



ANALISIS TINGKAT KESULITAN (*DIFFICULTY LEVEL*) SOAL PADA BUKU SEJARAH KEBUDAYAAN ISLAM KELAS 8 KURIKULUM 2013

R. Ahmad Nur Kholis (kholis3186@gmail.com)¹

¹ *STAINU Karangploso Malang*

(Received: November 2017 / Revised: Desember 2017 / Accepted: Desember 2017)

ABSTRACT

This study aims at describing the difficulty level of task items presented in chapter 1 and 2 in a book titled “Buku Guru Kurikulum 13 Mata Pelajaran Sejarah Kebudayaan Islam (SKI) Kelas 8”. This book is published by the Ministry of Religion Affair in 2015. The major data source analyzed in this study is the test result done by the eighth graders of MTs Nahdlatul Ulama Karangploso. The subject of this study is the task items contained in the topics about (1) the footsteps of Abbasite Dynasty, and (2) the Brilliance of Islamic Scholars in the Era of Abbasite Dynasty. Two kinds of analysis technique are employed in this study namely (1) Descriptive Statistic Analysis, also called Social Statistics, and (2) Proportion of correct. The result of the study shows that task items on both discussions in the book are 100% in difficulty level.

Keywords: Difficulty Level, SKI Subject

PENDAHULUAN

Guba & Lincoln mendefinisikan evaluasi sebagai suatu proses untuk menggambarkan evaluasi (orang yang dievaluasi dan menimbang makna dan nilainya).¹ Sax mendefinisikan evaluasi sebagai suatu proses di mana pertimbangan atau keputusan suatu nilai dibuat dari berbagai pengamatan, latar belakang serta latihan dari evaluator.² Arifin menyimpulkan dari dua rumusan ini bahwa evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan makna) dari pada sesuatu. Berdasarkan pertimbangan dan kriteria tertentu untuk menentukan keputusan.³ Hamdani Ihsan & Fuad Ihsan mendefinisikan bahwa evaluasi merupakan suatu proses menilai dan mengukur berjalannya sesuatu.⁴ Noor & Brookhart mendefinisikan evaluasi sebagai suatu proses penetapan nilai yang berkaitan dengan kinerja dan hasil karya siswa.⁵ Menurut Criffin & Nix evaluasi adalah *judgment* terhadap nilai atau implikasi dari hasil pengukuran. Menurut Tayler evaluasi adalah proses penentuan sejauh mana tujuan pendidikan telah tercapai.⁶ Dalam pendapat Kirkendall, dkk evaluasi adalah penentuan nilai atau manfaat dari suatu data kolektif.⁷ Stufflebeam berpendapat bahwasanya evaluasi adalah proses memperoleh, menyajikan, dan menggambarkan informasi yang berguna untuk menilai suatu alternatif pengambilan keputusan.⁸ Harun Rasyid & Mansur menyimpulkan dari berbagai pendapat di atas bahwasanya evaluasi merupakan proses mengumpulkan informasi untuk mengetahui pencapaian belajar kelas atau kelompok. Hasil evaluasi diharapkan dapat mendorong pendidik untuk mengajar lebih baik dan mendorong peserta didik untuk belajar lebih baik.⁹

Dari rumusan-rumusan definisi evaluasi tersebut dapat ditarik sebuah benang merah. Bahwasanya evaluasi selalu berkaitan dengan hal-hal sebagai berikut: (1) Tujuan (*goal*); (2) Proses (*process*); (3) Penilaian (*Assessment*); (4) Pengukuran (*Measurement*); dan (5) Kebijakan (*Decision*).

Dalam sebuah kegiatan pendidikan,--baik formal maupun non-formal,--evaluasi berfungsi sebagai pemberi informasi dan pertanggungjawaban kepada publik dan/atau *stakeholder*-nya tentang berbagai aspek yang terkait dengan pelaksanaan

¹ Zainal Arifin, *Evaluasi Pembelajaran*, (Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama, 2012), hlm. 8

² Ibid.

³ Ibid.

⁴ Hamdani Ihsan & Fuad Ihsan, *Filsafat Pendidikan Islam*, (Bandung: Pustaka Setia, 2008)

⁵ Harun Rasyid & Mansur, *Penilaian Hasil Belajar*, (Bandung: CV. Wacana Prima, 2008), hlm. 2

⁶ Ibid.

⁷ Ibid. hlm. 3

⁸ Ibid.

⁹ Ibid

dan hasil yang dicapai.¹⁰ Dalam proses pembelajaran di kelas evaluasi adalah sebuah metode pengumpulan data mengenai capaian yang diraih oleh siswa. Data-data yang didapatkan guru dari proses yang dinamakan evaluasi pembelajaran ini dapat dijadikan sebagai bahan kajian dan landasan untuk mengambil keputusan yang akan diambil.

Dalam sebuah aktivitas program pendidikan, kegiatan belajar-mengajar (akademik) merupakan porsi terbesar yang dievaluasi. Fokus pelaksanaan evaluasi dalam kebiasaannya ditekankan pada kinerja dan hasil belajar yang merupakan indikator keberhasilan proses pembelajaran. Kegiatan evaluasi ada kalanya dilakukan dalam rangka menyimpulkan tentang kinerja hasil atau untuk mendapatkan umpan balik yang dijadikan dasar bagi upaya-upaya perbaikan proses.¹¹

Kegiatan evaluasi dalam dalam proses pembelajaran pada hakikatnya adalah sebuah proses penelitian terhadap keberhasilan sebuah proses pembelajaran. Hal ini karena pada hakikatnya proses evaluasi tersebut telah menggunakan prinsip-prinsip penelitian. Mulai dari menemukan masalah, mengumpulkan data, menganalisis, dan kemudian menyimpulkan.¹²

Hanya saja, berbeda dengan penelitian yang diakhiri dengan kesimpulan dan saran, sebuah proses evaluasi diakhiri dengan pengambilan kebijakan (*making decision*) untuk memecahkan masalah yang dialami.¹³ Pengambilan keputusan ini dipilih oleh pemangku kebijakan dari beberapa alternatif yang diajukan.

Dalam proses inilah maka kita harus menyadari bahwa sebenarnya sebuah tes yang diberikan kepada siswa pada hakikatnya merupakan instrumen pengambilan data dalam sebuah penelitian yang menjelma menjadi evaluasi. Soal-soal ujian (tes) yang diberikan kepada peserta didik baik itu dibuat sendiri oleh guru maupun oleh negara pada intinya adalah untuk mengumpulkan data dan informasi tentang sejauh mana capaian peserta didik terhadap pembelajaran yang diberikan.

Dalam paradigma baru pembelajaran, digunakanlah istilah penilaian (*assessment*) dan bukannya ujian (*test*) lagi. Hal ini karena paradigma baru yang telah berkembang sejak awal tahun 2000-an ini memandang keberhasilan pembelajar diukur dari berbagai sudut pandang secara integral. Segi afektif, kognitif dan

¹⁰ R. Ibrahim & M. Ali, "Teori Evaluasi Pendidikan", dalam: M. Ali, R. Ibrahim, R. N.D. Sukmadinata, D. Sudjana & W. Rasjidin (ed.), *Ilmu dan Aplikasi Pendidikan*. (Bandung: PT. Imperial Bhakti Utama, 2007), hlm. 103-104

¹¹ Ibid.

