

BAB II **KAJIAN TEORETIK**

A. Deskripsi Konseptual

1. Penilaian

Penilaian merupakan salah satu kegiatan dalam proses pembelajaran. Penilaian pendidikan adalah proses pengumpulan, analisis, dan mengolah informasi untuk menentukan hasil belajar. Nurgiyantoro (2013: 5-7) mengatakan bahwa ada tiga istilah yang sering dipergunakan di bidang pendidikan secara bergantian. Istilah tersebut adalah penilaian (*evaluation*), pengukuran (*measurement*), dan tes (*test*). Gronlund (dalam Nurgiyantoro, 2013: 6) membedakan dengan jelas antara penilaian / evaluasi, pengukuran dan tes. Penilaian atau evaluasi adalah proses sistematis dalam pengumpulan, analisis, dan penafsiran informasi untuk menentukan seberapa jauh seorang peserta didik dapat mencapai tujuan pendidikan.

Pengukuran merupakan proses untuk memperoleh deskripsi angka (skor) yang menunjukkan tingkat capaian seseorang dalam suatu bidang tertentu. Tes merupakan sebuah instrumen atau prosedur yang sistematis untuk mengukur suatu sampel tingkah laku, misalnya untuk menjawab pertanyaan “seberapa baik (tinggi) kinerja seseorang” yang jawabnya suatu angka.

Penilaian merupakan salah satu kegiatan dalam pembelajaran. Penilaian pendidikan merupakan proses pengumpulan dan pengolahan

informasi untuk menentukan pencapaian hasil belajar peserta didik (Depdiknas, 2007: 1). Semua kegiatan pembelajaran harus selalu diikuti dengan kegiatan penilaian. Ada beberapa tujuan dari penilaian, di antaranya (1) untuk menilai hasil belajar peserta didik, (2) menilai kualitas pembelajaran yang telah dilakukan, dan (3) sebagai umpan balik kegiatan pembelajaran selanjutnya (Nurgiyantoro, 2013: 3-5).

Kegiatan penilaian memerlukan alat, yaitu yang dikenal sebagai alat penilaian. Secara garis besar alat penilaian dapat dibedakan menjadi dua macam, yaitu teknik tes dan teknik nontes. Keduanya digunakan untuk mendapatkan informasi atau data-data penilaian tentang subjek belajar yang dinilai secara berhasil guna apabila dipakai secara tepat.

Ada sejumlah teknik nontes yang dapat dipergunakan untuk memperoleh informasi hasil belajar atau informasi tentang peserta didik antara lain teknik kuisisioner, pengamatan, daftar cocok, wawancara, proyek, portofolio, dan lain-lain (Nurgiyantoro, 2013: 90). Kemudian teknik tes adalah pengumpulan informasi yang dilakukan lewat pemberian seperangkat tugas, latihan, atau pertanyaan yang harus dikerjakan oleh peserta didik. Teknik tes antara lain ulangan harian, ulangan tengah semester, ulangan akhir semester, ulangan kenaikan kelas, ujian sekolah dan lainnya (Nurgiyantoro, 2013: 105).

Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan pasal 1 menyatakan bahwa penilaian hasil belajar peserta didik jenjang pendidikan dasar dan menengah

dilaksanakan berdasarkan standar penilaian pendidikan yang berlaku secara nasional. Standar penilaian pendidikan yang dimaksud adalah yang berkaitan dengan mekanisme, prosedur, dan instrumen penilaian hasil belajar peserta didik.

Beberapa jenis penilaian yang ada dalam Permendiknas nomor 20 tahun 2007 adalah ulangan harian, ulangan tengah semester (UTS), ulangan akhir semester (UAS), ulangan kenaikan kelas (UKK), ujian sekolah, dan ujian nasional.

2. Analisis Butir Soal

Analisis butir soal adalah identifikasi jawaban benar dan salah tiap butir soal yang diujikan kepada peserta didik (Nurgiyantoro, 2013: 190). Melalui kerja analisis ini akan diketahui butir-butir soal mana saja yang dijawab benar oleh peserta tes atau sebaliknya, yaitu butir-butir soal mana saja yang dijawab salah oleh peserta tes. Berdasarkan jawaban benar dan salah itulah akan dihitung validitas, reliabilitas, indeks tingkat kesukaran, daya beda maupun efektivitas butir pengecohnya.

Analisis butir soal juga dapat sebagai estimasi kualitas butir soal, yaitu sejauh mana butir-butir soal itu berfungsi dengan baik. Alat tes yang didukung oleh butir-butir soal yang baik, efektif, dapat dipertanggungjawabkan kualitasnya. Analisis butir soal merupakan analisis hubungan antara skor-skor butir soal dengan skor keseluruhan,

membandingkan jawaban peserta didik terhadap butir soal dengan jawaban terhadap keseluruhan tes.

Ada keterkaitan antara reliabilitas dengan analisis butir soal. Dalam sebuah pengukuran, jika diketahui koefisien reliabilitasnya tinggi, maka keadaan lazimnya didukung dengan butir-butir soal yang layak baik dilihat dari indeks tingkat kesukaran maupun daya beda. Demikian sebaliknya, jika koefisien reliabilitasnya rendah, maka butir-butir soalnya kebanyakan tidak layak.

Secara keseluruhan hasil pengujian sebuah tes mungkin reliabilitasnya kurang tinggi, akan tetapi tidak semua butir soal yang ada tidak layak sehingga semuanya direvisi. Sebab tentunya terdapat sejumlah butir soal yang memenuhi kriteria kelayakan dan karenanya dapat dipertahankan. Untuk memilih butir-butir soal yang layak, dan sebaliknya perlu revisi, sekali lagi butir soal tes itu dapat ditentukan berdasarkan kerja analisis butir soal. Analisis butir soal akan memberikan jawaban terhadap maksud itu yang dapat dipertanggungjawabkan karena sanggup memberikan informasi secara rinci tentang keadaan masing-masing soal, yaitu indeks tingkat kesulitan, daya beda, kesahihan, reliabilitas dan juga efektivitas sebaran jawaban.

Jika suatu soal tes hasil belajar telah valid, maka soal tes hasil belajar itu juga reliabel. Perakitan soal seperti itu di dalam suatu tes hasil belajar tidak akan menimbulkan masalah (Waridjan, 1983: 381). Tetapi, sejumlah masalah akan timbul apabila soal tes yang tidak valid terakit di

dalam tes hasil belajar. Masalah itu dapat bersumber dari semua pengukuran tes hasil belajar yang menggunakan soal tes tidak valid sebagai alat ukur. Hal tersebut harus dianggap sebagai tantangan untuk memperbaiki sesuatu soal tes sedemikian rupa sehingga jenjang validitas soal tes itu memenuhi kriteria yang dipersyaratkan. Bahan-bahan masukan untuk memperbaiki suatu soal tes dapat diperoleh melalui analisis soal tes hasil belajar. Analisis soal tes yang praktis untuk kepentingan sehari-hari para guru ialah analisis indeks kesukaran, indeks daya beda dan juga efektivitas jawaban-jawaban soal yang disediakan.

Nurgiyantoro (2013: 190-191) menyatakan bahwa ada 2 (dua) teori pengukuran terkait dengan analisis butir soal, yaitu teori pengukuran klasik (*Classical Measurement Theory*) dan teori respon butir (*Item Response Theory*). Teori respon butir muncul sebagai reaksi dan koreksi terhadap sejumlah kelemahan teori pengukuran klasik. Salah satu kelemahan teori pengukuran klasik adalah adanya saling ketergantungan (interdependensi) antara peserta tes yang diuji dengan tingkat kesulitan butir-butir soal tersebut. Jika peserta tes termasuk kelompok pandai, butir-butir soal itu akan menjadi mudah. Demikian juga sebaliknya, butir-butir soal akan menjadi sulit jika kelompok tesnya adalah kelompok rendah. Idealnya, kesulitan butir-butir soal, sesuai dengan pandangan teori respon butir, bersifat konstan, tidak peduli kelompok manapun yang diuji.

