

**ANALISIS HUBUNGAN ANTARA VALIDITAS BUTIR, RELIABILITAS,
TINGKAT KESUKARAN, DAN DAYA PEMBEDA SOAL UJI COBA YANG
DISUSUN OLEH MAHASISWA UNTUK
INSTRUMEN PENELITIAN**

SKRIPSI

*Diajukan Sebagai Salah Satu Persyaratan Guna Memperoleh
Gelar Sarjana Pendidikan*



**Oleh:
LISA KARNELA
NIM.01851**

**PROGRAM STUDI PENDIDIKAN BIOLOGI
JURUSAN BIOLOGI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2012**

PENGESAHAN

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Program Studi Pendidikan Biologi Jurusan Biologi Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Negeri Padang

Judul : Analisis Hubungan antara Validitas Butir, Reliabilitas,
Tingkat Kesukaran, dan Daya Pembeda Soal Uji Coba
yang Disusun oleh Mahasiswa untuk Instrumen
Penelitian

Nama : Lisa Karnela

NIM/ TM : 01851 / 2008

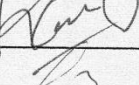
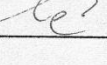
Program Studi : Pendidikan Biologi

Jurusan : Biologi

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 26 April 2012

Tim Penguji

	Nama	Tanda Tangan
1. Ketua	: Muhyiatul Fadilah, S.Si.,M.Pd.	1. 
2. Sekretaris	: Drs. Mades Fifendy, M. Biomed.	2. 
3. Anggota	: Drs. H. Rusdi Adnan	3. 
4. Anggota	: Drs. Anizam Zein, M.Si.	4. 

ABSTRAK

Lisa Karnela : Analisis Hubungan Antara Validitas butir, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Uji Coba yang Disusun oleh Mahasiswa untuk Instrumen Penelitian.

Kajian mengenai adanya keterkaitan antar faktor-faktor yang dianalisis secara empiris belum dilakukan secara komprehensif, seperti ada atau tidaknya hubungan antara tingkat kesukaran dengan daya pembeda, tingkat kesukaran dengan validitas butir, daya pembeda dengan validitas butir dan validitas butir dengan reliabilitas. Sementara analisis terhadap semua faktor ini telah dilakukan oleh berbagai pihak, termasuk peneliti tingkat mahasiswa demi memastikan kualitas instrumen melalui data kuantitatif. Penelitian ini bertujuan untuk mendapatkan informasi mengenai hubungan antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal yang disusun oleh mahasiswa untuk instrumen penelitian.

Penelitian ini merupakan penelitian deskriptif. Populasi penelitian ini adalah semua soal uji coba yang disusun dan dikembangkan sendiri oleh mahasiswa Biologi FMIPA UNP dengan jenis penelitian eksperimen dari tahun 2006-2011. Pada penelitian ini dilakukan penarikan sampel dengan teknik *proportionate stratified random sampling*. Data diolah dengan teknik analisis korelasi menggunakan program SPSS 16.

Dari hasil analisis data diperoleh bahwa: 1). Sebanyak 90,32% tingkat kesukaran tidak berkorelasi dengan daya pembeda, 2). 93,87% tingkat kesukaran tidak berkorelasi dengan validitas butir, 3). 100% daya pembeda berkorelasi dengan validitas butir, 4). 96,77% nilai r reliabilitas butir soal yang valid lebih besar dari nilai r semua butir soal. Dengan ini dapat disimpulkan: faktor yang harus dipertimbangkan dalam pengambilan keputusan untuk menentukan kualitas soal adalah faktor validitas, reliabilitas dan daya pembeda.

KATA PENGANTAR



Alhamdulillah, puji syukur kehadiran Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Hubungan Antara Validitas Butir, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda pada Soal Uji Coba yang Disusun oleh Mahasiswa untuk Instrumen Penelitian". Penulisan skripsi ini adalah untuk memenuhi salah satu syarat memperoleh gelar sarjana pendidikan pada program studi pendidikan Biologi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.

Dalam penelitian dan penulisan skripsi ini, penulis telah banyak mendapat bantuan, dorongan, petunjuk dan bimbingan dari berbagai pihak. Untuk itu, penulis mengucapkan terima kasih kepada yang terhormat:

1. Ibu Dra. Helendra, M.S., sebagai dosen pembimbing I yang telah menyediakan waktu untuk memberikan bimbingan, arahan, motivasi serta masukan dan saran.
2. Ibu Muhyiatul Fadilah, S.Si., M.Pd., sebagai dosen pembimbing II yang telah menyediakan waktu untuk memberikan bimbingan, arahan, motivasi serta masukan dan saran.
3. Bapak Drs. H. Rusdi Adnan, Bapak Drs. Mades Fifendy, M.Biomed., dan Bapak Drs. Anizam Zein, M.Si., sebagai dosen penguji.