¹² Suharsimi Arikunto & Cepi Safrudin Jabar, *Evaluasi Program Pendidikan*, (Jakarta: Bumi Aksara, 2014)

¹³ ibid

psikomotorik diamati secara menyeluruh. Penilaian inilah yang dimaksud dengan penilaian autentik. Dan pemberian tes bukanlah satu-satunya cara untuk menilai, melainkan ia merupakan bagian untuk mengukur kemampuan peserta didik. Dalam hal ini ia digunakan sebagai alat untuk mengukur kemampuan kognitif saja.

ANTARA TES DAN NON TES

Secara umum, alat penilaian (*instrumen*) dapat dikategorikan ke dalam dua bentuk yakni: 1) Tes; dan 2) Bukan Tes (*non-test*). Alat pengukuran yang termasuk ke dalam kategori non tes adalah: a) Kuesioner; b) Wawancara; c) Daftar Cocok (*check list*); d) Pengamatan atau Observasi ; e) Penugasan; f) Portofolio; g) Jurnal; h) Inventori; i) Penilaian diri (*Self Assesment*); j) Penilaian oleh teman (*peer assessment*). Sedangkan tes adalah sejumlah pertanyaan yang harus dijawab, atau pertanyaan-pertanyaan yang harus dipilih atau ditanggapi, tugas-tugas yang harus dikerjakan oleh orang yang diuji pada waktu tertentu. Tes merupakan sejumlah pertanyaan yang memiliki jawaban benar atau salah, pertanyaan yang membutuhkan jawaban atau diberi tanggapan untuk mengukur tingkat kemampuan seseorang dalam aspek tertentu.¹⁴

Dalam kaitannya dengan penyusunan sebuah tes, Badrun Kartowagiran memaparkan tentang langkah-langkah penyusunan tes secara umum, yaitu: (1) Penentuan tujuan tes; (2) penyusunan kisi-kisi; (3) Penulisan soal; (4) Penelaahan soal (Review dan Revisi Soal); (5) Uji coba soal; (6) perakitan soal menjadi perangkat tes. Penentuan tujuan dapat berupa tujuan khusus, yaitu melihat tingkat capaian suatu program. Kisi-kisi soal adalah deskripsi mengenai ruang lingkup soal yang akan diujikan serta memberikan perincian mengenai soal-soal yang diperlukan dalam tes tersebut. Penulisan soal adalah penjabaran jenis dan tingkat perilaku yang hendak diukur menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perinciannya dalam kisi-kisi. Telaah soal adalah kajian soal secara teoritis. Uji coba soal merupakan langkah untuk mengetahui kualitas soal yang akan diujikan dengan melihat secara empiris respons peserta tes. Perakitan soal adalah penyajian soal menjadi perangkat tes yang terpadu.¹⁵

Yahya Qohar dalam mengatakan bahwa sebuah evaluasi harus memenuhi 6 (enam) syarat, yaitu: (1) Reliabel; (2) Valid; (3) Obyektif; (4) Harus Diskriminatif; (4) Imperhensif; dan (5) Mudah Digunakan. Diskriminatif maksudnya ia mampu membedakan nilai dan ukuran siswa dalam capaiannya terhadap materi pelajaran

¹⁴ Sri Wening. *Materi Evaluasi Pembelajaran*. (Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta, 2012), hlm. 4

¹⁵ Badrun Kartowagiran, *Pengantar Teori Tes Klasik*, (Yogyakarta: Pascasarjana UNY & Dinas Pendidikan Propinsi Yogyakarta, 2009), hlm. 2-4

yang telah diberikan. Dan sejalan dengan pendapat ini, yakni bahwa sebuah evaluasi harus memenuhi syarat diskriminatif, maka sebuah tes haruslah dianalisis dari segi tingkat kesulitan dan daya pembedanya.¹⁶

Syamsudin mengatakan bahwa ada 4 (empat) cara yang dapat digunakan untuk menilai sebuah test yang baik, yaitu: (1) meneliti secara jujur soal-soal yang telah disusun; (2) mengadakan analisis soal (*item analysis*); (3) memeriksa validitas (*checking validity*); dan (4) memeriksa reliabilitas (*checking reliability*). Cara yang pertama adalah cara yang bersifat subyektif, sedang 3 (tiga) yang lain dapat dilaksanakan dengan secara kuantitatif dan terukur. Analisis soal dapat dilakukan dengan cara: (1) Tingkat kesukaran (*difficulty level*); (2) Daya Pembeda (*Discriminative test*); dan (3) Pola Jawaban soal.¹⁷

Sri Wening mengatakan bahwa sebuah tes dapat dikatakan layak sebagai alat pengukur jika memiliki beberapa persyaratan sebagai berikut: (1) *Valid* (Ketepatan); (2) *Reliabel* (Handal); (3) Obyektif; (4) Praktikabilitas; (5) Ekonomis. Valid berarti sebuah tes mampu mengukur apa yang seharusnya diukur. Sekumpulan soal ujian adalah valid untuk mengukur prestasi dan tidak valid jika digunakan untuk mengukur tingkat partisipasi. Bentuk validitas ada 3 (tiga) macam yaitu: (a) validitas isi (*content validity*); (b) validitas kriteria (*criterion-relation validity*); dan (c) validitas konstruk.¹⁸

Sedangkan berkaitan dengan pembelajaran, Arifin menyebutkan bahwa jenis-jenis validitas ada 5 (lima) yaitu: (1) Validitas permukaan (*Face Validity*); (2) Validitas isi (*content validity*); (3) Validitas empiris (*empirical validity*); (4) Validitas konstruk (*construct validity*); dan (5) Validitas faktor (*factorial validity*). Validitas permukaan adalah validitas yang paling sederhana, karena dalam validitas ini yang dilihat hanyalah sisi muka atau tampang dari instrumen itu sendiri. Validitas kurikuler berkenaan dengan pertanyaan apakah soal yang diberikan dalam tes telah sesuai dengan silabus yang ditentukan?. Validitas empiris adalah validitas yang berkenaan dengan pencarian korelasi. Validitas ini berkenaan dengan penggunaan analisis statistik. Validitas empiris ada 3 (tiga) macam, yaitu: (a) validitas prediktif (*predictive validity*); (b) validitas konkruen (*concruent validity*); dan (c) validitas sejenis (*congruent validity*). Validitas konstruk berkenaan dengan pertanyaan sejauh mana sebuah tes adalah benar-benar mengukur apa yang seharusnya diukur?. Validitas faktor berkenaan dengan homogenitas data yang diperoleh dengan

¹⁶ Syamsudin, "Pengukuran Daya Pembeda, Taraf Kesukaran dan Pola Jawaban Tes (Analisis Butir Soal)" dalam *Jurnal At-Tajdid*, Vol: 1, No.: 2 (2012), hlm. 188

¹⁷ *ibid*

¹⁸ Idrus Alwi, "Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran dan Daya Pembeda", dalam *Jurnal Ilmiah Faktor Exacta*, Vol.: 3, No. 2 (Juni 2010) hlm. 187

menggunakan instrumen tersebut meskipun ada beberapa faktor yang mempengaruhi hasil.¹⁹

Reliabel berarti sebuah instrumen harus dapat diandalkan sebagai alat pengukur. Jika sebuah instrumen diujikan secara berkali-kali dan hasilnya adalah sama atau mendekati sama, maka instrumen tersebut dapat dikatakan reliabel atau dapat diandalkan. Gronlund mengatakan bahwa ada 4 (empat) faktor yang mempengaruhi reliabilitas tes, yaitu: (1) Panjang tes (*length of test*); (2) Sebaran skor (*Spread of scores*); (3) Tingkat kesukaran (*difficulty level*); dan (4) Obyektivitas (*Objectivity*).²⁰

