

PENGEMBANGAN INSTRUMEN PENILAIAN TES TERTULIS BENTUK URAIAN UNTUK PEMBELAJARAN PAI BERBASIS MASALAH MATERI FIQH

Mochamad Zaenal Muttaqin¹ & Kusaeri²

Yayasan Pondok Pesantren Khozinul Ulum, Sidoarjo¹, Universitas Islam Negeri Sunan Ampel Surabaya²
myqinz@gmail.com¹, kusaeri@uinsby.ac.id²

Abstrak : Tujuan penelitian ini untuk menghasilkan instrumen penilaian tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah pada materi Fiqh. Tes tertulis yang dikembangkan didesain dengan mengacu pada Taksonomi Bloom edisi revisi. Penelitian ini merupakan penelitian pengembangan yang terdiri dari tujuh langkah, yaitu: (1) menyusun spesifikasi tes (2) menulis soal (3) menelaah soal (4) memperbaiki tes (5) melakukan uji coba (6) menganalisis butir soal (7) menafsirkan hasil uji coba. Uji coba instrumen dilakukan di MTSN 4 Sidoarjo, pemilihan subjek coba dilakukan dengan teknik sampel acak sederhana. Pengujian kualitas instrumen menggunakan bantuan *software excel*. Parameter butir dianalisis menggunakan teknik klasik yang meliputi: tingkat kesulitan dan daya pembeda soal. Validitas isi instrumen diperoleh dari penilaian pakar dengan menggunakan lembar validasi. Reliabilitas tes dianalisis menggunakan persamaan *Flanagan*. Penelitian ini menghasilkan enam butir soal tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah. Hasil validasi menunjukkan bahwa instrumen yang dihasilkan sangat valid dengan rata-rata total validitas sebesar 3,6. Butir-butir tes memiliki parameter tingkat kesulitan pada rentang 0,3 – 0,7 dengan indeks kesulitan terendah adalah 0,53 dan tertinggi adalah 0,70. Daya pembeda berada pada rentang 0,3-0,4 dengan indeks daya pembeda terendah adalah 0,24 dan tertinggi 0,36. Instrumen memiliki reliabilitas yang sangat tinggi yaitu 0,819.

Kata Kunci: *non objektif, pembelajaran berbasis masalah, Taksonomi Bloom.*

PENDAHULUAN

Indonesia merupakan negara dengan penduduk yang mayoritas beragama Islam. Namun, realita yang ada menandakan bahwa nilai-nilai Islam di Indonesia belum teraktualisasikan secara maksimal dalam perilaku kehidupan sehari-hari. Misalnya, tindak kenakalan remaja pada tahun 2015 masih relatif tinggi. Kenakalan remaja tersebut meliputi pencurian, kriminal, dan tawuran. Pelaku tindak kejahatan rata-rata dilakukan remaja dengan usia 17 tahun. Kasus kekerasan seksual juga cukup tinggi (Fauziah, 2016) . Bahkan terus meningkat setiap tahun. pada tahun 2014 terdapat 11 kasus

kekerasan seksual. Pada tahun 2015 peristiwa tersebut meningkat drastis menjadi 29 kasus. Pada tahun 2016 per April sudah tercatat 27 kasus (Willy, 2016). Banyaknya perbuatan-perbuatan yang melanggar norma agama Islam sangat memprihatinkan.

Berangkat dari berbagai realita tersebut maka perlu dilakukan sebuah perbaikan dari berbagai aspek. Perbaikan tersebut dilakukan agar tujuan utama dari pendidikan agama Islam bisa tercapai. Untuk memperbaiki pendidikan agama Islam, diperlukan upaya untuk meningkatkan kualitas pembelajaran. Salah satu aspek yang penting dalam suatu pembelajaran adalah penilaian. Ngalm Purwanto menyatakan bahwa untuk melaksanakan penilaian secara baik dan benar bukanlah hal yang mudah (Kusaeri, 2014).

Penilaian merupakan salah satu komponen penting dalam pembelajaran. Penilaian diharapkan memberikan umpan balik yang objektif terhadap apa yang telah dipelajari oleh peserta didik dan digunakan pula untuk mengetahui efektifitas pembelajaran” (Kusaeri, 2014). Menurut Kunandar penilaian hasil belajar bertujuan untuk mengukur keberhasilan pembelajaran yang dilakukan oleh guru sekaligus mengukur keberhasilan dalam penguasaan yang telah ditentukan (Kunandar, 2014). Menurut Van den Berg, model penilaian akan sangat berpengaruh pada peserta didik (Akbar, 2013). Dari paparan tersebut diketahui bahwa penilaian sangat penting bagi keberhasilan pembelajaran. Dengan penilaian tersebut seorang guru bisa melakukan refleksi dan evaluasi terhadap kualitas pembelajaran yang telah dilakukan. Sehingga dapat diketahui apakah pembelajaran tersebut berhasil atau tidak.

Hal yang berkaitan dengan penilaian telah dirumuskan dalam Permendikbud nomor 23 tahun 2016 tentang standar penilaian pendidikan. Standar Penilaian Pendidikan adalah kriteria mengenai lingkup, tujuan, manfaat, prinsip, mekanisme, prosedur, dan instrumen penilaian hasil belajar peserta didik yang digunakan sebagai dasar dalam penilaian hasil belajar peserta didik pada pendidikan dasar dan pendidikan menengah. Prinsip

penilaian hasil belajar harus professional, terbuka, edukatif, efektif, efisien, dan sesuai konteks sosial budaya, dan pelaporan hasil penilaian peserta didik secara objektif, akuntabel dan informatif (Permendikbud no 23 Tahun 2016).

Untuk mencapai standar proses penilaian yang telah ditetapkan oleh Kemendikbud bukanlah hal yang mudah. Instrumen penilaian yang digunakan dalam proses pembelajaran agama Islam hendaknya berkualitas. Sehingga mutu pembelajaran agama Islam tidak mengecewakan.

Pada dasarnya untuk melakukan sebuah penilaian dapat digunakan dua bentuk instrumen, yaitu tes dan non tes. Instrumen tes meliputi tes tertulis bentuk pilihan dan uraian, sedangkan non tes terdiri dari portofolio, kinerja, proyek, penilaian diri, penilaian jurnal dan tes lisan (Kusaeri, 2014). Dalam penelitian ini difokuskan pada penilaian tes tertulis bentuk uraian non objektif.