4. Bapak Ketua Jurusan Biologi FMIPA UNP yang telah memberikan kemudahan untuk penyusunan skripsi ini.
5. Bapak Drs. Armen, S. U., selaku penasehat akademik, yang telah menyediakan waktu untuk membimbing penulis.
6. Semua pihak yang telah membantu dalam menyelesaikan penulisan skripsi ini.

Semoga semua bimbingan, arahan, saran dan bantuan yang telah diberikan menjadi amal ibadah dan mendapat balasan dari Allah SWT.

Penulis telah berusaha semaksimal mungkin menyelesaikan skripsi ini. Namun jika terdapat kesalahan dan kekurangan, penulis mengharapkan kritikan dan saran dari pembaca demi kesempurnaan skripsi ini. Atas kritik dan saran yang diberikan penulis ucapkan terima kasih.

Padang, April 2012

Penulis

DAFTAR ISI

	Halaman
ABSTRAK.....	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vi
DAFTAR LAMPIRAN.....	vii
BAB I PENDAHULUAN	
A. Latar Belakang Masalah.....	1
B. Identifikasi masalah.....	6
B. Batasan Masalah.....	6
C. Rumusan Masalah.....	6
D. Pertanyaan Penelitian.....	7
E. Tujuan Penelitian.....	7
F. Manfaat Penelitian.....	8
BAB II TINJAUAN PUSTAKA	
A. Kajian Teori.....	9
1. Pengertian Evaluasi.....	9
2. Pengertian Tes.....	10
3. Fungsi Tes.....	11
4. Bentuk-bentuk Tes.....	12

5. Ciri-ciri Tes yang Baik.....	15
6. Analisis Butir Soal.....	16
B. Kerangka Konseptual.....	34
BAB III METODE PENELITIAN	
A. Jenis Penelitian.....	35
B. Definisi Operasional.....	35
C. Populasi dan Sampel.....	36
D. Variabel dan Data.....	37
E. Prosedur Penelitian.....	38
F. Teknik Analisis Data.....	39
BAB IV HASIL DAN PEMBAHASAN	
A. Hasil Penelitian	41
1. Deskripsi Data.....	41
2. Analisis Data.....	42
B. Pembahasan.....	44
BAB V PENUTUP	
A. Kesimpulan.....	55
B. Saran.....	55
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Kriteria reliabilitas soal.....	24
2. Kriteria indeks kesukaran.....	26
3. Kriteria daya pembeda soal.....	30
4. Kriteria daya pembeda soal.....	30
5. Jumlah sampel penelitian.....	36
6. Jumlah butir soal yang dianalisis pada setiap sampel penelitian.....	40
7. Nilai koefisien reliabilitas pada 31 sampel.....	40
8. Variasi nilai tingkat kesukaran dan daya pembeda pada sampel 4 tahun 2010.....	44
9. Distribusi koefisien tingkat kesukaran dan koefisien validitas butir pada sampel 4 tahun 2010.....	47
10. Distribusi koefisien daya pembeda dan validitas butir pada pada sampel 2 tahun 2010	49
11. Perbandingan nilai r semua butir dengan butir yang valid pada 31 sampel.....	52

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Daftar Judul Skripsi dari Instrumen Penelitian.....	59
2. Tabel Distribusi Jawaban Siswa Soal Uji Coba	63
3. Tabel Data Faktor Analisis	94
4. Analisis Korelasi	140
5. Interpretasi Analisis Data	171
6. Surat Permohonan Izin Penelitian	254
7. Surat Izin Melakukan Penelitian	255
8. Surat Keterangan Selesai Melakukan Penelitian	256

BAB I PENDAHULUAN

A. Latar Belakang Masalah

Proses pembelajaran terdiri dari berbagai unsur penting, yaitu: tujuan pembelajaran, pelaksanaan pembelajaran, dan pencapaian hasil belajar sesuai dengan tujuan pembelajaran. Untuk mengetahui hasil belajar telah tercapai atau belum perlu diadakan sebuah tes atau evaluasi.

Evaluasi merupakan cara yang dapat dilakukan untuk mengetahui hasil belajar dan efektifitas pelaksanaan proses pembelajaran. Melalui kegiatan evaluasi ini dapat ditentukan tindak lanjut dari hasil belajar siswa maupun perbaikan pelaksanaan pengajaran oleh guru.

Untuk melakukan suatu evaluasi dibutuhkan instrumen evaluasi. Instrumen merupakan sesuatu yang mempunyai kedudukan sangat penting, karena instrumen akan menentukan kualitas data yang dikumpulkan. Semakin tinggi kualitas instrumen, semakin tinggi pula kualitas hasil evaluasinya.

Salah satu instrumen yang digunakan untuk evaluasi adalah soal ujian, baik berupa soal objektif maupun essay. Soal objektif lebih banyak dipakai karena memiliki beberapa kelebihan, antara lain 1) lebih mudah membuatnya, 2) mampu mewakili materi lebih banyak, 3) memudahkan penilaian, dan 4) objektivitas guru dalam menilai lebih tinggi dibanding soal essay (Arikunto, 2008: 164).