Obyektif artinya bahwa sebuah instrumen penilaian haruslah bebas dari pandangan diri penilai. Praktikabilitas berarti sebuah instrumen haruslah mudah dioperasikan (dijalankan). Ekonomis berarti bahwa sebuah instrumen haruslah mudah dijangkau dalam hal pembiayaannya.²¹

Sedangkan Alwi menjelaskan bahwa persyaratan sebuah tes yang bermutu adalah harus memenuhi 5 (lima) persyaratan yaitu: (1) Valid; (2) Reliabel; (3) Memiliki taraf kesukaran yang baik; (4) memiliki daya pembeda yang baik; dan (5) Pengecohnya berfungsi.²²

Sebagai sebuah instrumen, soal-soal yang diberikan kepada peserta didik dalam sebuah tes haruslah diuji dari segala seginya. Rasyid & Mansur menjelaskan bahwa pengujian sebuah instrumen dalam bentuk soal ini ada 4 (empat) yaitu: (a) Uji Validitas; (b) Uji Reliabilitas; (c) Analisis Daya Pembeda; dan (d) Analisis Tingkat Kesulitan Soal. Dalam kaitan dengan ini semua, artikel ini akan membahas hal yang terakhir yang disebutkan itu, yakni: analisis tingkat kesulitan soal.²³

KEDUDUKAN EVALUASI DALAM PEMBELAJARAN

Arikunto & Jabar membedakan antara penilaian (*assessment*), pengukuran (*measurement*) dan evaluasi (*evaluation*). Penilaian yang merupakan kata benda dari 'nilai' berkaitan dengan hal yang bersifat kualitatif. Sedangkan pengukuran berkaitan dengan hal-hal yang bersifat kuantitatif. Adapun evaluasi adalah kegiatan pengumpulan data dan informasi tentang bekerja dan berjalannya sesuatu yang

¹⁹ Zainal *Evaluasi Pembelajaran ...*, hlm. 315-326

²⁰ *Ibid.* hlm. 327

²¹ Sri Wening, *Materi Evaluasi ...*, hlm. 4-6

²² Idrus Alwi, "Pengaruh Jumlah Alternatif..."; Abdul Kadir, "Menyusun dan Menganalisis Tes Hasil Belajar", dalam: *Jurnal At-Ta'dib*, Vol.: 8, No.: 2, (Juli-Desember, 2015), hlm. 72-80

²³ Harun Rasyid & Mansur, *Penilaian Hasil Belajar...*

mana data dan informasi tersebut selanjutnya digunakan untuk menentukan kebijakan.²⁴

Penilaian merupakan kegiatan melekat dari pengajaran. Kegiatan penilaian terjadi baik pada awal, proses, maupun pada akhir pembelajaran. Pada awal pembelajaran, penilaian dilakukan untuk tujuan diagnosa yakni untuk mengetahui tingkat kemampuan dan modal pertama yang dimiliki siswa atau juga berfungsi sebagai penentuan kebijakan dalam hal penempatan (*placement*) siswa pada kelompok belajar tertentu. Pada saat pembelajaran berlangsung, penilaian berfungsi untuk mengukur ketercapaian tujuan pembelajaran yang telah ditetapkan dan hasilnya digunakan sebagai *feedback* atas kegiatan pembelajaran yang dilakukan (*formative*). penilaian pada akhir belajar dilakukan untuk mengukur ketercapaian keseluruhan tujuan kurikulum yang telah ditetapkan pada jenjang pendidikan tertentu (*summative*) dan hasilnya digunakan sebagai laporan kepada siswa tentang hasil belajarnya, kepada guru, orang tua siswa, masyarakat dan pemerintah sebagai wujud akuntabilitas penyelenggaraan pendidikan.²⁵

Secara prinsipiil evaluasi merupakan suatu kegiatan penilaian yang bertujuan untuk mengukur tingkat efektivitas kegiatan dalam mencapai tujuan yang diharapkan. Secara fungsional kegiatan penilaian merupakan kegiatan mencari informasi yang akan dijadikan landasan menentukan kebijakan selanjutnya.²⁶

Berdasarkan fungsinya, evaluasi dapat diklasifikasikan menjadi dua bagian, yakni evaluasi proses dan evaluasi hasil pembelajaran. Evaluasi Proses merupakan kegiatan pengukuran yang dilaksanakan secara sistematis untuk memperoleh informasi tentang efektivitas aktivitas belajar mengajar. Sedangkan evaluasi hasil belajar menunjuk pada aktivitas penilaian terhadap tingkat kualitas hasil belajar yang dicapai peserta didik.²⁷ Di sinilah maka benar apa yang dikatakan bahwa evaluasi adalah sebuah proses pengukuran dan penilaian.²⁸

Siskandar dalam penelitiannya menyimpulkan bahwa efektivitas monitoring dan evaluasi merupakan salah satu faktor yang menentukan kesuksesan implementasi Kurikulum 2013 di samping Kelengkapan Infrastruktur, Kompetensi guru, Prinsip

²⁴ Suharsimi Arikunto & Cepi Sfrudin Jabar, *Evaluasi Program Pendidikan...*, hlm.1-2; lihat juga: Hamdani Ihsan & Fuad Ihsan, *Filsafat Pendidikan Islam...*

²⁵ Nunung Nuriyah, "Evaluasi Pembelajaran: Sebuah Kajian Teori", dalam: *Jurnal Edueksos*, Vol., 3 No 1 (2014), hlm.79

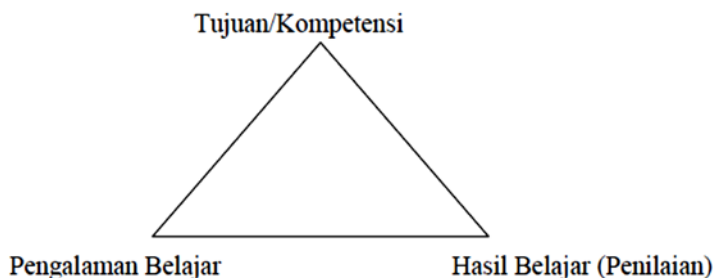
²⁶ Fajri Chairawati, "Evaluasi Pembelajaran pada Kelas Internasional", Fakultas Dakwah IAIN Ar-Raniry, dalam: *Jurnal Al-Bayan*, Vol., 20. No. 20 (Januari-Juni 2014), hlm. 16

²⁷ *ibid*

²⁸ Hamdani Ihsan & Fuad Ihsan, *Filsafat Pendidikan Islam...*

kepemimpinan, Fasilitas guru,²⁹ Lingkungan dan budaya sekolah, dan Efektivitas Monitoring dan Evaluasi Implementasi Kurikulum. Dan kesalahan dalam penerapan atau pelaksanaan penilaian (*Assessment*) merupakan salah satu faktor dari kegagalan implementasi kurikulum 2013 di samping kesalahan persepsi tentang Kurikulum 2013, dan kurang diterapkannya Rencana Pelaksanaan Pembelajaran, ruang kelas pembelajaran, penggunaan media IT, fasilitas mengajar dan manajemen madrasah.