Tes tertulis bentuk uraian merupakan seperangkat soal yang berupa tugas, pertanyaan yang menuntut peserta didik untuk mengorganisasikan dan menyatakan jawabannya menurut kata-kata sendiri. Jawaban tersebut dapat berbentuk mengingat kembali, menyusun, mengorganisasikan atau memadukan pengetahuan yang telah dipelajarinya dalam rangkaian kalimat atau kata-kata yang tersusun secara baik (Kunandar, 2014). Sedangkan menurut Asmawi Zaenul dan Noehi Nasution tes tertulis bentuk uraian adalah butir soal yang mengandung pertanyaan yang jawaban dari soal tersebut dilakukan dengan cara mengekspresikan pikiran peserta tes (Zaenul, 2005).

Berdasarkan sistem penskorannya, tes tertulis bentuk uraian diklasifikasikan menjadi dua, yaitu tes tertulis bentuk uraian objektif dan non objektif. Tes objektif memberi pengertian bahwa penskorannya dilakukan secara objektif, karena bentuk soalnya menuntut sekumpulan jawaban dengan pengertian atau konsep tertentu. Sementara bentuk uraian non objektif menuntut jawaban berupa pengertian atau konsep berdasarkan pendapat masing-masing peserta tes, sehingga penskorannya sangat sulit untuk

dilakukan secara objektif. Penskoran untuk tes tertulis bentuk uraian non objektif dinyatakan dalam bentuk rentangan (Kusaeri, 2014).

Eko Putro Widoyoko menambahkan bahwa penskoran tes uraian non objektif dipengaruhi oleh pemberi skor. Jawaban yang sama dapat memiliki skor yang berbeda oleh pemberi skor yang berbeda. Hal ini diakibatkan oleh beberapa hal, di antaranya: (a) Ketidakkonsistenan penilai (b) *Hallo effect* atau kesan guru terhadap peserta didik sebelumnya (c) Pengaruh urutan pemeriksaan dan (d) Pengaruh bentuk tulisan dan bahasa (Widoyoko, 2011).

Pedoman penskoran merupakan panduan atau petunjuk yang menjelaskan tentang: Batasan atau kata-kata kunci untuk melakukan penyekoran terhadap soal-soal bentuk uraian dan kriteria-kriteria jawaban yang digunakan untuk melakukan penyekoran terhadap soal-soal bentuk uraian non-objektif. Dengan pedoman atau rubrik penskoran, guru dapat mengoreksi jawaban peserta didik secara akurat. Pedoman penskoran hendaknya disusun segera setelah perumusan kalimat butir-butir soal untuk menjaga keobjektivitasan dari penilaian yang akan dilakukan (Kunandar, 2014).

Rubrik penskoran diklasifikasikan kedalam dua bentuk, yaitu rubrik penskoran analitik dan holistik. (a) Rubrik penskoran analitik adalah rubrik penskoran dengan cara mengidentifikasi jawaban dari berbagai aspek yang berbeda. Skor untuk masing-masing aspek diletakkan secara terpisah. (b) Rubrik penskoran holistik adalah rubrik penskoran dimana guru hanya memberikan skor tunggal berdasarkan pada keseluruhan jawaban peserta tes (Kusaeri, 2014).

Dalam penskoran analitik Djemari Mardapi menambahkan bahwa penskoran tersebut digunakan untuk soal ujian yang batas jawabannya sudah jelas dan terbatas. Misalnya soal mata pelajaran matematika dan fisika. Namun cara penskoran analitik juga bisa digunakan dalam bidang sosial dengan syarat batas jawabannya jelas dan komponen jawaban diberi skor (Mardapi, Pengukuran, Penilaian dan Evaluasi Pendidikan, 2012).

Materi pelajaran fiqh merupakan materi yang jelas. Sehingga batas jawaban dalam pelajaran fiqh juga jelas. Untuk menjamin keakuratan penskoran terhadap tes yang dilakukan dengan menggunakan pedoman penskoran analitik, karena pedoman penskoran analitik lebih detail bila dibandingkan dengan rubrik penskoran holistik.

Penilaian tes tertulis bentuk uraian non objektif sebaiknya digunakan bersamaan dengan metode pembelajaran yang bersifat merangsang kemampuan nalar peserta didik (Kunandar, 2014). Salah satu model pembelajaran yang memiliki kelebihan untuk merangsang kemampuan bernalar peserta didik adalah model pembelajaran berbasis masalah.

Pembelajaran berbasis masalah atau disingkat dengan PBM adalah suatu model pembelajaran yang melibatkan peserta didik untuk memecahkan suatu masalah. Masalah yang digunakan adalah permasalahan yang ada pada dunia nyata, agar peserta didik mampu untuk belajar cara berpikir kritis dan keterampilan dalam pemecahan masalah, serta untuk memperoleh pengetahuan dan konsep yang esensial dari materi pelajaran (Sudarman, 2007).

Kokom Komalasari menyatakan bahwa pembelajaran berbasis masalah merupakan strategi pembelajaran yang melibatkan peserta didik dalam memecahkan masalah, dengan mengintegrasikan berbagai konsep dan keterampilan dari berbagai disiplin ilmu. Strategi ini meliputi mengumpulkan informasi, dan mempresentasikan penemuan (Komalasari, 2013). Pembelajaran berbasis masalah mampu untuk menunjang pembangunan kecakapan diri sendiri, kolaboratif dan kemampuan berpikir analisis, evaluasi dan mencipta (Amir, 2010).

Untuk menunjang keberhasilan pembelajaran berbasis masalah, hendaknya menggunakan teknik penilaian yang tepat, agar kemampuan peserta didik dapat terukur. Tes tertulis bentuk uraian non objektif dapat digunakan untuk mengukur kemampuan hasil belajar peserta didik pada tingkat menganalisa, mengevaluasi dan mencipta. Atau dalam tingkatan

kemampuan berpikir Taksonomi Bloom kategori C4, C5, C6 (Masidjo, 1995). Karena dalam menjawab tes tertulis bentuk uraian non objektif peserta didik harus memulai dengan pengetahuan faktual yang dimilikinya dan mengorganisasikan fakta pilihannya dalam suatu susunan yang logis.