Persyaratan analisis untuk teori respon butir tidaklah mudah, terutama yang berkaitan dengan jumlah peserta didik yang harus jauh lebih

banyak daripada dalam teori klasik, demikian juga dalam hal kerja analisisnya. Oleh karena itu, teori pengukuran klasik lebih mudah dilakukan karena jumlah peserta didik yang dituntut lebih sedikit, misalnya cukup 30-40 peserta didik saja. Kerja analisisnya juga dapat dilakukan secara manual (dengan kalkulator) maupun dapat juga dengan program komputer.

Purwanto (1984: 118) menyatakan bahwa salah satu cara untuk memperbaiki proses belajar mengajar yang paling efektif ialah dengan jalan mengevaluasi tes hasil belajar yang diperoleh dari proses belajar mengajar itu sendiri. Dengan kata lain, hasil tes itu diolah sedemikian rupa sehingga dari hasil pengolahan itu dapat diketahui komponen-komponen manakah yang masih lemah.

Pengolahan tes hasil belajar dapat dilakukan dengan berbagai cara, antara lain, dengan membuat analisis soal (*item analysis*) dan dengan menghitung validitas dan keandalan tes. Selanjutnya Thorndike dan Hagen (dalam Purwanto, 1984: 118) menyatakan bahwa analisis terhadap soal-soal (*items*) tes yang telah dijawab oleh peserta didik mempunyai dua tujuan yang penting.

Pertama, jawaban-jawaban soal itu merupakan informasi diagnostik untuk meneliti pembelajaran dari kelas itu dan kegagalan-kegagalan belajarnya, serta selanjutnya untuk membimbing ke arah cara belajar yang lebih baik.

Kedua, jawaban-jawaban terhadap soal-soal yang terpisah dan perbaikan (*review*) soal-soal didasarkan atas jawaban-jawaban itu merupakan basis bagi penyiapan tes-tes yang lebih baik untuk waktu berikutnya.

Jadi, tujuan dari *items analysis* ialah mencari soal tes mana yang baik dan mana yang tidak baik, dan mengapa item atau soal itu dikatakan baik atau tidak baik. Dengan mengetahui soal-soal yang tidak baik itu selanjutnya dapat dicari kemungkinan sebab-sebab mengapa item itu tidak baik. Dengan membuat analisis soal, sedikitnya dapat diketahui tiga hal penting yang dapat diperoleh dari tiap soal itu, yaitu:

1. Bagaimana tingkat atau taraf kesukaran soal itu (*difficulty level of an item*);
2. Apakah soal itu mempunyai daya pembeda (*discriminating power*) sehingga dapat membedakan kelompok siswa yang pandai dengan kelompok siswa yang kurang pandai;
3. Apakah semua alternatif jawaban (*option*) menarik, ataukah ada yang tidak menarik sehingga tidak perlu dimasukkan de dalam soal.

Selanjutnya Daryanto (2010: 177-178) menyatakan analisis soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes. Adapun faedah mengadakan analisis butir soal, yaitu:

1. Membantu guru dalam mengidentifikasi butir-butir soal yang jelek.

2. Memperoleh informasi yang dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan selanjutnya.
3. Memperoleh gambaran keadaan sebuah butir soal secara empiris.

Dari penjelasan di atas, dapat diketahui bahwa analisis butir soal bertujuan untuk mengidentifikasi soal-soal yang baik, kurang baik dan soal yang jelek. Dengan analisis, diperoleh informasi tentang sebuah soal untuk dibuang atau direvisi. Identifikasi tersebut dapat dari tingkat validitas, reliabilitas, tingkat kesulitan, daya beda serta dapat juga untuk mengetahui efektifitas butir pengecoh khususnya untuk soal dalam bentuk pilihan ganda.

3. Ujian Sekolah SMA

Ujian sekolah /madrasah adalah kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan oleh satuan pendidikan untuk memperoleh pengakuan atas prestasi belajar dan merupakan salah satu persyaratan kelulusan dari satuan pendidikan. Mata pelajaran yang diujikan adalah mata pelajaran kelompok mata pelajaran ilmu pengetahuan alam dan sosial serta yang diujikan dalam ujian nasional dan aspek kognitif dan atau psikomotorik kelompok mata pelajaran agama dan akhlak mulia serta kelompok mata pelajaran kewarganegaraan dan kepribadian yang akan diatur dalam POS Ujian Sekolah/Madrasah (Depdiknas, 2007:3)

Selain itu, ujian sekolah juga digunakan sebagai pemetaan mutu satuan pendidikan, khususnya untuk mata pelajaran Bahasa Indonesia, matematika, IPA pada SMA/SMK yang dilakukan oleh Kementerian.

Selanjutnya dijelaskan bahwa instrumen penilaian dalam bentuk ujian sekolah harus memenuhi persyaratan substansi, konstruksi, dan bahasa, serta bukti validitas empirik. Hal ini bertujuan bahwa untuk mengetahui validitas empirik maka perlu menganalisis hasil dari peserta ujian sekolah.

Bahan ujian sekolah untuk ketiga mata pelajaran sebagaimana tersebut, terdiri atas 25 % paket soal yang ditetapkan oleh kementerian dan 75 % paket soal yang ditetapkan oleh provinsi dengan melibatkan pendidik dari satuan pendidikan yang merupakan perwakilan dari kabupaten/kota se-Jawa Barat. Untuk pengolahan hasil ujian sekolah dilaksanakan oleh tim pemindai dari kabupaten/ kota yang bersangkutan, kemudian hasilnya dikirim ke pemerintah provinsi. Selanjutnya oleh pemerintah provinsi dilakukan penyekoran hasil ujian sekolah masing-masing peserta.

Peserta didik dinyatakan lulus ujian sekolah apabila memenuhi kriteria kelulusan yang telah ditetapkan oleh satuan pendidikan berdasarkan perolehan nilai hasil ujian sekolah tersebut. Kriteria kelulusan ditetapkan melalui rapat pendidik pelaksana ujian sekolah yang mencakup (1) nilai minimal setiap mata pelajaran ujian sekolah, (2) nilai rata-rata minimal mata pelajaran ujian sekolah. Peserta didik dinyatakan lulus dari datuan pendidikan setelah, (1) menyelesaikan seluruh program

pembelajaran, (2) memperoleh nilai minimal baik pada penilaian akhir untuk mata pelajaran dan muatan local, dan (3) lulus ujian sekolah (POS US, 2014: 18-19).

4. Kriteria Tes yang Baik

Tes (dalam hal ini, soal ujian sekolah) sebagai salah satu alat pengukur hasil peserta didik diharapkan mampu memberikan informasi yang dapat dipertanggungjawabkan kebenarannya. Artinya, soal ujian sekolah tersebut dapat memberikan informasi tentang peserta didik sesuai dengan keadaan yang sesungguhnya (Nurgiyantoro, 2013: 149). Data penilaian yang baik adalah data yang sesuai dengan kenyataan yang sebenarnya dan data tersebut bersifat tetap, ajeg, dan dapat dipercaya. Arikunto (2009: 57) menyatakan bahwa suatu tes dapat dikatakan baik apabila memenuhi lima persyaratan, yaitu: validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis.

Permendiknas RI Nomor 20 tahun 2007 tentang standar penilaian pendidikan menyatakan bahwa penilaian hasil belajar peserta didik pada jenjang pendidikan dasar dan menengah didasarkan pada prinsip-prinsip sebagai berikut:

- a. Sahih, berarti penilaian didasarkan pada data yang mencerminkan kemampuan yang diukur.
- b. Objektif, berarti penilaian didasarkan pada prosedur dan kriteria yang jelas, tidak dipengaruhi subjektivitas penilaian.