Hubungan antara penilaian dan pembelajaran secara ringkas digambarkan Nuriyah sebagai berikut ini:³⁰



Gambar 1: Hubungan Penilaian dan Pembelajaran

TEORI TES, DARI KLASIK HINGGA MODERN

Sampai saat ini, ada dua teori tes yang masih digunakan dan terus dikembangkan, yaitu teori tes klasik disebut juga *classical test theory* (CTT) dan teori tes modern disebut juga disebut teori respons item atau *item response theory* (IRT). Kedua teori ini dibangun atas dasar yang berbeda.. Teori tes klasik didasarkan kepada model aditif yang menyatakan bahwa skor amatan (X) adalah hasil penjumlahan dari skor sebenarnya atau *true score* (T) dan kesalahan pengukuran (e). Jadi secara matematis model ini dirumuskan sebagai berikut: $X = T + e$, di mana X adalah skor amatan, T adalah skor sebenarnya, dan e adalah kesalahan pengukuran atau *error*. Pengukuran menurut teori klasik adalah pemberian angka kepada objek atau kejadian dengan

²⁹ Siskandar, "Evaluasi Implementasi Kurikulum 2013 di Madrasah Aliyah", dalam: *Cendekia*, Vol: 10. No. 2 (2016), hlm.117

³⁰ Nunung Nuriyah, *Evaluasi Pembelajaran...*, hlm.75

yang diatur dengan aturan tertentu. Angka merupakan sifat yang melekat pada sesuatu dan tak dapat dipisahkan.³¹

Ada 7 (tujuh) asumsi yang diajukan teori klasik ini, yaitu:³² (1) Asumsi yang paling utama dari teori klasik ini adalah apa yang telah dipaparkan di atas yang dirumuskan secara matematis dengan: $X = T + e$.; (2) Bahwa nilai rata-rata (*mean*) populasi dari nilai amatan adalah bersifat independen dan pasti sama dengan skor sebenarnya untuk setiap peserta tes untuk tes yang sama. Asumsi ini jika dirumuskan ke dalam bentuk matematis menjadi: $E(X)=T$. Dengan kata lain bahwa skor sebenarnya merupakan nilai skor rata-rata perolehan teoritis sekiranya dilakukan pengukuran secara berulang dengan menggunakan instrumen yang sama; (3) Tidak ada korelasi antara skor sebenarnya dan skor kesalahan, atau secara matematis disimbolkan: $\rho_{ET} = 0$; (4) Tidak ada korelasi antara tes pertama dengan kedua. Dengan kata lain bahwa peserta tes yang memiliki tingkat kesalahan tertinggi pada tes pertama belum tentu demikian pula pada tes kedua; (5) Skor kesalahan pada tes pertama tidak berkorelasi dengan kesalahan pada tes kedua; (6) Dua perangkat tes yang mengukur *trait* yang sama dapat dikatakan paralel jika: (a) memenuhi syarat nomor 1 sampai 5; dan (b) $\text{traitment}(T)_1 = \text{traitment}(T)_2$ serta $\text{varian}(\sigma)_1 = \text{varian}(\sigma)_2$; (7) Dua buah tes yang digunakan untuk mengukur *trait* yang sama dapat dikatakan setara (*equivalent*) jika skor tes dari para peserta pada tes pertama sama dengan perolehan skor pada tes kedua yang ditambahkan dengan suatu konstanta (C), atau jika dirumuskan secara matematis adalah: $X_1 = X_2 + C$. Karakteristik soal menurut teori klasik ini ada 5 (lima) yaitu: (1) Tingkat Kesulitan Tes; (2) Daya Pembeda; (3) Validitas; dan (4) Reliabilitas; dan (5) Efektivitas Distraktor.³³

Sedangkan teori tes modern mengatakan bahwa model-model yang diinginkan harus memenuhi sifat-sifat: (1) karakteristik butir tergantung pada kelompok peserta tes yang dikenai tes tersebut; (2) skor yang menyatakan kemampuan peserta tes tidak tergantung pada tes; (3) model dinyatakan dalam tingkatan (*level*) butir soal; (4) model tidak memerlukan tes paralel untuk menghitung koefisien reliabilitas; (5) model menyediakan ukuran yang tepat untuk setiap skor kemampuan. Model alternatif yang dapat memenuhi kesemua ciri-ciri tersebut adalah model pengukuran yang disebut teori respons butir (*Item Respon Theory*). Teori tes modern atau yang

³¹ Nonoh Siti Aminah, "Teori Pengukuran dalam Pendidikan" dalam *Jurnal dan Materi Pembelajaran Fisika (JPMF)*, Vol. 3, No 2 (2013), hlm.33-34

³² Ibid, hlm. 34, lihat juga: Suwanto, "Teori Tes Klasik & Teori Tes Modern" dalam *WIDYATAMA*, Vol. 20, No. 1 (2011), hlm. 69-70

³³ Nonoh, *Teori Pengukuran...*, hlm. 34-35. lihat juga Suwanto, *Teori Tes Klasik...*, hlm. 70-72; Junaidi Lababa, "Analisis Butir Soal dengan Teori Tes Klasik: Sebuah Pengantar" dalam *IQRA'*, Vol. 5 (januari 2008) hlm. 35

disebut juga dengan teori respons butir ini dikembangkan para ahli untuk menyempurnakan teori klasik yang dianggap memiliki beberapa kekurangan.³⁴

Sebagaimana juga teori klasik, teori modern ini juga memiliki beberapa asumsi yang diajukan, yaitu: (1) Unidimensionalitas; (2) Independensi lokal; dan (3) Fungsi karakteristik butir. Asumsi unidimensionalitas menyatakan bahwa yang diukur oleh sekumpulan butir soal dalam tes adalah satu. Asumsi ini dalam praktiknya sukar untuk dipenuhi karena terdapat banyak faktor yang dapat mempengaruhi hasil tes. Asumsi independensi lokal menyatakan bahwa sikap kemampuan yang mempengaruhi suatu tes adalah konstan, oleh karenanya respons peserta tes adalah independen secara statistik. Fungsi karakteristik butir menyatakan hubungan yang sebenarnya antara variabel yang tak terobservasi (yaitu kemampuan) dan variabel yang terobservasi (yaitu respons butir).³⁵

Pada dasarnya, masing-masing teori memiliki kelebihan dan kelemahannya masing-masing. Pada teori tes klasik, kelebihan adalah: (1) Menggunakan konsep yang sederhana dalam menentukan kemampuan tes; (2) Menggunakan konsep yang sederhana dalam menghitung koefisien validitas dan reliabilitas tes serta menghitung nilai parameter butir soal; (3) Dapat digunakan pada sampel kecil, misalnya pada tingkat kelas; (4) Sudah digunakan dalam praktiknya dalam kurun waktu yang cukup lama dan telah diketahui sejak lama oleh sebagian besar orang yang berkecimpung dalam psikologi dan pendidikan. Sedangkan kelemahan-kelemahannya adalah: (1) Kemampuan peserta tes dinyatakan dalam variabel yang deskriptif; (2) besarnya koefisien validitas dan koefisien reliabilitas suatu tes serta nilai parameter suatu butir soal tergantung kepada peserta tes yang dikenai suatu tes.³⁶

Sedangkan teori tes modern memiliki kelebihan-kelebihan sebagai berikut: (1) Sebagai sebuah penyempurnaan, maka landasan teorinya lebih baik; (2) Kemampuan peserta tes dinyatakan dalam secara kontinu; (3) Tidak diperlukan tes yang paralel dalam menentukan reliabilitas; (4) Nilai koefisien reliabilitas suatu tes dan nilai parameter suatu butir soal tidak tergantung pada peserta tes yang mengikuti suatu tes. Sedangkan kelemahan-kelemahannya adalah: (1) Memerlukan sampel yang besar untuk mendapatkan parameter yang stabil sehingga tidak dapat diterapkan dalam lingkup kelas; (2) Memerlukan aplikasi komputer yang handal untuk mendapatkan hasil penghitungan estimasi parameter yang akurat; (3) Belum diterima adanya oleh sebagian besar orang yang telah berkecimpung dalam dunia pendidikan dan psikologi, khususnya di Indonesia.³⁷

