BAB II

KAJIAN PUSTAKA

A. Kajian Teori

1. Tes

Ketika membahas tentang sebuah proses pembelajaran di sekolah, salah satu aspek yang tidak dapat dilupakan adalah ketercapaian tujuan pembelajaran. Tujuan pembelajaran yang akan dicapai dapat ditinjau dari beberapa aspek. Dari aspek waktu misalnya ketercapaian tujuan pembelajaran pada akhir proses pembelajaran harian, ketercapaian tujuan pembelajaran pada akhir semester (Ujian Akhir Semester) dan ketercapain tujuan pembelajaran pada akhir jenjang pendidikan (Ujian Akhir Sekolah atau Ujian Nasional). Untuk mengukur kertcapaian tujuan tersebut perlu dilakukan sebuah evaluasi pembelajaran dengan menggunakan seperangkat tes, dan yang paling sering digunakan adalah tes tulis *(pencil and paper test).*

Menurut Suharsimi (1997: 51-52) istilah tes diambil dari kata “testum” suatu pengertian dalam bahasa Perancis kuno yang berarti piring untuk menyisihkan logam-logam mulia. Ada pula yang mengartikan sebagai sebuah piring yang dibuat dari tanah. Seorang ahli bernama James Mc. Cattel, pada tahun1890 telah memperkenalkan pengertian tes ini kepada masyarakat melalui bukunya yang berjudul “*Mental Test and Measurement*”. Selanjutnya di Amerika Serikat tes ini berkembang dengan cepat sehingga dalam tempo yang tidak begitu lama masyarakat mulai menggunakannya.

Banyak ahli yang mulai mengembangkan tes ini untuk berbagai bidang, namun yang terkenal adalah sebuah tes inteligensi yang disusun oleh seorang Perancis bernama Binet, yang kemudian dibantu penyempurnaannya oleh *Simon* pada tahun1904, sehingga tes tersebut dikenal sebagai tes *Binet-Simon*.

Didorong oleh munculnya statistik dalam penganalisaan data dan informasi, maka akhirnya tes ini digunakan dalam berbagai bidang seperti tes kemampuan dasar, tes kelelahan perhatian, tes ingatan, tes minat, tes sikap dan sebagainya. Yang terkenal penggunaannya di sekolah hanyalah tes prestasi belajar.

Lee J. Cronbach dalam bukunya Essensials of Psycological testing (1970), mendefiniskan tes sebagai "*... a systematic procedure for observing a person's behavior and describing it with te aid of a numerical scale or a category system*" (Saifuddin Azwar, 2004 : 14). Dalam definisi dapat dilihat bahwa tes merupakan suatu prosedur yang sistematis, yaitu suatu proses yang dilakukan berdasarkan tujuan dan tata cara yang jelas. Ketika membicarakan konsep tes, ada beberapa hal yang tak terpisahkan antara lain :

a. Testee

   Testee adalah orang atau responden yang sedang mengerjakan tes. Orang-orang inilah yang akan dinilai atau diukur, baik mengenai kemampuan, minat, bakat, pencapaian dan sebagainya.

b. Testor

   Testor adalah orang yang diserahi untuk melaksanakan pengambilan tes terhadap para responden. Dengan kata lain, testor adalah subyek evaluasi (tapi

adakalanya hanya orang yang ditunjuk oleh subyek evaluasi untuk melaksanakan tugasnya).

c. Testing

Testing merupakan saat pada waktu tes itu dilaksanakan. Dapat juga dikatakan testing adalah saat pengambilan tes.

d. Pengukuran

Pengukuran adalah suatu prosedur pemberian angka (kuantitatif) terhadap atribut atau variabel sepanjang satu kontinum. Jadi, guna memberikan gambaran mengenai kecepatan laju sebuah kendaraan, kita memberikan suatu angka yang dapat mendeskripsikan kecepatan tersebut. Untuk itu kita perlu melakukan pengukuran kecepatan (Saifuddin Azwar, 2004 : 3).

e. Evaluasi

Evaluasi merupakan interpretasi terhadap hasil pengukuran yang disandarkan pada suatu norma atau suatu kriteria (Saifuddin Azwar, 2004 : 6). Dengan adanya norma dan kriteria, hasil yang sama dari suatu pengukuran dapat saja mendatangkan interpretasi yang berbeda. Sebagai contoh, laju kendaraan 40 km/jam akan lain sekali maknanya apabila kendaraan tersebut adalah sepeda dan apabila kendaraannya adalah sebuah mobil. Demikianlah, dengan evaluasi dapat dikatakan suatu atribut sebagai baik-buruk, cepat-lambat, jauh-dekat, tinggi-rendah, dan semacamnya.

Secara ringkas, karakteristik evaluasi adalah :

a. Merupakan pembanding antara hasil ukur dengan suatu norma atau suatu kriteria.

b. Hasilnya bersifat kualitatif.

c. Hasilnya dinyatakan secara evaluatif / kualitatif

Dari berbagai pemahaman di atas dapat disampaikan bahwa tes ialah himpunan pertanyaan yang harus dijawab atau pernyataan-pernyataan yang harus dipilih/ ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang dites (testee) dengan tujuan untuk mengukur suatu aspek (perilaku) tertentu dari orang yang dites. Dalam tes prestasi belajar, yang hendak diukur ialah tingkat kemampuan seorang siswa dalam menguasai bahan pelajaran yang telah diajarkan kepadanya. Jadi tes merupakan seperangkat alat atau instrumen yang diperlukan untuk memperoleh data tentang sesuatu objek.

Dari penjelasan sebelumnya telah disampaikan bahwa tes merupakan seperangkat alat atau instrumen yang diperlukan untuk memperoleh data tentang sesuatu objek. Agar data yang diperoleh tersebut merupakan data yang baik dan dapat dipertanggung jawabkan maka tes atau instrumen tersebut harus baik. Menurut Suharsimi sebuah tes yang dapat dikatakan baik sebagai alat pengukur jika memenuhi persyaratan validitas, reliabilitas, obyektifitas, praktis, dan ekonomis (Suharsimi Arikunto,1997 : 56).

a. Validitas

Sebuah tes disebut valid apabila tes itu dapat tepat mengukur apa yang hendak diukur.

b. Reliabilitas

Reliabilitas diambil dari kata *reliability* yang berasal dari *reliable* yang artinya dapat dipercaya. Tes dikatakan dapat dipercaya jika memberikan

hasil yang tetap apabila diteskan berkali-kali. Sebuah tes dikatakan reliabel apabila hasil-hasil tes tersebut menunjukkan ketetapan.

c. Obyektivitas

Dalam pengertian sehari-hari, obyektif berarti tidak adanya unsur pribadi yang mempengaruhi. Lawan dari obyektif adalah subyektif, artinya terdapat unsur pribadi yang masuk mempengaruhi. Sebuah tes dikatakan memiliki obyektivitas apabila dalam melaksanakan tes itu tidak ada faktor subyektif yang mempengaruhi. Hal ini terutama terjadi pada sistem skoringnya.

Apabila dikaitkan dengan reliabilitas maka obyektivitas menekankan ketetapan (*consistency*) pada sistem skoring, sedangkan reliabilitas menekankan ketetapan dalam hasil tes.

d. Praktikabilitas

Sebuah tes dikatakan memiliki praktikabilitas yang tinggi apabila tes tersebut bersifat praktis, atau mudah pengadministrasiannya.

e. Ekonomis

Yang dimaksud dengan ekonomis disini ialah bahwa pelaksanaan tes tersebut tidak membutuhkan ongkos/biaya yang mahal, tenaga yang banyak dan waktu yang lama.

Menurut Puslitbang Sisjian ciri ciri tes yang baik adalah sebagai berikut

a. Setiap alat ukur hanya mengukur satu dimensi atau aspek saja.
b. Kehandalan atau reliabilitas dari alat ukur.

Berdasarkan pendapat beberapa ahli di atas dua unsur penting dari persyaratan tes yang baik adalah aspek validitas dan aspek reliabiltas. Oleh sebab itulah sebelum melangkah lebih jauh menganalisis tentang kualitas suatu tes dua aspek besar ini harus dianalisis terlebih dahulu (Puslitbang Sisjian, 1999 : 16).

2. Bentuk-Bentuk Tes

Tes ialah himpunan pertanyaan yang harus dijawab atau pernyataan-pernyataan yang harus dipilih/ ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang dites (testee) dengan tujuan untuk mengukur suatu aspek (perilaku) tertentu dari orang yang dites. Dalam tes prestasi belajar, yang hendak diukur ialah tingkat kemampuan seorang siswa dalam menguasai bahan pelajaran yang telah diajarkan kepadanya. Tes untuk mengukur prestasi, baik untuk aspek pengetahuan maupun ketrampilan, dapat dibagi menjadi dua jenis, yaitu (Puslitbang Sisjian, 1999 :15) :

1. Tes kemampuan atau *power test*
2. Tes kecepatan atau speed test

Menurut Suharsimi berdasarkan bentuknya, tes dapat di bedakan menjadi dua macam yaitu tes subyektif dan tes obyektif. Tes subyektif pada umumnya berbentuk essay atau uraian. Tes bentuk essay adalah suatu tes yang memerlukan jawaban yang bersifat pembahasan atau uraian kata-kata. Soal tes bentuk essay ini menuntut kemampuan siswa untuk dapat mengorganisir, mengintepretasi, menghubungkan pengertian atau pemahaman konsep yang telah dimiliki. Tes essay menuntut siswa untuk dapat mengingat dan mengenal kembali dan

menuntut siswa untuk mempunyai daya kreatifitas yang tinggi (Suharsimi Arikunto,1997: 163).

