

KUALITAS BUTIR SOAL PENILAIAN AKHIR SEMESTER (PAS) YANG DISUSUN GURU MADRASAH

¹Eva Dwi Kumala Sari, ²Siti Sarah, ³Wathroh Mursyidi

¹STIT Al-Marhalah Al-Ulya, Bekasi

²STIT Al-Marhalah Al-Ulya, Bekasi

³STIT Al-Marhalah Al-Ulya, Bekasi

Email: kumalasarieva@gmail.com

Abstract

This study aims to analyze the quality of multiple choice questions at the Final Evaluation of Class IX Jurisprudence Subjects in MTs Negeri 3 Bekasi City. The method in this study uses quantitative descriptive methods. This study uses an evaluative research design carried out in the scope of class IX MTs, located in MTs Negeri 3 Bekasi City. The subjects of this study include the entire population, namely all multiple choice items at the end of semester assessment of fiqh subjects in class IX in MTs Negeri 3 Bekasi City in 2018/2019 school year. Data obtained using the multiple-choice item questionnaire technique used as an instrument for data collection to examine the items viewed from the aspects of material, construction, and language. The collected data is analyzed quantitatively. Quantitative analysis was carried out using the help of SPSS and Iteman computer programs, to determine the validity, Alpha-Cronbach reliability and item analysis which included Difficulty Index (ITK), Difference Power Index (IDB), and the effectiveness of the distractor. The analysis shows that: 1) multiple choice questions of class IX are sufficient to meet the validity standard of 65% which is declared valid; 2) the reliability index of multiple choice questions for class IX is high, that is 0.745; 3) The difficulty level of class IX multiple choice questions is quite good, which is 65% which is included in the criteria interval; 4) The difference in power of multiple choice questions in class IX is very good, that is, 90% who have decent power; 5) the effectiveness of class IX multiple choice questions is quite good at 60% with all the distractors already effective.

Keywords: *Item Analysis, Final Evaluation Semester, SPSS, ITEMAN*

1. Pendahuluan

Selama ini kegiatan analisis butir soal jarang dilakukan, tepatnya tidak diketahui secara pasti karena memang jarang dilakukan penelitian, sehingga belum dapat dipastikan apakah soal UASBN yang dibuat BSNP maupun oleh pemerintah daerah (guru) sudah memenuhi syarat-syarat tes yang baik

atau belum. Faktor kualitas tes yang belum diketahui, akan berpengaruh terhadap kemampuan siswa dalam mengerjakan tes. Kelemahan ini akan berdampak pada sulitnya menentukan kemampuan siswa yang sebenarnya. Dampak lainnya adalah ketidaktepatan dalam pengambilan keputusan berkenaan dengan hasil belajar siswa

dan pemetaan mutu pendidikan di sekolah dasar.

Hal inilah yang sering menjadi isu dikalangan para orang tua karena kecewa dengan nilai atau hasil belajar yang diperoleh oleh anak-anak mereka. Belum lagi sekarang ini kurikulum yang digunakan disekolah terutama kelas 1 dan 4 SD menggunakan tematik. Model pembelajaran tematik adalah model pendekatan yang dilaksanakan dengan mengintegrasikan berbagai materi ajar dengan karakteristik dan aspek materi yang saling berkaitan didalam satu kegiatan pembelajaran yang tersusun secara terencana dan sistematis. Sehingga para tua terkadang merasa kesulitan dalam memberikan pelajaran dirumah. Soal yang sering didapat siswa terkadang berbeda jauh dengan apa yang mereka pelajari di sekolah. Untuk itulah kualitas tes UASBN perlu mendapat perhatian yang serius untuk dapat meningkatkan mutu pendidikan, dengan kualitas soal yang baik diharapkan akan mampu mengambil keputusan yang bijak tentang hasil belajar siswa dan tidak mengecewakan para orang tua.

Pendampingan yang harus dilakukan oleh pihak-pihak yang berkompeten di bidang pengembangan instrumen dan analisa butir soal sangat dibutuhkan di tiap dinas pendidikan yang membuat soal UASBN. Dalam MGMP biasanya para guru berkumpul dari tiap perwakilan sekolah untuk membuat soal UASBN, karena porsi yang diberikan pusat untuk daerah sebesar 75%, hal inilah yang harus menjadi perhatian serius dalam

pendampingan pembuatan soal UASBN. Kesalahan sedikit saja dalam mengukur kemampuan atau kompetensi siswa maka akan sangat besar dampaknya bagi siswa dan bagi orang tua tentunya.

Evaluasi merupakan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja atau produktifitas suatu lembaga dalam melaksanakan programnya (Mardapi, 2012). Oleh karena itu, evaluasi merupakan salah satu subsistem yang penting dalam sistem pendidikan. Dalam Undang- Undang Nomor 20 Tahun 2003 tentang sistem Pendidikan Nasional, evaluasi diatur dalam Bab XVI pasal 57, 58, dan 59. Pelaksanaan evaluasi bertujuan untuk mengukur dan mengendalikan mutu pendidikan.