Kunandar juga menyatakan bahwa tes tertulis bentuk uraian non objektif dapat digunakan untuk mengukur kemampuan hasil belajar peserta didik pada tingkat C4, C5, C6. Karena tes tertulis bentuk uraian non objektif dapat menilai berbagai jenis kemampuan seperti: mengemukakan pendapat, berpikir kritis, berpikir kreatif dan pemecahan masalah (Kunandar, 2014).

Dalam taksonomi Bloom revisi juga diuraikan tentang klasifikasi dimensi pengetahuan dalam empat kategori, yaitu pengetahuan faktual, konseptual, prosedural dan metakognitif (Anderson & Krathwohl, 2001). Pengetahuan faktual berisikan pengetahuan tentang elemen dasar yang harus diketahui siswa untuk mengenal satu disiplin ilmu atau untuk menyelesaikan masalah didalamnya. Pengetahuan ini meliputi Pengetahuan tentang istilah dan pengetahuan tentang rincian dan unsur tertentu.

Agar soal tes tertulis bentuk uraian non objektif terjamin keakuratannya, maka soal harus memenuhi syarat-syarat sebagai berikut: (a) membatasi ruang lingkup dengan memilih materi atau bahan pelajaran yang esensial (b) menggunakan bahasa yang baik dan benar sehingga mudah difahami dengan baik oleh peserta didik (c) jangan mengulang pertanyaan pada materi yang sama (d) tuliskan rubrik penskoran sebelum menulis soal (e) menuliskan skor untuk masing-masing soal (f) rumusan soal harus jelas dan tegas (g) rumusan soal tidak boleh menggunakan kata yang menimbulkan penafsiran ganda (h) memiliki validitas yang tinggi (i) memiliki reliabilitas yang tinggi.

Berdasarkan latar belakang masalah tersebut maka pertanyaan penelitian dalam penelitian ini adalah “ Bagaimanakah desain instrumen penilaian tes tertulis bentuk uraian non objektif yang relevan untuk pembelajaran pendidikan agama Islam berbasis masalah? ”.

METODE PENELITIAN

Guna menghasilkan tes tertulis bentuk uraian non objektif yang berkualitas, maka penelitian ini menggunakan model penelitian pengembangan. Prosedur pengembangan yang digunakan mengacu pada prosedur pengembangan instrumen yang dikemukakan oleh Djemari. Teknik tersebut terdiri dari sembilan langkah (Mardapi, Teknik Penyusunan Instrumen Tes dan Non Tes, 2008). Dalam penelitian ini langkah-langkah pengembangan instrumen dimodifikasi sehingga terdiri dari tujuh langkah yaitu: (1) menyusun spesifikasi tes (2) menulis soal tes (3) menelaah soal tes (4) melakukan uji coba tes (5) memperbaiki tes (6) menganalisis butir soal tes (7) Menafsirkan hasil tes (Mardapi, Teknik Penyusunan Instrumen Tes dan Non Tes, 2008).

Langkah-langkah di atas dibagi menjadi dua tahapan, tahap perancangan dan tahap uji coba. Tahap perancangan mencakup langkah pertama sampai langkah keempat dan tahap uji coba mencakup langkah kelima sampai langkah ketujuh. Langkah-langkah pengembangan instrumen menurut Djemari Mardapi yang tidak digunakan dalam penelitian ini adalah merakit tes dan melaksanakan tes. Merakit tes tidak digunakan karena langkah tersebut dilakukan pada saat memperbaiki tes. Melaksanakan tes tidak digunakan karena langkah tersebut sama dengan tahap uji coba. Dengan asumsi bahwa hasil telaah yang dilakukan para ahli mampu menjamin kualitas instrumen yang dibuat, maka uji coba yang dilakukan dalam penelitian ini cukup sekali yaitu pada langkah kelima.

Dalam kegiatan uji coba ini dibutuhkan subjek coba. Adapun subjek coba dalam penelitian ini adalah peserta didik kelas VIII di MTSN 4 Sidoarjo. kelas VIII di MTSN 4 Sidoarjo terdiri dari 7 kelas yaitu kelas A sampai G. Karena seluruh individu yang menjadi anggota populasi memiliki peluang yang sama untuk dijadikan sampel, maka penentuan sampel dilakukan dengan cara pengambilan acak sederhana (*simple random sampling*) (Sukmadinata, 2004). Sehingga kelas yang terpilih secara acak adalah kelas VIII B. Tempat

pelaksanaan uji coba adalah di MTSN 4 Sidoarjo yang beralamatkan di desa Tlasi Kecamatan Tulangan kabupaten Sidoarjo.

Penelitian ini menggunakan instrumen pengumpulan data berupa lembar validasi dan instrumen penilaian tes tertulis. Lembar validasi berupa format penelaahan instrumen penilaian tes tertulis. Lembar validasi bertujuan untuk mengetahui ketepatan interpretasi instrumen penilaian tes tertulis bentuk uraian non objektif dengan peserta tes. Adapun aspek-aspek yang terkandung dalam lembar validasi tersebut adalah materi, konstruksi dan bahasa. Instrumen penilaian tes berupa uraian non objektif. Instrumen penilaian tes tersebut digunakan saat uji coba.

Analisis data dilakukan dengan cara analisis butir soal. Menurut Kusaeri analisis butir soal merupakan kegiatan yang penting untuk menghasilkan soal tes yang bermutu (Kusaeri, 2014). Menurut Nana Sudjana analisis butir soal adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh seperangkat pertanyaan yang memiliki kualitas yang memadai (Sudjana, 2009).

Dari beberapa definisi yang dikemukakan oleh para ahli, maka Tujuan utama dari kegiatan ini adalah untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek guna menghasilkan soal yang bermutu. Penelitian ini menggunakan analisis butir soal yang dilakukan secara kualitatif dan kuantitatif.

Analisis soal secara kualitatif adalah analisis yang dilakukan dengan cara meninjau segi validitas dalam lembar validasi yang aspeknya meliputi: materi, konstruksi dan bahasa (Sukiman, 2012). Teknik yang digunakan untuk analisis kualitatif dalam penelitian ini adalah teknik panel. Teknik panel adalah suatu teknik menelaah soal berdasarkan kaidah penulisan soal dengan cara beberapa penelaah menelaah soal ditempat terpisah, sehingga menghasilkan perbaikan dan komentar terhadap soal yang ditelaah.