Tes essay ini mempunyai kelemahan dan kelebihan, menurut referensi dari beberapa ahli dapat disampaikan beberapa kelemahan dan kelebihan tes bentuk essay adalah sebagai berikut:

a. Kelebihan tes bentuk essay adalah sebagai berikut :
    1) Mudah disiapkan dan disusun.
    2) Tidak memberi banyak kesempatan untuk berspekulasi atau untung-untungan.
    3) Mendorong siswa untuk berani mengemukakan pendapat serta menyusun dalam bentuk kalimat yang bagus.
    4) Memberi kesempatan kepada siswa untuk mengutarakan maksudnya dengan gaya bahasa dan caranya sendiri.
    5) Dapat diketahui sejauhmana siswa mendalami sesuatu masalah yang diteskan.
b. Kekurangan tes bentuk essay adalah sebagai berikut :
    1) Kadar validitas dan reliabilitasnya rendah karena sukar diketahui segi-segi mana dari pengetahuan siswa yang betul-betul telah dikuasai.
    2) Kurang representatif dalam hal mewakili seluruh scope bahan pelajaran yang akan di tes karena soalnya hanya beberapa saja (terbatas).
    3) Cara memeriksanya banyak dipengaruhi oleh unsur-unsur subyektif.
    4) Pemeriksaan lebih sulit sebab membutuhkan pertimbangan individual lebih banyak dari penilai.

5) Waktu untuk koreksinya lama dan tidak dapat diwakilkan kepada orang lain.

Tes obyektif adalah suatu bentuk tes yang dalam pemeriksaannya dapat dilakukan secara obyektif. Tes obyektif ini lebih hemat waktu, karena dalam waktu 60 menit, bisa diberikan 60 soal. Berdasarkan referensi dari berbagai sumber dapat disampaikan kelemahan dan kelebihan tes obyektif adalah sebagai berikut (Suharsimi Arikunto, 1997 : 164-166):

a. Kelebihan-kelebihan tes bentuk obyektif

1) Mengandung lebih banyak segi-segi yang positif, misalnya lebih representatif mewakili isi dan luas bahan, lebih obyektif, dapat dihindari campur tangannya unsur-unsur subyektif baik dari segi peserta didik maupun segi guru yang memeriksa.
2) Lebih mudah dan cepat cara memeriksanya karena dapat menggunakan kunci tes bahkan alat-alat hasil kemajuan teknologi.
3) Pemeriksaannya dapat diserahkan orang lain.
4) Dalam pemeriksaan, tidak ada unsur subyektif yang mempengaruhi.

b. Kelemahan-kelemahan tes bentuk obyektif

1) Persiapan untuk menyusunnya jauh lebih sulit daripada tes essay karena soalnya banyak dan harus teliti untuk menghindari kelemahan-kelemahan yang lain.
2) Soal-soalnya cenderung untuk mengungkapkan ingatan dan daya pengenalan kembali saja, dan sukar untuk mengukur proses mental yang tinggi.

3) Banyak kesempatan untuk main untung-untungan.

4) “Kerjasama” antar siswa pada waktu mengerjakan soal tes lebih terbuka.

Tes obyektif mempunyai berbagai macam bentuk antara lain tes benar salah, tes pilihan ganda, menjodohkan *(matching test)*, dan tes isian. Dari sekian banyak bentuk tes obyektif, tes obyektif yang digunakan dalam TUC ujian Nasional SMK dan Ujian Nasional SMK adalah tes bentuk Pilihan Ganda.

3. **Tes Pilihan Ganda *(Multiple Choice Test)***

a. Hakekat Tes Pilihan Ganda *(Multiple Choice Test)*

Menurut Suharsimi *multiple choice test* atau tes pilihan ganda terdiri atas suatu keterangan atau pemberitahuan tentang suatu pengertian yang belum lengkap. Untuk melengkapinya harus memilih satu dari beberapa kemungkinan jawaban yang telah disediakan. Atau multiple choice test terdiri atas bagian keterangan *(stem)* dan bagian kemungkinan atau alternatif *(options).*Kemungkinan jawaban (*option*) terdiri atas satu jawaban yang benar yaitu kunci jawaban dan beberapa pengecoh *(distractor)* (Suharsimi Arikunto,1997: 169).

Soal bentuk pilihan ganda adalah suatu soal yang jawabannya harus dipilih dari beberapa kemungkinan jawaban yang telah disediakan (Puslitbang sisjian, 1999:41). Secara umum setiap soal pilihan ganda terdiri dari pokok soal *(stem)* dan pilihan jawaban *option*. Pilihan jawaban terdiri atas kunci jawaban dan pengecoh *(distractor)*. Kunci jawaban ialah jawaban yang benar atau paling benar. Pengecoh merupakan jawaban yang tidak benar namun memungkinkan seseorang memilihnya apabila tidak menguasai bahannya.

Berdasarkan pemahaman uraian diatas dapat disampaikan bahwa soal pilihan ganda terdiri dari kalimat yang merupakan bagian pokok soal (*stem*) dan alternatif jawaban (*option).* Dalam penelitian ini alternatif jawaban terdiri dari lima *option* dengan satu jawaban yang benar (kunci jawaban) dan empat *option* alternatif jawaban yang salah (*distractor*).

b. Keunggulan dan Keterbatasan

Dari berbagai referensi secara umum dapat disampaikan bahwa keunggulan soal bentuk pilihan ganda adalah dapat diskor dengan mudah, cepat, serta objektif, dan dapat mencangkup ruang lingkup bahan/ materi yang luas dalam suatu tes untuk suatu kelas atau jenjang pendidikan. Bentuk ini sangat tepat untuk ujian yang pesertanya sangat banyak sedangkan hasilnya harus segera diumumkan, seperti Ujian Sekolah atau Ujian Nasional. Di samping keunggulan yang telah disebut diatas, soal bentuk pilihan ganda memiliki beberapa keterbatasan antara lain memerlukan waktu relatif lama untuk menulis soalnya, sulit membuat pengecoh yang homogen dan berfungsi dengan baik, dan terdapat peluang untuk menebak kunci jawaban.

c. Kaidah Penulisan Soal

Seperti halnya bentuk soal yang lain, penulisan soal pilihan ganda harus didasarkan pada spesifikasi soal yang terdapat kisi-kisi tes. Soal pilihan ganda dapat ditingkatkan mutunya apabila penulisannya, di samping berlandaskan kisi-kisi, juga mengikuti berbagai kaidah penulisan soal. Kaidah-kaidah penulisan soal merupakan petunjuk atau pedoman yang perlu diikuti penulis agar soal yang dihasilkan memiliki mutu yang baik. Soal yang mutunya baik dalah soal yang

mampu menjaring informasi yang diperlukan dan berfungsi secara optimal.Kaidah-kaidah penulisan soal pilihan ganda yang baik dapat dikaji dari tiga aspek yaitu aspek materi, konstruksi dan bahasa.

4. Analisis Soal Pilihan Ganda

Analisis soal dilakukan untuk mengetahui berfungsi tidaknya sebuah soal. Analisis pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif (*qualitatif control*) dan analisis kuantitatif (*quantitatif control*). Analisis kualitatif sering pula dinamakan sebagai validitas logis (*logical validity*) yang dilakukan sebelum soal digunakan untuk melihat berfungsi tidaknya sebuah soal. Analisis soal secara kuantitatif sering pula dinamakan sebagai validitas empiris (*empirical validity*) yang dilakukan untuk melihat lebih berfungsi atau tidaknya sebuah soal, setelah soal itu diujicobakan kepada sampel yang representatif.

a. Analisis Kualitatif

Analisis kualitatif sering juga disebut sebagai validitas logis (*logical validity*) yaitu berupa penelaahan yang dimaksud untuk menganalisis soal ditinjau dari segi teknis, isi, dan editorial. Analisis secara teknis dimaksudkan sebagai penelaahan soal berdasarkan prinsip-prinsip pengukuran dan format penulisan soal. Analisis secara isi dimaksudkan sebagai penelaahan khusus yang berkaitan dengan kelayakan pengetahuan yang dipertanyakan. Analisis secara editorial dimaksudkan sebagai penelaahan yang khususnya berkaitan dengan keseluruhan format dan keajegan editorial dari soal yang satu ke soal yang lainnya.

Analisis kualitatif lainnya dapat juga dikategorikan dari segi materi, konstruksi, dan bahasa. Analisis materi dimaksudkan sebagai penelaahan yang

berkaitan dengan substansi keilmuan yang ditanyakan dalam soal serta tingkat kemampuan yang sesuai dengan soal. Analisis konstruksi dimaksudkan sebagai penelaah yang umumnya berkaitan dengan teknik penulisan soal. Analisis bahasa dimaksudkan sebagai penelaah soal yang berkaitan dengan penggunaan bahasa Indonesia yang baik dan benar menurut EYD (Sumarna Surapranata, 2009 : 1-2).

b. **Analisis Kuantitatif**

Pemberian tes kepada peserta tes, akan menghasilkan informasi tentang soal tes, maupun peserta tes. Informasi tersebut antara lain diperoleh melalui analisis statistik yang salah satunya dapat digunakan sebagai landasan untuk melihat lebih berfungsinya sebuah soal. Untuk memperoleh informasi tersebut perlu dilakukan analisis kuantitatif. Hasil analisis yang dimaksud untuk mengetahui sejauh mana soal dapat membedakan antara peserta tes yang kemampuannya tinggi dalam hal yang didefinisikan oleh kriteria dengan peserta tes yang kemampuannya rendah. Dalam hal memilih kriteria yang akan digunakan orang menginginkan adanya ukuran yang baik untuk kemampuan ataupun keterampilan yang diukur oleh soal. Informasi lainnya adalah bagaimana soal dapat membedakan antara individu maupun antar kelompok.

Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal secara kuantitatif dimaksudkan meliputi parameter soal tingkat kesukaran, daya pembeda, dan reliabilitas. Khusus soal-soal pilihan ganda, dua tambahan parameter yaitu dilihat dari peluang untuk menebak atau menjawab soal benar dan berfungsi tidaknya pilihan jawaban, yaitu penyebaran semua alternatif jawaban

dari subyek-subyek yang dites. Salah satu tujuan dilakukannya analisis adalah untuk meningkatkan kualitas soal, yaitu apakah (1) suatu soal dapat diterima karena telah didukung oleh data statistik yang memadai, (2) diperbaiki, karena terbukti terdapat beberapa kelemahan, atau bahkan (3) tidak digunakan sama sekali karena terbukti secara empiris tidak berfungsi sama sekali (Sumarna Surapranata, 2009 : 10).