Penjabaran lebih lanjut tentang pelaksanaan evaluasi dinyatakan dalam Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan. Pada pasal 63 ayat (1) menyebutkan bahwa penilaian pendidikan pada jenjang pendidikan dasar dan menengah terdiri atas: (a) penilaian hasil belajar oleh pendidik, (b) penilaian hasil belajar oleh satuan pendidikan, dan (c) penilaian hasil belajar oleh pemerintah. Penilaian hasil belajar oleh pendidik dan satuan pendidikan merupakan bentuk evaluasi internal (*internal evaluation*). Berknaan dengan evaluasi oleh pemerintah, sifatnya sebagai evaluasi eksternal (*external evaluation*) yang sasarannya adalah peserta didik, termasuk didalamnya hasil belajar peserta didik.

Penilaian hasil belajar peserta didik pada jenjang Sekolah Dasar (SD), pada tahun pelajaran 2016/2017 berdasarkan keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor 39 Tahun 2007 ditetapkan sistem penilaiannya dalam bentuk Ujian Akhir Sekolah Berstandar Nasional (UASBN). Khusus penyelenggara Ujian Akhir Sekolah Berstandar Nasional (UASBN) tahun pelajaran 2016/2017, komposisi butir soal, sebanyak 25% ditetapkan oleh Badan Standar Nasional Pendidikan (BSNP) dan berlaku secara Nasional, serta 75% soal ditetapkan oleh penyelenggara UASBN tingkat propinsi (Peraturan Mendiknas, 2007).

UASBN jenjang SD dalam pelaksanaannya berpedoman pada Keputusan Badan Standar Nasional Pendidikan Nomor 983/BSNP/XI/2007 tentang Prosedur Operasi Standar (POS) Ujian Akhir Sekolah Berstandar Nasional untuk Sekolah Dasar/ Madrasah Ibtidaiyah/ Sekolah Dasar Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2016/2017. Bentuk tes yang digunakan diantaranya berupa tes tertulis (*paper and pencil test*). Tes tertulis merupakan teknik penilaian yang seringkali digunakan untuk menilai prestasi belajar siswa. Melalui tes prestasi belajar, dapat diperoleh informasi yang dapat menggambarkan kemampuan siswa (Bauer, 2000). Oleh karena itu, pengelolaan ujian dan mutu bahan ujian yang digunakan perlu mendapat perhatian agar hasil tes dapat mencerminkan kemampuan siswa yang sebenarnya.

Secara sederhana (Allen & Yen, 1979) menyebutkan tes sebagai "*a test is device for obtaining a sample of individual's behavior*". Hal yang hampir sama juga dikemukakan oleh (Gronlund, 1982) yang mendefinisikan tes sebagai instrumen atau prosedur sistematis untuk mengukur perilaku sampel. Ahli pengukuran yang lain, (Djaali, 2006) menyatakan tes adalah suatu cara atau alat 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.

Klasifikasi tes dikemukakan oleh beberapa pakar, diantaranya (Djaali, 2006) yang membedakan tes menjadi empat yaitu: 1) tes penempatan, 2) tes formatif, 3) tes diagnostik, dan 4) tes sumatif. Tes juga dapat dibedakan menjadi *achievement test* atau tes prestasi belajar dan *learning outcome test* atau tes hasil belajar (Balitbang, 1999). Dalam tes prestasi belajar, yang hendak diukur adalah tingkat kemampuan seorang siswa dalam menguasai bahan pelajaran yang diajarkan kepadanya. Oleh karenanya, kedudukan tes prestasi dalam pengambilan keputusan sangat penting. Meski demikian, menurut (Bauer, 2000), skor yang diperoleh siswa dalam tes prestasi belajar kemungkinan tidak sepenuhnya mencerminkan kualitas pembelajaran di sekolah. Hal ini dikarenakan kemampuan siswa tidak semata dipengaruhi oleh pembelajaran di sekolah, tetapi oleh banyak faktor.

Bentuk soal dalam tes prestasi belajar, secara umum dapat dikelompokkan menjadi dua kategori yaitu: 1) tes uraian, terdiri dari uraian bebas, uraian terbatas atau isian singkat, uraian berstruktur, dan 2) tes objektif, terdiri dari pilihan benar salah, pilihan ganda dan menjodohkan. Yang lebih banyak digunakan biasanya adalah tes pilihan ganda. Bentuk soal pilihan ganda, jawabannya harus dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Penggunaan tes pilihan ganda pada umumnya, dijumpai pada ujian berskala besar/massal karena sifatnya yang obyektif dan mudah penskorannya. Bentuk soal ini juga dianggap pilihan yang tepat untuk ujian akhir dimana bahan pelajaran yang hendak diujikan biasanya cukup banyak. Dilihat dari strukturnya, bentuk soal pilihan ganda terdiri dari pokok soal (*stem*) dan pilihan jawaban (*option*). Pilihan jawaban terdiri atas satu kunci jawaban dan yang lainnya pengecoh (*distractor*). Pokok soal (*stem*) dapat berupa pertanyaan atau pernyataan tidak lengkap *The multiple-choice items consist of two parts: the stem, which presents the problem, and the list of possible answer or options. In the standard form of the item, one of the options is the correct or best answer and the others are foils or distractor. The stem of the item may be presented either as a question or as an incomplete statement* (Thorndike, 2005).