Soal yang digunakan untuk mengevaluasi hasil belajar siswa ini hendaknya sesuai dengan standar yang telah ditentukan. Menurut Arikunto (2008: 57), “Sebuah soal yang dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu memiliki validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis”. Oleh karena itu, sebelum digunakan hendaknya soal tersebut dianalisis terlebih dahulu.

Analisis soal terdiri dari analisis logis (kualitatif) dan analisis empiris (kuantitatif). Analisis logis merupakan penelaahan yang dilakukan terhadap suatu instrumen dalam lingkup kebenaran/ketepatan dan kelayakan soal menurut keilmuan/rasio manusia. Analisis empiris merupakan penelaahan terhadap kebenaran/ketepatan dan kelayakan suatu instrumen berdasarkan respon nyata testee di lapangan. Analisis empiris membutuhkan ujicoba instrumen yang kemudian dilanjutkan dengan analisis data hasil uji coba secara kuantitatif dan dimaknai dengan mengacu pada kategori-kategori tertentu. Secara keseluruhan, analisis instrumen terdiri dari analisis validitas, analisis reliabilitas, penentuan tingkat kesukaran soal, daya pembeda, dan kualitas option (Arikunto, 2008: 65-66). Dari dua jenis analisis, analisis kuantitatif menjadi fokus dalam penelitian ini.

Permasalahan tentang analisis soal adalah penentuan tindakan lanjutan terhadap soal-soal yang telah diujicobakan terhadap sekelompok siswa tertentu. Tindakan lanjutan setelah analisis soal ada tiga jenis yaitu memakai butir soal tersebut, memperbaiki, dan membuang butir soal tersebut. Hal ini sering ditemui oleh para pembuat instrumen tes hasil belajar seperti mahasiswa tingkat akhir

yang akan melakukan penelitian di bidang pendidikan. Secara keilmuan, para mahasiswa telah mampu melakukan analisis soal menurut teori yang diacu. Namun, kesulitan berikutnya adalah menentukan butir-butir soal yang memenuhi syarat untuk dipakai, diperbaiki, dan dibuang dengan mempertimbangkan faktor validitas, reliabilitas, tingkat kesukaran, daya pembeda dan kualitas option secara bersamaan.

Validitas berarti kesahihan. Validitas tes menunjukkan ketepatan pengukuran. Arikunto (2005: 57-58) menyatakan bahwa sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur. Reliabilitas berhubungan dengan masalah kepercayaan. Suatu tes dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap (Arikunto, 2005: 86). Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Tingkat kesukaran ditentukan untuk setiap butir soal (Arikunto, 2008: 212). Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan.

Belum adanya patokan dasar dalam menentukan sebuah soal layak dipakai, diperbaiki atau dibuang menyebabkan bervariasinya dasar penentuan keterpakaian suatu butir soal yang menimbulkan keraguan bagi pembuat soal. Selain itu, penentuan keterpakaian butir soal masih mengacu pada satu atau dua faktor saja. Dengan arti, ada hasil analisis faktor lain yang tidak terpakai. Logikanya, keempat

faktor analisis tersebut memiliki peran dalam menentukan kualitas soal. Salah satu opini yang cukup logis diantara pakar evaluasi adalah bahwa butir soal tersebut sudah dapat dipakai jika telah memenuhi kriteria baik minimal 50 % dari kriteria kualitas soal yang dipakai.

Menurut Sajekti (1988: 38), butir-butir tes yang mengukur suatu ciri sebaiknya berkorelasi satu dengan yang lainnya. Hal ini berarti antara validitas butir, reliabilitas, tingkat kesukaran dan daya pembeda memiliki korelasi. Namun bagaimana pola dari korelasi tersebut belum diketahui.

Mahasiswa tingkat akhir yang melakukan penelitian eksperimen di bidang pendidikan seringkali membuat instrumen tes hasil belajar. Selain itu mahasiswa merupakan salah satu calon tenaga pendidik yang seharusnya benar-benar paham tentang analisis soal. Tapi pada kenyataannya setelah instrumen tersebut selesai disusun, divalidasi dan diujicobakan, mahasiswa juga ragu untuk menentukan apakah semua butir soal tersebut layak dipakai atau tidak. Hal tersebut dikarenakan belum adanya pola untuk menentukan keterpakaian suatu butir soal.

Mahasiswa memiliki pendapat yang beragam dalam menginterpretasikan opini ini. Mereka mencoba merancang kriteria penentuan keterpakaian butir soal, sehingga ditemukan sejumlah kriteria yang bervariasi dan bersifat relatif. Contoh kriteria yang dipakai antara lain : soal dapat dipakai jika 2 faktor telah mencapai kriteria baik, soal dapat dipakai jika hanya valid dan reliabel, soal hanya dapat dipakai jika daya pembeda minimal baik, soal hanya dapat dipakai jika tingkat kesukaran sedang, dan lain-lain.