³⁴ Suwanto, *Teori Tes Klasik...*, hlm. 75

³⁵ *Ibid*

³⁶ *Ibid*, hlm. 77

³⁷ *ibid*

Grondlund dalam Arifin mengatakan bahwa ada 3 (tiga) faktor yang mempengaruhi hasil tes, yaitu: (1) Faktor Instrumen evaluasi; (2) Faktor administrasi, evaluasi dan penskoran; dan (3) Faktor jawaban dari peserta didik. Dalam mengembangkan sebuah instrument evaluasi, seorang evaluator (dalam hal ini adalah guru) haruslah memperhatikan hal-hal yang mempengaruhi validitas instrumen dan berkaitan dengan prosedur penyusunan instrumen, seperti silabus, kisi-kisi soal, petunjuk pengerjaan soal dan pengisian jawaban, kunci jawaban, penggunaan kalimat yang efektif, bentuk alternatif jawaban, tingkat kesukaran, daya pembeda dan sebagainya. Faktor administrasi evaluasi dan penskoran berkaitan dengan alokasi waktu pengerjaan soal yang proporsional, menghindarkan kecurangan selama pengisian tes, menghindarkan kesalahan penskoran dan memperhatikan kondisi fisik dan psikis peserta didik dalam pengerjaan soal. Faktor jawaban peserta didik meliputi kecenderungan peserta didik untuk menjawab dengan cepat tapi kurang cermat dalam mencoba menjawab dengan tepat, keinginan melakukan coba-coba, dan penggunaan gaya bahasa tertentu dalam soal uraian.³⁸

TINGKAT KESULITAN BUTIR TES (*DIFFCULTY LEVEL*)

Pemberian tes adalah sebuah pendekatan yang diambil oleh para ahli yang menganut model pengukuran (*measurement*) klasik dalam evaluasi. Model ini berpandangan bahwa pengukuran adalah suatu kegiatan yang ilmiah dan dapat diterapkan dalam berbagai bidang termasuk pendidikan. Obyek dari pengukuran ini adalah hasil belajar peserta didik. Sedang tujuannya adalah untuk mengungkapkan perbedaan individual. Dan meskipun banyak yang mengkritik, telah terbukti secara nyata dalam praktiknya bahwa cara ini telah digunakan sejak dulu dan bahkan sampai sekarang.³⁹

Kemudian, mengingat bahwa tujuan dari pengukuran adalah untuk mengungkapkan perbedaan individual peserta didik dilihat dari hasil belajarnya, maka dalam penyusunan soal tes harus memperhatikan masalah daya pembeda dan tingkat kesukaran masing-masing soal dan daya pembeda dari sebuah tes. Daya pembeda adalah tingkat kemampuan soal untuk mengungkapkan perbedaan kemampuan individual peserta didik. Setiap soal yang dapat dijawab dengan benar oleh hampir semua siswa adalah soal dengan daya pembeda yang rendah. Demikian pula soal yang tidak dapat dijawab dengan benar oleh hampir semua siswa juga memiliki daya pembeda yang rendah pula.

³⁸ Zainal Arifin, *Evaluasi Pembelajaran ...*, hlm. 314-315

³⁹ R. Ibrahim & M. Ali, *Teori Evaluasi Pendidikan...*, hlm. 106-108

Tingkat kesukaran (*difficulty level*) suatu soal adalah proposisi atau persentase subyek yang menjawab butir tes tertentu dengan benar. Sedang angka yang menunjukkan sukar atau tidaknya butir soal dalam tes disebut dengan indeks (dilambangkan dengan p (proporsi)). Indeks ini terentang antara 0 sampai 1. Diharapkan dalam sebuah soal tes tersusun dengan tingkat kesukaran yang merata (proporsional) antara soal yang tingkat mudah, sedang dan sulit.⁴⁰ Artikel ini fokus membahas tentang tingkat kesulitan tes.

Tes yang baik adalah yang mengandung soal yang tidak terlalu mudah dan tidak terlalu sukar. Sebuah soal yang memiliki tingkat kesukaran terlalu tinggi atau terlalu rendah adalah soal yang tidak memiliki fungsi yang layak secara akademik. Karena soal yang terlalu mudah akan menyebabkan minat belajar siswa kurang menjadi menurun. Demikian pula soal yang terlalu sukar akan membuat peserta didik tidak tertarik untuk belajar dan bahkan bisa frustrasi.⁴¹

Dalam membicarakan tentang tingkat kesukaran atau tingkat kesulitan butir soal, maka seorang guru pembuat soal haruslah memahami dan memperhatikan butir soal dan peserta tes.⁴² Yang dimaksud dengan butir soal adalah materi pelajaran yang dibuat soal. Sedangkan yang dimaksud dengan peserta tes adalah kemampuan menjawab (*intake*) peserta didik terhadap materi. Bisa saja sebuah butir soal adalah mudah bagi sebagian peserta didik namun dianggap sulit bagi sebagian yang lain.

Asumsi yang digunakan dalam menyusun sebuah soal tes yang dengan kualitas yang baik adalah adanya keseimbangan dalam tingkat kesulitannya di samping dalam masalah validitas dan reliabilitasnya. Keseimbangan dalam tingkat kesulitan yang dimaksud adalah bahwa dalam sekian jumlah soal harus ada butir-butir soal yang bersifat sulit, sedang dan mudah secara proporsional.⁴³

Persoalan yang lain dalam menentukan tingkat kesulitan (demikian pula daya pembeda) soal adalah ukuran untuk menentukan apakah soal tersebut memiliki daya pembeda rendah, sedang atau tinggi. Apakah soal tersebut mudah, sedang atau sulit. Dalam hal ini, maka dapat dilaksanakan uji otoritas, yakni otoritas gurulah yang menentukan ukuran rendah-tingginya daya pembeda atau mudah sulitnya sebuah soal. Hal ini dengan mempertimbangkan: 1) Kemampuan subyek tes (*testee*) dalam menjawab butir soal; 2) sifat materi yang diujikan atau dinyatakan; 3) isi bahan yang dinyatakan sesuai dengan bidang keilmuannya, baik luasnya maupun

⁴⁰ Ibid, 107-108; Harun Rasyid & Mansur, *Penilaian Hasil Belajar...*, hlm. 239 dan 245

⁴¹ Syamsudin, *Pengukuran Daya Pembeda ...*, hlm. 193

⁴² Harun Rasyid & Mansur, *Penilaian Hasil Belajar...*, hlm. 239

⁴³ Ibid., hlm. 240

kedalamannya; 4) bentuk tes obyektif apakah pilihan ganda, benar-salah atau mencocokkan. Tipe benar salah adalah lebih mudah dari pada pilihan ganda.⁴⁴

Namun demikian penentuan otoritatif sebagaimana dijelaskan di atas tersebut belumlah bersifat secara mutlak. Sebab masih memerlukan pembuktian secara empirik. Karena bisa jadi sebuah soal dianggap sulit berdasarkan keputusan *judgement* seorang guru (pemangku otoritas) namun secara empirik adalah sebaliknya. Untuk itu diperlukan beberapa teknik komputasi dalam menentukannya.