Kualitas tes, termasuk tes pilihan ganda (dikotomi) dapat diungkap melalui analisis butir soal secara teoritis (telaah) dan analisis empiris. Analisis

butir soal secara kualitatif dilakukan untuk menilai butir soal ditinjau dari aspek materi, konstruksi, dan bahasa. Analisis secara kuantitatif menekankan pada analisis karakteristik butir soal secara empiris. Karakteristik butir soal antara lain meliputi indeks kesukaran, daya beda, dan distribusi respons.

Teori-teori di atas lah yang mendukung bahwa evaluasi dalam proses pembelajaran sangat penting dilakukan, dan kualitas tes yang digunakan pun harus sesuai dan terpercaya sehingga bisa menentukan kemampuan siswa yang sebenarnya dan tepat dalam pengambilan keputusan berkenaan dengan hasil belajar siswa dan pemetaan mutu pendidikan di sekolah.

Tetapi realita dilapangan banyak sekali kekurangan dalam membuat kualitas tes penilaian akhir semester (PAS) yang akan menjadi tolak ukur kemampuan siswa. UASBN biasanya dibuat oleh perwakilan para guru dari tiap sekolah, atau yang lebih sering dikenal MGMP. Disanalah mereka biasanya berdiskusi dan membuat soal untuk 75% soal UASBN, yang menjadi permasalahan biasanya adalah didalam MGMP tersebut tidak terdapatnya para ahli yang mengawasi pembuatan soal tersebut, disinilah maka sering terjadi kekeliruan baik secara bahasa maupun keterwakilan dari tiap indikator yang akan diukur.

Dari lima penelitian yang pernah dilakukan tentang kualitas soal diantaranya adalah penelitian yang ditulis oleh Atik Fitriatun Prodi Pendidikan Akuntansi Universitas

Negeri Yogyakarta Pada Tahun 2012/2013 dengan judul “Analisis Validitas, Reliabilitas, dan Butir Soal Latihan Ujian Nasional Ekonomi Akuntansi di MAN Maguwoharjo” (Fitriatun, 2016). Dari penelitian ini dapat disimpulkan bahwa hasil penelitiannya menunjukkan bahwa jika dilihat dari segi validitas empiris kriteria konsistensi internal merupakan soal yang kurang baik karena 23 butir soal (57,5%) termasuk kategori soal yang memiliki validitas rendah dan 13 butir soal (32,5%) kategori sangat rendah. Kemudian jika dilihat dari segi reliabilitas maka soal tersebut termasuk soal yang memiliki reliabilitas rendah yaitu 0,546. Lalu jika dilihat dari segi kesukaran soal maka soal tersebut merupakan soal yang cukup baik karena 21 butir soal (52,5%) termasuk kategori soal sedang dan perbandingan antar tingkat kesukaran soal mendekati proporsional. Selanjutnya dari segi daya pembeda maka soal tersebut merupakan soal yang cukup baik karena 1 butir soal (2,5%) termasuk kategori baik sekali, 11 butir soal (27,5%) termasuk kategori baik dan 16 butir soal (40%) termasuk kategori cukup. Selanjutnya jika dilihat dari fungsi pengecoh maka soal tersebut merupakan soal yang baik karena 10 butir soal (25%) memiliki pengecoh yang berfungsi sangat baik, 15 butir soal (37,5%) berfungsi baik dan 12 butir soal (30%) berfungsi cukup.

Selanjutnya penelitian yang dilakukan oleh Yolanda Putri Novyitasari Fakultas Bahasa dan Seni di Universitas Negeri Yogyakarta pada

Tahun 2013/2014 dengan judul “Analisis Butir Soal Pilihan Ganda Menggunakan Teori Pengukuran Klasik Pada Ulangan Umum Akhir Semester Genap Bahasa Prancis SMA Negeri 9 Yogyakarta” (Novyitasari, 2014). Dari penelitian ini dapat disimpulkan hasil dari penelitian menunjukkan bahwa soal pilihan ganda kelas X dan XI memenuhi validitas isi; kemudian indeks reliabilitas soal pilihan ganda kelas X tergolong sedang yakni sebesar 0,601, sedangkan indeks reliabilitas soal pilihan ganda kelas XI tergolong tinggi, yakni sebesar 0,795; kemudian ITK soal pilihan ganda kelas X tergolong tidak baik, yakni hanya sejumlah 17 butir soal (42,5%) yang masuk dalam interval kriteria. ITK soal pilihan ganda kelas XI juga tergolong tidak baik, yakni sejumlah 12 butir soal (30%) yang masuk dalam interval kriteria; selanjutnya IDB soal pilihan ganda kelas X tergolong kurang baik, yakni hanya 24 butir soal (60%) yang memiliki IDB layak. IDB soal pilihan ganda kelas XI tergolong kurang baik, yakni sejumlah 26 butir soal (65%) yang memiliki IDB layak; dan efektivitas distraktor pada soal pilihan ganda kelas X tergolong tidak baik, karena butir soal yang seluruh distraktornya efektif adalah sebanyak 9 butir soal (22,5%), sedangkan efektivitas distraktor pada soal pilihan ganda kelas XI tergolong tidak baik, karena butir soal yang seluruh distraktornya efektif adalah sebanyak 4 butir soal (10%); 6) berdasarkan kelayakan butir soal dilihat dari ITK dan IDB-nya, kualitas soal pilihan ganda kelas X dan pada kelas XI

