

BAB II

LANDASAN TEORI

A. Pengukuran dan Penilaian

Penilaian sangat erat hubungannya dengan pengukuran, karena sebelum adanya penilaian harus didahului dengan adanya pengukuran. Pengukuran diartikan sebagai pemberian angka kepada suatu atribut atau karakteristik tertentu yang dimiliki oleh orang, hal, atau objek tertentu menurut aturan atau formulasi yang jelas (Asmawi Zainul dan Noehi Nasution 2001: 5). Sedangkan menurut Wandt dan Brown pengukuran sebagai "*Measurement means the act or proses of axestaining the extent or quantity of something*" (Wandt dkk, 1957 : 1). Jadi menurut Wandt dan Brown pengukuran adalah suatu tindakan atau proses untuk menentukan luas atau kuantitas daripada sesuatu.

Dari perbagai definisi yang sering diungkapkan oleh para ahli, tentang pengukuran, maka menurut Asmawi Zainul dan Noehi Nasoetion (1992: 7), terdapat dua karakteristik pengukuran yang utama, yaitu (1) penggunaan angka atau skala tertentu, dan (2) menurut suatu aturan atau formula tertentu. Pada aspek pembelajaran, pengukuran hasil belajar bertujuan untuk mengetahui sejauh mana perubahan tingkah laku siswa setelah mengalami proses pembelajaran. Hasil pengukuran tersebut dapat berupa angka atau pernyataan-pernyataan yang

Untuk lebih memahami pengukuran hasil belajar, maka sebaiknya memahami empat karakteristik skala, yaitu (1) skala nominal, (2) skala ordinal, (3) skala interval, dan (4) skala rasio. Dalam pengukuran hasil belajar, skala yang sering digunakan adalah skala nominal dan skala ordinal (Asmawi Zainul dan Noehi Nasoetion, 1992: 7).

Penilaian merupakan komponen penting dalam sistem pendidikan karena mencerminkan perkembangan atau kemajuan hasil pendidikan (baca "mutu" pendidikan) dari satu waktu ke waktu lain. Di samping itu, berdasarkan penilaian, tingkat pencapaian prestasi pendidikan antara satu sekolah dengan sekolah lain atau satu wilayah dengan wilayah lain dapat dibandingkan (Depdikbud, 1994: 9).

Penilaian merupakan interpretasi dari hasil pengukuran dengan jalan membandingkan dengan suatu patokan atau kriteria. Penilaian dalam pendidikan dilakukan terhadap konsep dan proses yang dilakukan. Penilaian proses merupakan penilaian terhadap kegiatan dan kemajuan siswa pada saat berlangsungnya pembelajaran di kelas yang dilakukan selama pelaksanaan pembelajaran. Kegiatan ini memberikan umpan balik bagi guru untuk menentukan apakah pembelajaran dapat dilanjutkan atau justru diulang kembali, atau bagi siswa perlu bantuan guru atau perlu tambahan pelajaran (Depdikbud, 1994: 9).

Fungsi penilaian meliputi 4 hal. Pertama, fungsi formatif, hasil penilaian digunakan untuk memperbaiki hasil belajar dan kegiatan pembelajaran. Ketentuan pelaksanaan dalam belajar tuntas adalah jika 95% siswa mempunyai 75% atau

maka pelajaran dapat dilanjutkan. Kedua, fungsi sumatif, hasil penilaian digunakan untuk menentukan prestasi belajar siswa. Fungsi ketiga, adalah fungsi penempatan untuk mengelompokkan siswa berdasarkan kriteria tertentu. Fungsi terakhir adalah fungsi diagnostik. Fungsi diagnostik dimaksudkan untuk mengetahui kelemahan siswa yang sifatnya menetap pada diri siswa.

1. Jenis Penilaian

Menurut Umar (2000: 23), pelaksanaan penilaian di dalam kegiatan belajar mengajar ditujukan pada dua aspek, yaitu tingkat keberhasilan dan tingkat efisiensi pelaksanaannya. Karena penelitian ini mengenai analisis hasil Ujian Akhir Sekolah, maka lebih menitikberatkan pada aspek tingkat keberhasilan terhadap penilaian hasil belajar. Dalam buku Pedoman Penilaian Kurikulum 1994 (Depdikbud, 1993c: 13) penilaian terdiri atas:

- a. Penilaian Formatif, yaitu penilaian yang dilakukan pada setiap satuan pelajaran yang berguna untuk mengetahui sejauh mana Tujuan Instruksional Khusus telah tercapai. Penilaian formatif mempunyai dua tujuan, yaitu untuk membantu guru membuat perencanaan dan membantu murid mengenali segi-segi yang perlu ditangani.
- b. Penilaian Subsumatif, yaitu penilaian yang dilakukan sejumlah satuan pelajaran setelah selesai disampaikan. Penilaian subsumatif ini merupakan

adalah jenis penilaian yang dilakukan pada akhir suatu satuan pelajaran.

c. Penilaian Sumatif, yaitu penilaian yang dilakukan pada setiap akhir semester. Penilaian ini berguna untuk melihat sejauh mana program pengajaran dalam satu semester atau program untuk jangka waktu tertentu telah mencapai tujuannya. UAS menurut waktu pelaksanaannya bisa digolongkan penilaian sumatif, karena dilaksanakan pada akhir semester/akhir pelajaran. Dalam penelitian ini, yang dimaksudkan penilaian hasil belajar adalah penilaian yang dilaksanakan pada akhir tahun ajaran yang disebut Ujian Akhir Sekolah, disingkat UAS.

2. Penyusunan Butir Soal

Jika diinginkan hasil penilaian yang baik, maka harus diperhatikan unsur-unsur penting yang ada di dalam penilaian, diantaranya adalah penyusunan, alat penilaian, prosedur penilaian, dan suasana penilaian. Alat penilaian yang baik harus memenuhi syarat-syarat, yaitu: valid, reliabel, mempunyai tingkat kesukaran yang memadai, dan mempunyai daya beda yang baik, serta memiliki pengecoh yang berfungsi.

