikan dan peningkatan instrumen evaluasi di berbagai konteks, baik dalam lingkup pendidikan, industri, maupun penelitian ilmiah. Dengan demikian, upaya terus-menerus untuk meningkatkan kualitas instrumen asesmen menjadi esensial guna mendukung pengambilan keputusan yang akurat dan relevan.

DAFTAR PUSTAKA

- Arikunto, S. (2008). *Prosedur Penelitian Suatu Pendekatan*. Rineka Cipta.
- Arikunto, S. (2018). *Dasar-dasar evaluasi Pendidikan*. Rineka Cipta.
- Azhar, N., & Adri, M. (2008). *Uji Validitas dan Reliabilitas Paket Multimedia Interaktif*. 1–15.

-
- Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, 1(1), 34–44.
- Haryanto, Ruslijanto, H., & Mulyono, D. (2000). *Metode Penulisan dan Penyajian Karya Ilmiah: Buku Ajar Untuk Mahasiswa*. Penerbit Buku Kedokteran.
- Komarudin, & Sarkadi. (2017). *Evaluasi pembelajaran*. Laboratorium Sosial Politik Press.
- Lestari, K. E., & Yudhanegara, M. R. (2017). *Penelitian Pendidikan Matematika*. Refika Aditama.
- Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya Evaluasi Dalam Pembelajaran Dan Akibat Memanipulasinya. *Jurnal Pendidikan Dan Sains*, 2(2), 244–257.
<https://ejournal.stitpn.ac.id/index.php/bintang>
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG: Jurnal Pendidikan Dan Sains*, 3(2), 198–214.
<https://ejournal.stitpn.ac.id/index.php/bintang>
- Novikasari, I. (2016). *Uji Validitas Instrumen*. Institut Agama Islam Negeri Purwokerto.
- Prastika, Y. D. (2021). Pengaruh Validitas, Reliabilitas dan Tingkat Kesukaran Terhadap Kualitas Butir Soal Ekonomi Menggunakan Software Anates di SMKN 3 Bangkalan. *STKIP PGRI Bangkalan*, 1–11.
- Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistik*, 1(2), 130–139.
- Son, A. L. (2019). Instrumentasi kemampuan pemecahan masalah matematis: analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema Wiralodra*, 10(1), 41–52.
-

Zed, M. (2014). *Metode Penelitian*
Kepustakaan. Yayasan Pustaka
Obor Indonesia.

Zein, A., Fadillah, M., & Novianti, R.
(2013). Hubungan Antara
Validitas Butir , Reliabilitas ,
Tingkat Kesukaran Dan Daya
Pembeda Soal Ujian Semester
Genap Bidang Studi Biologi Kelas
Xi Sma / Ma Negeri Di Kota
Padang Tahun Pelajaran 2010 /
2011 *. *Semirata 2013 FMIPA*
Unila, 2009, 39–47.
<https://jurnal.fmipa.unila.ac.id/semirata/article/download/572/392>