tergolong tidak baik, karena keduanya hanya terdapat 12 butir soal (30%) yang dinyatakan layak.

Penelitian yang dilakukan oleh Galuh Werdiningsih Fakultas Ekonomi di Universitas Negeri Yogyakarta Tahun 2014/2015 dengan judul "Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Ekonomi Kelas XII IPS SMAN 2 Banguntapan" (Werdiningsih, 2015). Dari penelitian ini dapat disimpulkan hasil dari penelitian menunjukkan bahwa kualitas soal berdasarkan validitas, soal yang valid berjumlah 31 soal atau 77,5% sedangkan soal yang tidak valid berjumlah 9 soal atau 22,5%; kemudian kualitas soal berdasarkan reliabilitas, dapat disimpulkan butir soal Ulangan Akhir Semester gasal mata pelajaran Ekonomi kelas XII IPS SMAN 2 Banguntapan tahun ajaran 2014/2015 memiliki koefisien reliabilitas yang rendah yaitu hanya sebesar 0,6776; selanjutnya kualitas soal berdasarkan daya pembeda, butir soal yang memiliki daya pembeda jelek berjumlah 22 butir atau sebesar 55%, butir soal yang memiliki daya pembeda cukup berjumlah 15 butir atau sebanyak 37,5%, butir soal yang memiliki daya pembeda baik berjumlah 1 butir atau sebesar 2,5% dan butir soal yang memiliki daya pembeda negatif berjumlah 2 butir soal atau sebesar 5%; dan kualitas soal berdasarkan tingkat kesukaran, butir yang tergolong sukar berjumlah 5 butir atau 12,5%, butir soal yang tergolong sedang berjumlah 22 butir atau 55%, dan butir soal yang tergolong mudah berjumlah 13 butir atau 32,5%; dan

kualitas soal berdasarkan analisis efektivitas pengecoh, terdapat 1 (2,5%) butir soal berkualitas sangat baik, 2 (5%) butir soal berkualitas baik, 15 (37,5%) butir soal berkualitas cukup, sebanyak 11 (27,5%) butir soal berkualitas kurang baik dan 11(27,5%) berkualitas tidak baik; (6) Kualitas soal berdasarkan analisis butir soal secara bersama-sama, soal yang berkualitas sangat baik berjumlah 6 butir soal atau 15%, soal yang berkualitas baik berjumlah 11 butir atau 27,5%, soal yang berkualitas sedang berjumlah 10 butir atau 25%, soal yang berjumlah tidak baik berjumlah 11 butir atau 27,5%, dan soal yang berkualitas sangat tidak baik berjumlah 2 butir atau 5%.

Penelitian keempat yang dilakukan oleh Afnia Haryanti dari Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta dengan judul "Validitas dan Reliabilitas Butir Soal Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Mata Pelajaran Bahasa Arab Madrasah Aliyah Program Keagamaan Daerah Istimewa Yogyakarta Tahun Ajaran 2012/2013" (Haryanti, 2012) Dari penelitian yang dilakukan oleh Afnia Haryanti dapat disimpulkan bahwa: 1) Validitas isi butir-butir soal UAMBN mata pelajaran bahasa arab MA Program Keagamaan DIY tahun ajaran 2012/2013 terbukti kurang valid berdasarkan kesesuaiannya dengan Standar Kompetensi dan Kompetensi Dasar. 2) Validitas item butir-butir soal UAMBN mata pelajaran bahasa arab MA Program Keagamaan tahun 2012/2013 yang tidak valid ada 9 butir soal yaitu

butir soal 1, 2, 7, 15, 26, 28, 30, 35 dan 44 dan yang valid adalah sisanya berjumlah 41. 3) Reliabilitas soal UAMBN mata pelajaran bahasa arab MA Program Keagamaan tahun ajaran 2012/2013 DIY terbukti sangat tinggi dengan nilai 0,910.