- c. Adil, berarti penilaian tidak menguntungkan atau merugikan peserta didik karena berkebutuhan khusus serta perbedaan latar belakang agama, suku, budaya, adat-istiadat, status social ekonomi, dan gender.
- d. Terpadu, berarti penilaian oleh pendidik merupakan salah satu komponen yang tak terpisahkan dari kegiatan pembelajaran.
- e. Terbuka, berarti prosedur penilaian, kriteria penilaian, dan dasar pengambilan keputusan dapat diketahui oleh pihak yang berkepentingan.
- f. Menyeluruh dan berkesinambungan, berarti penilaian oleh pendidik mencakup semua aspek kompetensi dengan menggunakan berbagai teknik penilaian yang sesuai untuk memantau perkembangan kemampuan siswa.
- g. Sistematis, berarti penilaian dilakukan secara berencana dan bertahap dengan mengikuti langkah-langkah baku.
- h. Beracuan kriteria, berarti penilaian didasarkan pada ukuran pencapaian kompetensi yang ditetapkan.
- i. Akuntabel, berarti penilaian dapat dipertanggungjawabkan, baik dari segi teknik, prosedur, maupun hasilnya (Depdiknas 2007: 2).

Agar tes yang digunakan itu dapat diharapkan sesuai prinsipnya, maka soal tes harus benar-benar memenuhi beberapa kriteria. Kriteria itu antara lain validitas, reliabilitas, dan objektivitas, (Sutomo, 1983: 48-50). Gronlund (dalam Nurgiyantoro, 2013: 150) menyatakan bahwa secara esensial alat tes harus memiliki kualitas validitas (*validity*), reliabilitas

(*reliability*), dan kebergunaan (*usability*). Hal tersebut tidak berbeda dengan apa yang dikemukakan Tuckman (dalam Nurgiyanto 2013: 150) bahwa alat tes mesti dapat dipertanggungjawabkan dari segi kelayakan (*appropriateness*), validitas (*validity*), reliabilitas (*reliability*), ketertafsiran (*interpretability*), dan kebergunaan (*usability*). Untuk kriteria ketertafsiran dan kebergunaan biasanya dicakup dalam kriteria kepraktisan (*practicality*).

Instrumen penilaian yang digunakan oleh satuan pendidikan dalam bentuk ujian sekolah/ madrasah harus memenuhi persyaratan substansi, konstruksi, dan bahasa, serta memiliki bukti validitas empirik (Depdiknas, 2007: 2). Waridjan (1991: 341) menyatakan bahwa kriteria tes yang dapat dipertanggungjawabkan meliputi validitas, reliabilitas, dan analisis butir soal.

Berdasarkan uraian di atas, dalam kajian teoretis ini berturut-turut akan diuraikan beberapa hal yang berkaitan dengan kriteria tes yang baik dan tujuan penelitian ini, yakni tentang validitas, reliabilitas, tingkat kesulitan butir soal, daya beda dan efektivitas butir pengecoh (distraktor).

a. Kriteria Validitas

Sahih (*valid*), artinya bahwa alat tersebut (*tes*) hanya mengukur satu dimensi/ aspek saja, misal tes matematika, hanya mengukur pengetahuan matematika saja, tidak mengukur pengetahuan lain (Depdikbud, 1992: 10). Sahih, berarti penilaian didasarkan pada data yang mencerminkan kemampuan yang diukur (Depdiknas, 2007: 1). Selanjutnya instrumen (*tes*) dikatakan valid apabila instrumen tersebut

dapat dengan tepat mengukur apa yang hendak diukur atau ketepatan alat ukur.

Gronlund (dalam Nurgiyantoro, 2009: 150-151) menyatakan bahwa validitas menunjuk pada kelayakan interpretasi yang dibuat berdasarkan skor hasil tes yang berkaitan dengan penggunaan tertentu. Hal-hal yang perlu dipertimbangkan, (1) validitas menunjuk pada kelayakan interpretasi yang dibuat berdasarkan skor hasil tes yang berkaitan dengan penggunaan tertentu dan bukan terhadap instrumennya itu sendiri, (2) validitas adalah kadar (*matter of degree*), maka harus dihindari pemikiran bahwa sebuah tes itu valid atau tidak valid, (3) validitas berkaitan dengan penggunaan khusus karena tidak ada satu tes pun yang valid untuk semua tujuan.

Jadi, validitas merupakan dukungan bukti dan teori terhadap penafsiran sesuai dengan tujuan penggunaan tes (Mardapi dalam Nurgiyantoro, 2013: 152). Proses validasi merupakan pengumpulan bukti-bukti untuk menunjukkan saintifik penafsiran skor sebagaimana yang direncanakan validitas adalah penafsiran hasil skor tes dan bukan alat tes nya itu sendiri.

Thoha (1994: 109) menyatakan bahwa suatu alat ukur disebut memiliki validitas apabila alat ukur tersebut isinya layak mengukur objek yang seharusnya diukur dan sesuai kriteria tertentu. Artinya, ada kesesuaian antar alat ukur dengan fungsi pengukuran dan sasaran pengukuran.

Menurut beberapa ahli, di antaranya : Anderson (dalam Arikunto 1984: 59) menyatakan bahwa suatu tes dinyatakan valid apabila tes itu mengukur suatu yang dimaksud untuk diukur. Garet (dalam Waridjan 1991; 342) menyatakan validitas dari suatu tes atau alat ukur apapun tergantung pada kemurnian pengukuran terhadap apa yang dimaksud untuk diukur, demikian juga Cronbach (Dalam Waridjan 1991: 343) secara sempit menyatakan bahwa validitas atau penyahihan merupakan proses pengujian skor tes hasil belajar dibandingkan dengan beberapa hasil observasi lain sebagai kriteria. Penyelenggaraan tes bertujuan untuk meramalkan apakah kriteria lain telah tercapai oleh peserta tes dan jasa dinilai pada ketepacermatan ramalan itu. Purwanto (1984: 137) menyatakan validitas adalah kualitas yang menunjukkan hubungan antara suatu pengukuran dengan arti atau tujuan kriteria belajar atau tingkah laku.

Dari beberapa pendapat ahli di atas, masing masing mengemukakan pernyataan yang kesemuanya mengandung arti saling mengisi pertama menekankan pada tepatnya sasaran pengukuram, kedua menegaskan perlu murninya tindakan pengukuran yang terkahir bahwa penyahihan merupakan upaya untuk memperoleh hasil pengukuran yang murni melalui perbandingan hasil pengukuran-pengukuran lain yang di anggap sebagai kriteria.

Macam-macam validitas tes hasil belajar dapat di klasfikasikan dari beberapa pendapat (Arikunto 1984: 54), (Nurkancana, 1983: 124-

126), (Nurgiyantoro, 2013: 154-159), (Waridjan, 1991:344-347) sebagai berikut : (1) validitas logis atau validitas kurikuler terdiri atas isi dan validitas konstruk, (2) validitas empirik atau validitas statistik terdiri atas validitas sejajar dan validitas prediktif.

Istilah “validitas logis” berasal dari kata “logika”. Yang berarti penalaran dengan demikian maka validitas logis untuk sebuah instrumen evaluasi menunjuk kepada kondisi bagi sebuah instrumen yang memenuhi syarat valid berdasarkan hasil penalaran dan sudah dirancang secara baik sesuai dengan teori ketentuan yang berlaku (Arikunto, 1984: 65).

Tes hasil belajar yang telah dilakukan penganalisaan secara rasional ternyata memiliki daya ketepatan mengukur, disebut tes hasil belajar yang sudah memiliki logika, validitas logis sering disebut juga dengan validitas rasional (Sugiyono, 2010: 163-164).