Sebagai contoh, butir A memiliki data valid, reliabel, daya pembeda jelek, dan tingkat kesukaran sedang. Butir B memiliki data tidak valid, reliabel, daya pembeda jelek, dan tingkat kesukaran sedang. Terhadap data seperti ini, mahasiswa mencoba membuat ketentuan bahwa butir soal bisa diterima jika minimal dua faktor bernilai baik. Dengan demikian dapat disimpulkan butir A memiliki 3 faktor yang bernilai baik, dan butir B memiliki 2 faktor yang bernilai baik. Dengan demikian, kedua butir sama-sama dapat diterima. Namun, setelah dikaji ulang, timbul pertanyaan, “bagaimana bisa butir B yang tidak valid tetap diterima sebagai butir yang berkualitas sejajar dengan butir A yang valid? Pertanyaan seperti ini sering membingungkan mahasiswa. Dengan kata lain, mahasiswa merasa tidak yakin dengan ketentuan yang telah mereka ciptakan sendiri.

Kajian mengenai adanya keterkaitan antar faktor-faktor yang dianalisis secara empiris belum dilakukan secara komprehensif, seperti ada atau tidaknya hubungan antara tingkat kesukaran-daya pembeda, tingkat kesukaran dengan validitas butir, daya pembeda dengan validitas butir dan validitas butir dengan reliabilitas. Sementara analisis terhadap semua faktor ini telah dilakukan oleh berbagai pihak, termasuk peneliti tingkat mahasiswa demi memastikan kualitas instrumen melalui data kuantitatif.

Beberapa penelitian yang relevan telah dilakukan tentang analisis tes hasil belajar biologi. Elfira (2010) dengan penelitian yang berjudul “Analisis Soal Ujian Akhir Semester I Bidang Studi Biologi Kelas XII SMAN 2 Padang Panjang

Tahun Pelajaran 2009/2010". Tika (2011) dengan penelitian yang berjudul "Analisis Soal Tes Hasil Belajar Mata Pelajaran Biologi Semester II Kelas VIII SMP Negeri 1 Padang Panjang Tahun Pelajaran 2009/2010". Namun analisis hubungan antara faktor-faktor analisis soal belum dilakukan.

Berdasarkan latar belakang di atas penulis telah melakukan penelitian dengan judul "Analisis Hubungan Antara Validitas Butir, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Uji Coba yang Disusun oleh Mahasiswa untuk Instrumen Penelitian".

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, ada beberapa masalah yang dapat diidentifikasi, yaitu:

1. Belum jelasnya hubungan antara validitas butir dengan reliabilitas soal.
2. Belum jelasnya hubungan antara validitas butir dengan tingkat kesukaran soal.
3. Belum jelasnya hubungan antara validitas butir dengan daya pembeda soal.
4. Belum jelasnya hubungan antara tingkat kesukaran dengan daya pembeda soal.
5. Belum adanya acuan yang tepat dalam menentukan syarat sebuah soal dapat dipakai, diperbaiki, atau tidak dipakai dengan memperhatikan faktor analisis soal secara bersamaan.

C. Batasan Masalah

Masalah penelitian ini dibatasi pada analisis hubungan antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian.

D. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah “Apakah terdapat hubungan antara validitas butir, reliabilitas, tingkat kesukaran dan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian?”

E. Pertanyaan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, maka pertanyaan penelitian ini adalah:

1. Apakah terdapat hubungan antara validitas butir dengan reliabilitas soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian?
2. Apakah terdapat hubungan antara validitas butir dengan tingkat kesukaran soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian?
3. Apakah terdapat hubungan antara validitas butir dengan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian?
4. Apakah terdapat hubungan antara tingkat kesukaran dengan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian?

F. Tujuan penelitian

Tujuan dari penelitian ini adalah untuk:

1. Mengetahui ada atau tidaknya hubungan antara validitas butir dengan reliabilitas soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian.
2. Mengetahui ada atau tidaknya hubungan antara validitas butir dengan tingkat kesukaran soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian.
3. Mengetahui ada atau tidaknya hubungan antara validitas butir dengan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian.
4. Mengetahui ada atau tidaknya hubungan antara tingkat kesukaran dengan daya pembeda soal uji coba yang disusun oleh mahasiswa untuk instrumen penelitian.

G. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat untuk:

1. Mahasiswa peserta kuliah evaluasi, agar memiliki panduan tambahan dalam menganalisis butir soal.
2. Peneliti lainnya, sebagai landasan untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

A. Kajian Teori

1. Pengertian evaluasi

Evaluasi merupakan suatu tindakan atau proses untuk menentukan nilai dari sesuatu. Sejalan dengan hal tersebut, Sudijono (2008: 1) mengemukakan bahwa:

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*; dalam bahasa Arab: al-Taqdir; dalam bahasa Indonesia berarti: penilaian. Dengan demikian secara harfiah, evaluasi pendidikan (*educational evaluation*) dapat diartikan sebagai: penilaian dalam (bidang) pendidikan atau penilaian mengenai hal-hal yang berkaitan dengan kegiatan pendidikan.