5. **Validitas**

Validitas berasal dari kata *validity* yang mempunyai arti sejauhmana suatu tes atau skala dalam menjalankan fungsi pengukurannya. Pengukuran dikatakan mempunyai validitas yang tinggi apabila menghasilkan data yang secara akurat memberikan gambaran mengenai variabel yang diukur seperti dikehendaki oleh tujuan pengukuran tersebut. Akurat dalam hal ini berarti tepat dan cermat sehingga apabila tes menghasilkan data yang tidak relevan dengan tujuan pengukuran maka dikatakan sebagai pengukuran yang memiliki validitas rendah.

Di sini terkandung pengertian bahwa valid tidaknya suatu pengukuran tergantung pada kemampuan alat ukur tersebut dalam mencapai tujuan pengukuran yang dikehendaki dengan tepat. Suatu tes yang dirancang untuk mengungkap atribut A dan kemudian memang menghasilkan informasi mengenai atribut A, dikatakan sebagai alat ukur yang hasilnya valid. Sebaliknya, suatu tes yang dirancang untuk mengukur atribut A akan tetapi menghasilkan data A’ atau bahkan B, dikatakan sebagai alat ukur yang

fungsinya tidak valid untuk mengukur atribut A tetapi valid untuk mengukur atribut A' atau B.

Sisi lain yang terkandung dalam pengertian validitas adalah aspek kecermatan pengukuran. Suatu hasil ukur yang disebut valid, tidak sekedar merupakan data yang tepat menggambarkan aspek yand diukur akan tetapi juga memberikan gambaran yang cermat mengenai variabel yang diukur. Cermat berarti bahwa pengukuran itu mampu memberikan gambaran dan makna terhadap perbedaan angka yang sekecil-kecilnya yang diperoleh oleh individu yang berbeda.

Menggunakan alat ukur yang bertujuan untuk mengukur suatu aspek tertentu akan tetapi tidak dapat memberikan hasil ukur yang cermat dan teliti tentu akan menimbulkan berbagai kesalahan. Kesalahan itu dapat berupa hasil yang terlalu tinggi (overestimasi) atau yang terlalu rendah (underestimasi). Keragaman kesalahan ini dalam istilah statistika disebut sebagai varians kesalahan atau *error variance.* Tes akan menghasilkan data kuantitatif yang valid bila varians eror pengukurannya kecil (disebabkan eror pengukurannya kecil) sehingga angka yang dihasilkan dapat dipercaya sebagai angka yang "sebenarnya" (*true-scores*) atau angka yang mendekati keadaan sebenarnya.

Sebagaimana halnya dengan istilah reliabilitas, begitu pula istilah validitas haruslah diartikan sebagai validitas skor atau hasil pengukuran yang diperoleh oleh tes tersebut. Begitulah arti yang ditekankan oleh Cronbach bahwa dalam proses validasi tujuan sebenarnya tidak untuk melakukan validasi terhadap tes akan tetapi melakukan validasi terhadap interpretasi data yang

diperoleh oleh suatu prosedur tertentu (Cronbach, 1971). Bahkan pada tahun 1989 Messick (Saiffudin Azwar, 2014: 10)mengatakan bahwa semua bentuk validitas harus dianggap sabagai aspek dari validitas konstrak sebab konsep validitas mengacu pada interpretasi skor tes bukan pada tes itu sendiri. Tentang validitas konstrak sebagai payung atau wadah jenis validitas yang lainnya, pada tahun 1957 Loevinger (Saiffudin Azwar, 2014: 10) telah mengatakannya lebih dahulu.

Validitas adalah pertimbangan yang paling utama dalam mengevaluasi kualitas tes sebagai instrumen ukur. Konsep validitas mengacu kepada kelayakan, kebermaknaan, dan kebermanfaatan inferensi tertentu yang dapat dibuat berdasarkan skor hasil tes yang bersangkutan. Validitas tes itu sendiri adalah proses pengumpulan bukti-bukti yang dapat mendukung inferensi termaksud. Dengan demikian, proses tersebut bukan untuk memvalidasi tes melainkan memvalidasi inferensi mengenai penggunaan hasil ukur secara spesifik. Oleh karena itulah Messick (Saiffudin Azwar, 2014: 11) mengatakan bahwa validitas merupakan suatu *evolving property* dan validasi merupakan suatu *continuiting process.*

Makna istilah “validitas” memang telah lama bergeser dari validitas tes ke makna validitas interpretasi skor tes. Pada awalnya, validitas dikonsepkan sebagai suatu karakteristik yang dimiliki oleh tes. Pada tahun 1927 Kelley (Saiffudin Azwar, 2014: 11) mengatakan bahwa pertanyaan tentang validitas adalah pertanyaan mengenai apakah sauatu tes mengukur apa yang hendak diukurnya. Menurut Hughes (Saiffudin Azwar, 2014: 11) Suatu tes disebut

valid bila tes tersebut mengukur secara akurat apa dimaksudkan hendak diukurnya.

Pada perkembangan lebih lanjut, validitas lalu dipandang sebagai suatu karakteristik skor tes dan bukanlah karakteristik tes itu sendiri. Kemudian, makna validitas diartikan sebagai suatu karakteristik interpretasi skor tes, bukan karakteristik tes ataupun karakteristik skor tes. Pandangan inilah yang dianut antara lain oleh Messick. Dari perspektif epistemologi, perubahan makna ini sangat dapat dipahami. Apa yang harus didukung oleh bukti bukanlah skor tes melainkan interpretasi skor tes. Fokus permasalahan validitas tidak pada skor atau *item* tes melainkan pada inferensi yang dihasilkan oleh instrumen yang bersangkutan, yakni inferensi terhadap perilaku yang dapat diextrapolasi dari skor tes. Untuk dinyatakan valid, inferensi yang dibuat berdasarkan skor tes harus "layak, bermakna, dan bermanfaat" (Saiffudin Azwar, 2014: 11).

Sebagaimana telah diuraikan, pengertian validitas erat berkaitan dengan masalah tujuan suatu pengukuran. Oleh karena itu, tidak ada validitas yang berlaku umum untuk semua tujuan pengukuran. Suatu alat ukur dirancang hanya untuk satu tujuan yang spesifik sehingga hanya menghasilkan data yang valid untuk tujuan tersebut saja. Dengan demikian, predikat valid sebagaimana dinyatakan dalam kalimat "Tes ini valid" adalah belum lengkap dan bahkan menyesatkan. Pernyataan valid harus diikuti oleh keterangan yang menunjuk kepada tujuan ukur, yaitu valid untuk mengukur apa. Lebih jauh, keterangan tersebut harus menunjuk kepada pengertian valid bagi kelompok subjek yang

mana, sehingga suatu pernyataan validitas yang seharusnya dapat diilustrasikan oleh kalimat “tes ini menghasilkan skor yang valid sebagai ukuran IQ orang Indonesia dewasa” yang artinya bahwa tes tersebut dapat menghasilkan data yang valid bila digunakan untuk mengukur IQ dan dikenakan pada orang Indonesia yang sudah dewasa.

Dengan demikian jelaslah mengapa suatu alat ukur yang dikatakan sebagai valid guna pengambilan suatu keputusan dapat saja sangat tidak berguna dalam pengambilan keputusan lain dan bagi kelompok subjek yang lain.

Dalam teori skor-murni klasikal, pengertian validitas tersebut dapat dinyatakan sebagai sejauh mana besarnya skor-tampak X mendekati besarnya skor-murni T. Skor-tampak X tidak akan sama dengan skor-murni T kecuali apabila alat ukur yang bersangkutan memiliki validitas yang sempurna. Semakin skor-tampak mendekati skor-murni akan semakin tinggi validitasnya dan sebaliknya semakin rendah validitas berarti semakin jauh perbedaan skor tampak dan skor-murni (Saiffudin Azwar, 2014: 40).

Suatu alat ukur yang tinggi validitasnya akan memiliki error pengukuran yang kecil, artinya skor setiap subjek yang diperoleh oleh alat ukur tersebut tidak jauh berbeda dari skor yang sesungguhnya. Dengan demikian secara keseluruhan alat tes yang bersangkutan akan menghasilkan varians error yang kecil pula. Itulah yang dalam teori skor-murni klasikal diartikan sebagai validitas intrinsik, yang dirumuskan sebagai akar kuadrat perbandingan antara varians skor-murni dan varians skor tampak, yakni

$$r_{xy} = \sqrt{(s_t{}^2/s_x{}^2)}$$

Keterangan :

r_{xy} : validitas intrinsik
s_t : varians skor murni
s_x : varians skor tampak

Prosedur validitas tes adalah suatu estimasi terhadap tingkat validitas dengan prosedur tertentu. Dengan menggunakan teknik dan cara yang tepat dapat dilakukan prosedur estimasi guna melihat apa yang sesungguhnya diukur oleh tes dan seberapa cermat hasil ukurnya (Saifuddin Azwar, 2004:44).

Walaupun peneliti terbiasa melekatkan predikat valid bagi suatu tes akan tetapi hendaklah dipahami bahwa sebenarnya pengertian validitas menyangkut masalah hasil ukur bukan masalah alat ukurnya sendiri. Sebutan validitas tes hendaklah diartikan sebagai validitas hasil pengukuran yang diperoleh oleh tes tersebut. Itulah yang ditekankan oleh Cronbach (Saifudin Azwar, 2004 : 43-44) bahwa dalam proses validitas sebenarnya peneliti tidak bertujuan untuk melakukan validasi tes akan tetapi melakukan validasi terhadap interpretasi data yang diperoleh oleh prosedur tertentu.

Berdasarkan pendapat beberapa ahli tes dan pengukuran ada beberapa kategori atau tipe umum validitas, antara lain sebagai berikut :

a. Validitas isi

Validitas isi merupakan validitas yang diestimasikan lewat pengujian terhadap isi tes dengan analisis rasional atau lewat pengujian terhadap isi tes dengan analisis rasional atau lewat professional judgement. Pertanyaan yang dicari jawabannya dalam validasi ini adalah “ Sejauhmana *item-item* dalam tes

mencakup keseluruhan kawasan isi objek yang hendak diukur" atau "sejauhmana isi tes mencerminkan ciri atribut yang hendak diukur".

Validitas isi terbagi menjadi dua tipe, yaitu *face validity* (validitas muka) dan *logical validity* (validitas logik).