Validitas isi adalah pengujian validitas dilakukan atas isinya untuk memastikan apakah butir tes hasil belajar mengukur secara tepat keadaan yang ingin diukur (Purwanto, 1984: 120) suatu tes hasil belajar dapat di katakan valid apabila materi tes tersebut betul-betul merupakan bahan-bahan yang representatif maka butir-butir tes hasil belajar harus dilaksanakan pada perencanaan kisi-kisi. Pengujian validasi isi yang dilakukan dengan mencermati kesesuaian isi butir yang dituliskan dengan perencanaan yang dituangkan dalam kisi-kisi. Kriteria yang menjadi dasar pengujian validasi ini adalah kisi-kisi yang

direncanakan. Butir-butir tes hasil belajar dinyatakan valid (*logically valid*) apabila setelah mencermati isi butir-butir yang ditulis telah menunjukkan kesesuaian dengan kisi-kisi (Purwanto, 1984: 120).

Validitas isi menurut Gronlund dan Popham (dalam Nurgiyantoro, 2013: 156) adalah validitas yang pembuktiannya berdasarkan isi. Gronlund memaknai validitas isi sebagai proses penentuan sejauh mana suatu alat tes menunjukkan kerelevansian dan keterwakilan terhadap ranah tugas yang diukur.

Validasi konstruk adalah suatu tes di mana butir soal tersebut membangun setiap aspek berpikir seperti yang disebutkan dalam tujuan instruksional khusus (Thoha, 1994: 110) menurut Benjamin S bloom bahwa taksonomi (pengelompokan) tujuan pendidikan harus senantiasa mengacu kepada tiga jenis ranah (*domain*) yang melekat pada diri peserta didik, yaitu : (1) ranah proses berpikir (*cognitive domain*), (2) ranah nilai atau sikap (*affective domain*), (3) ranah keterampilan (*psychomotor domain*). Setiap domain disusun menjadi beberapa jenjang kemampuan, mulai dari hal yang sederhana sampai dengan hal yang kompleks, mulai dari yang mudah sampai dengan hal yang sukar, dan mulai dari hal yang konkret sampai dengan yang abstrak.

Validitas empiris adalah memiliki pengertian pengalaman, sehingga sebuah instrumen dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Dengan validitas empiris tidak dapat diperoleh hanya dengan jalan menyusun instrumen berdasarkan

ketentuan seperti halnya validitas logis, tetapi harus dibuktikan dengan hasil analisis yang dilakukan terhadap data hasil pengamatan di lapangan, terbukti bahwa tes hasil belajar itu dengan secara tepat dapat mengukur hasil belajar yang seharusnya diukur (Sudijono, 2010:167-168).

Ada dua acara untuk mengetahui tes hasil belajar itu sudah memiliki validitas empiris ataukah belum, yakni dari segi ketepatan meramalnya (*predictive validity*) dan ketepatan bandingnya atau “ada sekarang” (*concurrent validity*) (Sudijono. 2010: 167-168).

Validitas ramalan adalah ketepatan dari pada suatu alat penilaian tersebut dapat digunakan untuk meramalkan suatu ciri atau kriteria tertentu yang dapat diinginkan, dengan kata lain validitas ini mengandung ciri adanya relevansi dan keajegan atau ketepatan (*reliability*) (Sudjana, 1984: 15).

Validitas banding adalah ketepatan daripada suatu tes dilihat dari korelasinya terhadap kecakapan yang telah dimiliki saat ini secara riil. Perbedaan antara validitas ramalan dengan validitas bandingan adalah dari segi waktunya.

Pengujian validitas / kesahihan suatu alat tes dapat dilakukan secara keseluruhan dan secara per butir tes, pengujian secara keseluruhan, artinya alat tes dipandang sebagai satu keseluruhan tanpa memperhitungkan keadaan masing masing butir tes secara sendiri. Jika melalui pengujian itu ditemukan bahwa tes yang bersangkutan sah,

kesahihan itu menyangkut alat tes itu secara keseluruhan, dan belum tentu butir soalnya.

Validitas/kesahihan butir dari suatu tes adalah ketepatan mengukur yang dimiliki oleh sebutir soal (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir soal tersebut (Sudijono, 2010:182).

Apabila kita mau memperhatikan secara cermat, maka tes hasil-hasil belajar yang dibuat atau disusun oleh para pengajar baik pendidik, pengajar, baik pendidik, dosen staf pengajar lainnya, sebenarnya adalah merupakan kumpulan dari sekian banyak butir-butir soal dengan soal mana para penyusun tes ingin mengukur atau mengungkap hasil belajar yang telah dicapai oleh masing-masing individu peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Pernyataan itu mengandung makna bahwa sebenarnya setiap butir soal yang ada dalam tes hasil belajar itu adalah merupakan bagian tak terpisahkan dari tes hasil tersebut sebagai suatu totalitas.

Erat hubungan antara butir soal dengan tes hasil belajar, belajar sebagai suatu totalitas itu kiranya dapat dipahami dari kenyataan, bahwa semakin banyak butir-butir soal yang dapat dijawab dengan betul oleh testee, maka skor-skor total hasil tes tersebut akan semakin tinggi dengan betul oleh testee, maka skor-skor total hasil tes itu akan semakin rendah atau semakin menurun.

Kesahihan butir soal adalah analisis tingkat kesahihan butir-butir tes untuk mengukur tingkat kesahihan masing-masing butir soal, perlu dilakukan analisis butir soal tes yang bersangkutan. Analisis butir soal adalah analisis jawaban siswa terhadap butir-butir soal tes yang diujicobakan jawaban siswa terhadap butir-butir soal itu mungkin benar, mungkin salah. Benar atau salahnya jawaban siswa terhadap masing-masing butir soal itulah yang menjadi bahan analisis untuk mengukur tingkat kesahihan butir-butir yang bersangkutan.

Sebuah butir soal dapat dikatakan telah memiliki validitas yang tinggi atau dapat dikatakan valid jika skor-skor pada butir soal yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya (Sudijono, 2010: 184) artinya peserta didik yang pandai relatif dapat menjawab benar terhadap soal yang bersangkutan demikian sebaliknya.

Sebuah instrumen tes dikatakan memiliki validitas jika sesuai dengan kriteria tertentu, dalam arti memiliki kesejajaran antara hasil pengukuran dengan kriteria tersebut, cara yang digunakan untuk mengetahui kesejajaran adalah dengan cara mengorelasikan hasil pengukuran dengan kriteria. Kriteria yang digunakan sebagai patokan untuk menilai validitas sebuah instrumen dapat berupa hasil tes.

Teknik korelasi yang digunakan untuk mengetahui kesejajaran adalah teknik korelasi *product moment* dari Carl Person, rumus korelasi *product moment* ada dua macam, Yaitu :

- a. Korelasi *Product Moment* dengan deviasi atau simpangan baku.
- b. Korelasi *Product Moment* dengan angka kasar.

Tinggi rendahnya koefisien korelasi yang di peroleh menunjukkan tinggi rendahnya kadar kesahihan ramalan alat tes yang di uji kesahihannya. Menurut Purwanto (2009: 139) menyatakan bahwa penafsiran terhadap besar kecilnya koefisien korelasi dapat dengan mendasarkan diri pada ketentuan berikut :

- a. Koefisien 0.90 sampai dengan 1.00 = sangat tinggi
- b. Koefisien 0.70 sampai dengan 0.90 = tinggi
- c. Koefisien 0.40 sampai dengan 0.70 = cukup
- d. Koefisien 0.20 sampai dengan 0.40 = rendah
- e. Koefisien 0.00 sampai dengan 0.20 = sangat rendah

Dari tingkatan koefisien tersebut di atas dapat di jelaskan bahwa semakin besar koefisien korelasinya maka semakin tinggi pula tingkat validitas alat tes yang di uji.

Selain itu, untuk menafsirkan koefisien korelasi dapat menggunakan tabel. Penggunaan tabel akan lebih akurat bila dibandingkan dengan penafsiran koefisien korelasi seperti di atas, sebab berdasar ketentuan interval di atas anda angka kembar dalam dua interval (0,20; 0,40; 0,70;0,90) sehingga akan membingungkan bila kita memperoleh indeks angka tersebut. Penafsiran harga koefisien korelasi melalui tabel dilakukan dengan membandingkan harga r_{xy} hasil perhitungan dengan harga r_{xy} yang ada instrumen tes dapat dikatakan valid. Begitu juga sebaliknya apabila r_{xy} hitung lebih kecil dari r_{xy} tabel

$(r_h < r_i)$ berarti korelasi bersifat tidak signifikan, kesimpulan instrumen tes tidak valid.