Penelitian selanjutnya yang dilakukan oleh Desy Respitarini dari Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta dengan judul "Analisis Kualitas Soal Ulangan Akhir Semester Gasal Mata Pelajaran Pendidikan Agama Islam kelas X SMK Negeri 1 Purwosari Gunungkidul" (Respitarini, 2013) Hasil dari penelitiannya yaitu: 1) Soal pilihan ganda yang tergolong valid berjumlah 15 soal atau 37% dari jumlah keseluruhan dan soal yang tidak valid ada 25 butir soal atau 62.5. Sedangkan untuk bentuk soal uraian sudah bisa dikatakan baik karena seluruh soal atau 100% soal dari jumlah keseluruhan dapat dikatakan valid. 2) Tingkat reliabilitas soal pilihan ganda sebesar 0.60 dan soal uraian sebesar 0.61 ini berarti keseluruhan soal sudah reliabel. 3) Rata-rata tingkat kesukaran soal pilihan ganda tergolong sedang dengan indeks nilai 0.461 begitu pula dengan soal uraian. Hanya saja proporsi untuk setiap tingkat kesukaran soal belum proporsional karena mayoritas soal termasuk kategori sukar. 4) Rata-rata daya pembeda soal pilihan ganda tergolong cukup, dengan hasil Mean Item Tot dan Mean Biserial sebesar 0,254 dan 0,349 begitu pula dengan soal uraian. 5) Sejumlah 109 pengecoh atau 68% dinyatakan efektif dalam

menjalankan fungsinya dengan baik dan hanya terdapat 51 pengecoh atau 31.87% pengecoh yang tidak dapat berfungsi dengan baik.

Dari hasil kelima penelitian diatas maka dapat disimpulkan bahwa; kualitas tes masih dikategorikan kurang baik, hal ini tergambar dari banyaknya butir soal yang tidak memenuhi kriteria. Untuk itulah maka perhatian dari berbagai pihak yang kompeten dalam bidang penelaahan butir soal sangat diperlukan dalam rangka membuat kualitas tes PAS akan menjadi lebih baik. Diharapkan dengan meningkatnya kualitas tes PAS yang dilaksanakan di jenjang pendidikan dasar merupakan awal dari kemajuan mutu pendidikan nasional.

2. Metode Penelitian

Penelitian ini adalah penelitian kuantitatif non eksperimental, yaitu penelitian yang bertujuan untuk menjelaskan fenomena yang sudah ada dengan menggunakan data berupa angka-angka untuk mengetahui karakteristik individu atau kelompok (Arikunto, 2014). Adapun fenomena yang menjadi objek penelitian ini adalah butir-butir soal yang terdapat dalam soal UAS Fikih Madrasah Tsanawiyah Negeri 3 Bekasi tahun ajaran 2018/2019. Populasi yang dijadikan sebagai objek penelitian ini adalah butir-butir soal Fikih dalam UAS Madrasah Tsanawiyah 3 Bekasi tahun ajaran 2018/2019.

Sedangkan dalam penelitian ini tidak menggunakan sampel pada obyek, karena penulis ingin melihat

secara keseluruhan hasil UAS bidang studi Fikih yang berjumlah 40 tes uraian objektif dengan jenis pilihan ganda di Madrasah Tsanawiyah tahun ajaran 2018/2019.

Pada subyek, penulis menggunakan teknik *Non Probability Sampling* yaitu pada bagian sampling kuota dimana dalam pengambilan sampelnya, penulis memilih sejumlah responden dari wilayah tertentu sampai batas data yang diinginkan terpenuhi (Arikunto, 2014). Oleh karena itu, penulis mengambil semua subyek-subyek di dalam populasi karena penulis membutuhkan semua responden kelas IX untuk dianalisis, sehingga semua obyek memiliki hak yang sama kepada subyek untuk dipilih menjadi sampel. Sampel dalam penelitian ini terdiri dari 200 siswa yang dipilih dengan *simple random sampling*.

3. Hasil

Data penelitian ini diambil pada Penilaian Akhir Semester mata pelajaran Fikih kelas IX tahun ajaran 2018/2019 di MTs Negeri 3 Kota Bekasi pada Kamis, 6 Desember 2018. Masing-masing soal terdiri dari 40 soal pilihan ganda dan 5 soal uraian. Namun karena penelitian ini difokuskan pada soal-soal pilihan ganda, maka data yang diambil hanya berupa butir soal pilihan gandanya saja. Tabel 7 menunjukkan informasi mengenai jumlah responden yang dianalisis, yaitu berjumlah 200 responden. Kemudian tabel 9

menunjukkan bahwa soal tersebut memiliki kadar reliabilitas yang tinggi.

Tabel 8 Jumlah Subjek Penelitian

Case Processing Summary			
		N	%
Cases	Valid	200	100.0
	Excluded ^a	0	.0
	Total	200	100.0

a. Listwise deletion based on all variables in the procedure.

Tabel 9 Hasil Tes Reliabilitas

Cronbach's Alpha	N of Items
.745	40

Selanjutnya melihat validitas item pada tabel Item-Total Statistics dan pada kolom "Corrected Item-Total Correlation" menginformasikan bahwa apakah butir soal tersebut valid atau tidak. Caranya adalah dengan membandingkan nilai dalam kolom "Corrected Item-Total Correlation" dengan nilai 0,2. Kriteria nilai 0,2 ini didasarkan pada pendapat yang dikemukakan oleh Crocker dan Algina, Nunnally, Aiken, Mehrens and Lehman, dan Naga (Nisfiannoor, 2009). Butir soal dinyatakan valid apabila nilai pada kolom "Corrected Item-Total Correlation" lebih besar daripada 0,2.