1) Validitas muka

Validitas muka adalah tipe validitas yang paling rendah signifikasinya karena hanya didasarkan pada penilaian terhadap format penampilan *(appereance)* tes. Apabila penampilan tes telah meyakinkan dan memberi kesan mampu mengungkap apa yang hendak diukur maka dapat dikatakan bahwa validitas muka telah terpenuhi.

2) Validitas logik

Validitas logik disebut juga sebagai validitas sampling *(sampling validity).* Validitas tipe ini menunjuk pada sejauhmana isi tes merupakan representasi dari ciri-ciri atribut yang hendak diukur (Saifuddin Azwar, 2014: 42-44).

b. **Validitas Konstrak**

Menurut Allen dan Yen (Saifuddin Azwar, 2014: 45) validitas konstrak adalah tipe validitas yang menunjukkan sejauhmana tes mengungkap suatu treat atau konstrak teoretik yang hendak diukurnya. Pengujian validitas konstrak merupakan proses yang terus berlanjut sejalan dengan perkembangan konsep mengenai trait yang diukur. Walaupun pengujian validitas konstrak biasanya memerlukan teknik analisis statistika yang lebih kompleks daripada teknik-teknik yang dipakai pada

pengujian validitas empirik lainnya akan tetapi hasil estimasi validitas konstrak tidak dinyatakan dalam bentuk suatu koefisien validitas.

c. Validitas Berdasar Kriteria

Prosedur Pendekatan validitas berdasar kriteria menghendaki tersedianya kriteria eksternal yang dapat dijadikan dasar pengujian skor tes. Suatu kriteria adalah variabel perilaku yang akan diprediksikan oleh skor tes atau berupa suatu ukuran lain yang relevan.

Untuk melihat tingginya validitas berdasarkan kriteria dilakukan komputasi korelasi antara skor tes dengan skor kriteria. Koefisien ini merupakan validitas bagi tes yang bersangkutan, yaitu r_{xy} , dimana X melambangkan skor tes dan Y melambangkan skor kriteria.

Prosedur validasi berdasarkan kriteria menghasilkan dua macam validitas, yaitu validitas prediktif (*predictive validity*) dan validitas konkuren (*concurrent validity*).

1) Validitas Prediktif

Validitas prediktif sangat pentimg artinya bila tes dimaksudkan untuk berfungsi sebagai prediktor bagi perfomansi di waktu yang akan datang. Contoh situasi yang menghendaki adanya prediksi perfomansi ini antara lain adalah dalam bimbingan karir, dalam seleksi mahasiswa baru, dalam klasifikasi dan penempatan karyawan, dan semacamnya.

Prosedur validasi prediktif pada umumnya memerlukan waktu yang lama dan mungkin pula biaya yang tidak sedikit dikarenakan prosedur ini dapat pada dasarnya bukan pekerjaan yang dianggap selesai setelah

melakukan sekali prosedur analisis, melainkan lebih merupakan kontinuitas dalam proses pengembangan tes. Sebagaimana prosedur validasi yang lain, validasi prediktif pada setiap tahapnya haruslah diikuti oleh usaha peningkatan kualitas *item* tes dalam bentuk revisi, modifikasi, dan penyusunan *item-item* baru agar prosedur yang dilakukan itu mempunyai arti lebih signifikan dan bukan sekedar pengujian secara deskriptif saja.

2) Validitas Konkuren

Apabila skor tes dan skor kriterianya dapat diperoleh dalam waktu yang sama, maka korelasi antara kedua skor termaksud merupakan koefisien validitas konkuren. Suatu contoh dimana validitas konkuren layak diuji adalah apabila kita menyusun suatu skala *self-concept* yang baru. Untuk menguji validitas skala tersebut kita dapat menggunakan skala *self-concept* lain yang telah lebih dahulu teruji validitasnya, misalnya saja *Tennessee Self-Concept Scale* (TSCS). Contoh lain adalah dalam penyusunan suatu skala inteligensi. Kita dapat menguji validitas skala inteligensi yang kita susun dengan cara menghitung korelasi antara skor skala tersebut dengan skor pada tes inteligensi lain yang telah valid, misalnya Skala Wechsler. Disamping itu, estimasi validitas skala inteligensi tersebut dapat pula diperoleh lewat perhitungan koefisien korelasinya dengan skor pada variabel lain yang relevan, yaitu yang dapat dianggap sebagai indikator tingkat inteligensi. (Saifuddin Azwar, 2014 : 47-49).

d. Validitas soal

Daya pembeda soal (*item discrimination*) merupakan validitas soal. Tujuan validitas soal adalah untuk menentukan dapat tidaknya suatu soal tersebut membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada dalam kelompok itu. Validitas soal adalah daya pembeda dalam membedakan antara peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah. Validitas soal adalah daya pembeda soal-soal yang ditetapkan dari selisih proporsi yang menjawab dari masing-masing kelompok. Indeks ini menunjukkan kesesuaian antara fungsi soal dengan fungsi tes secara keseluruhan. Dengan demikian validitas soal ini sama dengan daya pembeda soal yaitu daya dalam membedakan antara peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah.

Angka yang menunjukkan besarnya validitas soal disebut indeks validitas soal yang besarnya berkisar antara -1 sampai dengan +1. Tanda negatif menunjukkan bahwa peserta tes yang kemampuannya rendah menjawab benar sedangkan peserta tes yang kemampuannya tinggi menjawab salah dengan demikian soal yang validitasnya negatif menunjukkan terbaliknya kualitas peserta tes. Setiap soal dapat dipandang sebagai bagian yang terpisah dari sebuah tes. Sebuah soal mungkin dapat membedakan kelompok peserta tes secara baik. Sebuah soal mungkin juga tidak dapat membedakan kelompok peserta tes (misal soal dengan p = 0 atau p = 1). Sebuah soal juga mungkin membedakan kelompok secara

terbalik, yaitu peserta tes yang tidak mampu dapat menjawab soal dengan benar dan peserta tes yang mampu menjawab salah. Salah satu tujuan analisis soal adalah untuk mencari soal-soal yang dapat mengukur kemampuan secara tepat.

Jika tes atau soal mengukur hal yang sama, dapat diharapkan bahwa setiap peserta tes yang mampu dapat menjawab soal dengan benar, dan peserta tes yang tidak mampu akan menjawab soal dengan salah. Dengan kata lain, soal-soal tersebut membedakan antara peserta tes yang mampu dengan peserta tes yang tidak mampu. Dalam bagian ini akan dibahas berbagai teknik menentukan indeks validitas.

Terdapat berbagai cara yang digunakan untuk menentukan validitas diantaranya dengan menggunakan:

1) Indeks diskriminasi
2) Indeks korelasi
3) Indeks keselarasan

Teknik korelasi terdiri atas :

1) Teknik point biserial
2) Teknik phi
3) Teknik biserial
4) Teknik tetrachoric

Sebagaimana alat ukur lainnya, korelasi di dalam validitas soal memiliki prediktor dan kriterium, prediktor dalam validitas soal adalah skor soal sedangkan kriteriumnya adalah skor total tes.

Korelasi biserial maupun korelasi point biserial adalah korelasi product moment yang diterapkan pada data, dimana variabel-variabel yang dikorelasikan sifatnya masing-masing berbeda satu sama lain. Variabel butir soal bersifat dikotomi sedangkan variabel skor total atau sub skor total bersifat kontinum. Variabel butir soal dinamakan dikotomi karena skor-skor yang terdapat pada butir soal hanya ada satu nol. Seperti halnya pada bentuk soal pilihan ganda, soal yang benar diberi angka satu (1) dan yang salah diberi angka nol (0). Variabel skor total atau sub skor total peserta tes bersifat kontinum atau nondikotomi yang diperoleh dari jumlah jawaban yang benar. Korelasi biserial ditentukan dengan menggunakan persamaan (Sumarna Surapranata, 2009 : 60-61):

$$r_{bis} = \frac{M_p \quad - \quad M_t}{SD} \quad \times \quad \sqrt{\frac{p}{q}}$$

r_{bis} : Koefisien korelasi biserial
M_p : rerata skor pada tes dari peserta tes yang memiliki jawaban benar
M_t : rerata skor total
S_t : standar deviasi skor total
p : proporsi peserta tes yang jawabannya benar pada soal (tingkat kesukaran)
q : 1 – p

6. Reliabilitas

Reliabilitas menurut Saifuddin Azwar merupakan penerjemahan dari kata *reliability*. Suatu pengukuran yang mampu menghasilkan data yang memiliki tingkat reliabilitas disebut sebagai pengukuran yang reliabel (*reliable*). Walaupun istilah reliabilitas mempunyai berbagai berbagai nama lain seperti konsistensi, keterandalan, keterpercayaan, kestabilan, keajegan, dan sebagainya, namun gagasan pokok yang terkandung dalam konsep reliabilitas adalah sejauhmana hasil suatu proses pengukuran dapat dipercaya (Saifuddin Azwar,2014 : 7).

Hasil suatu pengukuran akan dapat dipercaya hanya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, selama aspek yang diukur dalam diri subjek memang belum berubah. Dalam hal ini, relatif sama berarti tetap adanya toleransi terhadap perbedaan-perbedaan kecil yang biasanya terjadi di antara hasil beberapa kali pengukuran. Bila perbedaan yang terjadi sangat besar dari waktu ke waktu maka hasil pengukuran tersebut tidak dapat dipercaya dan dikatakan sebagai tidak reliabel. Pengukuran yang hasilnya tidak reliabel tentu tidak dapat dikatakan akurat terhadap konsistensi menjadi syarat bagi akurasi.

Pengertian reliabilitas alat ukur dan reliabilitas hasil ukur, oleh sebagian orang, dianggap sama saja dan sering dipertukarkan. Sebenarnya perbedaan makna kedua istilah itu perlu diperhatikan. Konsep reliabilitas dalam arti reliabilitas alat ukur erat berkaitan dengan masalah eror pengukuran (*error of measurement*). Eror pengukuran sendiri menunjuk pada sejauhmana inkonsistensi hasil ukur terjadi apabila pengukuran dilakukan ulang pada kelompok subjek yang

sama. Konsep reliabilitas dalam arti reliabilitas hasil ukur erat berkaitan dengan eror dalam pengambilan sampel subjek (*sampling error*) yang mengacu kepada inkonsistensi hasil ukur apabila pengukuran dilakukan ulang pada kelompok sampel subjek yang berbeda dari suatu populasi yang sama. Berkaitan dengan hal itu, Thompson (Saefuddin Azwar, 2014: 8) mengatakan bahwa hal yang lebih penting untuk dipahami adalah bahwa estimasi terhadap reliabilitas merupakan fungsi dari skor yang diperoleh melalui tes, bukanlah fungsi dari tesnya itu sendiri.