Selain mencari validitas secara keseluruhan perlu juga dicari validitas per butir instrumen (soal). Jika validitas instrumen rendah, maka perlu diketahui butir-butir instrumen mana yang menyebabkan instrumen keseluruhan tersebut jelek, untuk keperluan inilah perlunya mencari validitas butir instrumen.

Sebuah butir soal dikatakan valid apabila memiliki sumbangan yang besar terhadap skor total. Dengan kata lain, sebuah butir soal dikatakan mempunyai validitas tinggi jika skor pada butir soal itu mempunyai kesejajaran dengan skor total. Kesejajaran ini dapat diartikan dengan korelasi, sehingga untuk mengetahui validitas butir soal digunakan rumus korelasi *product moment*, seperti yang dibahas sebelumnya atau dapat juga menggunakan salah satu cara yang terkenal yaitu rumus korelasi *point biserial* (Arikunto, 2009: 78).

Penafsiran harga korelasi butir instrumen dilakukan dengan membandingkan r_{xy} hasil perhitungan dengan r_{xy} yang ada dalam tabel harga kritik *product moment* sehingga dapat diketahui signifikan tidaknya korelasi tersebut. Apabila r_{xy} hitung lebih besar atau sama dengan r_{xy} tabel ($r_h \geq r_i$) berarti korelasi bersifat signifikan, artinya butir soal dapat dikatakan valid, begitu juga sebaliknya, apabila r_{xy} hitung lebih kecil dari r_{xy} tabel ($r_h < r_i$) berarti korelasi tidak signifikan, kesimpulan instrumen butir tes tidak valid.

b. Kriteria Realibilitas

Reliabel (terandal), artinya bahwa alat tes tersebut harus dapat memberikan hasil pengukuran yang tepat, cermat dan ajeg (kurang lebih sama walau diberikan pada waktu yang berbeda) (Depdikbud, 1992:11). Nurkencana (1986:131) menyatakan bahwa suatu tes dikatakan tes yang realibilitas apabila tes tersebut menunjukkan hasil yang mantap. Sedangkan Nurgiyantoro (2013:165) menyatakan realibilitas tes adalah jika sebuah tes diujicobakan lebih dari satu kali kepada subjek yang sama dapat menghasilkan data yang kurang lebih sama. Alat tersebut dapat mengukur secara konsisten, secara ajeg.

Arikunto (2009: 86) menyatakan, realibilitas berhubungan dengan masalah kepercayaan. Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Pengertian realibilitas tes berhubungan dengan masalah ketetapan hasil tes. Atau seandainya hasilnya berubah ubah, perubahan yang terjadi dapat dikatakan tidak berarti.

Pengertian konsisten (ajeg) dalam kepercayaan tes berhubungan dengan hal hal (a) tes dapat memberikan hasil yang relatif tetap terhadap sesuatu yang diukur, (b) jawaban siswa terhadap butir butir tes secara relatif tetap, dan (c) hasil tes diperiksa oleh siapapun juga akan menghasilkan skor yang kurang lebih sama. Ketiga hal tersebut merupakan sesuatu yang akan mempengaruhi tinggi atau rendahnya tingkat kepercayaan tes (Nurgiyantoro, 2013:166).

Purwanto (2009:139) menyatakan, keandalan (*reliability*) adalah ketetapan atau ketelitian suatu alat evaluasi. Suatu tes atau alat evaluasi dapat dikatakan andal (*reliable*) jika ia dapat dipercaya, konsisten, atau stabil dan produktif. Jadi, yang dipentingkan di sini adalah ketelitiannya, sejauh mana tes atau alat tersebut dapat dipercaya kebenarannya.

Instrumen tes dikatakan dapat dipercaya (*reliable*) jika memberikan hasil yang tetap, ajek (konsisten) apabila diteskan berkali-kali. Jika kepada siswa diberikan tes yang sama pada waktu yang berlainan, maka setiap siswa akan tetap berada dalam urutan (*ranking*) yang sama atau ajek dalam kelompoknya.

Perlu dikemukakan bahwa suatu pengukuran terhadap kemampuan tertentu yang dilakukan dua kali dalam kondisi subjek yang sama tidak akan menghasilkan data yang persis sama. Dengan melakukan pengukuran, kita bermaksud mendapatkan skor yang sesungguhnya yang dicapai seorang peserta didik.

Dari beberapa pendapat di atas, dapat disimpulkan bahwa reliabilitas alat penelitian adalah ketetapan atau keajegan alat tersebut dalam menilai apa yang dinilainya. Artinya, kapan pun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama. Tes hasil belajar dikatakan ajek apabila hasil pengukuran kesamaan hasil pada saat yang berlainan waktunya terhadap peserta didik yang sama.

Beberapa faktor yang mempengaruhi realibilitas, di antaranya :

- 1) Luas tidaknya suatu sampling yang diambil makin luas suatu sampling, berarti makin andal.
- 2) Perbedaan bakat dan kemampuan murid yang dites Makin variabel kemampuan peserta tes, berarti makin tinggi keandalan koefisien tes. Tes yang diberikan kepada beberapa tingkat kelas yang berbeda lebih tinggi keandaiannya daripada yang hanya diberikan kepada beberapa kelas yang sama karena tingkat kelas yang berbeda akan menghasilkan *achievement* yang lebih luas.

3) Suasana dan kondisi testing

Suasana ketika berlangsung seperti tenang, gaduh, banyak gangguan, pengetes, yang marah-marah dapat mengganggu pengerjaan tes sehingga dengan demikian mempengaruhi pula hasil dan keandalan tes (Purwanto, 1984:141).

Keandalan suatu tes dinyatakan dengan *coefisient of reliability* (r), yaitu dengan jalan mencari korelasi, misalnya dengan cara:

1) Dengan metode dua tes

Dua tes yang paralel dan setaraf (ekuivalen) diberikan kepada sekelompok peserta didik. Hasil kedua tes tersebut kemudian dicari korelasinya, Dalam hal ini dapat juga digunakan metode Pearson dan metode Spearman.

2) Dengan metode satu tes

Sebuah tes diberikan dua kali kepada sekelompok peserta didik yang sama, tetapi dalam waktu yang berbeda. Kedua hasil tes itu kemudian dicari korelasinya.

3) Metode *split half*

Satu tes diberikan kepada sekelompok peserta didik. Kemudian dilihat skor masing-masing bagian paruhan tes tersebut dan dicari korelasinya.

Ada beberapa cara yang dapat dipergunakan untuk mencari taraf reliabilitas suatu tes khususnya untuk penelitian ini adalah reliabilitas internal. Karena dalam penelitian ini bentuk soalnya pilihan ganda atau instrumen skornya diskrit, yang skor jawaban/ responnya hanya dua, yaitu 1 (satu) dan 0 (nol). Dengan kata lain hanya ada dua jawaban, yaitu benar atau salah. Jawaban benar diberi skor 1 (satu) sedangkan jawaban salah diberi skor 0 (nol). Untuk instrumen yang skornya diskrit, tingkat reliabilitasnya dapat dicari dengan menggunakan (1) teknik belah dua (*split half*), (2) rumus Flanagan, (3) rumus Rulon, (4) rumus Kruder Richardson (KR) 20 dan 21, serta (5) rumus Hoyt (Widoyoko, 2012: 147).