Analisis soal secara kuantitatif adalah proses penelaah butir soal melalui informasi dari jawaban peserta tes guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori klasik. Aspek yang diperhatikan

dalam teori klasik adalah tingkat kesukaran butir, daya pembeda dan penyebaran pilihan jawaban (untuk soal bentuk pilihan ganda). Sehingga dalam penelitian ini menggunakan dua aspek penelaahan soal, yaitu tingkat kesukaran soal dan daya pembeda soal ditambah dengan uji reliabilitas instrumen (Kusaeri & Suprananto, 2012).

HASIL PENELITIAN

Pada tahap perancangan instrumen penilaian tes tertulis bentuk uraian non objektif dilakukan empat langkah yaitu: menyusun spesifikasi tes, menulis soal tes, menelaah soal tes, dan memperbaiki tes. Spesifikasi tes berisi uraian yang menunjukkan keseluruhan karakteristik yang harus dimiliki suatu tes. Dalam penelitian ini tujuan tes yang digunakan adalah tes formatif, tes tersebut bertujuan untuk mengetahui, sudah sejauh manakah peserta didik telah terbentuk setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu (Rosana, 2011).

Kedua, menyusun kisi-kisi tes. Setelah proses penentuan tujuan tes, kegiatan berikutnya adalah menyusun kisi-kisi tes. Kisi-kisi yang disusun berdasarkan materi tentang ibadah puasa dengan KD “Menganalisis ketentuan ibadah puasa”. Kisi-kisi tes disajikan dalam bentuk matriks yang berisi komponen: kompetensi dasar, indikator, teknik penilaian dan bentuk instrumen

Ketiga, Memilih bentuk tes. Spesifikasi tes berfungsi sebagai petunjuk praktis bagi penyusun tes dalam merencanakan isi materi yang akan diujikan, bentuk tes dan panjang

tes. Bentuk tes yang dikembangkan dalam penelitian ini adalah tes tertulis bentuk uraian non objektif.

Setelah penyusunan kisi-kisi soal, langkah selanjutnya adalah menulis butir-butir soal. Kompetensi dasar yang dipilih dalam penelitian ini didasarkan pada Kurikulum 2013, yaitu “Menganalisis ketentuan ibadah puasa.” Setelah itu kompetensi dasar dijabarkan kedalam beberapa indikator sesuai dengan level tujuan pembelajaran yang terdapat dalam perjenjangan Taksonomi Bloom edisi revisi. Indikator dijabarkan dalam Tabel 1 berikut:

Tabel 1. Indikator Pembelajaran.

No	Indikator	level
3.3.1	Menyebutkan pengertian puasa beserta syarat, rukun, sunah, hal-hal yang makruh dan membatalkan puasa.	C1
3.3.2	Menyebutkan ketentuan puasa Ramadhan dan puasa sunnah	C1
3.3.3	Mengelompokkan antara syarat, rukun, dan sunnah puasa.	C2
3.3.4	Membandingkan antara ketentuan puasa Ramadhan dan puasa sunnah.	C4
3.3.5	Menemukan hal-hal yang membatalkan puasa dalam kehidupan sehari-hari.	C5
3.3.6	Memberi penilaian mengenai sempurna dan tidaknya puasa, berdasarkan ketentuan yang ada.	C5
3.3.7	Merumuskan amalan yang perlu dilakukan agar puasa yang dikerjakan lebih baik.	C6

Dalam penulisan butir soal, taksonomi yang diacu adalah Taksonomi Bloom edisi revisi. Taksonomi Bloom edisi revisi terdiri dari enam tingkatan berfikir yaitu C1, C2, C3, C4, C5, dan C6. Selain memperhatikan tingkatan berfikir, Taksonomi Bloom edisi revisi juga tidak bisa dilepaskan dari dimensi pengetahuan yang meliputi pengetahuan faktual, konseptual, prosedural dan metakognitif.

Sehingga level tes yang dibuat adalah sebagai berikut: mengingat pengetahuan faktual, Memahami pengetahuan konseptual, menganalisa pengetahuan prosedural, menganalisa pengetahuan konseptual, mengevaluasi pengetahuan prosedural., menciptakan pengetahuan metakognitif.

Salah satu tahapan untuk menghasilkan tes yang baik adalah melakukan penelaahan tes. Tes yang telah disusun diserahkan pada ahli untuk ditelaah. Penelaahan dilakukan melalui teknik panel yaitu dengan cara beberapa penelaah menelaah tes ditempat terpisah, sehingga menghasilkan perbaikan dan komentar terhadap tes yang ditelaah (Kusaeri, 2014). Kegiatan tersebut melibatkan dua orang pakar pendidikan agama Islam. Kegiatan penelaahan tes secara panel dilakukan pada tanggal 29 agustus 2016 di ruang guru MTSN 4 Sidoarjo, dengan peserta Ahsan Bisri, S,Ag dan Muh. Ali Mashudi, S.Pd.I.

Sebelum menelaah tes, kedua pakar disamakan persepsinya mengenai instrumen yang dikembangkan. Penelaahan butir tes didahului dengan penetapan level tes berdasarkan perjenjangan taksonomi Bloom edisi revisi dan kesesuaian antara instrumen tes tertulis bentuk uraian non objektif dengan pembelajaran agama Islam berbasis masalah. Para pakar setuju dengan model tes uraian yang diajukan. Secara lisan pakar menyatakan bahwa “ tes uraian yang disusun bisa digunakan sebagai alat ukur yang sesuai dengan pembelajaran berbasis masalah.”

Berdasarkan hasil validasi instrumen penilaian tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah, instrumen yang telah disusun dinyatakan valid. Adapun indikatornya adalah rata-rata total validitas dari instrumen yang disusun mencapai angka 3,6. Jika dibandingkan dengan kriteria total kevalidan maka instrumen yang telah disusun berada pada kategori sangat valid.

Setelah instrumen dibuat maka selanjutnya diuji cobakan kemudian dianalisis dengan menggunakan teknik klasik. Menurut Kusaeri dan Suprananto, aspek yang diperhatikan dalam teori klasik adalah tingkat kesukaran soal, daya pembeda dan reliabilitas (Kusaeri & Suprananto, 2012).