Menurut Daryanto (2005: 6) “Di dalam istilah asingnya penilaian adalah *evaluation*. Dari kata *evaluation* inilah diperoleh kata Indonesia evaluasi yang berarti menilai”. Sejalan dengan itu, Mahrens dan Lehmann dalam Purwanto (1994: 3) mengemukakan: “Dalam arti luas, evaluasi adalah suatu proses merencanakan, memperoleh, dan menyediakan informasi yang sangat diperlukan untuk membuat alternatif-alternatif keputusan”.

Sesuai dengan pengertian tersebut maka setiap kegiatan evaluasi atau penilaian dalam pembelajaran merupakan suatu proses sistematis yang sengaja direncanakan untuk memperoleh informasi atau data, berdasarkan data tersebut kemudian dibuat suatu keputusan sampai sejauh mana tujuan-

tujuan pembelajaran atau penguasaan materi pelajaran oleh siswa. Salah satu teknik untuk mengevaluasi hasil belajar itu adalah dengan tes.

2. Pengertian tes

Tes merupakan sebuah alat yang digunakan untuk mengukur apa yang hendak diukur. Sejalan dengan itu, Arikunto (2008: 32) mengemukakan bahwa tes adalah serentetan pertanyaan atau latihan atau alat lain yang digunakan untuk mengukur keterampilan, pengetahuan, intelegensi, kemampuan, atau bakat yang dimiliki oleh individu atau kelompok.

Menurut Nurkancana (1986: 24), Tes adalah suatu cara mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang dikerjakan oleh anak atau sehingga menghasilkan suatu nilai tentang tingkahlaku prestasi anak tersebut, yang dapat dibandingkan dengan nilai yang dicapai oleh anak-anak lain atau dengan nilai standar yang ditetapkan”.

Menurut Sudijono (2008: 66), “ *Test* adalah alat atau prosedur yang dipergunakan dalam rangka pengukuran dan penilaian; *testing* berarti saat dilaksanakannya atau peristiwa berlangsungnya pengukuran dan penilaian; *tester* artinya orang yang melaksanakan tes, atau pembuat tes, atau eksperimenter, yaitu orang yang sedang melakukan percobaan (eksperimen); sedangkan *testee* (*mufrad*) dan *testees* (*jama`*) adalah pihak yang sedang dikenai tes (peserta tes atau peserta ujian), atau pihak yang sedang dikenai percobaan (*tercoba*)”.

Berdasarkan uraian di atas maka dapat disimpulkan tes merupakan sebuah alat yang berisikan pertanyaan yang harus dijawab oleh siswa sehingga mampu menghasilkan suatu nilai tentang prestasi anak tersebut. Tes tersebut disusun dan diujukan kepada siswa. Untuk mengetahui hasilnya, maka nilai tersebut dibandingkan dengan standar yang telah ditetapkan.

Dalam dunia pendidikan, kemampuan siswa dalam menjawab pertanyaan atau latihan ini dapat dinyatakan dalam bentuk angka-angka sebagai indeks prestasi. Dimana angka-angka tersebut akan menunjukkan tingkat pencapaian hasil belajar siswa.

3. Fungsi tes

Menurut Sudijono (2008: 67), secara umum ada dua fungsi yang dimiliki oleh tes, yaitu:

- a. Sebagai alat pengukur terhadap peserta didik. Dalam hubungan ini tes berfungsi mengukur tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah mereka menempuh proses belajar mengajar dalam jangka waktu tertentu.
- b. Sebagai alat pengukur keberhasilan program pengajaran, sebab melalui tes tersebut akan dapat diketahui sudah seberapa jauh program pengajaran yang telah ditentukan, telah dapat tercapai.

Berdasarkan pendapat Sudijono, tes tidak hanya untuk mengukur perkembangan peserta didik, tetapi juga diperlukan untuk melihat ketercapaian program pengajaran.

4. Bentuk-bentuk tes

Daryanto (2005:28) mengemukakan bahwa “Pada umumnya teknik evaluasi yang digunakan dapat digolongkan menjadi 2 macam, yaitu teknik tes dan teknik non tes”. Sedangkan menurut Thoha (2003:55-59):

Tes dapat dibagi menjadi dua golongan berdasarkan cara mengajukan dan menjawab pertanyaan yaitu :

- a. Tes tertulis (*Pencil and Paper Test*), yakni jenis tes dimana *tester* dalam mengajukan butiran-butiran pertanyaan atau soalnya dilakukan secara tertulis dan *testee* memberikan jawaban dengan cara tertulis juga.
- b. Tes lisan (*Nonpencil and Paper Tes*), yakni tes dimana *tester* di dalam mengajukan butiran-butiran pertanyaan atau soalnya dilakukan secara lisan dan *testee* memberikan jawaban dengan cara lisan juga.