Pengukuran merupakan proses untuk memperoleh skor perorangan sehingga *attribute* yang diukur benar-benar menggambarkan kemampuan mereka. Menurut Sumarna Surapranata reliabilitas atau keajegan suatu skor adalah hal yang sangat penting dalam menentukan apakah tes telah menyajikan pengukuran yang baik. Hal yang penting dalam keajegan atau reliabilitas skor adalah adanya pengambilan keputusan tentang peserta tes (Sumarna Surapranata, 2009 : 86).

Sedangkan pengertian reliabilitas tes menurut Suharsimi adalah berhubungan dengan masalah ketetapan hasil tes atau suatu tes dapat dikatakan memiliki reliabilitas yang tinggi jika tes tersebut dapat memberikan hasil yang tetap (Suharsimi Arikunto,1997 : 83).

Berdasarkan pendapat dari beberapa ahli diatas reliabilitas dalam penelitian ini diartikan sebagai konsistensi estimasi data yang diperoleh berdasarkan instrumen tes matematika pada TUC Ujian Nasional.

Reliabilitas terkait sangat erat dengan data hasil tes. Data atau skor hasil tes dalam situasi dan kondisi yang berbeda hasilnya dapat berbeda pula. Sumber keajegan dan ketidakajegan menurut Thorndike (Sumarna Surapranata, 2009: 87) adalah sebagai berikut:

Tabel 1. Enam Faktor Penyebab Terjadinya Perbedaan Skor

1	**Karakteristik umum yang permanen peserta tes** a. Kemampuan yang dimiliki peserta didik dalam menghadapi tes b. Kemampuan umum dan teknik yang digunakan ketika mengambil tes c. Kemampuan umum untuk memahami petunjuk tes
2	**Karakteristik khusus yang permanen peserta tes** a. Khusus yang berkaitan dengan tes secara keseluruhan 1) Kemampuan peserta didik yang berkaitan dengan atribut yang diukur dalam sebuah tes 2) Pengetahuan dan kemampuan khusus yang berkaitan dengan soal 3) Keajegan respon peserta didik terhadap pilihan jawaban. b. Khusus yang berkaitan dengan soal 1) Pengetahuan khusus yang berkaitan dengan fakta atau konsep khusus 2) Pengetahuan dan kemampuan yang berkaitan dengan soal
3	**Karakteristik umum yang temporer seperti :**

	a. Kesehatan b. Kelelahan c. Motivasi d. Gangguan emosi e. Kemampuan umum dan teknik yang digunakan ketika mengambil tes f. Pemahaman mekanisme tes g. Faktor panas, cahaya, ventilasi, dan lain sebagainya
4	**Karakteristik khusus yang temporer seperti :** a. Khusus yang berkaitan dengan tes secara keseluruhan 1) Pemahaman terhadap petunjuk khusus 2) Trik atau teknik-teknik mengatasi tes 3) Pengalaman / latihan menghadapi tes terlebih lagi dalam tes psikomotor 4) Kebiasaan menghadapi sebuah tes b. Khusus yang berkaitan dengan soal 1) Fluktuasi ingatan yang dimiliki peserta didik 2) Hal-hal yang berkaitan dengan perhatian dan keakuratan
5	**Faktor penyelenggaraan** a. Waktu, bebas dari gangguan, dan petunjuk yang jelas b. Pengawasan c. Penskoran

6	**Faktor yang tidak pernah diperhitungkan** a. Keberuntungan karena faktor menebak b. Mengingat soal yang telah dilihatnya

Pengukuran yang baik tentunya akan tetap membedakan kemampuan peserta didik. Dalam kenyataannya, pengukuran itu tidak pernah sempurna. Teori reliabilitas dikembangkan untuk menentukan ketetapan tersebut.

Setiap pengukuran, baik dalam sains maupun dalam penilaian kelas selalu mengandung kesalahan pengukuran. Konsep dasar reliabilitas juga menggunakan kesalahan pengukuran tersebut. Setiap skor yang diperoleh seorang peserta tes terdiri atas tiga hal, skor amatan yang sering pula disebut sebagai skor perolehan (*observed test score*) skor yang sebenarnya (*true score*) yaitu skor yang sesuai dengan kemampuan peserta tes yang sebenarnya, dan kesalahan pengukuran, yaitu faktor-faktor yang mempengaruhi ketidakajegan suatu pengukuran adalah situasi yang mempengaruhi perolehan skor. Secara umum konsep dasar tersebut dinyatakan dalam persamaan matematis sebagai berikut (Sumarna Surapranata, 2009: 88)

$$\mathbf{X = T + \boldsymbol{e}}$$

X : Skor perolehan (*observed test score*)

T : Skor sebenarnya (*true score*)

e : Kesalahan pengukuran (*Error of measurement*)

Kesalahan pengukuran merupakan selisih antara skor amatan (perolehan) dengan skor sebenarnya.

$$\boldsymbol{e} = \mathbf{X} + \mathbf{T}$$

Atas dasar konsep ini reliabilitas tes dikembangkan (Suryana Surapranata, 2009: 88). Berbagai hasil telah diturunkan dari teori ini yang pada akhirnya mengandung beberapa implikasi terhadap pengukuran (penilaian berbasis kelas khususnya). Seperti yang tertera pada tabel Teori Thorndike, sumber-sumber kesalahan perolehan skor yang boleh jadi memberikan kontribusi terhadap ketidakajegan skor. Teori reliabilitas ditentukan berdasarkan varian antar skor perolehan yang merupakan penjumlahan dari varian skor sebenarnya dengan varian kesalahan pengukuran yaitu:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Keterangan :

σ_X^2 : varian antar skor perolehan

σ_T^2 : varian antar skor perolehan

σ_E^2 : varian kesalahan pengukuran

Persamaan ini menunjukkan bahwa skor tes ditentukan oleh dua hal, yaitu variabilitas skor sebenarnya dan variabilitas kesalahan pengukuran. Jika kesalahan pengukuran itu memberikan kontribusi yang signifikan, maka tes menjadi tidak ajeg. Dengan kata lain, apabila varian kesalahan pengukuran relatif tidak berarti, maka pengukuran menjadi ajeg.

Koefisien reliabilitas (r_{xx}) menyajikan indeks relatif pengaruh skor sebenarnya dan skor kesalahan pengukuran terhadap skor yang diperoleh. Persamaan umum reliabilitas diturunkan dari perbandingan antara varian skor sebenarnya dengan varian skor perolehan

$$r_{XX} = \frac{\sigma_T^2}{\sigma_X^2}$$

$$r_{XX} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}$$

Tujuan utama mengestimasi reliabilitas adalah untuk menentukan seberapa besar variabilitas yang terjadi akibat adanya kesalahan pengukuran dan seberapa besar variabilitas skor tes sebenarnya. Menurut teori klasik, sebagaimana telah diuraikan di atas, reliabilitas dihubungkan dengan pengertian adanya ketepatan suatu tes dalam pengukurannya. Nunnaly (1970), Allen dan Yen (1979) dan Anastasi (1986) (Suryana Surapranata, 2009: 89) menyatakan bahwa reliabilitas adalah kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda atau dari satu pengukuran ke pengukuran lainnya. Jadi reliabilitas dapat dinyatakan sebagai tingkat keajegan atau kemantapan hasil dari hasil dua pengukuran terhadap hal yang sama. Hasil pengukuran itu diharapkan akan sama apabila pengukuran itu diulangi.

Reliabilitas memiliki dua keajegan. Keajegan yang pertama adalah keajegan internal, yakni tingkat sejauhmana butir soal itu homogen baik dari segi tingkat kesukaran maupun bentuk soalnya keajegan yang kedua yaitu keajegan

eksternal yakni tingkat sejauhmana skor dihasilkan tetap sama sepanjang kemampuan orang yang diukur belum berubah (Sumarna Surapranata, 2009: 89).

Perbedaan skor dari satu pengukuran ke pengukuran lainnya dapat saja terjadi. Menurut Allen dan Yen (Suryana Surapranata, 2009: 90) perbedaan skor dari satu pengukuran ke pengukuran lain ini terjadi karena adanya *standard error of measurement* atau standar kesalahan pengukuran. Oleh karena itu koefisien reliabilitas harus benar-benar diperhitungkan lebih dahulu standar kesalahan pengukurannya itu. Untuk itu dalam pengukuran perlu diidentifikasi sumber-sumber utama yang menyebabkan terjadinya kesalahan pengukuran tersebut. Nunnaly (Suryana Surapranata, 2009:90) menyebutkan bahwa sumber kesalahan pengukuran itu antara lain (1) variasi dalam tes itu sendiri, (2) struktur sampel yang dipilih, (3) variasi di antara tes yang sedang digunakan. Menurut Crocker dan Algina (Suryana Surapranata, 2009: 90) sumber utama kesalahann pengukuran disebabkan oleh (1) perubahan dalam kelebihan waktu yang diuji, (2) isi bentuk ke bentuk, dan (3) sampel isi soal yang cacat. Sumber kesalahan pengukuran bisa juga terjadi karena pengaruh teknik pemilihan sampel dan situasi yang ada pada masing-masing individu yang dijadikan sampel.

Perbedaan pengertian reliabilitas sangat bergantung kepada bagaimana indeks reliabilitas dihitung. Paling tidak terdapat empat konsep reliabilitas yaitu (1) *paralel* atau *ekuivalen,* (2) *test-retest* atau *stabilitas,* (3) *split-half* atau *belah-dua* dan (4) *internal consistency*. Sebagian orang berpendapat bahwa metode *split-half* atau *belah dua* merupakan bagian dari metode keajegan internal (*internal consistency*) sehingga pembagian metode menjadi tiga bagian yaitu

(1)*ekuivalen,* (2) *stabilitas,* dan (3) *internal consistency.* Kedua teknik ini pada prinsipnya sama. Tabel berikut menunjukkan bentuk reliabilitas dan prosedur untuk memperolehnya (Sumarna Surapranata, 2009: 91).