Tingkat kesukaran soal adalah peluang menjawab benar suatu soal pada tingkat tertentu, yang besarnya berkisar dari 0 sampai 1. Semakin besar indeks tingkat kesukaran suatu soal maka semakin mudah soal itu. Untuk

mendapatkan hasil tingkat kesukaran soal, dapat dilakukan dengan langkah-langkah sebagai berikut (Kusaeri, 2014): (a), menghitung jumlah skor setiap butir yang diperoleh setiap peserta didik (b) menghitung total skor maksimal setiap butir yang seharusnya diperoleh peserta tes (c) membagi perhitungan langkah pertama dengan langkah kedua.

Instrumen penilaian tes tertulis bentuk uraian non objektif yang dikembangkan didominasi oleh soal dengan tingkat kesukaran mendekati atau sama dengan 0,60. Butir-butir tes yang memiliki tingkat kesukaran mendekati atau sama dengan 0,60 adalah butir tes nomor 2, 3, 4 dan 6. Dengan hasil tersebut maka diketahui bahwa instrumen tes yang dikembangkan berada pada kategori sedang atau diterima.

Daya pembeda soal adalah kemampuan sebuah soal membedakan antara peserta didik yang pandai dan kurang. Untuk dapat menghitung daya pembeda soal dilakukan langkah-langkah sebagai berikut: (Kusaeri, 2014) (a) mengurutkan peserta tes berdasarkan skor dari yang teratas sampai terendah (b) menentukan kelompok atas dengan mengambil sebanyak 27% peserta dengan urutan teratas dan kelompok bawah sebanyak 27% peserta dengan urutan terbawah (c) menghitung tingkat kesukaran soal kelompok atas dan bawah (d) menghitung selisih tingkat kesukaran soal kelompok atas dan bawah.

butir-butir soal memiliki indeks daya pembeda soal yang bervariasi. Indeks daya pembeda terendah adalah soal nomor 1 dengan daya pembeda sebesar 0,24. Indeks daya pembeda tertinggi adalah soal nomor 6 dengan daya pembeda sebesar 0,36.

Reliabilitas soal pada instrumen tes ini menggunakan model belah dua (*split half methods*). Metode belah dua dilaksanakan dengan cara satu kali pelaksanaan tes. Metode ini sering disebut dengan *single test- single trial method*. Metode ini memungkinkan untuk mengestimasi reliabilitas tanpa harus menyelenggarakan tes dua kali. Untuk mengukur tingkat reliabilitas soal yang dikembangkan digunakan persamaan *Flanagan* (Surapranata, 2009).

Menurut Sumarna Surapranata faktor yang mempengaruhi ketidakajegan reliabilitas dengan penggunaan metode belah dua adalah karena dalam merespon tes itu sendiri. Sehingga yang paling berpengaruh pada reliabilitas tes adalah tingkat kesukaran soal. Ada banyak opsi untuk membagi tes menjadi dua bagian (Surapranata, 2009). Dengan mempertimbangkan tingkat kesukaran tes dari masing-masing butir soal, diputuskan bahwa bagian pertama terdiri dari soal nomor 1, 4, 5. Bagian kedua terdiri dari soal nomor 2, 3, 6.

Berdasarkan perhitungan reliabilitas terhadap instrumen yang dikembangkan, didapatkan hasil 0,819 atau dibulatkan menjadi 0,82. Jika dibandingkan dengan tabel interpretasi indeks reliabilitas, maka instrumen yang dikembangkan memiliki indeks reliabilitas yang sangat tinggi yaitu 0,82.

PEMBAHASAN

Penyusunan butir soal tes dalam penelitian ini dimulai dari merumuskan tujuan pembelajaran yang jelas. Tujuan pembelajaran yang jelas akan sangat membantu agar penilaian yang dilakukan benar-benar mengukur apa yang telah diajarkan kepada peserta didik (Kusaeri, 2014). Para ahli pendidikan telah sepakat bahwa untuk merumuskan tujuan pembelajaran yang baik, hendaknya mengacu pada salah-satu klasifikasi (taksonomi) tujuan pembelajaran (Rosana, 2011). Dalam penelitian ini taksonomi tujuan pembelajaran yang digunakan adalah Taksonomi Bloom edisi revisi.

Menurut Brookhart, kemampuan berpikir tingkat tinggi adalah (1) berpikir tingkat tinggi berada pada bagian atas taksonomi kognitif Bloom, meliputi kemampuan menganalisis, mengevaluasi, dan mengkreasikan, (2) tujuan pengajaran di balik taksonomi kognitif yang dapat membekali peserta didik untuk melakukan transfer pengetahuan, (3) mampu berpikir artinya peserta didik mampu menerapkan pengetahuan dan keterampilan yang mereka kembangkan selama bel-ajar pada konteks yang baru (Brookhart, 2010).

Pembelajaran berbasis masalah atau disingkat dengan PBM adalah suatu model pembelajaran yang melibatkan peserta didik untuk memecahkan suatu masalah (Sudarman, 2007). Menurut Taufiq Amir, pembelajaran berbasis masalah mampu untuk menunjang pembangunan kecakapan diri sendiri, kolaboratif dan kemampuan berpikir analisis, evaluasi dan mengkreasi (Amir, 2010). Menurut Masidjo, untuk menunjang keberhasilan pembelajaran berbasis masalah, hendaknya menggunakan teknik penilaian yang tepat, agar kemampuan peserta didik dapat terukur (Masidjo, 1995).

Jadi untuk mengukur tingkat keberhasilan pembelajaran berbasis masalah pada peserta didik, hendaknya digunakan instrumen penilaian yang tepat. Karena model penilaian akan sangat berpengaruh pada peserta didik. Menurut Van den Berg seperti dikutip Sa'dun Akbar kurikulum memiliki potensi yang kaya untuk mengembangkan keterampilan berpikir tingkat tinggi peserta didik (Akbar, 2013).