Menurut Arikunto (2008: 162-177), tes dapat dibedakan atas dua bentuk, yaitu:

a. Tes subjektif

Pada umumnya berbentuk esai (uraian). Tes bentuk esai adalah sejenis tes kemajuan belajar yang memerlukan jawaban yang bersifat pembahasan atau uraian kata-kata. Ciri-ciri pertanyaannya didahului dengan kata-kata seperti; uraikan, jelaskan, mengapa, bagaimana, bandingkan, simpulkan, dan sebagainya.

Soal-soal bentuk esai biasanya jumlahnya tidak banyak, hanya sekitar 5-10 buah soal dalam waktu kira-kira 90 s.d. 120 menit.

b. Tes objektif

Tes objektif adalah tes yang dalam pemeriksaannya dapat dilakukan secara objektif. Dalam penggunaan tes objektif, jumlah soal yang diajukan jauh lebih banyak daripada tes esai. Kadang-kadang untuk tes yang berlangsung selama 60 menit dapat diberikan 30-40 soal.

Macam-macam tes objektif:

1) Tes benar-salah (*true-false*)

Soalnya berupa pernyataan-pernyataan (*statement*). *Ststatement* tersebut ada yang benar dan ada yang salah. *Testee* bertugas untuk menandai masing-masing pernyataan itu dengan

melingkari huruf B jika pernyataan itu betul menurut pendapatnya dan melingkari huruf S jika pernyataannya salah.

2) Tes pilihan ganda (*multiple choice test*)

Multiple choice test terdiri atas suatu keterangan atau pemberitahuan tentang suatu pengertian yang belum lengkap. Dan untuk melengkapinya harus memilih satu dari beberapa kemungkinan jawaban yang telah disediakan. Atau *multiple choice test* terdiri atas bagian keterangan (*stem*) dan bagian kemungkinan jawaban atau alternatif (*option*). Kemungkinan jawaban (*option*) terdiri atas satu jawaban yang benar, yaitu kunci jawaban dan beberapa pengecoh (*distractor*).

3) Menjodohkan (*matching test*)

Matching test dapat kita ganti dengan istilah mempertandingkan, mencocokkan, memasangkan, atau menjodohkan. *Matching test* terdiri atas satu seri pertanyaan dan satu seri jawaban. Masing-masing pertanyaan mempunyai jawaban yang tercantum dalam seri jawaban. Tugas murid adalah mencari dan menempatkan jawaban-jawaban sehingga sesuai atau cocok dengan pertanyaannya.

4) Tes isian (*completion test*)

Completion test biasa kita sebut dengan istilah tes isian, tes menyempurnakan, atau tes melengkapi. *Completion test* terdiri atas kalimat-kalimat yang ada bagian-bagiannya yang dihilangkan. Bagian yang dihilangkan atau yang harus diisi oleh murid ini adalah merupakan pengertian yang kita minta dari murid.

Dalam penulisan butir soal pilihan ganda ada beberapa petunjuk yang perlu diperhatikan, seperti yang diungkapkan oleh Masidjo (1995:78-79), sebagai berikut:

- a. *Option* seharusnya berisi informasi atau pernyataan yang dirumuskan secara jelas, komplit, dan terfokus pada prinsip atau konsep yang disepakati oleh kebanyakan ahli dalam bidang materi ajar yang diujikan.
- b. *Option* hanya berisi informasi yang diperlukan dan tidak berlebihan.
- c. *Option* disusun dalam pernyataan atau kata-kata yang tepat, tidak membingungkan, dan dengan tata bahasa yang baik dan benar.
- d. *Option* harus terhindar dari situasi atau kata-kata yang memerlukan interpretasi dan penerapan dari pengikut ujian.

- e. *Option* tidak boleh mengandung masalah yang masih diperdebatkan oleh para ahli.
- f. Alternatif pilihan jawaban harus konsisten tata bahasanya serta logikanya dengan stem.
- g. Kunci jawaban harus merupakan pilihan terbaik dari semua alternatif (*option*) yang ada.
- h. Pengecoh, penyesat (*distractor*) harus masuk akal, tetapi jelas bukan merupakan pilihan yang terbaik dan benar.
- i. Alternatif pilihan jawaban yang disediakan harus homogen dalam arti fokus, isi, tata bahasa dan panjangnya kalimat.
- j. Hindari membuat "*clue*" atau petunjuk arah (jawaban yang benar maupun yang salah) pada tata bahasa *option* dan alternatif jawaban.
- k. Setiap butir soal dalam tes harus merupakan butir soal yang independen, tidak tergantung dan tidak dikaitkan dengan butir soal lainnya.
- l. Jumlah alternatif pilihan jawaban untuk setiap butir soal tidak harus sama jumlahnya (misalnya semua harus empat atau lima alternatif pilihan).
- m. Tingkat kesukaran pilihan ganda ini dapat diperluas sesuai dengan kebutuhan. Karena itu jumlah alternatif dapat ditingkatkan. Makin banyak alternatif yang tersedia tingkat kesukaran makin tinggi. Demikian pula jumlah alternatif yang benar dapat lebih dari satu, bahkan bisa semua benar atau semua tidak benar. Yang perlu diperhatikan di sini adalah apabila alternatif yang benar lebih dari satu maka alternatif itu harus seimbang, demikian pula kalau salah. Selain itu petunjuk pengerjaan harus jelas dan terperinci.