Secara keseluruhan sebuah tes mungkin tidak terpercaya, akan tetapi tentunya tidak semua butir soal yang ada perlu direvisi. Sebab, tentunya terdapat sejumlah butir soal yang telah memenuhi kriteria keiayakan dan karenanya dapat dipertahankan. Untuk memilih butir-butir soal yang layak, dan yang perlu direvisi, dapat ditentukan

berdasarkan kerja analisis butir soal. Analisis butir soal akan memberikan jawab terhadap maksud itu secara dapat dipertanggungjawabkan karena ia sanggup memberikan informasi secara terinci tentang “keadaan” masing-masing butir soal, yaitu berdasarkan tingkat kesulitan (*item difficulty*) dan daya beda (*item discriminability*). Sebuah butir soal dinyatakan layak jika indeks tingkat kesulitan dan daya bedanya memenuhi standar yang ditentukan.

Sebuah tes yang dinyatakan tepercaya melalui suatu teknik pengujian, perlu juga dianalisis butir-butir soalnya. Sebab, belum tentu semua butir soal ada layak, mungkin terdapat sejumlah butir yang kurang layak. Hal itu pun dapat diketahui secara pasti melalui analisis butir soal.

Berdasarkan analisis butir soal, akan dapat diperbaiki dan ditingkatkan tingkat kepercayaan sebuah tes, yang telah dipergunakan sebelumnya. Analisis butir soal merupakan analisis hubungan antara skor-skor butir soal dengan skor keseluruhan, membandingkan jawaban siswa terhadap suatu butir soal dengan jawaban terhadap keseluruhan tes (Nurgiyantoro, 2013:190), menilai tes sebagai alat pengukuran, karena suatu alat tes jika tidak diuji, efektivitas pengukuran tidak dapat ditentukan secara memuaskan.

Analisis butir soal selain untuk mengetahui validitas dan realibilitas juga dilakukan untuk mengetahui tingkat kesukaran butir soal, daya beda dan efektivitas butir pengecoh (distraktor).

c. Tingkat Kesulitan Butir Soal

Nurgiyantoro (2013:194) menyatakan, indeks tingkat kesulitan (ITK) adalah indeks yang menunjukkan seberapa mudah atau sulit suatu butir soal bagi peserta tes yang diuji. Selanjutnya Oller (dalam Nurgiantoro, 2013:194) mengemukakan bahwa tingkat kesulitan (*item difficulty*) adalah pernyataan tentang seberapa mudah atau sulit sebuah butir soal bagi siswa yang dikenai pengukuran. Oller sendiri menggunakan istilah *item facility* karena hal yang sebenarnya dimaksud adalah seberapa besar butir soal memberi fasilitas atau kemudahan bagi siswa.

Daryanto (2010: 179) menyatakan bahwa indeks kesukaran adalah bilangan yang menunjukkan dan mudahnya soal. Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk memecahkannya karena di luar jangkauannya.

Butir soal yang baik adalah yang tingkat kesulitannya cukupan, tidak terlalu mudah atau terlalu sulit. Butir soal yang terlalu mudah atau terlalu sulit sama tidak baiknya karena keduanya tidak dapat membedakan antara siswa kelompok tinggi dan kelompok rendah. Butir soal yang demikian tidak memberikan informasi apa-apa tentang perbedaan prestasi antara tiap individu (Nurgiyantoro, 2013:195).

Tingkah kesukaran soal dipandang dari segi kesanggupan peserta didik atau kemampuan peserta didik waktu menjawab, bukan dilihat dari sudut pendidik sebagai pembuat soal. Persoalan yang penting melakukan analisis tingkat kesukaran soal yaitu penentuan proporsi dan kriteria soal yang termasuk mudah, sedang, dan sukar (Sudjana, 1984:135).

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang peserta didik untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menjadikan peserta didik putus asa dan tidak mempunyai semangat untuk mencoba karena diluar jangkauannya (Arikunto, 2009: 207).

Tingkat kesulitan butir soal dinyatakan dengan sebuah indeks yang berkisar antara 0,0 sampai dengan 0,1. Indeks 0,0 berarti butir soal yang bersangkutan sangat sulit karena tidak seorang siswa pun dapat menjawabnya. Sebaliknya, indeks 1,0 berarti butir soal yang bersangkutan sangat mudah karena semua siswa dapat menjawabnya dengan betul. Oller (dalam Nurgiyantoro 2013:128) mengemukakan bahwa suatu butir soal yang dinyatakan layak jika indeks tingkat kesulitannya berkisar 0,15 sampai dengan 0,85. Indeks di luar itu berarti butir soal terlalu mudah atau terlalu sulit.

Purwanto (2009:124) berpendapat bahwa derajat kesukaran (kesulitan) yang baik adalah berkisar 0,24 sampai dengan 0,76. Item

yang mempunyai derajat kesukaran di bawah 0,24 berarti bahwa item tersebut terlalu sulit. Sebaliknya item yang mempunyai derajat kesukaran di atas 0,76 berarti item tersebut terlalu mudah. Sedangkan Nurgiyantoro (2013: 196) menyatakan bahwa indeks tingkat kesulitan yang dapat ditoleransi berkisar 0,20 sampai dengan 0,80. Nurkencana (1986: 140) menyatakan, derajat kesukaran yang baik berkisar antara 0,25 sampai dengan 0,75. Selanjutnya Nurgiyantoro (2013: 195) membagi ITK soal menjadi tiga kategori yaitu: soal berkategori sulit 0,20 – 0,40; soal berkategori sedang 0,41 – 0,60 dan berkategori mudah 0,61 – 0,80.

Ada beberapa dasar pertimbangan dalam menentukan proporsi jumlah soal kategori mudah, sedang, dan sukar. Pertimbangan pertama adalah adanya keseimbangan, yakni jumlah soal sama untuk ketiga kategori mudah, sedang, dan sukar jumlahnya seimbang. Pertimbangan yang kedua proporsi jumlah soal untuk ketiga kategori tersebut didasarkan atas kurva normal. Artinya, sebagian besar soal berada dalam kategori sedang, sebagian lagi termasuk ke dalam kategori mudah dan sukar dengan proporsi seimbang.

Perbandingan antara soal mudah-sedang-sukar bisa dibuat 3-4-3. Artinya, 30% kategori mudah, 40% soal kategori sedang, dan 30% soal kategori sukar. Perbandingan lain yang sejenis dengan proporsi di atas 3-5-2. Artinya, 30% soal soal kategori mudah, 50% soal kategori sedang dan 20% kategori sukar (Sudjana, 1984: 135).

Dari beberapa pendapat tersebut, peneliti memilih perhitungan tingkat kesukaran butir soal yang dapat diterima berkisar antara 0,20 sampai dengan 0,80. Indeks diluar itu berarti soal terlalu sulit atau sebaliknya, soal terlalu mudah. Proporsi untuk kategori mudah-sedang-sukar adalah 3-5-2.

d. Daya Pembeda Butir Soal

Daya pembeda (*item discriminability*) maksudnya adalah seberapa besar suatu butir soal dapat membedakan antara siswa kelompok tinggi dengan kelompok rendah. Butir soal yang baik adalah yang dapat membedakan antara kedua kelompok tersebut secara layak. Hal itu berdasarkan logika bahwa siswa dari kelompok tinggi seharusnya dapat menjawab dengan betul yang lebih banyak dari pada kelompok rendah (Nurgiyantoro, 2013: 197).

Purwanto (2013: 120) menyatakan bahwa daya pembeda suatu soal tes ialah kemampuan soal itu untuk membedakan siswa-siswa yang termasuk kelompok pandai (*upper group*) dengan siswa-siswa yang termasuk kelompok kurang (*lower group*).

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh atau berkemampuan rendah (Daryanto, 2010: 183).

Daya pembeda soal adalah kemampuan suatu butir tes soal hasil belajar untuk dapat membedakan (mendiskriminasi) antara peserta didik yang berkemampuan tinggi, dengan peserta didik yang kemampuannya rendah (Sudijono, 2010: 385).