Tabel 2. Metode untuk Menentukan Reliabilitas

Bentuk Reliabilitas	**Prosedur untuk memperoleh**
a. *Test-retest methods* (Stabilitas) Produk Momen dan Korelasi intra kelas	Sajikan tes yang sama sebanyak dua kali kepada peserta tes yang sama dalam waktu yang berbeda dan tentukan korelasi
b. *Paralel* (Ekuivalen) Produk Momen dan Korelasi intra kelas	Sajikan dua tes yang sama kepada peserta tes yang sama dalam waktu yang relatif tidak lama (misalnya dua minggu). Korelasikan kedua skor tersebut untuk mencari reliabilitasnya.
c. *Split-half methods* (belah dua) Persamaan *split-half* dan Spearman-Brown	Sajikan satu kali tes lalu dibelah dua, gunakan persamaan untuk mengkorelasikan keduanya
d. Internal consistency 1) Koefisien alpha	Berikan sekali tes, gunakan persamaan

2) Kuder Richardson (KR-20)	Berikan sekali tes, gunakan persamaan
3) Kuder Richardson(KR-21)	Berikan sekali tes, gunakan persamaan

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 hubungannya rendah, maka tes itu dikatakan tidak reliabel.

Besar kecilnya reliabilitas suatu tes ditentukan oleh besar kecilnya nilai korelasi hasil tes yang dinamakan indeks reliabilitas.Untuk mengestimasi reliabilitas banyak formula yang dapat digunakan. Pada tahun 1986 Crocker dan Algina (Suryana Surapranata, 2009: 91) memberikan pendekatan untuk mengestimasi reliabilitas dengan memperhatikan sumber kesalahan utama melalui penggunaan koefisien reliabilitas, ekuivalensi dan keajegan internal. Pada tahun 1954 Guilford (Suryana Surapranata, 2009: 91) memberikan beberapa modifikasi yang dilakukan oleh Tucker unuk memperbaiki dan menyederhanakan ketidakakuratan formula Kuder-Richardson yang penggunaan analisis varian serta formula khusus. Pada umumnya untuk menentukan estimasi reliabilitas khususnya dalam bidang pengukuran prestasi belajar digunakan internal keajegan seperti formula Cronbach alpha ataupun Kuder-Richarson. Akan tetapi dalam prakteknya penggunaan formula itu memungkinkan adanya usaha-usaha tertentu untuk meninggikan koefisien reliabilitasnya dengan cara mengubah pola susunan skor,

sehingga untuk menghindari hal itu maka penggunaan analisis varian lebih dapat dipertanggung jawabkan.

Koefisien reliabilitas soal pilihan ganda lebih mudah dimanfaatkan dalam pengambilan keputusan prestasi belajar. Gronlund (Suryana Surapranata, 2009: 92) menyebutkan bahwa untuk pengambilan keputusan individu, koefisien reliabilitasnya harus tinggi. Tinggi rendahnya koefisien reliabilitas dipengaruhi oleh beberapa faktor. Crocker dan Algina (Suryana Surapranata, 2009: 92) menyebutkan bahwa faktor itu antara lain panjang suatu tes, kecepatan, homogenitas belahan dan tingkat kesukaran soal. Menurut Alen dan Yen dan Crocker dan Algina tingkat kesukaran itu memegang peranan paling dominan. Hasil penelitian Aiken (Suryana Surapranata, 2009: 92) menunjukkan bahwa pengaruh bahwa pengaruh tingkat kesukaran memegang peranan paling besar pada koefisien reliabilitas. Hal ini disebabkan karena menyangkut variasi jumlah soal yang dapat dijawab benar. Semakin sukar soal-soal dalam perangkat tes akan semakin besar pula variasi skor yang yang diperoleh belahan. Dengan demikian maka akan semakin besar pula reliabilitas tes tersebut. Sebaliknya semakin rendah tingkat kesukaran suatu soal semakin kecil pula reliabilitasnya. Untuk itu harus dihindari banyaknya terkaan yang dilakukan peserta tes dan diusahakan menyesuaikan pengetahuan peserta tes dengan materi tes yang akan diujikan kepada mereka. Alen dan Yen, Nunnaly, dan Gronlund menyatakan bahwa penambahan panjang tes akan menaikkan koefisien reliabilitas sepanjang soal yang digunakan untuk menambah tes itu memiliki kualitas yang sama baik dengan soal-soal lainnya (Sumarna Surapranata, 2009 : 92).

Kriteria penetapan batas minimal koefisien reliabilitas yang digunakan dalam penelitian ini adalah mengacu pendapatnya Nunnally (1978: 245) yaitu koefisien reliabilitas sebesar 0,7 atau lebih dianggap memenuhi syarat. Dengan demikian batas minimal koefisien reliabilitas adalah 0,7. Selain Nunnally, Fraenkel dan Wallen menyatakan " *For research purposes, a useful role of thumb is that reliability should be at least 0.70 and preferably higher".*

7. **Tingkat kesukaran soal**

Soal yang baik adalah soal yang tidak terlalu mudah atau 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 mencoba lagi karena di luar jangkauannya.

Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran *(difficulty index).* Besarnya indeks kesukaran antara 0,00 sampai dengan 1,00. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,00 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soal terlalu mudah.

Di dalam istilah evaluasi, indeks kesukaran ini diberi simbol P, singkatan dari kata "proporsi". Dengan demikian maka soal dengan P = 0,70 lebih mudah jika dibandingkan dengan P = 0,20. Sebaliknya soal dengan P = 0,30 lebih sukar daripada soal dengan P = 0,80.

Melihat besarnya bilangan indeks ini maka lebih cocok jika bukan disebut sebagai indeks kesukaran tetapi indeks kemudahan atau indeks fasilitas, karena

semakin mudah soal itu, semakin besar pula bilangan indeksnya. Akan tetapi telah disepakati bahwa walaupun semakin tinggi indeksnya menunjukkan soal yang semakin mudah, tetapi tetap disebut indeks kesukaran (Suharsimi Arikunto, 2009 : 211-212).

Rumus mencari P adalah :

$$P = \frac{B}{JS}$$

P : indeks kesukaran

B : banyaknya siswa yang menjawab soal itu dengan betul.

JS : jumlah seluruh siswa peserta tes

Pada analisis butir soal secara klasikal, tingkat kesukaran (P) dapat diperoleh dengan beberapa cara antara lain :

a. Skala kesukaran linear
b. Skala bivariat
c. Indeks Davis
d. 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 (Bahrul Hayat dan Suprananto, 1999: 121):

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

Keterangan :

P : proporsi menjawab benar pada butir soal tertentu

∑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 tersebut juga dinamakan kesukaran rata-rata. Indeks kesukaran ini paling banyak dipergunakan untuk menentukan tingkat kesukaran soal.

Tingkat kesukaran ini mengandung banyak kelemahan antara lain : tingkat kesukaran sebenarnya merupakan ukuran kemudahan soal karena makin tinggi indeks tingkat kesukaran (P) maka makin mudah soalnya dan sebaliknya makin rendah tingkat kesukaran (P) makin sulit soalnya. Dengan demikian ini lebih tepat dinamakan indeks atau tingkat kemudahan (*easiness*). Hal ini dapat lebih mempunyai hubungan yang langsung antara besarnya angka (indeks) dengan istilah, yaitu makin tinggi nilai P makin mudah soal tersebut. Besarnya tingkat kesukaran berkisar antara 0 sampai 1. Tingkat kesukaran dikategorikan menjadi tiga bagian seperti nampak pada tabel berikut ini (Bahrul Hayat dan Suprananto, 1999: 121):

Tabel 3.Kategori Tingkat Kesukaran Soal

Proportion Correct (P)	**Kategori Soal**
P > 0,70	Mudah

0,30≤ P ≤ 0,70	Sedang
P < 0,30	Sukar

Suatu soal kadang dikategorikan keadaan ekstrim sukar, yaitu apabila P mendekati nol dan ekstrim mudah apabila P mendekati satu. Apabila akan memasukkan soal berdasarkan analisis tingkat kesukaran ke dalam bank soal, beberapa hal yang perlu diperhatikan antara lain (Bahrul Hayat dan Suprananto, 1999: 123) :

a. Soal yang ekstrim mudah atau soal yang ekstrim sukar tidak memberikan informasi yang berguna bagi sebagian besar peserta tes. Oleh sebab itu, soal seperti ini kemungkinan distribusi jawaban pada alternatif jawaban ada yang tidak memenuhi syarat.
b. Sekalipun soal terlalu sukar atau terlalu mudah, apabila setiap pengecoh (distribusi jawaban) pada soal tersebut menunjukkan kondisi: distribusi jawaban merata atau distribusi jawabannya logis, dan daya pembedanya negatif (kecuali kunci) maka soal-soal tersebut masih memenuhi syarat untuk diterima.
c. Sekalipun soal sukar atau terlalu mudah, namun apabila memiliki daya pembeda dan statistika pengecoh memenuhi kriteria, maka soal tersebut dapat dipilih dan diterima sebagai salah satu alternatif untuk disimpan dalam bank soal.

d. Apabila soal ekstrim mudah atau ekstrim sukar, serta daya pembeda dan statistik pengecohnya belum memenuhi kriteria, maka soal tersebut perlu direvisi dan diujicoba lagi.

8. Daya Pembeda

Daya pembeda atau 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 pembeda adalah untuk melihat kemampuan butir soal untuk butir soal dalam membedakan antara peserta didik yang berkemampuan tinggi dengan peserta didik yang berkemampuan rendah. Ada beberapa cara yang digunakan untuk menentukan daya pembeda, antara lain dengan menggunakan (Bahrul Hayat dan Suprananto, 1999 :hlm 124).

a. Daya pembeda
b. Indeks korelasi
c. Indeks keselarasan

(Bahrul Hayat dan Suprananto, 1999 :hlm 124)

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah).