Bentuk soal yang digunakan dalam penilaian berdasarkan pada tujuan yang akan dicapai. Penyusunan soal merupakan penyusunan butir soal karena butir soal adalah yang membangun suatu soal. Biasanya digunakan bentuk soal objektif (pilihan ganda) dan uraian dalam berbagai penilaian

a. Butir Soal Objektif.

Soal objektif adalah butir soal yang telah mengandung kemungkinan jawaban yang harus dipilih atau dikerjakan oleh peserta tes. Kemungkinan jawaban dibuat oleh pembuat soal, peserta tes hanya memilih jawaban yang disediakan pembuat soal.

Secara umum, butir soal bentuk objektif dibagi menjadi tiga tipe, yaitu : (1) banar-salah (*true false*), (2) menjodohkan (*matching*), dan (3) pilihan ganda (*multiple choice*) (Asmawi Zainul dan Noehi Nasoetion, 1992: 61).

Oleh sementara pendidik, soal pilihan ganda dianggap paling bermanfaat dan paling luwes di antara semua jenis tes karena dapat digunakan untuk menguji sebagian terbesar mata pelajaran. Bentuk dari soal pilihan ganda adalah suatu *stem*. Stem dapat berupa pertanyaan ataupun pernyataan. Stem diikuti oleh alternatif-alternatif atau pilihan-pilihan jawaban. Tidak ada batas mengenai jumlah alternatif itu, tetapi umumnya empat sampai lima alternatif, dimana hanya ada satu jawaban yang benar.

Ada beberapa kaidah penulisan butir soal pilihan ganda yang harus diperhatikan, yaitu (Depdikbud, 2000: 33-35):

1). Soal harus sesuai dengan indikator.

2). Pilihan jawaban harus logis dan masuk akal, di antara pilihan jawaban

- 3). Setiap soal harus mempunyai satu jawaban yang benar atau yang paling benar.
- 4). Pokok soal harus dirumuskan secara jelas dan tegas.
- 5). Rumusan pokok soal dan pilihan jawaban harus merupakan pernyataan yang diperlukan saja.
- 6). Pokok soal jangan memberi petunjuk ke arah jawaban banar.
- 7). Pokok soal jangan mengandung pernyataan yang bersifat negatif ganda.
- 8). Pilihan jawaban harus homogen dan logis ditinjau dari segi materi.
- 9). Panjang rumusan pilihan jawaban harus relatif sama.
- 10). Pilihan jawaban jangan mengandung pertanyaan, “semua pilihan jawaban di atas salah”, atau “semua pilihan jawaban di atas benar”.
- 11). Pilihan jawaban yang berbentuk angka atau waktu harus disusun berdasarkan urutan besar kecilnya nilai angka tersebut.
- 12). Gambar, grafik, tabel, diagram dan sejenisnya yang terdapat pada soal harus jelas dan berfungsi.
- 13). Butir soal jangan bergantung pada jawaban soal sebelumnya

b. Butir Soal Essay (Uraian)

Tes essay (uraian) adalah suatu bentuk tes yang terdiri dari suatu pertanyaan atau suatu suruhan yang menghendaki jawaban yang berupa uraian-uraian yang relatif panjang. Bentuk bentuk pertanyaan

suruhan yang meminta kepada murid-murid untuk menjelaskan, membandingkan, menginterpretasikan dan mencari perbedaan. Semua bentuk pertanyaan atau suruhan tersebut mengharapkan agar murid-murid menunjukkan pengertian mereka terhadap materi yang dipelajari (Wayan Nurkencana dan Sumartana, 1983: 41-42).

Peserta tes dengan tipe soal uraian ini bebas untuk menjawab dengan kata-katanya sendiri. Dilihat dari model jawaban peserta tes, maka segera akan kelihatan bahwa pemberian skor terhadap jawaban soal tidak mungkin dilakukan secara objektif. Penggunaan bentuk soal uraian lebih ditekankan pada ulangan umum dan ujian akhir (Depdikbud, 1994: 23).

Ada beberapa kaidah penulisan butir soal uraian yang harus diperhatikan, yaitu (Depdikbud, 2000: 41-43) :

- 1). Soal harus sesuai dengan indikator.
- 2). Batasan pertanyaan dan jawaban yang diharapkan (ruang lingkup) harus jelas.
- 3). Isi materi sesuai dengan petunjuk pengukuran.
- 4). Isi materi yang ditanyakan sudah sesuai dengan jenjang, jenis sekolah, atau tingkat kelas.
- 5). Rumusan kalimat soal atau pertanyaan harus menggunakan kata-kata tanya.
- 6). Rumusan kalimat soal harus komunikatif.
- 7). Buatlah petunjuk yang jelas tentang cara menjawab soal.

kesiapan, respon terpimpin, mekanisme, respon yang kompleks, penyesuaian dan mencipta.

Dalam konteks evaluasi hasil belajar, maka ketiga domain atau ranah itulah yang harus dijadikan sasaran dalam setiap kegiatan evaluasi hasil belajar, yaitu: (1) Apakah peserta didik sudah dapat memahami semua bahan atau materi pelajaran yang telah diberikan mereka? (2) Apakah peserta didik sudah dapat menghayatinya? (3) Apakah materi pelajaran yang telah diberikan itu sudah dapat diamalkan secara kongkret dalam praktek atau dalam kehidupan sehari-hari? (Anas Sudijono, 2005: 49).

4. Aspek Kognitif

Dalam penelitian ini digunakan klasifikasi dari Bloom, karena klasifikasi tersebut sangat rinci dan banyak dipakai dalam dunia pendidikan di Indonesia.