Artinya, bila soal tersebut diberikan kepada anak yang mampu, hasilnya menunjukkan prestasi yang tinggi, dan bila diberikan kepada siswa yang lemah, hasilnya rendah. Tes dikatakan tidak memiliki daya pembeda apabila tes tersebut, jika diujikan kepada anak berprestasi tinggi, hasilnya rendah, tetapi bila diberikan kepada yang lemah, hasilnya lebih tinggi. Atau bila diberikan kepada kedua kategori siswa tersebut, hasilnya sama saja. Dengan demikian, tes yang tidak memiliki daya pembeda, tidak akan menghasilkan gambaran hasil yang sesuai dengan kemampuan siswa yang sebenarnya. Sungguh aneh bila pandai tidak lulus, tetapi anak bodoh lulus dengan baik tanpa dilakukan manipulasi oleh si penilai atau di luar faktor kebetulan (Sudjana, 1984: 141).

Daya pembeda dihitung berdasarkan perbedaan jumlah jawaban betul untuk tiap butir soal antara kelompok tinggi dan kelompok rendah. Jika terjadi kelompok rendah menjawab betul lebih banyak dari pada kelompok tinggi, butir soal yang bersangkutan kurang baik karena menyalahi logika. Lebih jauh, hal ini berarti bahwa butir soal tersebut tidak tepercaya. Butir soal yang demikian, sebagai konsekuensinya perlu direvisi atau diganti.

Besar kecilnya daya pembeda suatu butir soal dinyatakan dengan suatu indeks yang berkisar antara -1,00 sampai dengan 1,00. Indeks yang semakin besar atau mendekati 1,00 berarti butir soal yang bersangkutan semakin baik, sebab semakin nyata perbedaan antara

kelompok tinggi dan kelompok rendah. Indeks negatif berarti siswa kelompok rendah justru menjawab dengan betul lebih banyak dari pada kelompok tinggi.

Nurgiyantoro (2013:197) menyatakan bahwa butir soal yang baik, indeks daya bedanya paling tidak harus mencapai 0,25 atau bahkan 0,30. Butir soal yang indeks daya pembedanya kurang dari 0,25 dianggap tidak layak, dan perlu direvisi atau diganti. Indeks yang kurang dari 0,25 berarti butir soal yang bersangkutan kurang mampu membedakan antara kelompok tinggi dan kelompok rendah. Sedangkan Nurkencana (1986: 140) mengemukakan bahwa indeks daya pembeda yang ideal adalah 0,40 ke atas. Namun untuk ulangan-ulangan harian, masih dapat ditolerir daya beda sebesar 0,20.

Purwanto (2009:124) menyatakan bahwa daya beda soal itu adalah 0 (nol) atau negatif (minus), maka soal itu perlu direvisi atau diperbaiki. Daya beda dapat juga dicari dengan menggunakan tabel koefisien "*biserial*" dengan mencari r_{bis} dari tabel. Sebuah butir soal dinyatakan layak jika indeks tingkat kesulitan maupun daya pembeda dapat memenuhi standar yang telah ditetapkan. Indeks tingkat kesulitan mungkin telah memenuhi persyaratan, tetapi jika indeks daya pembeda rendah, di bawah standar yang ditentukan, butir soal yang bersangkutan tetap dinyatakan kurang layak. Demikian pula sebaliknya, hal ini biasanya cukup berat dipenuhi terutama yang berkaitan dengan tuntutan indeks daya pembeda. Akan tetapi, jika butir-butir soal tes memenuhi

persyaratan tersebut, tingkat keterpercayaan tes akan menjadi lebih tinggi.

e. Efektivitas Butir Pengecoh

Pengecoh (*Distractor*) adalah pilihan yang bukan merupakan kunci jawaban. Pengecoh diadakan untuk menyesatkan peserta didik agar tidak memilih kunci jawaban (Waridjan, 1984: 387).

Tes objektif bentuk *multiple choice*, pada tes/soal tersebut untuk setiap butir dilengkapi dengan beberapa kemungkinan jawaban, atau yang sering dikenal dengan istilah *option* atau alternatif.

Alternatif atau *option* itu jumlahnya berkisar antara tiga sampai dengan lima buah, dan dari kemungkinan-kemungkinan jawab yang terpasang pada setiap butir soal itu, salah satu di antaranya adalah merupakan jawaban betul, sedangkan sisanya adalah merupakan jawaban salah. Jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (Pengecoh).

Tujuan utama dari pemasangan distraktor pada setiap butir soal itu adalah agar dari sekian banyak testee yang mengikuti tes hasil belajar ada yang tertarik untuk memilihnya, sebab mereka menyangka bahwa *distractor* yang mereka pilih merupakan jawaban betul. Jadi mereka terkecoh, menganggap bahwa distraktor yang terpasang pada soal itu sebagai kunci jawaban soal, padahal bukan. Semakin banyak testee yang berkecoh, maka dapat dinyatakan bahwa *distractor* yang di pasang itu makin dapat menjalankan fungsinya dengan sebaik-baiknya (Sudijono, 2010: 409).

Penentuan revisi terhadap suatu butir soal tidak semata-mata berdasarkan indeks tingkat kesulitan dan daya pembeda saja, melainkan juga bagaimana sebaran distribusi frekuensi jawaban pada alternative yang disediakan. Dengan kata lain, kita perlu juga menganalisis butir – butir pengecoh (distraktor) untuk tiap butir soal (Nurgiyantoro, 2013:201).

Daryanto (2010: 192) menyatakan bahwa pola jawaban soal adalah distribusi test dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban diperoleh dengan menghitung banyaknya test yang memilih pilihan jawaban A, B, C, D atau E. Dari pola jawaban soal dapat diketahui apakah butir pengecoh berfungsi dengan baik atau tidak. Sebuah pengecoh dapat dikatakan berfungsi dengan baik apabila pengecoh mempunyai daya Tarik yang besar bagi pengikut-pengikut tes yang kurang memahami konsep atau kurang menguasai materi.

Dasar pemikiran analisis distraktor tidak berbeda halnya dengan analisis daya pembeda suatu butir soal, harus ada perbedaan frekuensi jawaban antara siswa kelompok tinggi dan kelompok rendah. Untuk setiap alternatif jawaban yang betul, kelompok-kelompok tinggi harus memilih secara lebih banyak karena besarnya selisih jawaban betul inilah yang harus memilih secara lebih banyak karena besarnya selisih jawaban betul inilah yang akan menentukan besar kecilnya indeks daya pembeda. Sebaliknya, alternatif-alternatif jawaban yang merupakan

distraktor, kelompok rendah harus memilih secara lebih banyak. Di samping itu. Agar semua opsi dalam butir soal efektif, penyusunan opsi-opsi salah harus dilakukan sedemikian rupa sehingga tidak terlalu terlihat mencolok.

Jika terjadi penyimpangan terhadap hal-hal tersebut, suatu butir soal disarankan untuk revisi. Kegiatan revisi tidak harus mencakup butir soal dengan seluruh alternatif jawabannya, melainkan cukup pada alternatif yang mengalami penyimpangan saja.

B. Penelitian yang Relevan

Penelitian yang relevan dengan penelitian ini adalah:

1. Penelitian oleh supriyadi (2007) tentang “*Analisis Butir Tes Prestasi Hasil Belajar*” yang dimuat dalam jurnal Himmah Volume VIII Nomor 24 September – Desember 2007. Analisis yang dilakukan oleh peneliti hanya pada tiga hal, yaitu: indeks tingkat kesukaran, indeks daya beda dan reliabilitas tes. Kesimpulan dari penelitian ini menyatakan tingkat tinggi kesukaran, indeks daya beda dan reliabilitas tes masih rendah karena masih di bawah kriteria yang ditentukan.
2. Ramadhani (2013) tentang “*Studi Analisis Butir-Butir Soal Objektif Berbentuk Multiple Choice Mata Pelajaran PAI dalam Ujian Sekolah Berstandar Nasional pada Siswa Kelas IX SMP Negerri Petarukan Tahun 2011/2012*”. Analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektifitas fungsi distraktor (Pengecoh), digunakan untuk

mengetahui kualitas tes mata pelajaran PAI dalam ujian sekolah berstandar nasional pada kelas IX di SMP Negeri 2 Petarukan tahun ajaran 2011/2012

Hasil dari penelitian ini menunjukkan bahwa validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas fungsi distraktor (pengecoh) masih rendah.