Angka yang menunjukkan besarnya daya pembeda disingkat dengan D. Seperti halnya indeks kesukaran, menurut Suharsimi daya pembeda ini berkisar antara 0,00 sampai 1,00. Hanya bedanya, indeks kesukaran tidak mengenal

tanda negatif tetapi pada daya pembeda ada tanda negatif. Tanda negatif pada daya pembeda digunakan bila suatu soal "terbalik" menunjukkan kualitas *testee* (Suharsimi Arikunto, 1997: 215).

Rumus untuk menentukan daya pembeda (D) adalah sebagai berikut (Suharsimi Arikunto, 1997: 219):

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Keterangan:

J	: jumlah peserta tes.
J_A	: banyaknya peserta kelompok atas.
J_B	: banyaknya peserta kelompok bawah.
B_A	: banyaknya peserta kelompok atas yang menjawab soal itu dengan benar.
B_B	: banyaknya peserta kelompok bawah yang menjawab soal itu dengan benar
$P_A = \frac{B_A}{J_A}$	: Proporsi peserta kelompok atas yang menjawab benar.
$P_B = \frac{B_B}{J_B}$	: Proporsi peserta kelompok bawah yang menjawab benar.

Butir- butir soal yang baik adalah butir soal yang mempunyai daya pembeda 0,40 – 0,70. Sedangkan nilai – nilai P yang dianjurkan oleh penulis soal adalah antara 0,30 – 0,70 namun harus diingat bahwa soal-soal itu tidak berarti mempunyai daya pembeda yang tinggi (Suharsimi Arikunto, 1997: 223).

Tabel 4.Klasifikasi daya pembeda

Skor D	Kategori
$0,00 < D \leq 0,20$	Jelek *(poor)*
$0,20 < D \leq 0,40$	Cukup *(satisfactory)*
$0,40 < D \leq 0,70$	Baik *(good)*
$0,70 < D \leq 1,00$	Baik sekali *(Excellent)*
Untuk Skor D Negatif tidak baik	

Salah satu tujuan analisis kuantitatif soal adalah untuk menentukan dapat tidaknya suatu soal membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada dalam kelompok itu. Indeks yang digunakan dalam membedakan antara peserta tes yang berkemampuan rendah adalah indeks daya pembeda (*item discrimination*). Indeks daya pembeda soal-soal yang ditetapkan dari selisih proporsi yang menjawab dari masing-masing kelompok. Indeks ini menunjukkan kesesuaian antara fungsi soal dengan fungsi tes secara keseluruhan. Dengan demikian validitas soal ini sama dengan daya pembeda soal yaitu daya dalam membedakan antara peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah.

Angka yang menunjukkan besarnya daya pembeda berkisar antara -1 sampai dengan +1. Tanda negatif menunjukan bahwa peserta tes yang berkemampuan rendah dapat menjawab benar sedangkan peserta tes yang berkemampuan tinggi menjawab salah. Dengan demikian soal yang indeks daya pembedanya negatif menunjukkan terbaliknya kualitas peserta tes.

Setiap soal dapat dipandang sebagai bagian yang terpisah dari sebuah tes. Sebuah soal mugkin dapat membedakan kelompok peserta tes secara baik. Sebuah soal mungkin juga tidak dapat kelompok peserta tes (misal soal dengan p = 0 atau p= 1). Sebuah soal juga mungkin membedakan kelompok secara terbalik, yaitu peserta tes yang tidak mampu dapat menjawab soal dengan benar sedangkan peserta tes yang mampu menjawab salah. Salah satu tujuan analisis soal adalah untuk mencari-cari soal yang dapat mengukur kemampuan secara tepat.

Jika tes atau soal mengukur hal yang sama, dapat diharapkan bahwa setiap peserta tes yang mampu dapat menjawab soal dengan benar, dan peserta tes yang tidak mampu akan menjawab soal dengan salah. Dengan kata lain, soal-soal tersebut membedakan (*discriminate*) antara peserta tes yang mampu dengan peserta tes yang tidak mampu. Dalam bagian ini akan dibahas berbagai teknik menentukan indeks daya pembeda.

Indeks daya pembeda dihitung atas dasar pembagian kelompok menjadi dua bagian, yaitu kelompok atas yang merupakan kelompok peserta tes yang berkemampuan tinggi dengan kelompok bawah yang berkemampuan rendah, kemampuan tinggi ditunjukkan dengan perolehan skor yang tinggi dan kemampuan rendah ditunjukkan dengan perolehan skor yang rendah. Menurut Crocker dan Algina daya pembeda didefinisikan sebagai selisih antara proporsi jawaban benar pada kelompok atas dengan proporsi jawaban benar pada kelompok bawah. Pembagian kelompok ini dapat dilakukan dengan berbagai macam metode bergantung pada keperluannya. Menurut Kelley, Crocker dan

Algina, yang paling stabil dan sensitif serta paling banyak digunakan adalah dengan menentukan 27% kelompok atas, 27% kelompok bawah.

Metode kelompok ekstrim dapat juga digunakan untuk menghitung daya pembeda soal. Jika tes diberikan kepada peserta tes yang cukup besar, daya pembeda soal dapat ditentukan dengan membandingkan dengan kelompok yang memiliki skor tinggi (misalnya 25% teratas) dengan kelompok yang memiliki skor yang rendah (misalnya 25% terbawah). Jika soal dapat membedakan dengan baik kedua kelompok tersebut, maka kebanyakan peserta tes pada kelompok atas akan dapat menjawab benar dan kebanyakan peserta tes pada kelompok bawah akan menjawab salah. Indeks daya pembeda (D) ditentukan berdasarkan gambaran sederhana tersebut. Tahapan pertama dalam menghitung indeks daya pembeda adalah menentukan kelompok atas dan kelompok bawah. Umumnya, para ahli tes membagi kelompok ini menjadi 27% atau 33% kelompok atas dan 27 atau 33% kelompok bawah (Cureton, 1957). Untuk berbagai macam keperluan, pembagian kelompok dapat 50% kelompok atas dan 50% kelompok bawah atau berkisar antara 25% sampai dengan 35% kelompok atas dan kelompok bawah (Sumarna Surapranata, 2009 : 23-24).

Hubungan antara tingkat kesukaran dan daya pembeda

Tingkat kesukaran berpengaruh langsung pada daya pembeda soal. Jika setiap orang memilih benar jawaban (p = 1), atau jika setiap orang menjawab

soal (p = 0), maka soal tidak dapat digunakan untuk membedakan kemampuan peserta tes.

Daya pembeda soal pilihan ganda

Menurut Engelhart (Sumarna Suryapranata, 2009: 31), daya pembeda ditentukan dengan melihat kelompok atas dan kelompok bawah berdasarkan skor total. Perhitungan D sangatlah sederhana, dan menyajikan informasi yang dapat membedakan masing-masing kelompok berdasarkan kemampuan mereka.

Daya pembeda menurut indeks daya pembeda ini dapat dicari dengan menggunakan persamaan:

$$D = \frac{JK_a \quad - \quad JK_b}{nK_a}$$

$$D = \frac{JK_a \quad - \quad JK_b}{nK_b}$$

D : daya pembeda (validitas)

JK_a : jumlah peserta tes menjawab benar pada kelompok atas

JK_b : jumlah peserta tes menjawab benar pada kelompok bawah

nK_a : jumlah peserta tes pada kelompok atas

nK_b : jumlah peserta tes kelompok bawah

Prosedur yang telah digambarkan diatas untuk menghitung D sangatlah sederhana. Namun demikian, sekalipun sederhana, metode untuk mengitung D dapat digunakan formula sebagai berikut :

$$D = \frac{\sum A}{n_A} - \frac{\sum B}{n_B}$$

D : daya pembeda

∑A : jumlah peserta tes yang menjawab benar pada kelompok atas

∑B : jumlah peserta tes yang menjawab benar pada kelompok bawah

n_A : jumlah peserta tes kelompok atas

n_B : jumlah peserta tes kelompok bawah

Dalam kebanyakan kasus, jumlah peserta tes kelompok atas sama dengan jumlah peserta tes kelompok bawah, $n_A = n_B = n$. Dengan demikian maka persamaan daya pembeda menjadi :

$$D = \frac{\sum A - \sum B}{n}$$

9. Analisis Pengecoh

Analisis pola jawaban soal dalam tes pilihan ganda disebut juga dengan analisis *distracor* (pengecoh). Yang dimaksud pola jawaban soal di sini adalah distribusi testee dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih jawaban a, b, c, d atau e atau yang tidak memilih pilihan manapun *(blank o)*. Dalam istilah evaluasi disebut omit, disingkat O.

Dari pola jawaban soal dapat ditentukan apakah pengecoh berfungsi sebagai pengecoh dengan baik atau tidak. Pengecoh yang tidak dipilih sama sekali oleh *testee* berarti bahwa pengecoh itu jelek, terlalu menyolok menyesatkan. Sebaliknya sebuah pengecoh dapat dikatakan berfungsi dengan baik apabila pengecoh tersebut mempunyai daya tarik yang besar bagi pengikut–pengikut tes yang kurang memahami konsep atau kurang menguasai bahan (Suharsimi Arikunto, 1997 : 225).

Berdasarkan beberapa kajian teori yang telah disampaikan, secara umum dapat disampaikan bahwa dengan melihat pola jawaban soal dapat diketahui beberapa hal antara lain taraf kesukaran soal, daya pembeda soal dan baik tidaknya kualitas distraktor.

10. Matematika

Ada tiga kemampuan dasar yang harus di miliki oleh seorang individu agar mampu untuk terus belajar dan mengembangkan potensi dirinya. Ketiga kemampuan tersebut adalah kemampuan membaca, menulis dan berhitung *(calistung)*. Pada perkembangan ketiga kemampuan ini yang pertama kali di kembangkan pada seorang anak adalah kemampuan berhitung. Sebagai contoh dalam kehidupan keluarga seorang anak balita sudah diajari oleh orang tuanya untuk menghitung dengan media jari tangan. Dalam perkembangannya kemampuan berhitung ini merupakan dasar dari kemampuan seorang anak untuk mempelajari matematika.

Menurut James dalam Erman Suherman, dkk (2003 : 16) menyatakan bahwa matematika adalah ilmu tentang logika, bentuk, susunan, dan konsep-

konsep yang berhubungan satu dengan yang lainnya dengan jumlah yang banyak yang terbagi kedalam tiga bidang yaitu aljabar, analis, geometri. Matematika tumbuh dan berkembangan karena proses berpikir, oleh karena itu logika adalah dasar untuk terbentuknya matematika.