Masing-masing tes yang digunakan memiliki keunggulan dan kelemahan. Sebaiknya dalam pemakaian divariasikan guna saling menunjang satu sama lain. Pada penelitian ini, tes yang akan dianalisis berbentuk objektif.

Tes objektif memiliki kelemahan-kelemahan dan kelebihan-kelebihan tersendiri. Kelebihan dari tes objektif seperti yang dikemukakan oleh Arikunto (2008:164) yaitu :

- 1) Mengandung lebih banyak segi-segi yang positif, misalnya lebih *representative* memiliki isi luas bahan, lebih objektif, dapat dihindari campur tangannya unsure-unsur subjektif baik dari segi siswa maupun dari segi guru
- 2) Lebih mudah dan cermat cara pemeriksaannya karena dalam menggunakan kunci tes bahkan alat-alat hasil kemajuan teknologi
- 3) Pemeriksaannya dapat diserahkan pada orang lain
- 4) Dalam pemeriksaannya, tidak ada unsur subjektif yang mempengaruhinya.

Dari uraian tersebut, dapat disimpulkan bahwa tes objektif lebih representatif mewakili dari segi mata pelajaran, pemeriksaannya lebih cepat dan mudah serta dapat dilakukan oleh orang lain tanpa dipengaruhi oleh unsur subjektif.

Tes objektif juga memiliki kelemahan seperti yang dikemukakan oleh Arikunto (2008:165) sebagai berikut :

- a. Persiapan untuk menyusun jauh lebih sulit dari pada tes essay karena soalnya banyak dan harus teliti untuk menghindari kelemahan-kelemahan yang lain.
- b. Soal-soalnya cenderung untuk mengungkapkan ingatan dan daya pengenalan kembali saja dan sukar untuk mengukur proses mental yang tinggi.
- c. Banyak kesempatan untuk main untung-untungan.
- d. “Kerjasama” antara siswa pada waktu mengerjakan soal tes lebih terbuka.

5. Ciri-ciri tes yang baik

Baik buruknya suatu tes dapat ditinjau dari beberapa segi, yaitu: validitas, reliabilitas, daya beda soal, tingkat kesukaran soal, dan kualitas *option* dari soal tersebut. Sejalan dengan itu, Arikunto (2008: 57-58) mengemukakan: “Sebuah tes dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu

memiliki validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis”.

6. Analisis butir soal

Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban siswa untuk membuat keputusan tentang setiap penilaian. Tujuan penelaahan adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum soal digunakan. Di samping itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada siswa apakah mereka sudah/belum memahami materi yang telah diajarkan. Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuannya di antaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru.

Dalam melaksanakan analisis butir soal, para penulis soal dapat menganalisis secara kualitatif dan kuantitatif. Analisis kualitatif pada prinsipnya dilaksanakan berdasarkan kaidah penulisan soal (tes tertulis, perbuatan, dan sikap), yang bertujuan untuk menilai kualitas suatu perangkat soal. Penelaahan ini biasanya dilakukan sebelum soal digunakan/diujikan. Analisis kualitatif dilakukan melalui analisis validitas logis.

Analisis kuantitatif merupakan analisis untuk menilai suatu perangkat

soal melalui interpretasi pengolahan sejumlah data berupa angka, dalam kaitan dengan ciri-ciri statistiknya atau prosedur peningkatan secara *judgment* dan prosedur peningkatan secara empirik. Analisis kualitatif mencakup pertimbangan validitas isi dan konstruk, sedangkan analisis kuantitatif mencakup penghitungan validitas butir soal, reliabilitas, tingkat kesulitan dan daya diskriminasi (pembeda) soal.

a. Analisis validitas butir

Validitas merupakan syarat yang terpenting dalam alat evaluasi karena suatu alat evaluasi dikatakan mempunyai validitas yang tinggi (valid) jika tes itu mengukur apa yang sebenarnya diukur Purwanto (1994:138). Hal ini senada dengan Arikunto (2008:56) “tes akan valid apabila tes tersebut dapat mengukur apa yang hendak diukur”. Sedangkan Sudjana (2008:12) mengemukakan bahwa “validitas berkenaan dengan ketetapan alat penilaian terhadap konsep yang dinilai sehingga betul menilai”.