Tes tertulis bentuk uraian non objektif dapat digunakan untuk mengukur kemampuan hasil belajar peserta didik pada tingkat C4, C5, C6. Karena tes tertulis bentuk uraian non objektif dapat menilai berbagai jenis kemampuan seperti: mengemukakan pendapat, berpikir kritis, berpikir kreatif dan pemecahan masalah (Kunandar, 2014). Sehingga butir soal uraian non objektif dalam penelitian ini didesain berdasarkan perjenjangan Taksonomi Bloom edisi revisi dengan memperhatikan kemampuan menganalisis, mengevaluasi, dan mengkreasi yang dimiliki oleh peserta didik.

Dalam penulisan butir soal, materi awal cukup melibatkan aspek pengetahuan dan pemahaman (C1 dan C2). Karena sangat tidak mungkin untuk menguasai materi pembelajaran yang lebih tinggi tanpa menguasai materi pembelajaran yang lebih rendah. Selanjutnya untuk topik inti disusun berdasarkan level yang lebih kompleks yaitu level C4, C5, C6. Kategori pada Taksonomi Bloom edisi revisi disusun menjadi sebuah hierarki kumulatif. Artinya, penguasaan kategori yang lebih kompleks mensyaratkan penguasaan semua kategori yang dibawahnya (Kusaeri, 2014).

Berdasarkan penelitian yang dilakukan oleh Darodjat, Darmiyati Zuchdi dan Zamroni bahwa butir soal yang disusun mulai dari yang mudah mampu mengurangi rasa panik peserta tes. Sehingga peserta tes mampu merespon butir soal dengan baik. Samritin dalam disertasinya menyatakan, tes yang dirakit dimulai dari butir tes yang mudah dapat mengurangi kecemasan peserta tes (Sumritin, 2014).

Oleh karena itu, dalam penyusunan butir soal diperlukan soal dengan level mengingat dan memahami. Dengan demikian peserta didik mampu untuk mengorganisasikan penguasaan materi tingkat rendah sampai pada penguasaan materi yang lebih kompleks yaitu pada tingkatan berpikir level menganalisa, mengevaluasi dan mengkreasi. Sehingga peserta didik tidak panik dan mampu merespon soal dengan baik.

Untuk menjamin keakuratan soal tes tertulis bentuk uraian non objektif, maka soal harus memenuhi syarat-syarat sebagai berikut: (a) membatasi ruang lingkup dengan memilih materi atau bahan pelajaran yang esensial (b) menggunakan bahasa yang baik dan benar sehingga mudah difahami dengan baik oleh peserta didik (c) jangan mengulang pertanyaan pada materi yang sama (d) tuliskan rubrik penskoran sebelum menulis soal (e) menuliskan skor untuk masing-masing soal (f) rumusan soal harus jelas dan tegas (g) rumusan soal tidak boleh menggunakan kata yang menimbulkan penafsiran ganda (h) memiliki kriteria parameter butir soal yang baik (i) memiliki reliabilitas yang tinggi (Kunandar, 2014).

Pembuatan instrumen tes tertulis bentuk uraian harus disertai pedoman penskoran yang disebut rubrik. Rubrik penskoran yang digunakan dalam penelitian ini adalah rubrik penskoran analitik. Penggunaan rubrik penskoran analitik dimaksudkan agar penyekoran yang dilakukan lebih teliti.

Instrumen penilaian yang telah dihasilkan kemudian direvisi. Revisi terhadap instrumen dilakukan dua kali. Revisi pertama dilakukan berdasarkan saran para pakar. Pada revisi pertama menghasilkan perubahan besar terhadap instrumen. Revisi tersebut terdiri dari penghapusan salah satu

indikator dan revisi konstruksi atau redaksi soal. Revisi kedua dilakukan pada saat melakukan uji keterbacaan. Revisi tersebut hanya menghasilkan perubahan redaksi dari soal.

Setelah soal disusun maka soal diuji cobakan terhadap subjek coba. Kemudian dianalisis berdasarkan kevalidan, tingkat kesukaran, daya pembeda dan reliabilitas soal.

Validitas

Menurut Anas Sudijono, salah satu ciri tes hasil belajar yang baik adalah memiliki validitas. Suatu tes hasil belajar dengan validitas yang tinggi dapat dikatakan handal dan tidak perlu diragukan ketepatan dalam mengukur hasil belajar peserta didik (Sudijono, 2011). *Wainer & Braun* juga berpendapat bahwa tes yang baik harus memiliki karakteristik validitas agar dapat menyajikan informasi yang tepat tentang kondisi siswa yang mengikuti tes (Kusaeri & Suprananto, 2012). Keakuratan soal tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah yang dikembangkan dalam penelitian ini telah terjamin kevalidannya.

Kevalidan instrumen tes yang dikembangkan dibuktikan dengan penilaian para pakar. Berdasarkan penilaian pakar mengenai kesesuaian instrumen dengan level Taksonomi Bloom, dinyatakan bahwa tingkatan level instrumen yang dikembangkan telah sesuai dengan level Taksonomi Bloom edisi revisi.

Berdasarkan hasil penilaian pakar dengan menggunakan lembar validasi, dihasilkan rata-rata total validitas sebesar 3,62. Jika dibandingkan dengan tabel kriteria rata-rata total validitas, maka instrumen yang dikembangkan berada pada kategori sangat valid. Karena instrumen yang dikembangkan dinyatakan valid maka instrumen dapat digunakan untuk tahap selanjutnya, yaitu untuk tahap uji coba dalam mengukur kemampuan peserta didik.

Tingkat Kesukaran Soal

Tingkat kesukaran butir soal adalah peluang menjawab benar suatu soal pada tingkat kemampuan tertentu. Dari hasil analisis tingkat kesukaran soal, didapatkan indeks tingkat kesukaran soal yang rata-rata berada pada rentang 0,3-0,7. Menurut Djemari mardapi Mardapi, butir soal yang baik memiliki kisaran indeks kesulitan 0,3 – 0,7. Butir soal yang memiliki tingkat kesulitan di bawah 0,3 dianggap terlalu sulit dan butir soal yang memiliki tingkat kesulitan di atas 0,7 dianggap terlalu mudah. Kriteria indeks daya beda butir soal yang boleh digunakan adalah $\geq 0,3$ (Mardapi, Teknik Penyusunan Instrumen Tes dan Non Tes, 2008). Dengan demikian, instrumen yang dikembangkan berada pada kategori baik.

Butir soal nomor 1 memiliki tingkat kesukaran soal sebesar 0,70. Ini menandakan bahwa soal tersebut tergolong dalam soal yang mudah. Hal ini dapat dilihat dari banyaknya jumlah peserta yang mampu menjawab soal tersebut. Butir soal nomor 1 tidak direvisi, karena soal dirancang agar mudah untuk dikerjakan. Supaya peserta didik tidak panik saat mengerjakan soal selanjutnya.

Butir soal dengan indeks kesukaran soal terendah adalah butir soal nomor 5 dengan hasil sebesar 0,54. Hal ini menandakan bahwa soal tersebut adalah soal paling sulit diantara soal-soal yang lain. Secara keseluruhan, butir soal yang dikembangkan berada pada kategori sedang yaitu pada rentang 0,60. Rentang 0,60 adalah indeks tingkat kesukaran soal yang sedang tetapi mendekati mudah. Hal ini dipengaruhi oleh kualitas subjek coba. Subjek coba yang digunakan dalam penelitian ini adalah kelas dengan kriteria keagamaan, yang artinya subjek coba menguasai materi pembelajaran. Apabila instrumen diuji cobakan di kelas dengan kriteria selain agama, maka dipastikan tingkat kesukaran soal berada pada kategori sedang atau ideal.

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Menurut Sukiman “Butir soal yang digunakan untuk keperluan ulangan atau ujian semester memiliki tingkat kesukaran yang sedang”. Indeks tingkat

kesukaran butir soal yang baik antara 0,3- 0,7. Jadi dapat disimpulkan bahwa instrumen tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah yang dikembangkan dalam penelitian ini bisa digunakan untuk keperluan tes formatif atau ulangan harian (Sukiman, 2012).

Daya Pembeda Soal

Daya pembeda soal adalah kemampuan butir soal dalam membedakan siswa yang pandai dan siswa yang kurang pandai (Kunandar, 2014). berdasarkan hasil analisis daya pembeda soal, instrumen yang dikembangkan memiliki dua kriteria utama. Soal dengan daya pembeda dibawah 0,3 dan soal diatas 0,3. Soal dengan indeks daya pembeda dibawah 0,3 adalah soal nomor 1 dan 3 dengan indeks daya pembeda 0,24 dan 0,29 atau 33%. Soal dengan indeks daya pembeda diatas 0,3 adalah soal nomor 2, 4, 5, 6 atau 77%.

Tindak lanjut butir soal sesudah dianalisis daya pembedanya sebagai berikut: (Sudijono, 2011) (a) butir soal yang memiliki daya pembeda baik disimpan (b) butir soal dengan daya pembeda rendah, ada dua kemungkinan tidak lanjut yaitu: ditelusuri untuk kemudian diperbaiki dan selanjutnya digunakan kembali dalam tes hasil belajar mendatang guna mengetahui daya pembedanya meningkat atau tidak atau dibuang (c) butir item yang angka indeks diskriminasinya bertanda negatif, sebaiknya dibuang karena kualitas butir soalnya sangat jelek.

Jika ditafsirkan dengan menggunakan tabel kriteria indeks daya pembeda soal, maka soal nomor 1 dan 3 berada pada kategori kurang baik. Soal nomor satu memiliki indeks daya pembeda soal yang paling rendah, hal ini dikarenakan soal nomor 1 memang dirancang agar mudah untuk dikerjakan, sehingga memiliki indeks daya pembeda dibawah 0,3. Namun soal nomor satu tetap dipertahankan karena tujuannya adalah untuk mengurangi rasa cemas peserta tes.

Hasil penelusuran soal nomor tiga yang memiliki indeks daya pembeda soal dibawah 0,3 ditemukan, bahwa ada dua orang peserta tes dari kelompok atas yang mendapatkan skor rendah yaitu 4. Oleh karena itu diputuskan

bahwa soal nomor 3 tetap dipertahankan, karena faktor yang mempengaruhi indeks daya pembeda soal nomor 3 tidak terlalu signifikan.

Jadi disimpulkan bahwa instrumen tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah dilihat dari segi daya pembeda soal memiliki kualitas baik. Sehingga tes berfungsi sebagaimana mestinya, yaitu mampu membedakan antara peserta didik yang pandai dan kurang pandai. Hal ini dibuktikan bahwa rata-rata soal memiliki indeks daya pembeda pada rentang 0,3-0,4 atau memuaskan.

Reliabilitas

Reliabilitas adalah derajat konsistensi pengukur seperangkat soal. Reliabilitas soal pada instrumen tes ini menggunakan model belah dua (*split half methods*). Menurut Sumarna Surapranata metode belah dua dapat mengatasi kelemahan yang terdapat pada metode tes ulang dan tes paralel. Metode ini memungkinkan untuk mengestimasi reliabilitas tanpa harus menyelenggarakan tes dua kali. Dengan demikian beberapa kelemahan seperti *reactivity effect* dan khususnya pengaruh waktu terhadap perolehan skor sebenarnya dapat diminimalisasi. Dengan demikian ketidakajegan perolehan skor bukan karena penyelenggaraan tes tetapi karena dalam merespon tes itu sendiri (Surapranata, 2009).

Berdasarkan perhitungan indeks reliabilitas yang dilakukan dengan menggunakan persamaan *flanagan* didapatkan hasil 0,819 atau dibulatkan menjadi 0,82. Jika dibandingkan dengan tabel interpretasi indeks reliabilitas, maka instrumen yang dikembangkan memiliki indeks reliabilitas yang sangat tinggi yaitu 0,82. Disimpulkan bahwa indeks reliabilitas instrumen yang dikembangkan berada pada kategori sangat tinggi atau instrumen yang dikembangkan ajeg bila digunakan untuk mengukur kemampuan peserta didik.

Ketika berdiskusi dengan guru mata pelajaran Fiqh, dihasilkan fakta bahwa guru-guru setuju untuk menggunakan soal-soal model uraian bentuk uraian dengan mengacu pada perjenjangan Taksonomi Bloom edisi revisi.

Guru-guru beralasan bahwa jika model pembelajarannya sudah bagus namun penilaiannya kurang, maka tidak bisa mengukur kemampuan peserta didik yang sesungguhnya. Menyusun soal berdasarkan perjenjangan Taksonomi Bloom edisi revisi juga tidak begitu sulit, karena sudah banyak panduan untuk menyusunnya. Yang diperlukan hanya latihan dan pembiasaan untuk menyusun soal yang baik sesuai dengan perjenjangan Taksonomi Bloom edisi revisi.