Ranah kognitif adalah ranah yang mencakup kegiatan mental (otak). Menurut Bloom, segala upaya yang menyangkut aktivitas otak adalah termasuk dalam ranah kognitif. Dalam ranah kognitif itu terdapat enam jenjang proses berfikir, mulai dari jenjang terendah sampai dengan jenjang yang paling tinggi (Anas Sudijono, 2005: 50). Keenam tingkatan dari yang

terendah hingga tingkat tertinggi yang dijabarkan dalam tabel berikut ini (Anas Sudijono, 2005: 50).

- c. Aplikasi (C₃)
- d. Analisis (C₄)
- e. Sintesis (C₅)
- f. Evaluasi (C₆)

Pengetahuan (*knowledge*) adalah kemampuan seseorang untuk mengingat-ingat kembali (*recall*) atau mengenali kembali tentang nama, istilah, ide, gejala, rumus-rumus dan sebagainya, tanpa mengharapkan kemampuan untuk menggunakannya. Pengetahuan atau ingatan ini adalah merupakan proses berpikir yang paling rendah (Anas Sudijono, 2005: 50). Kata operasional yang dapat digunakan untuk pertanyaan dalam aspek ini, antara lain adalah mendefinisikan, mendiskripsikan, mendaftarkan, menjodohkan, mengidentifikasi, menyebutkan dan menyatakan (Sukardjo, 1994: 19).

Pemahaman (*comprehension*) adalah kemampuan seseorang untuk mengerti atau memahami sesuatu setelah sesuatu itu diketahui dan diingat. Dengan kata lain, memahami adalah mengetahui tentang sesuatu dan dapat melihat dari berbagai segi. Pemahaman merupakan jenjang kemampuan berfikir yang setingkat lebih tinggi dari ingatan atau hafalan (Anas Sudijono, 2005: 50). Ada tiga (3) aspek bentuk pertanyaan dalam aspek ini, yaitu translasi, interpretasi, dan eksplorasi. Soal termasuk translasi, jika dalam soal tersebut terkandung perintah untuk menyatakan kembali dengan perkataan sendiri atau menuliskan suatu istilah dengan istilah lain suatu ide yang telah

diterimanya. Soal termasuk interpretasi, jika soal tersebut mengandung maksud agar siswa dapat melakukan translasi dengan mengidentifikasi serta memahami ide mayor yang ada di dalamnya dan menuntut mengerti saling hubungan satu dengan lainnya, dapat dalam bentuk tabel, kalimat, chart, grafik, dan sebagainya. Soal termasuk eksplorasi, jika dalam soal tersebut menuntut siswa bisa keluar dari batas-batas data atau informasi yang diberikan, membuat aplikasi yang benar, dan memperluas data atau informasi. Kata-kata operasional dalam tingkat ini, antara lain membedakan, menjeneralisasikan, memberikan contoh, menulis kembali, memperkirakan (Sukardjo, 1994.: 19).

Aplikasi (*application*) adalah kesanggupan seseorang untuk menerapkan atau menggunakan ide-ide umum, tata cara ataupun metode-metode, prinsip-prinsip, rumus-rumus, teori-teori dan sebagainya, dalam situasi yang baru dan kongkret. Aplikasi atau penerapan ini adalah merupakan proses berpikir setingkat lebih tinggi ketimbang pemahaman (Anas Sudijono, 2005: 51). Kata-kata operasional dalam tingkatan ini, antara lain mengubah, mendemonstrasikan, menemukan, memanipulasi, memodifikasi, mengoperasikan, meramalkan, menyiapkan, menghasilkan, menghubungkan, menunjukkan, memecahkan, dan menggunakan (Sukardjo, 1994.: 18).

Analisis (*analysis*) adalah kemampuan seseorang untuk merinci atau menguraikan suatu bahan atau keadaan menurut bagian-bagian yang lebih kecil dan mampu memahami hubungan di antara bagian-bagian atau faktor

faktor yang satu dengan faktor-faktor lainnya. Jenjang analisis adalah setingkat lebih tinggi ketimbang jenjang aplikasi (Anas Sudijono, 2005: 51). Kata-kata operasional dalam tingkatan ini, antara lain merinci, menyusun, membedakan, mengidentifikasi, mengilustrasikan, menyimpulkan, menunjukkan, menghubungkan, memilih, memisahkan, dan membagi (Sukardjo, 1994.: 18).

Sintesis (*Synthesis*) adalah kemampuan berpikir yang merupakan kebalikan dari proses berpikir analisis. Sintesis merupakan suatu proses yang memadukan bagian-bagian atau unsur-unsur secara logis, sehingga menjelma menjadi suatu pola yang berstruktur atau berbentuk pola baru. Jenjang sintesis kedudukannya setingkat lebih tinggi ketimbang jenjang analisis (Anas Sudijono, 2005: 51). Kata-kata operasional dalam tingkatan ini, antara lain, mengkategorikan, mengkombinasikan, menciptakan, membuat desain, menjelaskan, memodifikasi, mengorganisir, menyusun, membuat rencana, mengatur kembali, merekonstruksi, merevisi, dan menulis kembali (Sukardjo, 1994.: 19).

Evaluasi (*Evaluation*) adalah merupakan jenjang berpikir paling tinggi dalam ranah kognitif menurut Taksonomi Bloom. Soal termasuk dalam aspek evaluasi jika dalam pertanyaan tersebut terkandung maksud agar siswa mampu membuat pendugaan atau penilaian terhadap materi, pekerjaan, fenomena, atau tingkah laku berdasarkan suatu kriteria tertentu (Anas Sudijono, 2005: 51). Kata-kata operasional dalam tingkatan ini, antara lain

menilai, membandingkan, menyimpulkan, mengkritik, mendiskripsikan, membedakan, memutuskan, menafsirkan, menghubungkan, dan membantu (Sukardjo, 1994.: 19). Dalam menganalisis butir soal, biasanya ketiga tingkatan yang terakhir, yaitu Analisis, Sintesis, dan Evaluasi dijadikan satu.