Dari 40 butir soal yang dianalisis, terdapat 14 butir soal (35%) yang dinyatakan tidak valid karena nilai pada kolom "Corrected Item-Total Correlation" lebih kecil daripada 0,2. Diantaranya yaitu butir 1, 4, 5, 9, 10, 12, 22, 24, 25, 27, 30, 34, 35, dan 36. Selebihnya terdapat 26 butir soal (65%) adalah butir soal yang dinyatakan valid. Hasil uji validitas dapat dilihat pada tabel 10.

Tabel 10 Hasil Uji Validitas Butir Soal

Butir	Jumlah	Persentase	Ket
2, 3, 6, 7, 8, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 26, 28, 29, 31, 32, 33, 37, 38, 39, 49	26	65%	Valid
1, 4, 5, 9, 10, 12, 22, 24, 25, 27, 30, 34, 35, 36	14	35%	Drop
Total	40	100%	

Butir soal yang tingkat kesukarannya tergolong sulit adalah butir soal yang indeksinya berkisar antara 0,20 – 0,40. Sedangkan butir soal yang tergolong sedang adalah butir soal yang indeksinya berkisar antara 0,41 – 0,60. Dan butir soal yang tergolong mudah adalah butir soal yang indeksinya antara 0,61 – 0,80. Dengan demikian, butir soal yang indeksinya

berkisar antara 0,00 – 0,19 dianggap tidak layak karena butir soal tersebut terlalu sulit, dan butir soal yang indeksinya berkisar antara 0,81 – 1,00 dianggap tidak layak juga karena butir soal tersebut terlalu mudah.

Tabel 11 merupakan tabel hasil analisis butir soal pilihan ganda jika dilihat dari indeks tingkat kesukarannya.

Tabel 11 Indeks Tingkat Kesukaran Butir Soal

Kategori	Butir	Jumlah	Persentase	Keterangan
Sangat sulit (0,00 – 0,19)	-	0	0%	-
Sulit (0,20 – 0,40)	23, 30, 33, 36,	4	10%	diterima
Sedang (0,41 – 0,60)	6, 8, 9, 11, 12, 25, 27, 29, 39, 40	10	25%	diterima
Mudah (0,61 – 0,80)	7, 10, 14, 15, 17, 18, 20, 24, 28, 31, 32, 35	12	30%	diterima
Sangat mudah (0,81 – 1,00)	1, 2, 3, 4, 5, 13, 16, 19, 21, 22, 26, 34, 37, 38	14	35%	ditolak
Total		40	100%	

Berdasarkan pada tabel 11, dapat diketahui bahwa soal pilihan ganda yang berkategori baik ditinjau dari ITK-nya berjumlah 26 butir soal (65%), dengan rincian bahwa butir soal nomor 23, 30, 33 dan 36 berkategori sulit, butir soal nomor 6, 8, 9, 11, 12, 25, 27, 29, 39 dan 40 berkategori sedang, dan butir soal nomor 7, 10, 14, 15, 17, 18, 20, 24, 28, 31, 32 dan 35 berkategori mudah. Selanjutnya, butir soal yang berkategori tidak baik berjumlah 14 butir soal (35%),

butir soal tersebut tergolong sangat mudah.

Analisis selanjutnya merupakan analisis Indeks Daya Beda (IDB). Besarnya IDB berkisar antara 0,00 sampai 1,00. Butir soal yang memiliki IDB sebesar 0,20 sudah dapat dinyatakan layak. Dengan demikian, butir soal dengan IDB kurang dari 0,20 berarti tidak layak karena butir soal tersebut tidak mampu membedakan kemampuan peserta tes kelompok tinggi dan kelompok rendah.

Tabel 12 Indeks Daya Beda

Kategori	Butir	Jumlah	Persentase	Keterangan
0,00 – 0,19	4, 22, 25, 30	4	10%	Soal ditolak
0,20 – 0,29	1, 5, 9, 10, 12, 27, 33, 35, 36	9	22,5%	Soal diperbaiki
0,30 – 0,39	2, 6, 7, 13, 14, 15, 16, 17, 19, 21, 23, 24, 26, 28, 31, 32, 34, 38, 40	19	47,5%	Soal diterima dengan perbaikan
0,40 – 1,00	3, 8, 11, 18, 20, 29, 37, 39	8	20%	Soal baik
Total		40	100%	

Berdasarkan tabel IDB pada tabel 12, dapat diketahui bahwa soal pilihan ganda yang berkategori baik ditinjau dari IDB-nya berjumlah 36 butir soal (90%), yang berarti bahwa butir-butir soal tersebut mampu membedakan kemampuan antara peserta tes kelompok tinggi dan kelompok rendah dengan baik. Butir-butir soal tersebut yakni butir soal nomor 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 34, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39 dan 40.