Apabila dibandingkan dengan penelitian ini untuk lingkup analisisnya sama yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas fungsi distraktor (pengecoh). Perbedaannya terletak pada mata pelajaran yang diteliti, kalau penelitian yang dilakukan oleh Ramadhani adalah mata pelajaran Pendidikan Agama Islam dengan lingkup satu sekolah dan jenjang sekolahnya SMP. Sedangkan penelitian ini menyangkut wilayah kabupaten dan jenjang sekolahnya adalah SD dengan mata pelajaran yang dianalisis adalah mata pelajaran Bahasa Indonesia.

3. Sulistyawati (2013) tentang “*Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Bahasa Arab Kelas VIII MTs Negeri Piyungan Bantul Tahun Pelajaran 2012/ 2013*”. Penelitian ini bertujuan untuk mengetahui kualitas butir soal ulangan akhir semester mata pelajaran Bahasa Arab kelas VIII MTs Negeri Piyungan Bantul Tahun Pelajaran 2012/2013, ditinjau dari analisis: validitas, reliabilitas, daya pembeda, fungsi pengecoh, derajat kesukaran, dan kesesuaian butir soal dengan standar kompetensi dan kompetensi dasarnya. Perbedaan penelitian ini hanya pada jenjang sekolah dan ruang lingkup wilayahnya. Kalau penelitian ini

juga membahas kesuaian standar kompetensi dan kompetensi dasarnya, tetapi pada penelitian yang penulis lakukan tidak membahasnya karena kisi-kisi sudah dibuat oleh Balitbang Kementrian Pendidikan dan Kebudayaan berdasarkan standar kompetensi dan kompetensi dasar yang ada pada permendiknas Nomor 22 Tahun 2006 sehingga untuk kajian mengenai hal tersebut tidak perlu penulis lakukan.

4. Daniati (2012) tentang “*Analisis Validitas dan Reliabilitas Soal Ulangan Kenaikan Kelas Mata Pelajaran Bahasa Arab Kelas VIII Mts Muhammadiyah Wates, Kulon Progo Tahun Pelajaran 2010/2011*”. Kesimpulan dari penelitian tersebut adalah soal ulangan kenaikan kelas tersebut memiliki validitas sangat rendah hal tersebut dapat dilihat dari 40 soal tidak ada satupun yang valid dan tingkat reliabilitasnya juga rendah, sedangkan untuk tingkat kesulitan, daya beda serta efektifitas butir pengecoh yang tidak diteliti.
5. Syahabuddin (2010) dengan judul penelitian “*Analisis Kualitas Butir Soal Mata Pelajaran Bahasa Arab Kelas V di Madrasah Ibtidaiyah Tempel, Sleman, Yogyakarta*”. Hasil penelitiannya adalah untuk tingkat validitas, reliabilitas, tingkat kesukaran memiliki kriteria dapat diterima sedangkan untuk daya pembeda dan efektifitas butir pengecoh tidak teliti.
6. Mustaghfiri, Muhammad (2011) “*Analisis Kualitas Soal UAMBN Mata Pelajaran Al-Quran Hadist Tahun Pelajaran 2009/2010*”. Hasil Penelitiannya adalah hanya menjelaskan dari 50 soal pilihan ganda ada 32

soal yang sudah bisa dikatakan baik tanpa dirinci dari faktor mana apa saja sedangkan 12 soal dikatakan tidak memenuhi kriteria.

7. Barri, Nur (2011) dengan judul “Analisis Tes Multiple Choice Buatan KKMTs Mata Pelajaran Sejarah Kebudayaan Islam Kelas VII Semester II di MTs Negeri Gondong Sragen tahun 2010/2011”. Penelitian ini bertujuan mengetahui validitas butir tes mata pelajaran sejarah kebudayaan islam buatan KKMTs kelas VII semester II memiliki validitas sedang yaitu 57, 5% kedua reliabilitas tinggi kesukaran terdapat 67.5% termasuk dalam kategori sedang/ cukup, keempat dilihat dari daya pembeda menunjukkan 42.5%, fungsi distraktor termasuk memiliki distraktor yang baik yaitu, sebesar 76%.
8. Makruf (2009) dengan judul “*Studi Deskriptif Analisis Instrumen Tes Pilihan Ganda Mata Pelajaran Akidah Akhlak Buatan Pendidik KKM Rayon 2 Semester 1 Peserta Didik Kelas VII MTs Miftahusalam 1 Wonosalam Demak*”. Penelitian ini memiliki validitas sedang 62.5%, untuk reliabilitas sebesar 0,71% menyatakan bahwa butir soal termasuk reliabilitas tinggi, tingkat kesukaran 37,5% butir tes yang terlalu mudah, daya pembeda menunjukkan 10% tes menunjukkan daya pembedanya lemah sekali, kemudian distraktornya 74,17% berfungsi dengan baik.
9. Kinanthi, Purwi (2006) dengan judul “*Analisis Soal Tes Buatan Tes Pendidikan Agama Islam Dalam Ujian Sekolah di SMPN 2Banjarnegara Tahun 2005/2006* “. Hasil penelitian ini menunjukkan bahwa: hasil analisis validitas yang sedang yaitu dengan 77,78% dari soal yang valid,

analisis reliabilitas tes PAI memiliki reliabilitas yang tinggi yaitu koefisien korelasi $r_{11}=0,810$, tingkat kesukaran yang mudah 62,2% daya pembeda yang sedang atau cukup yaitu sebesar 42,22% dan fungsi distraktor yang telah berfungsi dengan baik yaitu sebesar 53,3 %.

Dari beberapa penelitian yang telah dilakukan sangat jelas bahwa penelitian yang penulis lakukan yakni sama-sama meneliti tentang kualitas butir soal. Meskipun secara kuantitatif tulisan-tulisan yang membahas tentang instrumen tes sudah banyak, akan tetapi nampaknya dari beberapa tulisan yang penulis cari belum ada yang spesifik membahas tentang analisis butir soal *Ujian Sekolah Mata Pelajaran Bahasa Indonesia Tahun Pelajaran 2016/2017 di SMA Negeri 3 Banjar Kota Banjar Jawa Barat*.

Perbedaan penelitian yang penulis lakukan dengan penelitian tersebut di atas adalah soal, waktu dan tempat penelitian. Semua kriteria soal di teliti, serta secukupnya wilayah dan subjeknya juga lebih luas dan lebih banyak dibandingkan dengan beberapa penelitian yang disebutkan di atas.

C. Kerangka Pikir

Dari uraian yang telah dikemukakan oleh penulis di atas maka dapat dikatakan bahwa, soal dilakukan baik (memenuhi syarat) dapat dilihat dari beberapa hal. Hal tersebut adalah kesahihan atau sering disebut dengan apa yang akan diukur. Keterpercayaan (reliabilitas). Mengacu bahwa suatu alat ukur (tes) akan menghasilkan sesuatu yang kurang lebih sama apabila dikenakan pada satu individu meskipun berbeda pelaksanaannya. Indeks

tingkat kesulitan, mengacu bahwa suatu alat tes tidak terlalu sukar dan juga tidak terlalu mudah. Indeks daya beda, menunjukkan bahwa sebuah butir soal harus bisa membedakan antara siswa yang berkemampuan tinggi dan siswa berkemampuan rendah. Dan efektivitas butir pengecoh, mengisyaratkan bahwa semua pilihan jawaban (opsi) harus ada yang memilih dan jawaban betul lebih banyak dipilih oleh siswa yang berkemampuan tinggi.

Oleh karena itu butir soal dapat dikatakan baik (memenuhi kriteria) apabila memenuhi kriteria validitas reliabilitas, tingkat kelayakan (indeks tingkat kesulitan, Indeks daya beda, dan efektivitas butir pengecoh) yang dipersyaratkan.