Perbandingan persentase masing-masing komponen dalam aspek kognitif belum ada ketentuan baku. Namun, menurut Sukardjo (1994.: 20) untuk kelas tiga SMA suatu soal hasil belajar memenuhi kriteria cukup baik bila komposisi perbandingan antara aspek pengetahuan : aspek pemahaman : aspek aplikasi : aspek analisis, sintesis, evaluasi memiliki perbandingan sebagai berikut 40% : 30% : 20% : 10% atau 4 : 3 : 2 : 1.

B. Analisis Butir Soal

Analisis adalah suatu pemecahan sebuah komunikasi ke dalam unsur-unsur atau bagian-bagian sedemikian rupa sehingga hirarki ide-idenya menjadi jelas atau hubungan antara ide-ide yang dinyatakan dibuat menjadi eksplisit (Subino, 1987: 21). Analisis butir soal bertujuan untuk mengetahui validitas isi yang berupa validitas empiris maupun validitas isi yang berupa kebenaran konsep, tingkat kesukaran, daya beda, fungsi distraktor, reliabilitas, serta kebenaran konstruksinya yang dilakukan dengan cara memperbaiki, menyeleksi, mengganti, atau merevisi.

Analisis butir soal teori tes klasik dilakukan dengan menggunakan bantuan komputer program *Item and Test Analysis* yaitu *Itaman Versi 3.0*

- 8). Buatlah pedoman penskoran segera setelah soal ditulis.
- 9). Rumusan butir soal menggunakan bahasa yang sederhana.

3. Taksonomi Tujuan Pendidikan.

Taksonomi adalah ilmu tentang klasifikasi secara umum, dan juga klasifikasi spesifik mengenai suatu hal dengan aturan-aturan tertentu, contohnya taksonomi tujuan pengajaran (Viviane, 1977: 97). Taksonomi tujuan pengajaran membantu guru untuk menetapkan arah dan tujuan yang akan dicapai dalam kegiatan pembelajaran. Taksonomi tujuan pengajaran disusun menurut prinsip struktur yang makin lama makin sukar. Hal ini dapat dipahami karena secara psikologi, mengingat adalah proses berpikir yang lebih sederhana daripada memahami.

Bloom, seperti dikutip oleh Sukardjo (1994: 17) membagi tujuan pendidikan/pengajaran menjadi tiga aspek yaitu:

- a. Aspek kognitif, berkaitan dengan daya pikir, penalaran atau pengetahuan. Meliputi pengetahuan, pemahaman, aplikasi, analisis, sintesis dan evaluasi.
- b. Aspek afektif, berkaitan dengan perasaan atau kesadaran, sikap senang atau tidak. Meliputi penerimaan, penilaian, pengorganisasian, dan pemeranan (karakterisasi).
- c. Aspek psikomotorik, terutama berkaitan dengan ketrampilan fisik, ketrampilan motorik, atau ketrampilan tangan. Terdiri dari persepsi,

Analisis ini akan menghasilkan butir dan skala statistik perangkat tes. Statistik butir yang meliputi; tingkat kesukaran butir, daya pembeda, dan efektivitas distraktor, sedangkan skala statistik perangkat tes, antara lain : rerata, median, keandalan, kemencengan, dan kesalahan baku pengukuran. Menurut Azwar (1996: 131) analisis pada teori tes klasik bisa dilanjutkan sampai pada efektivitas distraktor, disamping tingkat kesukaran dan daya pembeda. Mardapi (1999: 7), mengemukakan bahwa parameter butir pada teori klasik adalah tingkat kesukaran butir, daya pembeda, dan tebakan. Hal ini penting untuk diketahui, agar pembuat butir tes tidak sembarangan memasukkan pilihan jawaban.

Menurut Hasan dan Zainul (1993: 133), analisis butir soal dapat dilakukan secara kuantitatif maupun kualitatif. Analisis secara kuantitatif akan mendapatkan karakteristik butir soal menyangkut tingkat kesukaran dan daya beda. Analisis kualitatif dilakukan untuk mendapatkan spesifikasi butir soal meliputi validitas isi dan kesesuaian dengan tujuan.

1. Validitas

Validitas dapat didefinisikan sebagai tingkat ketepatan yang dimiliki alat penilaian untuk mengukur sesuatu terhadap kelompok tertentu. Suatu alat pengukur dapat dikatakan alat pengukur yang valid apabila alat pengukur tersebut dapat mengukur apa yang hendak diukur secara tepat. Menurut Wayan Nurkencana dan Sumartana (1983: 127-129) validitas suatu tes dapat ditinjau dari beberapa segi, yaitu:

a. Validitas Ramalan

Validitas ramalan artinya ketepatan (kejituan) daripada suatu alat pengukur ditinjau dari kemampuan tes tersebut untuk meramalkan prestasi yang dicapainya kemudian. Suharsimi Arikunto (1996: 66), menyatakan bahwa sebuah tes dikatakan memiliki validitas prediksi atau validitas ramalan, apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan datang.

b. Validitas Bandingan.

Validitas bandingan artinya kejituan daripada suatu tes dilihat dari korelasinya terhadap kecakapan yang telah dimilikinya saat kini secara riil. Validitas ini lebih umum dikenal dengan validitas empiris. Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman. Jika ada istilah “sesuai” tentu ada dua hal yang dipasangkan. Dalam hal ini hasil tes dipasangkan dengan hasil pengalaman. Pengalaman selalu mengenai hal yang telah lampau sehingga data pengalaman tersebut sekarang sudah ada. Dalam membandingkan hasil sebuah tes maka diperlukan suatu kreterium atau alat banding. Maka hasil tes merupakan sesuatu yang dibandingkan (Suharsimi Arikunto, 1996: 65).

c. Validitas Isi.