Selanjutnya, butir soal yang berkategori tidak baik berjumlah 4 butir (10%), yakni butir soal nomor 4, 22, 25

dan 30, yang berarti bahwa butir-butir soal tersebut tidak mampu membedakan kemampuan antara peserta tes kelompok tinggi dan kelompok rendah dengan baik.

Distraktor dianggap tidak efektif jika dipilih oleh kurang dari 0,025 (2,5%) peserta tes atau *point biser* bernilai positif, dan sebaliknya, distraktor dianggap efektif jika dipilih oleh minimal 0,025 (2,5%) dari peserta tes atau *point biser* bernilai negative (Rustam, 2018).

Berikut ini merupakan pengelompokan butir soal berdasarkan efektifitas distraktor (pengecoh).

Tabel 13 Efektifitas Distraktor

Butir soal	Distraktor		Kategori
	Jumlah	Persentase	
1, 3, 6, 7, 8, 9, 11, 12, 14, 15, 17, 18, 20, 24, 25, 28, 29, 30, 31, 33, 35, 36, 39, 40.	24	60%	Layak
2, 4, 5, 10, 13, 16, 19, 21, 22, 23, 26, 27, 32, 34, 37, 38.	16	40%	Direvisi
-	-	-	Dibuang
Total	40	100%	

Berdasarkan tabel diatas, diketahui bahwa ada 24 butir soal yang seluruh distraktornya sudah efektif, yakni nomor 1, 3, 6, 7, 8, 9, 11, 12, 14, 15, 17, 18, 20, 24, 25, 28, 29, 30, 31, 33, 35, 36, 39, dan 40. Selanjutnya, terdapat 16 butir soal yang memiliki distraktor yang tidak efektif, yakni butir soal nomor 2, 4, 5, 10,

13, 16, 19, 21, 22, 23, 26, 27, 32, 34, 37, dan 38. Selanjutnya, tidak terdapat satu butir soal pun yang seluruh distraktornya tidak efektif. Dengan demikian, 40% butir soal harus direvisi distraktornya, sedangkan 60% sisanya sudah baik dan tidak perlu direvisi distraktornya.

4. Kesimpulan

Hasil penelitian yang telah dilakukan pada soal Penilaian Akhir Semester ganjil berbentuk pilihan ganda pada mata pelajaran Fiqih kelas IX di MTs Negeri 3 Kota Bekasi tahun ajaran 2018/2019 adalah sebagai berikut:

Sesuai dengan kriteria butir soal yang baik dari segi validitas pada soal pilihan ganda kelas IX, terdapat 14 butir soal (35%) yang dinyatakan tidak valid, yaitu butir soal nomor 1, 4, 5, 9, 10, 12, 22, 24, 25, 27, 30, 34, 35, dan 36. Selebihnya terdapat 26 butir soal (65%) adalah butir soal yang dinyatakan valid, yakni butir soal nomor 2, 3, 6, 7, 8, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 26, 28, 29, 31, 32, 33, 37, 38, 39, dan 49. Dengan demikian, validitas soal pilihan ganda kelas IX di MTs Negeri 3 Kota Bekasi pada mata pelajaran Fiqih tergolong cukup baik karena 65% soal tersebut dinyatakan valid dan 35% soal tersebut dinyatakan drop.

Reliabilitas Alpha-Cronbach soal kelas IX tergolong tinggi, yakni sebesar 0,745. Kadar reliabilitas ini dipengaruhi oleh beberapa hal, yakni banyak sedikitnya butir soal, tingkat kesukaran butir soal serta banyak sedikitnya responden yang menjawab soal. Kadar reliabilitas ini hanya menunjuk kekonsistenan soal, bukan pada kesahihannya. Soal yang memiliki reliabilitas yang tinggi belum tentu kadar validitasnya juga tinggi. Artinya, alat ukur yang ajek membedakan kemampuan peserta tes belum tentu merupakan alat ukur yang benar-benar tepat dalam mengukur kemampuan peserta tes tersebut.

Ditinjau dari ITK-nya, soal pilihan ganda kelas IX tergolong cukup baik, karena soal yang memenuhi kelayakan dari segi ITK-nya adalah sebanyak 65%, selebihnya sebanyak 35% soal pilihan ganda kelas IX dinyatakan tidak baik.

Ditinjau dari IDB-nya soal pilihan ganda kelas IX tergolong sangat baik, karena soal yang memenuhi kelayakan dari segi IDB adalah sebanyak 36 butir soal (90%), selebihnya sebanyak 4 butir soal (10%) soal pilihan ganda kelas IX dinyatakan soal tersebut ditolak.

Ditinjau dari efektifitas distraktornya, butir soal pilihan ganda kelas IX tergolong cukup baik, karena butir soal yang seluruh distraktornya efektif adalah sebanyak 24 butir (60%), selebihnya sebanyak 14 butir soal (40%) dinyatakan soal tersebut perlu direvisi, karena distraktornya tidak efektif.