Untuk mengukur tingkat kesulitan soal (selanjutnya disimbolkan dengan: p , merujuk pada: proporsi) digunakan beberapa cara, yaitu: 1) proporsi menjawab benar; 2) skala kesukaran linear; 3) indeks davis; 4) skala bivariat.⁴⁵

PEMBAHASAN

Analisis dalam penelitian ini menggunakan 2 (dua) metode yaitu: (1) Analisis statistik deskriptif atau yang disebut juga dengan statistik sosial⁴⁶ pada data yang berupa persentase; (2) metode analisis tingkat kesulitan (*difficulty level*) tes dengan teknik analisis proporsi menjawab benar sebagaimana dijelaskan dalam Rasyid & Mansur (2008:241). Analisis statistik deskriptif dalam penelitian ini digunakan untuk dengan mudah mengetahui seberapa besar persentase siswa mampu menjawab dengan benar sebuah butir tes. Juga untuk mengetahui seberapa besar persentase soal dalam berbagai kriterianya (mudah, sedang dan sulit). Sedangkan analisis proporsi menjawab benar adalah untuk mengetahui dan dapat memastikan mengenai seberapa sulit soal tersebut bagi siswa.

Penelitian mengenai tingkat kesulitan tes dalam artikel ini dilakukan pada 20 (dua puluh) soal dalam bentuk pilihan ganda (*multiple choice*) pada latihan Bab 1 dan 2 pada Buku Guru Kurikulum 2013 mata pelajaran Sejarah Kebudayaan Islam (SKI) Kelas 8 (Sembilan) yang diterbitkan oleh Kementerian Agama Jakarta Revisi tahun 2016. Dua bab yang dimaksud adalah pembahasan tentang: (1) Jejak Peradaban Dinasti Abbasiyah; dan (2) Kecemerlangan Ilmuwan Muslim Dinasti Abbasiyah. Soal latihan dari kedua bab tersebut terdiri dari 10 (sepuluh) butir soal yang berasal dari Bab 1, dan 10 (sepuluh) soal dari Bab 2. Butir soal disusun secara berurutan yakni 10 (sepuluh) soal secara berurutan adalah soal yang diambil dari Bab 1, dilanjutkan 10 (sepuluh) soal yang diambil dari Bab 2. Subyek penelitian ini adalah

⁴⁴ *ibid*

⁴⁵ *Ibid.*, hlm.: 241

⁴⁶ George Ritzer, *Handbook Teori Sosial*, (Jakarta: Nusamedia, 2001)

siswa kelas 8 (Delapan) MTs Nahdlatul Ulama Kepuharjo Karangploso Malang sebanyak 40 (empat puluh) siswa.

Terkait dengan penelitian ini, peneliti menggunakan cara mencari proporsi menjawab benar dengan rumus:

$$p_i = \frac{\sum x_i}{S_{mi}N}$$

di mana:

p_i = tingkat kesukaran butir i atau proporsi menjawab benar butir i

$\sum x_i$ = banyaknya peserta tes yang menjawab benar butir i

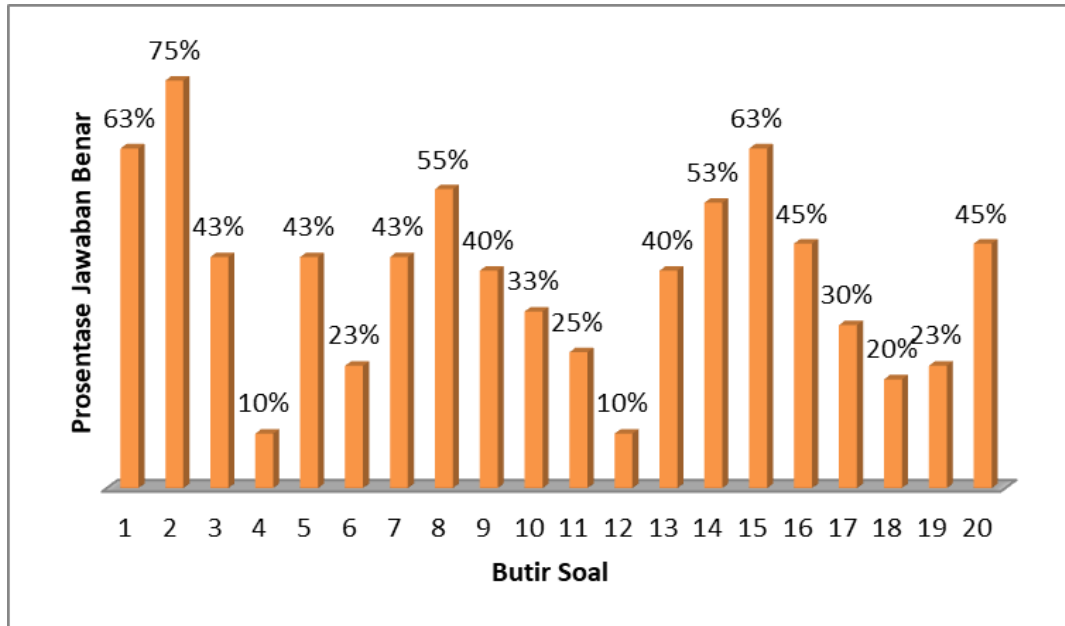
S_{mi} = skor maksimum butir soal

N = jumlah peserta tes

Berdasarkan hasil penghitungan persentase jawaban benar siswa terhadap setiap butir soal ($\sum x_i$), dapat dipaparkan data sebagai berikut: (1) sebesar 63% (25 siswa) dari keseluruhan siswa menjawab benar pada butir soal nomor 1; (2) sebesar 75% (30 siswa) dari keseluruhan siswa menjawab benar pada butir soal nomor 2; (3) sebesar 43% (17 siswa) dari keseluruhan siswa menjawab benar pada butir soal nomor 3; (4) sebesar 10% (4 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 4; (5) sebesar 43% (17 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 5; (6) sebesar 23% (9 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 6; (7) sebesar 43% (17 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 7; (8) sebesar 55% (22 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 8; (9) sebesar 40% (16 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 9; (10) sebesar 33% (13 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 10; (11) sebesar 25% (10 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 11; (12) sebesar 10% (4 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 12; (13) sebesar 40% (16 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 13; (14) sebesar 53% (21 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 14; (15) sebesar 63% (25 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 15; (16) sebesar 45% (18 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 16; (17) sebesar 30% (12 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 17; (18) sebesar 20% (8 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 18; (19) sebesar 23% (9 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal

nomor 19; (20) sebesar 45% (18 siswa) dari keseluruhan siswa mampu menjawab benar pada butir soal nomor 20.

Jika disajikan dalam bentuk diagram batang, maka data-data tersebut akan terlihat sebagaimana berikut ini:



Gambar 2: Grafik persentase jawaban Siswa terhadap setiap butir soal

Skor maksimum butir soal dalam penelitian ini adalah 5 (lima) hal ini menunjukkan bahwa setiap siswa yang mampu menjawab benar satu butir soal maka ia mendapat skor 5. Penentuan skor ini diambil peneliti dengan menghitung 100 dibagi 20 soal ($100/20$). Dengan asumsi bahwa jika salah satu siswa dapat menjawab setiap butir soal dengan benar maka mendapatkan nilai 100. Maka skor masing-masing soal adalah dengan cara membagikan angka 100 tersebut dengan jumlah butir soal.