Validitas isi artinya kejituan daripada suatu tes ditinjau dari isi tes tersebut. Validitas isi dari suatu tes hasil belajar adalah validitas yang diperoleh setelah dilakukan pengalihan, penelusuran atau pengujian

terhadap isi yang terkandung dalam tes hasil belajar tersebut. Validitas isi adalah validitas yang ditilik dari segi isi tes itu sendiri sebagai alat pengukur hasil belajar yaitu: sejauh mana tes hasil belajar sebagai alat pengukur hasil belajar peserta didik, isinya telah dapat mewakili secara representatif terhadap keseluruhan materi atau bahan pelajaran yang seharusnya diteskan (diujikan) (Anas Sudijono, 2005: 164). Validitas isi sangat penting bagi tes prestasi belajar, sebab validitas isi menunjuk kepada sejauh mana suatu tes mengukur sampel bahan ajar dan atau perubahan perilaku hasil belajar secara representatif. Artinya, sejauh mana butir-butir soal yang dikembangkan, telah menguji apa yang seharusnya diuji.

Oleh karena materi yang diajarkan itu pada umumnya tertuang dalam Garis-garis Besar Program Pengajaran (GBPP) yang merupakan penjabaran dari kurikulum yang telah ditentukan, maka validitas isi sering disebut validitas kurikuler. Dalam praktek, validitas isi dari suatu tes hasil belajar dapat diketahui dengan jalan membandingkan antara isi yang terkandung dalam tes hasil belajar, dengan tujuan instruksional khusus yang telah ditentukan untuk masing-masing mata pelajaran. Jika penganalisisan secara rasional itu menunjukkan hasil yang membenarkan tentang telah tercerminnya tujuan instruksional khusus itu di dalam tes hasil belajar, maka tes hasil belajar yang sedang diuji validitas isinya itu

dapat dinyatakan sebagai tes hasil belajar yang telah memiliki validitas isi (Anas Sudijono, 2005: 165).

Allen dan Yen (1979: 95), mengatakan bahwa validitas isi terdiri dari dua bagian, yaitu validitas muka (*face validity*) dan validitas logik (*logical or sampling validity*). Suatu tes memiliki validitas muka, jika penampilan tes tersebut telah meyakinkan dan memberi kesan mampu mengungkap atribut yang hendak diukur. Validitas logik menunjuk pada sejauh mana isi tes merupakan representasi dari ciri-ciri atribut yang hendak diukur.

d. Validitas Susunan/Konstruksi.

Validitas susunan/konstruksi artinya kejituan daripada suatu tes ditinjau dari susunan tes tersebut. Secara terminologis, suatu tes hasil belajar dapat dinyatakan sebagai tes yang telah memiliki validitas konstruksi, apabila tes hasil belajar tersebut telah dapat dengan secara tepat mencerminkan suatu konstruksi dalam teori psikologis. Validitas konstruksi dari suatu tes hasil belajar dapat dilakukan penganalisisannya dengan jalan melakukan pencocokan antara aspek-aspek berpikir yang terkandung dalam tes hasil belajar tersebut, dengan aspek-aspek berpikir yang dikehendaki untuk diungkap oleh tujuan instruksional khusus. Dengan demikian, maka tes hasil belajar tersebut dapat dinyatakan sebagai tes hasil belajar yang valid dari segi susunannya atau telah memiliki validitas konstruksi (Anas Sudijono, 2005: 167).

Sedangkan menurut Suharsimi Arikunto (1993: 64), validitas ada 2 yaitu validitas logis yang diketahui dari hasil pemikiran dan validitas empiris yang dapat diketahui dari hasil pengalaman. Validitas logis terdiri dari validitas isi (*content validity*) dan validitas konstruk (*construk validity*).

2. Tingkat Kesukaran Butir Soal

Sax (1980: 193), menyatakan bahwa tingkat kesukaran (p) butir soal adalah proporsi peserta tes yang menjawab benar. Tingkat kesukaran (p) butir soal dapat ditentukan dengan beberapa cara, antara lain: (1) skala kesukaran, (2) skala bivariat, dan (3) proporsi menjawab benar.

Cara yang paling mudah dan paling umum digunakan adalah skala rata-rata atau proporsi menjawab benar atau *proportion correct* (p), yaitu jumlah peserta tes yang menjawab benar pada soal yang dianalisis dibandingkan dengan peserta tes seluruhnya. Persamaan yang digunakan untuk menentukan tingkat kesukaran (p) ini adalah (Hayat dkk., 1997: 17):

$$p = \frac{\sum B}{N}$$

p = proporsi menjawab benar pada butir soal tertentu

$\sum B$ = banyaknya peserta tes menjawab benar

N = jumlah peserta tes yang menjawab.

Tingkat kesukaran (p) sebenarnya merupakan nilai rata-rata dari kelompok peserta tes. Oleh karena itu, tingkat kesukaran sebenarnya adalah rerata dari suatu distribusi skor kelompok dari suatu soal. Oleh karena tingkat kesukaran ini dihitung atas dasar nilai rata-rata, maka tingkat kesukaran tersebut juga dinamakan kesukaran rata-rata. Indeks kesukaran rata-rata ini paling banyak dipergunakan untuk menentukan tingkat kesukaran soal.

Besarnya tingkat kesukaran (p) berkisar antara 0 sampai dengan 1. Tingkat kesukaran dikategorikan menjadi tiga bagian seperti tampak pada tabel 1 berikut ini (Hayat dkk., 1997: 18):

Tabel 1: Kreteria Tingkat Kesukaran Butir Soal.