Jika dilihat secara keseluruhan berdasarkan Tingkat Kesukaran Soal, Indeks Daya Beda serta Efektifitas Distraktor, maka dapat disimpulkan bahwa soal Penilaian Akhir Semester pada Mata Pelajaran Fiqih Kelas IX yang harus dibuang ialah sebanyak 6 butir (15%), yaitu butir soal nomor 1, 4, 5, 16, 22, dan 33. Sedangkan untuk soal yang perlu direvisi sebanyak 17 butir soal (42,5%), yaitu soal nomor 2, 3, 9, 10, 12, 13, 19, 21, 25, 26, 27, 30, 34, 35, 36, 37, dan 38. Dan untuk soal yang secara keseluruhan dapat diterima dengan baik sebanyak 17 butir soal (42,5), yaitu butir soal nomor 6, 7, 11, 14, 15, 17, 18, 20, 23, 24, 28, 29, 31, 32, 39 dan 40.

Melihat realitas dilapangan bahwa kualitas tes PAS di madrasah khususnya soal-soal keagamaan, yang ditetapkan

oleh penyelenggara tingkat provinsi masih rendah atau kurang baik. Kegiatan peningkatan kemampuan dan keterampilan guru madrasah dalam bentuk pendidikan dan pelatihan analisis butir soal (Diklat), workshop, dan lokakarya sebaiknya diprogramkan oleh pihak dinas provinsi maupun oleh Dinas kabupaten/kota. Kerjasama pihak Dinas Pendidikan Nasional (DIKNAS), LPMP, dan Perguruan

Tinggi yang berkompeten dalam penyediaan bank soal sangat diperlukan. Apabila bank soal belum bisa diadakan oleh daerah, seyogyanya dalam proses penyusunan soal untuk keperluan PAS dimasa yang akan datang dapat melibatkan ahli bidang studi, ahli bahasa, dan ahli pengukuran/ psikometri sebagai tim pendamping/ penelaah untuk menjaga mutu soal yang dihasilkan.

5. Referensi

- Allen, & Yen. (1979). *Introduction to Measurement Theory*. California: Wadsworth Inc.
- Arikunto, S. (2014). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Balitbang, T. P. (1999). *Manual ITEMAN*. Jakarta: PUSISJIAN Balitbang Depdiknas.
- Bauer, S. (2000). Sholud achivment test be used to judge school quality. Education policy analysis archives. 1-18.
- Djaali. (2006). *Hasil Belajar Evaluasi dalam Evaluasi Pendidikan: Konsep dan aplikasi*. Jakarta: Uhamka Press.
- Fitriatun, A. (2016). Analisis Validitas, Reliabilitas, Dan Butir Soal Latihan Ujian Nasional Ekonomi Akutansi di MAN Maguwaharjo. *Jurnal Kajian Pendidikan Akuntansi Indonesia Edisi 8*, 1.
- Gronlund, N. (1982). *Constructing achievement test.*(3rd ed). New York: Prentice Hall, Inc., Englewood Cliffs.
- Haryanti, A. (2012). *Validitas dan Reliabilitas Butir Soal Ujian Akhir Madrasah Berstandar Nasional (UAMBN) Mata Pelajaran Bahasa Arab Madrasah Aliyah Program Keagamaan Dareah Istimewa Yogyakarta Tahun Ajaran 2012/2013*. Retrieved from Universitas Islam Negeri Sunan Kalijaga : <http://digilib.uin-suka.ac.id/>
- Mardapi, D. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan* . Yogyakarta: Nuha Medika.
- Nisfiannoor, M. (2009). *Pendekatan Statistika Modern Untuk Ilmu Sosial*. Jakarta: Salemba Humanika.
- Novitasari, Y. P. (2014). *Analisis Butir Soal Pilihan Ganda Menggunakan Teori Pengukuran Klasik Pada Ulangan Umum Akhir Semester Genap Bahasa Prancis SMA Negeri 9 Yogyakarta*. Retrieved from Universitas Negeri Yogyakarta: <https://eprints.uny.ac.id/>
- Respitarini, D. (2013). *Analisis Kualitas Soal Ulangan Akhir Semester Gasal Mata Pelajaran Pendidikan Agama Islam kelas X SMK Negeri 1 Purwosari Gunungkidul*. Retrieved from Universitas Islam Negeri Sunan Kalijaga : <http://digilib.uin-suka.ac.id/>
- Rustam, d. (2018). *Statistika Pengukuran Pendidikan. Analisis menggunakan SPSS Iteaman dan Lisrel*. Bogor: PT Ilham Sejahtera Persada.

- Thorndike, R. (2005). *Measurement and evaluation in psychology and education* (7th ed). New Jersey: Pearson Education, Inc.
- Werdiningsih, G. (2015). *Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Ekonomi Kelas XII IPS SMAN 2 Banguntapan*. Retrieved from Universitas Negeri Yogyakarta: <https://eprints.uny.ac.id/>