Temuan dilapangan menunjukkan bahwa peserta didik dipaksa berfikir lebih keras untuk menjawab soal-soal. Karena soal-soal yang dikembangkan tidak hanya bersifat mengingat dan memahami materi. Soal-soal juga dirancang agar peserta didik mampu menganalisa, mengevaluasi dan mengkreasikan hal baru dari apa yang telah mereka pelajari. Dengan proses tersebut maka kemampuan peserta didik yang dihasilkan dari dampak pembelajaran berbasis masalah mampu tersalurkan dengan baik dan tidak sia-sia.

PENUTUP

Berdasarkan hasil analisis dalam penelitian ini, diperoleh kesimpulan sebagai berikut: Instrumen tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah, didesain dengan mengacu pada perjenjangan tingkatan berfikir Taksonomi Bloom edisi revisi. Instrumen tes tertulis bentuk uraian non objektif yang dihasilkan dalam penelitian ini terdiri dari satu butir soal level C1, satu butir soal level C2, satu butir soal level C4, dua butir soal level C5, dan satu butir soal level C6. Butir soal dirakit mulai dari soal yang mudah, sehingga peserta didik tidak panik ketika mengerjakan soal.

Instrumen tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah yang dihasilkan dalam penelitian telah memenuhi kriteria soal yang baik. Bukti bahwa instrumen tes yang dikembangkan telah memenuhi kriteria yang baik adalah sebagai berikut: (1) valid berdasarkan penilaian para ahli dengan indeks rata-rata total validitas

sebesar 3,62 (2) memiliki reliabilitas sebesar 0,82 (3) butir-butir tes memiliki parameter tingkat kesulitan dan daya beda pada rentang 0,3-0,7 .

Berdasarkan simpulan yang telah disampaikan maka penelitian ini memiliki beberapa implikasi antara lain : Hasil penelitian ini dapat dijadikan oleh guru untuk menyusun instrumen penilaian tes tertulis bentuk uraian non objektif untuk pembelajaran berbasis masalah, khususnya bagi guru fiqh. Hal ini akan membuat guru menjadi terbiasa untuk menyusun soal-soal yang dapat meningkatkan kemampuan peserta didik dalam memecahkan masalah. Hasil penelitian ini juga memberikan implikasi bagi peserta didik. Yaitu peserta didik lebih berpengalaman untuk mengerjakan soal yang menuntut kemampuan peserta didik untuk memecahkan masalah dalam mata pelajaran fiqh. Dengan demikian penilaian yang dilakukan tidak hanya menuntut kemampuan mengingat peserta didik. Soal-soal tersebut akan membantu peserta didik untuk menguasai materi yang disampaikan secara lebih mendalam. Hasil penelitian ini juga memiliki implikasi bagi pihak sekolah. Salah satu implikasi dari penelitian ini adalah meningkatnya kualitas pembelajaran yang diselenggarakan didalam sekolah tersebut. Sehingga menghasilkan peserta didik yang kritis dan berkualitas.

Saran bagi guru yang ingin menilai kemampuan peserta didik yang melakukan proses pembelajaran dengan metode pembelajaran berbasis masalah. Supaya menggunakan instrumen yang dihasilkan dalam penelitian ini, karena instrumen yang dihasilkan sudah valid dan reliabel. Bagi guru-guru yang ingin mengembangkan instrumen penilaian tes tertulis bentuk uraian non objektif untuk pembelajaran agama Islam berbasis masalah, supaya mengikuti langkah-langkah pengembangan instrumen yang digunakan dalam penelitian ini agar diperoleh instrumen yang baik.

REFERENSI

- Akbar, S. (2013). *Instrumen Perangkat Pembelajaran*. Bandung: PT Remaja Rosda Karya.
- Amir, T. (2010). *Inovasi Pendidikan Melalui Problem Based Learning*. Jakarta: Prenada Media Group.
- Anderson, L., & Krathwohl. (2001). *A Taxonomy for Learning Teaching and Assessing*. New York: Addison Wesley Longman.
- Brookhart. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom*. Alexandria: ASDS.
- Fauziah, L. (2016, November 23). *Kekerasan Anak Justru Banyak Terjadi di Desa*. Retrieved from Metrotvnews: <http://www.metrotvnews.com>
- Komalasari, K. (2013). *Pembelajaran Kontekstual Konsep dan Aplikasi*. Bandung: Revika Aditama.
- Kunandar. (2014). *Penilaian Autentik Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013*. Jakarta: Raja Grafindo Persada.
- Kusaeri, & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.
- Kusaeri, K. (2014). *Acuan dan Teknik proses dan Hasil Belajar dalam Kurikulum 2013*. Yogyakarta: Ar-Ruqq Media.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendekia.
- Mardapi, D. (2012). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- Masidjo. (1995). *Penilaian Pencapaian Hasil Belajar Siswa di Sekolah*. Yogyakarta: Kanisius.
- Rosana, D. (2011). *Model Evaluasi UT*. Yogyakarta: Universitas Terbuka.
- Sudarman. (2007). *Problem Based Learning: Suatu Model Pembelajaran untuk Mengembangkan dan Meningkatkan Kemampuan Memecahkan Masalah. Pendidikan Inovatif*, 1-20.
- Sudijono, A. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Sudjana, N. (2009). *Penilaian Hasil dan Proses Belajar*. Bandung: PT Remaja Rosda Karya.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.
- Sukmadinata, N. S. (2004). *Metode Penelitian Pendidikan*. Bandung: Remaja Remaja Karya.
- Sumritin. (2014). *Pengembangan Instrumen Penilaian Kemampuan Highr Order Thinking Siswa SMP dalam Mata Pelajaran Matematika*. Yogyakarta: UNY.
- Surapranata, S. (2009). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Bandung: Remaja Rosda Karya.
- Widoyoko, E. P. (2011). *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Willy, A. (2016). *Kasus Pencabulan Anak*. Siduarjo: Jawa Pos.

Zaenul, A. (2005). *Penilaian Hasil Belajar*. Jakarta: Pusat Pengembangan Aktivitas Instruksional Ditjen DIKTI.