<i>Proportion Correct</i> (p)	Kategori Soal
$p > 0,70$	Mudah
$0,30 < p < 0,70$	Sedang
$p < 0,30$	Sukar

Azwar (1996: 135), menyatakan bahwa nilai p yang ideal adalah 0,5. Menurut Allen dan Yen (1979: 122), indeks kesukaran yang sedang merupakan butir yang baik, yaitu butir soal yang memiliki tingkat kesukaran 0,3 sampai dengan 0,7.

Sedangkan untuk menyusun suatu naskah ujian, menurut Zainul, A dan Nasution, N (2001, 177), sebaiknya digunakan butir soal yang tingkat

kesukarannya berimbang sebagai berikut: sukar 25%; sedang 50%; dan mudah 25%.

3. Daya Beda

Daya beda suatu soal berfungsi untuk menentukan dapat tidaknya suatu soal membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan dari pengujian daya beda adalah untuk melihat kemampuan butir soal dalam membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah (Hayat dkk., 1997: 19). Hubungan antara daya beda dan tingkat kesukaran menurut Sax (1980: 92), daya beda yang tinggi terjadi pada butir soal dengan tingkat kesukaran yang optimal, tetapi tingkat kesukaran yang optimal tidak menjamin daya beda butir akan tinggi. Besarnya indeks daya beda butir soal dalam program Iteman ditandai dengan nilai yang berada di bawah tulisan *point biserial* dan *biserial*.

Ada beberapa cara yang digunakan untuk menentukan daya pembeda, antara lain dengan menggunakan (Hayat dkk., 1997: 19): (1) indeks diskriminasi, (2) indeks korelasi, (3) indeks keselarasan.

Daya beda biasanya disimbolkan dengan D (huruf kapital), langkah-langkah untuk menentukan daya beda adalah (Oemar Hamalik, 1989: 67):

Mengurutkan langkah-langkah dari kelas tertinggi ke kelas terendah sesuai

- b. Mengalikan N dengan 27%, hasil pembulatan diperoleh adalah n.
- c. Menghitung n kelompok atas (lembar jawaban dengan skor tertinggi di hitung dari atas) dan n kelompok bawah (lembar jawaban dengan skor terendah dihitung dari bawah).
- d. Menentukan proporsi butir soal yang dijawab dengan benar untuk masing-masing kelompok. Kelompok atas (P_H) dan kelompok bawah (P_L) dengan cara membagi jumlah jawaban yang benar dengan n.

Daya beda butir soal (D), merupakan selisih proporsi butir soal yang dijawab dengan benar antara kelompok atas (P_H) dengan kelompok bawah (P_L).

$$D = (P_H - P_L)$$

Untuk menentukan keputusan soal diterima, direvisi atau ditolak maka menggunakan kriteria parametrik soal yang digambarkan dalam tabel 2 berikut (Pakpahan, 1990: 76) :

Tabel 2: Kriteria Parametrik Daya Beda Butir Soal.

Parameter	Koefisien	Keputusan
Daya Beda	> 0,30	Diterima
	0,10 - 0,29	Direvisi
	< 0,10	Ditolak

Daya beda yang ideal adalah sebesar mungkin mendapat angka 1, berarti butir soal tersebut semakin mampu membedakan antara mereka yang menguasai bahan yang diujikan dan mereka yang tidak. Semakin kecil indeks daya beda (semakin mendekati 0) berarti tidak jelaslah fungsi butir yang bersangkutan dalam membedakan subjek yang menguasai bahan pelajaran dan yang tidak (Azwar, 1996: 139)

4. Efektifitas Pengecoh (*distraktor*)

Dalam tes objektif bentuk multiple choice, setiap butir soal telah dilengkapi dengan beberapa kemungkinan jawaban atau yang sering dikenal dengan istilah option atau alternatif jawaban. Option atau alternatif jawaban jumlahnya berkisar antara tiga sampai lima buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir item tersebut, salah satu diantaranya adalah merupakan jawaban yang benar (= kunci jawaban), sedangkan sisanya adalah merupakan jawaban yang salah. Menurut Anas Sudijono (2005: 409) jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (distraktor = pengecoh). Untuk penelitian ini, jumlah option atau alternatif jawaban terdiri dari lima buah.

Tujuan utama dari pemasangan distraktor pada setiap butir item itu adalah agar dari sekian banyak peserta tes yang mengikuti tes hasil belajar ada yang tertarik atau terangsang untuk memilihnya, sebab mereka menyangka

bahwa distraktor yang mereka pilih merupakan jawaban yang betul (Anas Sudijono, 2005: 410).

Azwar (1996: 141), menyatakan bahwa tingkat kesukaran dan daya beda tidaklah cukup untuk menilai kualitas butir, tetapi bagaimana jawaban-jawaban informasi peserta tes tersebar pada pilihan jawaban yang tersedia juga harus diperhatikan agar fungsi suatu butir dapat dipenuhi secara maksimal. Efektifitas distraktor untuk suatu butir dianalisis dari distribusi jawaban terhadap butir yang bersangkutan, pada setiap alternatif jawaban yang disediakan. Efektifitas distraktor ditelaah untuk mengetahui apakah pilihan jawaban yang bukan kunci jawaban telah berfungsi sebagaimana mestinya. Artinya apakah distraktor tersebut dipilih oleh lebih banyak atau semua peserta tes dari kelompok yang berkemampuan rendah sedangkan peserta tes dari kelompok yang berkemampuan tinggi hanya sedikit atau tidak ada yang memilihnya.

Untuk menentukan berfungsi tidaknya pengecoh, diadakan analisis butir soal. Cara yang mudah untuk menentukan berfungsi tidaknya pengecoh, dapat dilihat pada print out computer hasil analisis program IteMan *Version 3.00 Micro CAT (tm) Testing*, dengan melihat tanda “minus” dan “plus” pada kolom *Prop. Endorsing* dan *Point Biser*. Jawaban yang baik, ialah jika kunci jawaban positif dan distraktor negatif.

5. Reliabilitas

Reliabilitas adalah kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari suatu pengukuran ke pengukuran yang lainnya. Jadi reliabilitas dapat dikatakan sebagai tingkat konsistensi atau kemantapan hasil dari hasil dua pengukuran terhadap hal yang sama. Hasil pengukuran itu diharapkan akan sama apabila pengukuran itu diulangi. Dengan perangkat tes yang reliabel, apabila tes itu kita berikan dua kali pada orang yang sama, tetapi dalam selang waktu yang berbeda, sepanjang tidak ada perubahan kemampuan, maka skor yang diperoleh akan konstan (Hayat dkk., 1997: 22).

Reliabilitas dan validitas merupakan suatu hal yang penting dalam suatu tes. Reliabilitas mendukung validitas. Suatu tes, mungkin saja reliabel tetapi belum tentu valid. Sebaliknya, tes yang valid sudah pasti reliabel. Reliabilitas memiliki dua konsistensi. Konsistensi pertama, adalah konsistensi internal yakni tingkat sejauhmana soal itu homogen baik dari tingkat segi kesukaran maupun bentuk soalnya. Konsistensi yang kedua, yaitu konsistensi eksternal yakni tingkat sejauh mana skor yang dihasilkan tetap sama sepanjang kemampuan orang yang diukur belum berubah. Apabila hasil skor tes pertama sama dengan hasil skor tes kedua, maka tes dikatakan memiliki reliabilitas yang tinggi atau terdapat korelasi yang tinggi antara hasil tes pertama dengan hasil tes kedua. Kalau antara hasil tes pertama dan kedua

tidak terdapat hubungan atau hubungan rendah, maka tes itu dikatakan tidak reliabel (Hayat dkk., 1997: 22).

Reliabilitas dalam arti konsistensi atau homogenitas tes, merupakan koefisien korelasi yang menunjukkan seberapa jauh suatu perangkat tes homogen, untuk mengukur suatu mata pelajaran atau bidang studi yang sama. Reliabilitas yang paling baik, bila dicapai angka koefisien 1,00. Dalam pengukuran angka koefisien biasanya kurang dari 1,00, yang disebabkan oleh sifat soal, situasi pada saat pengukuran, keadaan subjek, dan sebagainya (Izzak Latunussa, 1988: 36).

Untuk mengukur keandalan suatu tes, dapat digunakan beberapa teknik statistik, antara lain : (1) metode tes ulang; (2) metode tes paralel; dan (3) metode konsistensi internal (Allen dan Yen, 1979: 76). Rumus untuk menentukan besarnya indeks keandalan suatu tes menurut Spearman-Brown dapat dituliskan sebagai berikut (Mehrens & Lehmann, 1973: 112):

$$r_{xx} = \frac{2 r_{1/2/2}}{1 + r_{1/2/2}}$$

$r_{1/2/2}$ = korelasi antara skor belahan 1 dan belahan 2.

Formula indeks keandalan suatu tes menurut Cronbach-alpha adalah :

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right]$$

dimana n = jumlah butir soal

$\sum S_i^2$ = jumlah varians skor butir

S_x^2 = varians skor total

Pada penelitian ini hanya menggunakan satu rumus untuk menghitung koefisien keandalan melalui pendekatan konsistensi internal yang digunakan dalam program IteMan, yaitu koefisien Alpha atau Cronbach's Alpha. Ukuran besarnya koefisien keandalan suatu tes yang dikemukakan para ahli berbeda-beda. Ebel (1979: 275) menyatakan bahwa alat ukur (tes prestasi) yang mempunyai koefisien keandalan mendekati 0,90 sudah baik. Kapien dan Succuzzo (1982) dalam Mardapi (1999:14), menyatakan bahwa tes yang memiliki indeks keandalan 0,70 sampai 0,80 sudah termasuk baik. Untuk menentukan tinggi rendahnya reliabilitas soal, pada penelitian kali ini digunakan rumus sebagaimana pada tabel 3 berikut (Izzak Latunussa, 1988:

Tabel 3: Kreteria Tingkat Reliabilitas Butir Soal.

No.	Rentang Nilai	Keputusan
1	$rh < 0,200$	Tidak Reliabel
2	$0,200 < rh < 0,399$	Rendah
3	$0,400 < rh < 0,699$	Sedang
4	$0,700 < rh < 0,899$	Tinggi
5	$rh > 0,900$	Sangat Tinggi

C. Ujian Akhir Sekolah (UAS)

Ujian Akhir Sekolah (UAS) dalam sistem pendidikan di Muhammadiyah khususnya di lingkungan Majelis Pendidikan Dasar dan Menengah Pimpinan Wilayah Muhammadiyah Daerah Istimewa Yogyakarta, dijadikan salah satu bentuk pengukuran dan penilaian hasil belajar pada akhir jenjang pendidikan dasar dan menengah. Demikian juga SMA Muhammadiyah 3 Yogyakarta sebagai lembaga pendidikan Muhammadiyah, tiap tahun menyelenggarakan UAS, yang pelaksanaannya bersamaan dengan Ujian Akhir Nasional (UAN).

UAS bisa diidentikkan dengan UAN, dimana untuk UAS yang diujikan mata pelajaran ISMUBA, meliputi bidang studi Aqidah, Ibadah, Akhlak, Tarikh, Al-Qur'an/Hadits, Kemuhimmadiyahan dan Bahasa Arab. Paket soal UAS

dibuat oleh Majelis Pendidikan Dasar dan Menengah PWM DIY yang dilaksanakan secara serentak untuk sekolah-sekolah Muhammadiyah se Daerah Istimewa Yogyakarta.