Jumlah peserta tes (*testee*) pada penelitian ini adalah 40 orang peserta didik yang terdiri dari siswa kelas 8 (delapan). Sehingga jika angka-angka tersebut dimasukkan ke dalam rumus maka akan menjadi sebagaimana berikut ini:

Analisis tingkat kesulitan butir soal pertama:

$$p_i = \frac{25}{5 \times 40} = 0.13$$

Analisis tingkat kesulitan butir soal kedua:

$$p_i = \frac{30}{5 \times 40} = 0.15$$

Demikian seterusnya untuk setiap butir soal.

Peneliti menghitung masing-masing tingkat kesulitan ke 20 (dua puluh) soal tersebut dengan menggunakan bantuan aplikasi Microsoft Excel 2010 yang dijalankan di bawah sistem operasi Windows 10. Penghitungan dimulai dengan melakukan analisis butir soal dan memaparkan proporsi jawaban benar siswa terhadap setiap butir soal. Data ini dipaparkan dalam bentuk persentase.

Penelitian dilanjutkan dengan menentukan kriteria tingkat kesulitan setiap soal tes. Yang mana kriteria ini didasarkan pada apa yang dijelaskan oleh Rasyid & Mansur (2008) yakni: (a) $p \leq 0.30$ = butir soal sulit; (b) $0.30 < p \leq 0.70$ = butir soal sedang; dan (c) $p > 0.70$ = butir soal mudah.

Hasil penghitungan proporsi menjawab benar menunjukkan nilai dari setiap butir soal tersebut adalah: (1) tingkat kesukaran butir soal nomor 1 adalah 0.13. Angka ini menunjukkan butir soal dengan kriteria sulit; (2) tingkat kesukaran butir soal nomor 2 adalah 0.15. Angka ini menunjukkan butir soal dengan kriteria sulit; (3) tingkat kesukaran butir soal nomor 3 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit; (4) tingkat kesukaran butir soal nomor 4 adalah 0.02. Angka ini menunjukkan butir soal dengan kriteria sulit; (5) tingkat kesukaran butir soal nomor 5 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit; (6) tingkat kesukaran butir soal nomor 6 adalah 0.05. Angka ini menunjukkan butir soal dengan kriteria sulit; (7) tingkat kesukaran butir soal nomor 7 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit; (8) tingkat kesukaran butir soal nomor 8 adalah 0.11. Angka ini menunjukkan butir soal dengan kriteria sulit; (9) tingkat kesukaran butir soal nomor 9 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit; (10) tingkat kesukaran butir soal nomor 10 adalah 0.07. Angka ini menunjukkan butir soal dengan kriteria sulit; (11) tingkat kesukaran butir soal nomor 11 adalah 0.05. Angka ini menunjukkan butir soal dengan kriteria sulit; (12)

tingkat kesukaran butir soal nomor 12 adalah 0.02. Angka ini menunjukkan butir soal dengan kriteria sulit; (13) tingkat kesukaran butir soal nomor 13 adalah 0.08. Angka ini menunjukkan butir soal dengan kriteria sulit; (14) tingkat kesukaran butir soal nomor 14 adalah 0.11. Angka ini menunjukkan butir soal dengan kriteria sulit; (15) tingkat kesukaran butir soal nomor 15 adalah 0.13. Angka ini menunjukkan butir soal dengan kriteria sulit; (16) tingkat kesukaran butir soal nomor 16 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit; (17) tingkat kesukaran butir soal nomor 17 adalah 0.06. Angka ini menunjukkan butir soal dengan kriteria sulit; (18) tingkat kesukaran butir soal nomor 18 adalah 0.04. Angka ini menunjukkan butir soal dengan kriteria sulit; (19) tingkat kesukaran butir soal nomor 19 adalah 0.05. Angka ini menunjukkan butir soal dengan kriteria sulit; dan (20) tingkat kesukaran butir soal nomor 20 adalah 0.09. Angka ini menunjukkan butir soal dengan kriteria sulit.

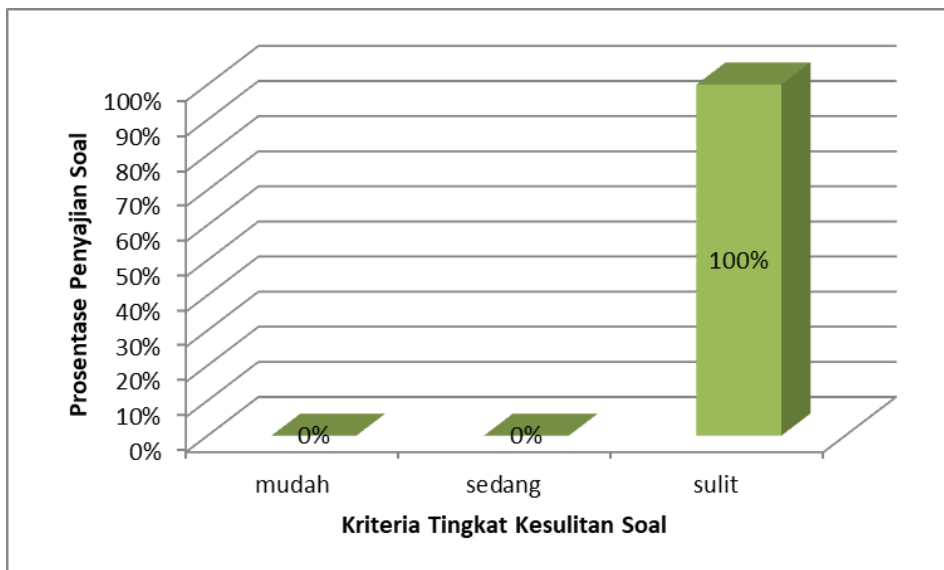
Paparan data tersebut di atas jika disajikan dalam bentuk tabel maka akan terlihat sebagai berikut:

Tabel 1.
Hasil penghitungan Tingkat Kesulitan Soal (Difficulty Level Test)

No. Soal	N	Sm	$\sum x$	Persentase	<i>P</i>	Kriteria
1	40	5	25	63%	0.13	Sulit
2	40	5	30	75%	0.15	Sulit
3	40	5	17	43%	0.09	Sulit
4	40	5	4	10%	0.02	Sulit
5	40	5	17	43%	0.09	Sulit
6	40	5	9	23%	0.05	Sulit
7	40	5	17	43%	0.09	Sulit
8	40	5	22	55%	0.11	Sulit
9	40	5	16	40%	0.08	Sulit
10	40	5	13	33%	0.07	Sulit
11	40	5	10	25%	0.05	Sulit
12	40	5	4	10%	0.02	Sulit
13	40	5	16	40%	0.08	Sulit
14	40	5	21	53%	0.11	Sulit
15	40	5	25	63%	0.13	Sulit
16	40	5	18	45%	0.09	Sulit

17	40	5	12	30%	0.06	Sulit
18	40	5	8	20%	0.04	Sulit
19	40	5	9	23%	0.05	Sulit
20	40	5	18	45%	0.09	Sulit

Berdasarkan hasil analisis statistik sosial (Ritzer, 2001), dan dengan menggunakan kriteria sebagaimana yang dijelaskan dalam Rasyid & Mansur (2008) maka dapat dinyatakan bahwa ke dua puluh soal yang disajikan dalam buku guru SKI Kelas 8 MTs adalah masuk dalam kriteria sulit. Jika disajikan ke dalam bentuk diagram batang maka data tersebut akan terlihat sebagai berikut ini:



Gambar 3: Persentase Penyajian Soal

Data hasil analisis statistik di atas disajikan berdasarkan persentase yang bersifat general. Namun jika data umum (persentase) tersebut diuraikan berdasarkan tingkat kesulitannya, yakni diurutkan dari soal yang tingkat kesulitannya paling rendah menuju soal yang tingkat kesulitannya paling tinggi adalah: (1) butir soal nomor 4 dan 12 dengan memiliki dan persentase yang sama yaitu: p : 0.02 dan persentase peserta didik yang menjawab benar sebesar: 10% (4 peserta didik); (2) butir soal nomor 18 dengan nilai p : 0.04 dan persentase peserta didik yang menjawab benar sebesar: 20% (8 peserta didik); (3) butir soal nomor 6 dan 19 dengan nilai p : 0.05 dan persentase peserta didik yang menjawab benar sebesar: 23% (9 peserta didik);

(4) butir soal nomor 11 dengan nilai p : 0.05 dan persentase peserta didik yang menjawab benar sebesar: 25% (10 peserta didik); (5) butir soal nomor 17 dengan nilai p : 0.06 dan persentase peserta didik yang menjawab benar sebesar: 30% (12 peserta didik); (6) butir soal nomor 10 dengan nilai p : 0.07 dan persentase peserta didik yang menjawab benar sebesar: 33% (13 peserta didik); (7) butir soal nomor 9 dan 13 dengan nilai p : 0.08 dan persentase peserta didik yang menjawab benar sebesar: 40% (16 peserta didik); (8) butir soal nomor 3, 5 dan 7 dengan nilai p : 0.09 dan persentase peserta didik yang menjawab benar sebesar: 43% (17 peserta didik); (9) butir soal nomor 16 dan 20 dengan nilai p : 0.09 dan persentase peserta didik yang menjawab benar sebesar: 45% (18 peserta didik); (10) butir soal nomor 14 dengan nilai p : 0.11 dan persentase peserta didik yang menjawab benar sebesar: 53% (21 peserta didik); (11) butir soal nomor 8 dengan nilai p sama dengan butir soal nomor 14 yaitu: 0.11 akan tetapi dengan persentase peserta didik yang menjawab benar sebesar: 55% (22 peserta didik); (12) butir soal nomor 1 dan 15 dengan nilai p : 0.13 dan persentase peserta didik yang menjawab benar sebesar: 63% (25 peserta didik); (13) butir soal nomor 2 dengan nilai p : 0.15 dan persentase peserta didik yang menjawab benar sebesar: 75% (30 peserta didik).

Jika disajikan dalam bentuk tabel maka uraian tersebut adalah sebagai berikut:

Tabel 2
Data soal diurutkan berdasarkan tingkat kesulitannya

No	no. soal	N	Sm	$\sum x$	Persentase	p	Kriteria
1	4	40	5	4	10%	0.02	sulit
	12	40	5	4	10%	0.02	sulit
2	18	40	5	8	20%	0.04	sulit
3	6	40	5	9	23%	0.05	sulit
	19	40	5	9	23%	0.05	sulit
4	11	40	5	10	25%	0.05	sulit
5	17	40	5	12	30%	0.06	sulit
6	10	40	5	13	33%	0.07	sulit
7	9	40	5	16	40%	0.08	sulit
	13	40	5	16	40%	0.08	sulit
8	3	40	5	17	43%	0.09	sulit
	5	40	5	17	43%	0.09	sulit
	7	40	5	17	43%	0.09	sulit
9	16	40	5	18	45%	0.09	sulit
	20	40	5	18	45%	0.09	sulit
10	14	40	5	21	53%	0.11	sulit

11	8	40	5	22	55%	0.11	sulit
12	1	40	5	25	63%	0.13	sulit
	15	40	5	25	63%	0.13	sulit
13	2	40	5	30	75%	0.15	sulit

KESIMPULAN

Berdasarkan pembahasan di atas dapat disimpulkan bahwa soal mata pelajaran Sejarah Kebudayaan Islam (SKI) untuk kelas 8 MTs pada pembahasan: (1) Jejak Peradaban Dinasti Bani Abbasiyah; dan (2) Kecemerlangan Ilmuwan Muslim Dinasti Bani Abbasiyah, pada Buku Guru yang diterbitkan oleh Kementerian Agama tahun 2015 disajikan dengan proporsi: 100% soal dalam kategori sulit (sukar).

REFERENCES

- Alwi, Idrus. 2010. Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran dan Daya Pembeda. Dalam: *Jurnal Ilmiah faktor Exacta*, Vol. 3, No. 2, Juni 2010.
- Aminah, Nonoh Siti. 2013. Teori Pengukuran dalam Pendidikan. Dalam: *Jurnal dan Materi Pembelajaran Fisika (JPMF)*, Vol. 3, No. 2 2013.
- Arifin, Zaenal. 2012. *Evaluasi Pembelajaran*. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama
- Arikunto, Suharsimi & Jabar, Cipi Safrudin. 2014. *Evaluasi Program Pendidikan, Pedoman Teoritis bagi Mahasiswa dan Praktisi Pendidikan*. Jakarta: Bumi Aksara.
- Chairawati, Fajri. 2014. Evaluasi Pembelajaran pada Kelas Internasional Fakultas Dakwah IAIN Ar-Raniry. Dalam: *Jurnal Al-Bayan*. Vol. 20, No. 29. Januari-Juni 2014.

- Ibrahim, R. & Ali, M. 2007. Teori Evaluasi Pendidikan. Dalam: Ali, M., Ibrahim, R., Sukmadinata, N.D., Sudjana, D., & Rasjidin, W. (ed). *Ilmu dan Aplikasi Pendidikan*. Bandung: PT. Imperial Bhakti Utama
- Ihsan, Hamdani & Ihsan, Fuad. 2008. *Filsafat Pendidikan Islam*. Bandung: Pustaka Setia
- Kadir, Abdul. 2015. Menyusun dan Menganalisis Tes Hasil Belajar. Dalam: *Jurnal At-Ta'dib*, Vol.: 8, No. 2
- Kartowagiran, Badrun. 2009. *Pengantar Teori Tes Klasik*. Yogyakarta: PASCARJANA UNY & Dinas Pendidikan Propinsi Yogyakarta.
- Lababa, Junaidi. 2008. Analisis Butir Sola dengan Teori Tes Klasik: Sebuah Pengantar. Dalam: *IQRA'*, Vol.: 5
- Nuriyah, Nunung. 2014. Evaluasi Pembelajaran: Sebuah Kajian Teori. Dalam: *Jurnal Edueksos*, Vol.: 03 No 1
- Rasyid, Harun & Mansur. 2008. *Penilaian Hasil Belajar*. Bandung: CV. Wacana Prima.
- Ritzer, George (ed). 2001. *Handbook Teori Sosial*. Jakarta: Nusamedia
- Siskandar. 2016. Evaluasi Implementasi Kurikulum 2013 di Madrasah Aliyah. Dalam: *Cendekia*, Vol: 10, No.: 2.
- Suwarto. 2011. Teori Tes Klasik & Teori Tes Modern. Dalam: *WIDYATAMA*, Vol, 20, No. 1.
- Syamsudin. 2012. Pengukuran Daya Pembeda, Taraf Kesukaran, dan Pola Jawaban Tes (Analisis Butir Soal). Dalam: *Jurnal At-Tajdid*, Vol: 1, No.: 2.
- Wening, Sri. 2012. *Materi Evaluasi Pembelajaran*. Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta.