



Hendrayadi¹
 Martin Kustati²
 Rezki Amelia³

ANALISIS ULANGAN HARIAN MATA PELAJARAN PAI DI SMA NEGERI 10 PADANG TAHUN PELAJARAN 2023/2024 (TELAAH TERHADAP RELIABILITAS, DAYA BEDA DAN TINGKAT KESUKARAN MENGGUNAKAN SOFTWARE ANATES)

Abstrak

Penelitian ini akan dibahas mengenai penelitian deskriptif kuantitatif yang bertujuan untuk mengetahui kualitas dan kelayakan butir soal penilaian tengah semester di SMAN 10 Padang. Kualitas dan kelayakan butir soal dapat dilihat dari validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kualitas distraktor. Subjek penelitian adalah siswa SMAN 10 Padang. Data dikumpulkan melalui penyebaran soal Penilaian Tengah Semester yang diisi oleh siswa, kemudian diolah dengan menggunakan program Anates ver. 4.0.9. Hasil uji coba menunjukkan bahwa butir soal sudah cukup layak untuk diujikan namun masih perlu perbaikan. Hal ini dikarenakan masih terdapat beberapa soal dengan kualitas yang kurang baik.

Kata Kunci: Realibilitas, Tingkat Kesukaran, Anates

Abstract

This research will discuss quantitative descriptive research that aims to determine the quality and feasibility of midterm assessment items at SMAN 10 Padang. The quality and feasibility of items can be seen from the validity, reliability, difficulty level, differentiating power, and quality of distractors. The research subjects were students of SMAN 10 Padang. The data were collected through the distribution of Midterm Assessment questions filled in by students, then processed using Anates ver. 4.0.9. The results of the trial showed that the items were suitable enough to be tested but still needed improvement. This is because there are still some questions with poor quality.

Keywords: Reliability, Level of Difficulty, Anates

PENDAHULUAN

Keberhasilan sebuah pendidikan tergantung pada tata kelola organisasi yang baik. Keberlangsungan suatu proses pembelajaran yang berkualitas sangat menentukan kualitas generasi yang dihasilkan. Terlebih lagi pendidikan abad XXI di era digital saat ini menuntut seluruh komponen terkait bersinergi guna mewujudkan pendidikan yang mampu bersaing dengan dunia global. Pendidik sebagai ujung tombak dalam mencetak generasi berkualitas harus mampu menguasai kompetensi terkait. Dalam proses pembelajaran pendidik tidak hanya mempersiapkan materi atau konten pembelajaran melainkan harus menyiapkan berbagai program pembelajaran seperti perangkat pembelajaran. Ada empat unsur dalam kegiatan belajar mengajar, yakni tujuan-bahan-metode dan alat serta penilaian. Tujuan sebagai arah dari proses belajar mengajar pada hakikatnya adalah rumusan tingkah laku yang diharapkan dapat dikuasai oleh siswa setelah menerima atau menempuh pengalaman belajarnya. Bahan adalah seperangkat pengetahuan ilmiah yang dijabarkan dari kurikulum untuk disampaikan atau dibahas dalam proses belajar mengajar agar sampai kepada tujuan yang telah ditetapkan. Metode dan alat adalah cara atau tehnik yang digunakan untuk mencapai tujuan. Sedangkan penilaian adalah

^{1,2,3}Progam Pascasarjana (S.3), Pendidikan Islam, UIN Imam Bonjol Padang
 email: hendrayadimpdi@gmail.com

upaya atau tindakan untuk mengetahui sejauh mana tujuan yang telah sebagai alat untuk mengetahui keberhasilan proses dan hasil belajar siswa (Nana Sudjana, 1995).

Pada umumnya hasil belajar dapat dikelompokkan menjadi tiga aspek yaitu ramah kognitif, psikomotorik dan afektif. Secara eksplisit ketiga aspek tersebut dipisahkan satu sama lain. Apapun jenis mata pelajarannya selalu mengandung tiga aspek tersebut namun memiliki penekanan yang berbeda. Perangkat pembelajaran seperti rencana pelaksanaan pembelajaran (RPP) menjadi rambu-rambu pendidik dalam melaksanakan proses pembelajaran sampai dengan proses penilaian peserta didik di kelas. Komponen RPP sendiri telah mengalami berbagai perubahan, terakhir menjadi RPP satu lembar sesuai surat edaran Kemendikbud Nomor 14 Tahun 2019. Komponen inti dalam RPP terbaru yakni RPP satu lembar minimal memuat identitas sekolah, tujuan pembelajaran, langkah-langkah kegiatan pembelajaran, dan evaluasi atau penilaian. Terkait dengan penilaian, seorang pendidik harus mengedepankan asas transparansi, objektivitas, validitas, dan akuntabilitas guna memberikan rasa keadilan kepada peserta didik. Untuk mengukur kecerdasan atau kecakapan peserta didik tidak hanya dilakukan dengan menilai sikap dan keterampilan melainkan dengan melakukan penilaian terhadap kognitif atau pengetahuan peserta didik. Penilaian berusaha menentukan apakah tujuan pendidikan tercapai atau tidak, sehingga evaluasi memiliki peran yang sangat besar dunia pendidikan. Evaluasi yang efektif dan efisien dilakukan secara kontinu serta menyeluruh diharapkan para pendidik dapat memperoleh gambaran secara utuh tentang prestasi dan kemajuan proses serta hasil belajar yang dicapai oleh peserta didik.

Setiap mata pelajaran menekankan evaluasi yang berkesinambungan, mengacu pada tujuan dan hasil yang komperhensif. Adapaun diantara ranah penilain dalam pendidikan adalah ranah kognitif. Penilaian kognitif merupakan kegiatan yang dilakukan untuk mengukur penguasaan pengetahuan peserta didik. Penilaian pengetahuan dapat dilakukan dengan berbagai teknik. Dalam hal ini seorang pendidik dapat memilih teknik penilaian yang sesuai dengan karakteristik kompetensi dasar, indikator, atau tujuan pembelajaran yang akan dinilai pada mata pelajaran tertentu. Pendekatan penilaian yang digunakan dalam penilaian kompetensi pengetahuan adalah penilaian autentik, yaitu penilaian yang dilakukan dengan mengukur kemampuan peserta didik dalam menerapkan pengetahuan yang telah dipelajari dalam situasi yang nyata.

METODE

Desain penelitian yang digunakan adalah deskriptif. Penelitian ini dilakukan dengan cara mengumpulkan hasil lembar jawab siswa, kunci jawaban soal ujian dan soal ujian semester mata pelajaran pendidikan agama Islam (PAI), setelah semuanya terkumpul selanjutnya adalah menganalisis butir soal tersebut. Sedangkan untuk pendekatan yang dipakai adalah pendekatan kuantitatif. Dimana data angka dari pengumpulan lembar jawab dimasukan kedalam program ANATES untuk dianalisis.

HASIL DAN PEMBAHASAN

Reliabilitas (Reliability)

Reliabilitas diterjemahkan dari kata reliability. Pengukuran yang memiliki reliabilitas tinggi maksudnya adalah pengukuran yang dapat menghasilkan data yang reliabel. Reliabilitas dikenal dengan beberapa nama lain seperti keterpercayaan, keterandalan, keajegan, konsistensi, kestabilan, dan sebagainya. Ide pokok dalam konsep reliabilitas adalah sejauhmana hasil suatu pengukuran dapat dipercaya. Hasil ukur dapat dipercaya apabila dalam beberapa kali pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, kalau aspek yang diukur dalam diri subjek memang belum berubah. Nama lain dari reliabilitas adalah ketetapan, kepercayaan, kemantapan, keterandalan, keajegan atau kestabilan. Menurut Anne Anastasi dan Susan Urbina (2007: 63), reliabilitas tes adalah kestabilan skor yang diperoleh dari orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda. Suatu tes dikatakan reliabel jika hasil tes sama pada seseorang walaupun telah diadakan pengesanan ulang pada waktu yang berbeda. T. Raka Joni (2006: 36) mengatakan bahwa suatu tes dikatakan reliabel sebuah tes adalah ketetapan hasil tes. Sebuah reliabilitas tes memberi nilai yang sama kepada individu yang sama pada setiap hasil tes.

Berdasarkan teori-teori yang telah dikemukakan di atas, dapat disimpulkan bahwa reliabilitas suatu tes adalah keajegan suatu tes. Tes dikatakan reliabel apabila menunjukkan hasil yang sama walaupun dilakukan tes berulang ulang. Reliabilitas dinyatakan dengan angka dan

dikenal sebagai koefisiensi reliabilitas. Makin tinggi koefisien reliabilitas sebuah tes makin tinggi pula reliabilitas tes tersebut. Pengertian relatif menunjukkan bahwa ada toleransi terhadap perbedaan-perbedaan kecil diantara hasil pengukuran. Bila perbedaan itu besar dari waktu ke waktu, maka hasil pengukuran itu tidak dapat dipercaya atau tidak reliabel. Reliabilitas tes dapat diestimasi melalui salah satu pendekatan umum, yaitu: pendekatan tes ulang, pendekatan tes seajar dan pendekatan konsistensi internal (Azwar, 2007: 180).

Sedangkan indeks daya pembeda soal adalah perbedaan persentase dari 27 % siswa yang mendapat nilai tinggi (kelompok atas) dan 27 % siswa yang mendapat nilai rendah (kelompok bawah). Soal yang mempunyai indeks daya pembeda antara 0,15 sampai 0,20 atau lebih tinggi menunjukkan daya pembeda yang baik. Berdasarkan pendapat-pendapat yang telah dikemukakan di atas, daya pembeda soal adalah kemampuan soal untuk membedakan kelompok peserta tes berkemampuan tinggi dan kelompok peserta tes yang berkemampuan rendah. Nilai daya pembeda dinyatakan melalui indeks daya pembeda. Makin tinggi atau makin besar indeks daya pembeda soal, makin besar soal tersebut dapat membedakan antara kelompok tinggi dan kelompok rendah. Untuk menghitung daya pembeda soal mengambil dari 27 % kelompok tinggi dan 27 % dari kelompok rendah. Nilai indeks daya pembeda soal berkisar antara -1 sampai +1.

Menurut Gronlund, ada empat faktor yang dapat mempengaruhi reliabilitas, yaitu :

1. Tingkat kesukaran, dalam penilaian yang menggunakan pendekatan penilaian acuan norma, baik untuk soal yang mudah maupun sukar, cenderung menghasilkan tingkat reliabilitas yang rendah. Hal ini disebabkan antara hasil tes yang mudah dengan hasil tes yang sukar keduanya dalam satu sebaran skor yang terbatas. Untuk tes yang mudah, skor akan berada dibagian atas dan akhir dari skala penilaian. Bagi kedua tes (mudah dan sukar), perbedaan antar peserta didik kecil sekali dan cenderung tidak dapat dipercaya. Tingkat kesukaran soal yang ideal untuk meningkatkan koefisien reliabilitas adalah soal yang menghasilkan sebaran skor berbentuk genta atau kurva normal.
2. Panjang tes, yaitu banyaknya soal tes. Ada kecenderungan, semakin panjang suatu tes akan lebih tinggi tingkat reliabilitas suatu tes, karena semakin banyak soal, maka akan semakin banyak sampel yang diukur dan proporsi jawaban yang benar semakin banyak, sehingga faktor tebakan akan semakin rendah.
3. Sebaran skor, besarnya sebaran skor akan membuat tingkat reliabilitas menjadi lebih tinggi, Karena koefisien reliabilitas yang lebih besar diperoleh ketika peserta didik tetap pada posisi yang relative sama dalam satu kelompok pengujian ke pengujian berikutnya. Dengan kata lain, peluang selisih dari perubahan posisi dalam kelompok dapat memperbesar koefisien reliabilitas.
4. Objektivitas, menunjukkan skor tes kemampuan yang sama antara peserta didik yang satu dengan peserta didik lainnya. Peserta didik memperoleh hasil yang sama dalam mengerjakan suatu tes. Jika peserta didik memiliki tingkat kemampuan yang sama, maka akan memperoleh hasil tes yang sama pada saat mengerjakan tes yang sama. Objektivitas prosedur tes yang tinggi akan memperoleh reliabilitas hasil tes yang tidak dipengaruhi oleh prosedur penskoran.

Konsep reliabilitas mendasari kesalahan pengukuran yang mungkin terjadi pada suatu proses pengukuran atau pada nilai tunggal tertentu, sehingga menimbulkan perubahan pada susunan kelompoknya. Misalnya, guru mengetes peserta didik dengan instrumen tertentu dan mendapat nilai 70. Kemudian pada kesempatan yang berbeda dengan instrumen yang sama, guru melakukan tes kembali, ternyata peserta didik tersebut mendapat nilai 75. Artinya, tes tersebut tidak reliabel, karena terjadi kesalahan pengukuran. Tes yang reliabel adalah apabila koefisien reliabilitasnya tinggi dan kesalahan baku pengukurannya rendah.

Reliabilitas berkaitan dengan sebuah kepercayaan, bilat es tersebut memiliki kepercayaan tinggi maka tes dapat menunjukkan hasil yang tetap (Huriaty & Gazali, 2021). Kategori dalam mengolah hasil analisis soal menggunakan software Anates adalah sebagai berikut:

Tabel 1: Nilai Koefisien Reliabilitas

No	Nilai Koefisien	Reliabilitas Soal
1	0,000 sampai 0,199	Kategori Sangat Rendah
2	0,200 sampai 0,399	Kategori Rendah

3	0,400 sampai 0,599	Kategori Cukup
4	0,600 sampai 0,799	Kategori Tinggi
5	0,800 sampai 1,00	Kategori Sangat Tinggi

Tes yang memiliki konsistensi reliabilitas tinggi adalah akurat, reproducibel, dan generalized terhadap kesempatan testing dan instrumen tes lainnya. Secara rinci faktor yang mempengaruhi reliabilitas skor tes di antaranya:

- a. Semakin banyak jumlah butir soal, semakin ajek suatu tes.
- b. Semakin lama waktu tes, semakin ajek.
- c. Semakin sempit range kesukaran butir soal, semakin besar keajegan.
- d. Soal-soal yang saling berhubungan akan mengurangi keajegan.
- e. Semakin objektif pemberian skor, semakin besar keajegan. Ketidaktepatan pemberian skor. Menjawab besar soal dengan cara menebak.
- f. Semakin homogen materi semakin besar keajegan.
- g. Pengalaman peserta ujian dan Salah penafsiran terhadap butir soal.
- h. Menjawab soal dengan buru-buru/cepat.
- i. Kesiapan mental peserta ujian.
- j. Adanya gangguan dalam pelaksanaan tes.
- k. Jarak antara tes pertama dengan tes kedua.
- l. Mencontek dalam mengerjakan tes.
- m. Posisi individu dalam belajar dan Kondisi fisik peserta ujian.

Ada 3 cara yang dapat dilakukan untuk menentukan reliabilitas skor tes, yaitu:

- a. Keajegan pengukuran ulang: kesesuaian antara hasil pengukuran pertama dan kedua dari sesuatu alat ukur terhadap kelompok yang sama.
- b. Keajegan pengukuran setara: kesesuaian hasil pengukuran dan 2 atau lebih alat ukur berdasarkan kompetensi kisi-kisi yang lama.
- c. Keajegan belah dua: kesesuaian antara hasil pengukuran belahan pertama dan belahan kedua dari alat ukur yang sama.

Daya Beda (Discrimination Power)

Dali S. Naga (2002: 67) mengatakan bahwa daya pembeda soal adalah kemampuan soal dengan skornya dapat membedakan peserta tes dari kelompok tinggi dan kelompok rendah. Dengan kata lain makin tinggi daya pembeda soal makin banyak peserta dari kelompok tinggi yang dapat menjawab soal dengan benar dan makin sedikit peserta tes dari kelompok rendah yang dapat menjawab soal dengan benar. Agar dapat diterima maka nilai D (discrimination: daya pembeda soal) adalah 0,30 atau lebih. Sedangkan untuk dapat dinyatakan cukup memuaskan adalah 0,40 ke atas. Pengertian daya pembeda (DP) dari sebuah butir soal adalah menyatakan seberapa jauh kemampuan butir soal tersebut mampu membedakan antara testee yang mengetahui jawabannya dengan benar dengan testee yang tidak dapat menjawab soal tersebut (testee yang menjawab salah). Dengan kata lain daya pembeda butir soal adalah kemampuan butir soal itu untuk membedakan antara testee yang pandai atau berkemampuan tinggi dengan testee yang berkemampuan rendah. Menurut Charles D. Hopkins dan Richard L. Antes (1999:159), indeks daya pembeda soal mengukur bagaimana baiknya sebuah soal membedakan tingkat kemampuan siswa. Indeks daya pembeda soal bernilai - 1,00 sampai +1,00. Makin tinggi nilai D sebuah soal, makin baik soal tersebut membedakan siswa yang berkemampuan tinggi dari siswa yang berkemampuan rendah.

Soal yang mempunyai nilai $D = 0,40$ dianggap sangat efektif dalam membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah. Soal yang mempunyai nilai D antara 0,20 sampai 0,39 dianggap memuaskan. Soal yang mempunyai nilai D lebih rendah dari 0,39 harus diperbaiki. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi atau ditolak. Kedua, untuk mengetahui seberapa jauh masing-masing soal dapat mendeteksi atau membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru. Apabila suatu soal tidak dapat membedakan kedua kemampuan siswa itu maka butir soal itu dapat dicurigai kemungkinannya: a) Kunci jawaban butir soal itu tidak tepat. b) Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar. c) Kompetensi yang diukur tidak jelas. d) Pengecoh tidak berfungsi. e) Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak dan f) Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya.

Derajat daya pembeda (DP) suatu butir soal dinyatakan dengan indeks diskriminan yang bernilai -1,00 sampai dengan 1,00. Apabila indeks diskriminasi soal makin mendekati nilai 1,00 ini berarti daya pembeda soal tersebut akan semakin baik, begitu juga sebaliknya, jika indeks diskriminasi suatu soal mendekati nilai 0,00 maka daya pembeda soal tersebut sangat jelek.

Indeks diskriminasi butir soal bernilai negatif (antara 0,00 sampai -1,00) ini berarti kelompok testee kurang mampu banyak yang menjawab benar, sebaliknya banyak testee yang pintar menjawab salah. Sedangkan jika suatu butir soal memiliki indeks diskriminasi 0,00 berarti bahwa soal tersebut tidak memiliki daya pembeda, artinya baik peserta didik pandai maupun yang kurang mampu menjawab benar soal tersebut. Berikut ini rumus untuk menentukan daya pembeda suatu soal.

Soal dinyatakan memiliki daya beda bila dapat membedakan kelompok unggul dan kelompok asor. Beberapa kategori klasifikasi daya pembeda ada pada tabel berikut:

Tabel 2. Kategori Daya Pembeda

No	Besarnya Daya Beda	Keterangan
1	0,00 - 0,20	Jelek (poor) yaitu tidak adanya daya pembeda yang bernilai Baik
2	0,20 - 0,40	Cukup (Satisfactory) yaitu daya pembeda cukup
3	0,40 - 0,70	Baik (Good) yaitu daya pembeda baik
4	0,70 - 1,00	Baik sekali (Excellent) yaitu adanya daya pembeda yang baik sekal

Sebaliknya, Bila terdapat koefisien daya beda bernilai negatif, maka soal dikategorikan tidak baik sebaiknya tidak digunakan (Anggreini & Darmawan, 2017). Daya pembeda jelek disebabkan indeks kesulitan yang terlalu tinggi atau terlalu rendah, demikian juga bila soal terlalu sulit atau mudah, dan adanya faktor pengecoh (Darus et al.,2021).

Tingkat Kesukaran (Difficulty Level)

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 - 1,00 (Aiken 1994: 66). Semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki TK= 0,00 artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki TK= 1,00 artinya bahwa siswa menjawab benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal. Pada prinsipnya, skor rata-rata yang diperoleh peserta didik pada butir soal yang bersangkutan dinamakan tingkat kesukaran butir soal itu. Rumus ini dipergunakan untuk soal selected response item, yaitu (Nitko, 1996: 310). Saifudin Azwar (2006:129) mengatakan bahwa tingkat kesukaran butir soal adalah proporsi antara banyaknya peserta tes yang menjawab butir soal dengan benar dengan banyaknya peserta tes. Hal ini berarti makin banyak peserta tes yang menjawab butir soal dengan benar maka makin besar indeks tingkat kesukaran, yang berarti makin mudah butir soal itu.

Sebaliknya makin sedikit peserta tes yang menjawab butir soal dengan benar maka soal tersebut makin sukar. Sedangkan menurut HJX. Fernades (2004:60), tingkat kesukaran butir soal diukur oleh persentase siswa yang menjawab soal dengan benar. Jika soal mudah maka indeks kesukaran lebih tinggi. Soal dengan nilai p yang mendekati 0 adalah soal yang sangat sulit, sedangkan soal dengan nilai p mendekati 1 adalah soal yang sangat mudah. Indeks tingkat kesukaran yang sangat baik adalah 0,3 sampai 0,7. Bistok Sirait (2009:301) berpendapat bahwa tinggi rendahnya tingkat kesukaran suatu butir soal dapat disebabkan oleh kerumitan (kompleksitas) pokok soal dan kondisi pilihan jawaban yang disediakan. Dalam arti keluasan pertanyaan soal sering membingungkan peserta tes dan alternative jawaban yang homogen atau kalimat soal sulit dipahami atau mempunyai pengertian ganda bagi peserta tes. Menurut Anastasi dan Susan Urbina (2007:128), tingkat kesukaran soal berkaitan dengan persentase peserta yang menjawab soal dengan benar. Semakin mudah butir soal, makin besarlah persentasenya. Jika tingkat kesukaran 70 % (p=0,70), soal tersebut dianggap lebih mudah dibandingkan jika tingkat kesukaran soalnya 15 % (p=0,15).yang paling baik adalah soal yang mempunyai tingkat kesukaran 0,50. Wayan Nurkancana (2002:161) mengemukakan tentang