Sedangkan penyelenggaraan UAN untuk SMA dengan sistem rayon tingkat propinsi, dan dapat dibentuk sub rayon bila memenuhi ketentuan yang berlaku. Adapun mata pelajaran yang diujikan untuk jurusan IPA meliputi Matematika, Bahasa Indonesia, dan Bahasa Inggris, sedangkan untuk jurusan IPS meliputi Ekonomi, Bahasa Indonesia dan Bahasa Inggris. Panitia UAN tingkat propinsi menerima dan memilih paket soal UAN dari panitia tingkat pusat. Paket soal kemudian digandakan sesuai dengan keperluan.

UAN merupakan usaha penilaian terakhir yang dilakukan untuk mengungkapkan hasil belajar siswa secara keseluruhan selama belajar di sekolah tersebut (B. Suryosubroto, 1991: 144). UAN dilaksanakan sebagai salah satu usaha untuk menyukseskan pendidikan dan bertujuan untuk (B. Suryosubroto, 1991: 148):

- a. Menciptakan standar mutu pendidikan dasar dan menengah.
- b. Mempercepat peningkatan dan pemerataan mutu pendidikan dasar dan menengah.
- c. Mengetahui tingkat pencapaian tujuan kurikuler pendidikan yang diselenggarakan oleh sekolah.
- d. Mendorong agar kegiatan pembelajaran dilaksanakan berdasarkan kurikulum, buku, dan alat peraga yang telah ditetapkan.

pembentukan kader tersebut, maka dalam sekolah-sekolah Muhammadiyah untuk semua jenis dan jenjang, bidang studi Al-Islam, Kemuhammadiyah dan Bahasa Arab, selanjutnya disebut (ISMUBA) merupakan ciri khusus pendidikan Muhammadiyah yang menjadi inti kurikulum (Majelis Dikdasmen, 2002: 1).

Bidang studi Ismuba adalah upaya sadar dan terencana dalam menyiapkan peserta didik untuk mengenal, memahami, menghayati, hingga mengimani, bertaqwa, dan berakhlak mulia dalam mengamalkan ajaran agama Islam dari sumber utamanya kitab suci Al Qur'an dan Al-Hadits. Bidang Ismuba, secara keseluruhan terliput dalam lingkup: Aqidah, Ibadah/Mu'amalah, Akhlak, Al-Qur'an/Al-Hadits, Tarikh, Kemuhammadiyah, dan Bahasa Arab. Bidang studi tersebut sekaligus menggambarkan bahwa ruang lingkup Ismuba mencakup perwujudan keserasian, keselarasan dan keseimbangan hubungan manusia dengan Allah SWT., diri sendiri, sesama manusia, makhluk lainnya maupun lingkungannya (*hablun minallah wa hablun minannas*), dan mengantarkan terbentuknya kader persyarikatan Muhammadiyah yang memiliki kemampuan dan kepribadian yang kuat (Majelis Dikdasmen, 2002: 2).

2. Fungsi dan Tujuan

Bidang studi di SMA berfungsi untuk : (a) Pengembangan keimanan dan ketakwaan kepada Allah SWT, serta akhlak mulia peserta didik,

kemuhammadiyah dan bahasa Arab seoptimal mungkin, yang telah ditanamkan lebih dahulu dalam lingkungan keluarga atau jenjang pendidikan di bawahnya; (b) Penanaman nilai ajaran Islam sebagai pedoman mencapai kebahagiaan hidup di dunia dan akhirat; (c) Penyesuaian mental peserta didik terhadap lingkungan fisik dan sosial melalui bidang studi Ismuba; (d) Perbaikan kesalahan-kesalahan, kelemahan-kelemahan peserta didik dalam keyakinan, pengamalan ajaran Islam dalam kehidupan sehari-hari sesuai faham agama menurut Muhammadiyah; (e) Pencegahan peserta didik dari hal-hal negatif budaya asing maupun lingkungan yang akan dihadapinya sehari-hari; (f) Pengajaran tentang ilmu pengetahuan keagamaan secara umum (alam nyata dan nir-nyata), sistem dan fungsionalnya, kemuhammadiyah dan bahasa Arab; serta (g) Penyaluran siswa untuk mendalami pendidikan agama ke lembaga pendidikan yang lebih tinggi (Majelis Dikdasmen, 2002: 3).

Bidang studi Ismuba di SMA bertujuan untuk menumbuhkan dan meningkatkan keimanan, melalui pemberian dan pemupukan pengetahuan, penghayatan, pengamalan serta pengalaman peserta didik tentang agama Islam menurut paham Muhammadiyah, Kemuhammadiyah, dan Bahasa Arab sehingga menjadi manusia muslim yang terus berkembang dalam hal keimanan, ketaqwaannya kepada Allah SWT, serta berakhlak mulia dalam kehidupan pribadi, bermasyarakat, berbangsa dan bernegara, memahami dan menghayati serta ikut berperan serta dalam gerakan persyarikatan Muhammadiyah serta memiliki kemampuan dasar berbahasa Arab untuk

dapat melanjutkan pada jenjang pendidikan yang lebih tinggi (Majelis Dikdasmen, 2002: 3). Untuk lebih jelasnya, mengenai kurikulum ISMUBA SMA Muhammadiyah 3 Yogyakarta, dapat dilihat pada lampiran 1.

Untuk dapat mencapai tujuan, maka diperlukan upaya pengorganisasian materi pembelajaran yang rasional dan menyeluruh. Pengorganisasian materi pembelajaran tersebut, mencakup tiga tahap kegiatan, yaitu perencanaan, pelaksanaan, dan penilaian. Perencanaan terdiri dari perencanaan per satuan waktu dan perencanaan tiap satuan bahan ajar. Pelaksanaan terdiri dari langkah-langkah pembelajaran di dalam atau di luar kelas, mulai dari pendahuluan, penyajian dan penutup. Sedangkan penilaian merupakan proses yang dilakukan terus menerus sejak perencanaan, pelaksanaan dan setelah pelaksanaan pembelajaran tiap pertemuan, satuan bahan ajar, maupun satuan waktu.