Menurut Herman Hudojo (2005: 11), matematika adalah ilmu pengetahuan yang abstrak dan terstruktur secara urut dan logis. Matematika berkenaan dengan ide-ide, struktur-struktur dan hubungannya diatur dalam urutan yang logis dan berkenaan dengan konsep-konsep yang abstrak

Dibawah ini terdapat beberapa definisi matematika yang dibentuk oleh beberapa pakar yang diungkapkan oleh Reys, et al (1998 : 2).

a. Matematika adalah studi atau kajian tentang pola dan hubungan.
b. Matematika adalah suatu cara berpikir.
c. Matematika adalah seni, digolongkan dengan tata urutan dan kejelasan didalamnya.
d. Matematika adalah suatu bahasa, menggunakan istilah dan simbol tertentu dengan hati-hati.
e. Matematika adalah suatu alat.

Soedjadi (2000 : 5) juga menguraikan beberapa karakteristik matematika, yaitu memiliki objek kajian abstrak; bertumpu pada kesepakatan; berpola pikir deduktif; memiliki simbol yang kosong dari arti; memperhatikan semesta pembicaraan. dan konsisten dalam sistemnya.

Menurut Erman Suherman (2001: 54-55) matematika sekolah adalah matematika yang diajarkan pada pendidikan formal, yaitu di Pendidikan

Dasar (SD dan SLTP) dan Pendidikan Menengah (SLTA dan SMK). Matematika sekolah ini terdiri dari materi-materi atau konsep- konsep yang berhubungan dengan kehidupan sehari-hari. Hal ini menunjukkan bahwa matematika sekolah tetap memiliki ciri-ciri yang dimiliki matematika, yaitu memiliki objek kajian yang abstrak serta berpola pikir deduktif konsisten. Fungsi mata pelajaran matematika adalah sebagai alat, pola pikir dan sebagai ilmu atau pengetahuan.

Berdasarkan pemahaman di atas dapat disampaikan bahwa matematika adalah suatu pengetahuan yang mengkaji pola, hubungan serta objek abstrak dengan menggunakan istilah dan simbol tertentu berdasarkan kesepakatan, dengan menggunakan pola pikir deduktif.

11. Tes Matematika yang Baik

Tes matematika merupakan salah satu tes hasil belajar yang sahih untuk menilai penguasaan siswa terhadap kompetensi yang ditetapkan maka dalam penyusunannya perlu memperhatikan prinsip-prinsip sebagai berikut :

a. Tes hasil belajar hendaknya mengukur secara jelas hasil belajar yang telah ditetapkan sesuai tujuan instruksional, artinya tes mengukur perilaku yang hendak diukur (valid). Oleh karena itu hasil belajar harus dirumuskan secara jelas dan cermat sehingga dapat mengukur tujuan instruksional yang diharapkan.
b. Tes hasil belajar hendaknya dapat mengukur sampel yang representatif dari hasil belajar dan materi pelajaran yang tercakup dalam program pengajaran atau instruksional.

c. Tes hasil belajar hendaknya mencakup jenis-jenis pertanyaan yang paling sesuai untuk mengukur hasil belajar yang diinginkan.

d. Tes hasil belajar hendaknya direncanakan agar hasilnya sesuai dengan tujuan dan fungsinya.

e. Reliabilitas tes hasil belajar diusahakan setinggi mungkin dan hasil ukurnya harus ditafsirkan dengan hati-hati.

Kelima prinsip diatas merupakan dasar atau pedoman bagi dalam menyusun tes matematikan yang sesuai (Lambas dkk, 2004 : 7).

B. Kajian Penelitian Yang Relevan

Penelitian yang relevan dengan penelitian ini antara lain :

1. Penelitian yang dilakukan oleh Muslikah Purwanti pada tahun 2013 "Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010". Hasil penelitian menunjukkan bahwa : (1) butir soal pilihan ganda valid 19 butir (63,33%) soal tidak valid 11 butir (36,67%), soal uraian valid 3 butir (75%) tidak valid 1 butir (25%) ; (2) soal pilihan ganda indeks reliabilitas menunjukkan angkka 0,66, uraian sebesar 0,50 sehingga tidak reliabel; (3) bentuk soal pilihan ganda 4 butir (13,33%) kategori sukar, 9 butir soal (30%) sedang dan 16 butir (56,67%) mudah, bentuk soal uraian 4 butir (100%) kategori soal tingkat sedang; (4) butir soal pilihan ganda dengan daya pembeda jelek 7 butir (23,33%), cukup 7 butir (23,33%), baik 10 butir (33,33%), baik sekali 6 butir (20%), bentuk soal uraian dengan daya pembeda jelek 1 butir (25%), cukup 1 butir (25%), dan baik sekali 2 butir (50%); (5) bentuk soal pilihan ganda yang termasuk soal

dengan pengecoh atau *distractor* yang berkualitas 3 butir (10%) memiliki pengecoh sangat baik, 10 butir (33,33%) baik, 11 butir (36,67%) cukup, 4 butir (13,33%) kurang baik, dan 2 butir (6,67%) tidak baik.

2. Penelitian yang dilakukan oleh Lilis Tri Ariyana pada tahun 2011 "Analisis Butir Soal Ulangan Akhir Semester Gasal IPA kelas IX SMP di Kabupaten Grobogan". Hasil analisis kuantitatif seluruh sekolah menunjukkan bahwa : 2% sangat sukar; 20% sukar; 70% sedang; 4% mudah dan 4% sangat mudah. Daya pembeda baik 26%, cukup baik 62%, jelek 10% dan sangat jelek 2%. Efektifitas pengecoh 82% berfungsi. Reliabilitas secara keseluruhan adalah 0,711 artinya soal memiliki keajegan tinggi. Analisis kualitatif soal pilihan ganda seuai dengan materi, konstruksi dan bahasa tetapi terdapat beberapa soal yang perlu diperbaiki. Soal uraian sesuai dengan aspek penelaahan tetapi aspek konstruksi perlu diperbaiki. Berdasarkan penelitian disimpulkan bahwa soal valid logis karena sesuai dengan soal standar tapi perlu perbaikan aspek kostruksi pada beberapa soal. Soal memiliki tingkat kesukaran sedang, daya pembeda baik, efektifitas pengecoh berfungsi serta reliabel dengan kategori tinggi.

3. Penelitian yang dilakukan Dewi Rahmayati, Nani Suwarni, Dedy Miswar pada tahun 2011 "Analisis Butir Soal Ujian Semester Ganjil Mata Pelajaran Geografi Siswa Kelas X SMA Bina Mulya Bandar Lampung Tahun Pelajaran 2011/2012". Hasil penelitian menunjukkan bahwa : (1) Dilihat dari validitas tiap butir soal terdapat 11 butir atau 22% butir soal dikatakan valid dan 39 atau 78% yang tidak valid. (2) Ditinjau dari tingkat kesukaran butir soal

terdapat 26 atau 52% butir soal yang termasuk kategori baik dan 24 atau 48% butir soal yang tidak baik. (3) Ditinjau dari daya pembeda terdapat 12 atau 24% butir soal memiliki daya pembeda baik dan 38 atau 76% butir soal yang jelek dan negatif. (4) Dilihat dari pola jawaban diketahui bahwa secara umum semua pengecoh (*distractor*) pada butir soal tersebut sudah berfungsi dengan baik**.**

C. Kerangka Berpikir

Keberhasilan pelaksanaan sebuah proses pembelajaran dapat diketahui dari sejauh mana ketercapaian tujuan pembelajaran yang telah dirumuskan. Begitu pula dalam proses pembelajaran matematika di SMK Kabupaten Purworejo, ketercapaian tujuan pembelajaran matematika selama tiga tahun dapat diketahui dari hasil ujian nasional matematika. Untuk membiasakan dan mengetahui kemampuan peserta didik dalam menghadapi atau mengerjakan soal-soal dalam UN Mata Pelajaran Matematika, pihak yang berwenang di Kabupaten Purworejo menyelenggararakan TUC UN selama beberapa kali. Salah satunya adalah mata pelajaran Matematika.

Untuk menghasilkan data yang baik dan dapat dipergunakan untuk mengetahui sejauh mana kemampuan dan kesiapan siswa dalam menghadapai UN dibutuhkan sebuah instrumen (tes) uji coba yang baik pula. Instrumen (tes) yang baik dapat di ketahui dari data hasil tes yang telah dilakukan. Oleh sebab itulah itulah dalam pelaksanaan TUC UN Matematika di Kabupaten Purworejo perlu dilakukan analisis validitas dan reliabilitas untuk mendeskripsikan kualitas tes matematika yang dipergunakan. Bentuk tes

matematika yang digunakan adalah tes tertulis *(pencil and paper test)* dengan bentuk soal pilihan ganda. Dikarenakan bentuk soal yang digunakan adalah soal pilihan ganda maka perlu dilakukan pula analisis tingkat kesukaran soal dan indeks daya pembeda. Selain itu diperlukannya analisis kesalahan jawaban yang dapat digunakan untuk pertimbangan perbaikan peserta dalam persiapan Ujian Nasional.

D. Pertanyaan Penelitian

Berdasarkan kajian teori, hasil penelitian yang relevan dan kerangka pikir yang telah disampaikan pada uraian di atas perntanyaan penelitian yang akan di jawab dalam penelitian ini adalah :

1. Bagaimanakah validitas soal tes uji coba ujian nasional matematika SMK kelompok teknik industri di wilayah Kabupaten Purworejo tahun 2016?
2. Bagaimanakah reliabilitas soal tes uji coba ujian nasional matematika SMK kelompok teknik industri di wilayah Kabupaten Purworejo tahun 2016?
3. Bagaimanakah tingkat kesukaran soal tes uji coba mata pelajaran matematika di SMK kelompok teknik industri di wilayah Kabupaten Purworejo tahun 2016?
4. Bagaimanakah daya pembeda butir soal tes uji coba ujian nasional matematika SMK kelompok teknik industri di wilayah Kabupaten Purworejo tahun 2016?
5. Bagaimanakah analisis kemungkinan kesalahan siswa pada butir soal yang bermasalah?