klasifikasi tingkat kesukaran soal yaitu jika nilai p (prosentase tingkat kesukaran soal) = 0,81 – 1,00 butir soal mudah sekali, jika nilai p= 0,61- 0,80 butir soal mudah, jika nilai p= 0,41 – 0,60 butir soal sedang, jika nilai p= 0,21 – 0,40 butir soal sukar, jika nilai p= 0,00 – 0,20 butir soal sukar sekali. Nana Sujana (2002:135) mengatakan bahwa terdapat tingkat kesukaran mudah, sedang dan sukar.

Tingkat kesukaran yang baik adalah 0,25 sampai 0,75. Soal yang mempunyai tingkat kesukaran di bawah 0,25 berarti soal itu sukar, sebaliknya soal yang mempunyai tingkat kesukaran di atas 0,75 adalah soal itu terlalu mudah Berdasarkan teori-teori yang telah dikemukakan, tingkat kesulitan soal adalah seberapa mudah dan seberapa sulitnya suatu soal bagi siswa. Tingkat kesukaran dinyatakan dengan persentase siswa yang menjawab soal dengan benar. Makin besar persentase siswa yang menjawab soal dengan benar, makin mudah soal itu. Sebaliknya makin kecil persentase siswa yang menjawab soal dengan benar, makin sukar soal itu. Dengan demikian soal yang dijawab benar oleh 85 % siswa, dinyatakan mempunyai tingkat kesukaran 0,85, tentunya soal ini lebih mudah dari soal yang mempunyai tingkat kesukaran 20 % (0,20). Soal yang baik adalah soal dengan taraf kesukaran tidak sukar dan tidak mudah. Soal yang telah dianalisis dapat dinyatakan dalam klasifikasi sukar, sedang atau mudah dengan menyesuaikan kriteria tertentu. Kriteria untuk mudah yaitu bila hasil analisis diperoleh dalam rentang 0,71-1,00 sedang berada pada rentang 0,31-0,7, dan soal dinyatakan sukar ditunjukkan dalam rentang 0-0,30 (Anggreini & Darmawan, 2017). Tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes.

Tabel 3: Range Tingkat Kesukaran

No	Range Tingkat Kesukaran	Kategori	Keterangan
1	0,7-1,0	Mudah	Ditolak/direvisi
2	0,3-0,7	Sedang	Diterima
3	0,0-0,3	Sulit	Ditolak/direvisi

Tingkat Kesukaran soal ditentukan dengan rumus sebagai berikut:

$$\text{Proporsi Indeks Kesukaran} = \frac{\text{Banyaknya peserta yang menjawab Benar}}{\text{Jumlah peserta yang mengikuti Tes}}$$

Atau dengan Rumus berikut:

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

P = Proporsi (Indeks Kesukaran)

B = Jumlah siswa yang menjawab benar

N = Jumlah peserta tes

Contohnya: Dari 50 Orang peserta yang mengikuti tes dan menjawab soal nomor 5, 40 orang diantaranya menjawab benar, dan 10 orang menjawab salah. Maka akan diperoleh hasil sebagai berikut:

$$P = \frac{40}{50} = 0,80$$

Tingkat kesukaran butir soal memiliki 2 kegunaan, yaitu kegunaan bagi guru dan kegunaan bagi pengujian dan pengajaran (Nitko, 1996: 310-313). Kegunaannya bagi guru adalah: (1) sebagai pengenalan konsep terhadap pembelajaran ulang dan memberi masukan kepada siswa tentang hasil belajar mereka, (2) memperoleh informasi tentang penekanan kurikulum atau mencurigai terhadap butir soal yang bias. Adapun kegunaannya bagi pengujian dan pengajaran adalah:

- a. Pengenalan konsep yang diperlukan untuk diajarkan ulang
- b. Tanda-tanda terhadap kelebihan dan kelemahan pada kurikulum sekolah,
- c. Memberi masukan kepada siswa
- d. Tanda-tanda kemungkinan adanya butir soal yang bias
- e. Merakit tes yang memiliki ketepatan data soal.

Analisis Tes Menggunakan Software Anates

Anates merupakan salah satu software dalam pemrograman komputer yang bertujuan untuk menganalisis soal. Program komputer khususnya berguna untuk guru dan pemerhati evaluasi pendidikan (P. I. Sari & Yudha, 2020). Software Anates Versi 4 dengan basis operating sistem Windows pilihan ganda dan uraian yaitu sebuah aplikasi untuk menganalisis soal pilihan ganda dan uraian. Anates dikembangkan oleh dosen Psikologi UPI yaitu Bapak Drs. Karno To,

M.Pd dan seorang konsultan komputer yaitu Bapak Yudi Wibisono, S.T (P. I. Sari & Yudha, 2020; A. I. C. Sari & Herawati, 2014). Anates versi 4 mampu dalam menganalisis soal tes misalnya : penghitungan skor (asli maupun dibobot), penghitungan reliabilitas tes, pengelompokan subyek kelompok atas atau bawah, penghitungan daya pembeda, penghitungan tingkat kesukaran soal, penghitungan korelasi skor total dan skor butir, penentuan disktraktor atau pengecoh(A. I. C. Sari & Herawati, 2014).

Menganalisis validitas dilakukan dengan tujuan memperoleh keakuratan tes melalui Anates V4 dengan korelasi point biserial (rpbi). Reliabilitas adalah untuk mengetahui keajegan tes terhadap yang akan dinilai dan menggunakan anates metode Spearman Brown.. Analisis tingkat kesukaran digunakan dalam mengetahui letak kesukaran soal. Soal yang terlalu mudah kurang dapat merangsang siswa dalam memecahkannya, sedangkan soal yang sangat sukar dapat menurunkan semangat siswa untuk mencoba. Sedangkan daya pembeda dianalisis dengan tujuan mengetahui perbedaan siswa yang sudah paham kompetensi dan yang belum (Kurniawan, 2015). Manfaat yang bisa diperoleh dalam memvalidasi butir soal adalah 1) memberikan bantuan kepada pengguna tes untuk memperoleh pengetahuan kualitas tes yang digunakan, 2) menentukan soal yang efektif, 3) soal dapat disesuaikan dengan materi yang sedang dibahas 4) meningkatkan validitas dan reliabilitas, 5) menambah pengetahuan keterampilan dalam pembuatan soal(Ariany & Al-Ghifari, 2018).

Output software Anates adalah berupa validitas, reliabilitas, tingkat kesukaran, daya beda soal, dan pengecoh soal. Fitur dalam software Anates juga mempermudah pengguna bila ingin menambah soal atau subyek baru(Ariany & Al-Ghifari, 2018). Level kesukaran soal adalah mengukur level soal yang diuji ada pada level mudah atau sukar sehingga memudahkan dalam menentukan proporsi soal di kategori soal mudah, sedang, dan sukar (A. I. C. Sari & Herawati, 2014). Daya pembeda yaitu suatu kemampuan soal dalam menentukan perbedaan siswa unggul dan siswa asor. Indeks diskriminasi ini ada pada kisaran 0,00 hingga 1,00. Indeks dalam diskriminasi tidak mendeteksi bilangan negative. Sedangkan untuk pengecoh atau distraktor dapat menjalankan fungsinya sebagai pengecoh dengan baik bila pengecoh atau distractor itu minimal dipilih 5% dari seluruh peserta tes. Reliabilitas biasa disebut keajegan atau derajat konsistensi (Wardany et al., 2015).

UCAPAN TERIMA KASIH

Tim peneliti mengucapkan terima kasih kepada UIN Imam Bonjol Padang serta SMAN 10 Padang, yang telah memberi dukungan terhadap terlaksananya penelitian ini. Tim peneliti juga mengucapkan terima kasih kepada Pengelola Jurnal Review Pendidikan dan Pengejaraan (JRPP) Universitas Pahlawan Tuanku Tambusai yang telah memberi review dan masukan atas terbitnya artikel ini.

SIMPULAN

Kegiatan dalam pembelajaran dapat dievaluasi dengan tes. Hasil tes siswa merupakan gambaran informasi yang dapat diberikan untuk mengetahui kemajuan siswa. Untuk dapat memberikan tes atau soal yang baik, maka diperlukan analisis hasil tes. Analisis yang dapat digunakan oleh guru sehingga lebih mudah dalam memperoleh hasil adalah menggunakan software Anates generasi. Aplikasi yang menggunakan input hasil ulangan harian peserta didik yaitu soal berbentuk pilihan ganda untuk kemudian diolah dan menghasilkan data berupa reliabilitas, adanya kelompok unggul(atas) dan adanya kelompok asor (bawah), adanya daya pembeda, tingkat atau level kesukaran, kualitas pengecoh soal, dan korelasi atau hubungan bobot skor dengan skor total. Reliabilitas soal menunjukkan angka 0,32 yang berarti tidak layak sehingga harus dibuang atau dilakukan revisi sehingga kualitas memiliki reliabilitas yang baik. Adanya pembagian kelompok untuk menentukan reliabilitas soal ini didasarkan pada pembagian kelomk ganjil dan genap. Daya Pembeda dalam kegiatan ulangan harian umum semester ganjil dengan 50 butir soal yang diujikan masih dalam kategori sangat rendah, demikian juga tingkat daya beda juga berada pada posisi sangat rendah sebesar 86 persen (. Sedangkan tingkat kesukaran soalnya dari 50 butir yang diujikan, 86 persen dengan kategori sangat rendah dan hanya terdapat14 persen dari 50 butir soal yang tidak dengan indeks 0,00.

DAFTAR PUSTAKA

Aiken, Lewis R. (1994). *Psychological Testing and Assessment*,(Eight Edition), Boston: Allyn and Bacon.

- Anastasi, Anne dan Susan Urbina. Tes Psikologi. Jakarta: Prenhallindo, 2007.
- Anggreini, D., & Darmawan, C. A. (2017). Analisis Kualitas Soal Try Out Ujian Nasional Dengan Menggunakan Aplikasi Program Anates. JP2M (Jurnal Pendidikan Dan Pembelajaran Matematika), 2(1), 20.
- Azwar, Saifuddin. Tes Prestasi. Yogyakarta: Pustaka Pelajar, 2006
- Azwar, Saifuddin. Reliabilitas, Validitas, Interpretasi dan Komputasi. Yogyakarta: Liberty, 2006.
- Azwar, S. Sikap Manusia Teori dan Pengukurannya. Edisi 2. Yogyakarta: Pustaka Pelajar, 2007
- Crocker, L dan J. Algina. Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston Inc, 1996.
- Darus, M. F., Imami, A. I., & Abadi, A. P. (2021). Analisis Soal Dalam Buku Matematika Kelas Vii Semester 1 Berdasarkan Kriteria Dari Higher Order Thinking Skills (Hots). JPMI (Jurnal ..., 4(4), 777–788. <https://doi.org/10.22460/jpmi.v4i4.777-788> Wayan Eka Mahendra, ANALISIS BUTIR SOAL, Disampaikan pada Workshop Peningkatan Kompetensi Evaluasi Pembelajaran Guru SLB N 1 Buleleng, Singaraja 4 - 7 September 2019
- Hopkins, Charles D dan Antes, Richard L. Classroom Measurement and Evaluation. Illionis, F.E. Peacock, 1999.
- Huriaty, D., & Gazali, R. Y. (2021). Analisis Instrumen Ulangan Akhir Semester Ganjil Mata Pelajaran Matematika Kelas Vii Smpn 1 Hulu Sungai Tengah. 1(4)
- Joni, T. Raka. Pengukuran dan Penilaian Pendidikan. Surabaya: Karya Anda, 2006.
- Naga, Dali. S. Pengantar Teori Skor pada Pengukuran Pendidikan. Jakarta: Gunadarma, 2002.
- Holland. PW & Thaycr. DT (1988). Test Validity. New Jersey: Lawrence Erlbaum Associates, Publishers
- Nana Sudjana, Penilaian Proses Hasil Belajar Mengajar, Bandung, Remaja Rosdakarya, 1995
- Nitko, Anthony J. (1996). Educational Assessment of Students, Second Edition. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.
- Nurkencana, Wayan dan Supartana. Evaluasi Hasil Belajar. Surabaya: Usaha Nasional, 2002
- Sirait, Bistok. Bahan Pengajaran Untuk Mata Kuliah Evaluasi Hasil Belajar Siswa. Jakarta: Depdikbud, 2009
- Tilaar, A. L. F., & Hasriyanti, H. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I), 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>