Secara garis besar Arikunto (2008:65-66) menyatakan bahwa ada 2 macam validitas yaitu validitas logis dan validitas empiris :

1) Validitas logis

Validitas logis artinya logika, validitas logis untuk sebuah instrumen menunjukkan pada kondisi bagi sebuah tes yang memenuhi persyaratan valid berdasarkan hasil penalaran. Ada dua macam validitas logis :

a) Validitas isi (*content validity*)

Validitas isi adalah kejituan dari pada suatu tes ditinjau dari isi tes tersebut. Untuk menilai apakah suatu tes memiliki validitas isi atau tidak dapat dilakukan dengan jalan membandingkan materi tes dengan materi yang terdapat pada kurikulum, maka validitas isi sering disebut validitas kurikuler.

b) Validitas susunan (*construct validity*)

Validitas susunan artinya kejituan dari pada suatu tes ditinjau dari susunan tes tersebut. Cara menentukan validitas susunan dapat dilakukan dengan melihat kemampuan apa yang diukur dalam tes tersebut.

2) Validitas empiris

Sebuah instrumen dapat dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Ada 2 bentuk validitas empiris :

a) Validitas “ada sekarang” (*concurrent validity*)

Sebuah tes dikatakan memiliki validitas *concurrent* jika hasilnya sesuai dengan pengalaman.

b) Validitas prediksi (*predictive validity*)

Sebuah tes memiliki validitas prediksi atau validitas ramalan apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan datang.

Validitas empiris merupakan salah satu dari analisis kuantitatif, dengan artian hasil ujicoba soal berupa data angka yang memerlukan pengolahan

secara statistic. Validitas empiris menekankan pada penggunaan suatu criteria data sebagai acuan. Untuk itu, di lapangan validitas empiris lebih dikenal dengan validitas item atau validitas butir soal. Prinsip yang mendasari validitas butir adalah korelasi antara dua variabel, yaitu korelasi antara skor setiap butir soal/item (1, 2, 3, ..., n) dengan skor total (variabel bebas berkorelasi dengan variabel terikat).

Validitas butir suatu tes dihitung dengan menggunakan beberapa rumus, salah satu rumus yang paling sering digunakan adalah Rumus Korelasi *Product Moment* dari *Karl Pearson*.

Rumus :

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

r_{xy} = koefisien korelasi antar variabel x dan variabel y /korelasi skor setiap butir dan skor total.

Koefisien korelasi yang diperoleh diinterpretasi dengan mengacu pada standar tertentu untuk mengetahui soal tersebut memiliki korelasi atau tidak. Ada dua pendapat dalam menentukan keputusan:

- a. Membandingkan r_{xy} dengan r_{tabel} . r_{tabel} adalah nilai korelasi Product Moment yang telah ditentukan dalam Tabel nilai r (biasanya terlampir pada buku statistic). Nilai r yang dipilih adalah nilai yang terdapat pada pertemuan jika $r_{xy} > r_{tabel}$ berarti X memiliki korelasi signifikan dengan Y, sehingga soal tersebut dinyatakan valid. Sebaliknya, jika

jika $r_{xy} < r_{\text{tabel}}$, berarti X tidak memiliki korelasi signifikan dengan Y, sehingga soal dinyatakan tidak valid.

b. Membandingkan nilai r_{xy} kedalam kategori yang telah ditentukan oleh beberapa ahli, yaitu :

- Antara 0,81 – 1,00 : sangat tinggi
- Antara 0,61 – 0,80 : tinggi
- Antara 0,41 – 0,60 : cukup
- Antara 0,21 – 0,40 : rendah
- Antara 0,00 – 0,20 : sangat rendah (dimodifikasi dari Surapranata, 2005: 51)

b. Reliabilitas

Reliabilitas (keandalan) adalah ketetapan atau ketelitian suatu alat evaluasi. Sebuah soal yang reliable menunjukkan hasil pengukuran yang sama jika soal yang sama diujikan kepada peserta yang sama, dengan kemampuan yang masih sama, tetapi pada waktu yang berbeda. Reliabilitas mencerminkan seberapa besar instrumen tersebut dapat dipercaya. Untuk itu, reliabilitas sangat penting dalam menentukan apakah tes tersebut telah menyajikan pengukuran yang terbaik. Bagaimana jika sebuah tes yang diberikan menunjukkan hasil yang tinggi pada tes pertama, dan menunjukkan hasil yang rendah pada tes kedua?

Arikunto (2008:83) mengemukakan :

Suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Maka pengertian reliabilitas tes berhubungan dengan masalah ketetapan hasil atau dapat juga dikatakan tes yang reliabel apabila hasil pengukurannya tetap.

Dalam hal ini, Sudjana (2008:148) mengemukakan :

“Sebuah tes hasil belajar dapat dinyatakan reliabel (*reliable*) apabila hasil-hasil pengukuran yang dilakukan menggunakan tes tersebut secara berulang kali terhadap subjek yang sama, senantiasa menunjukkan hasil tetap sama atau sifatnya stabil, pengujian dilaksanakan pada waktu yang berlainan dengan selang waktu yang tidak terlalu lama dan juga terlalu singkat, bisa juga dilakukan dengan membandingkan hasil pengujian dari tes yang setara”.

Ada beberapa hal menurut Arikunto (2008:87-89) yang mempengaruhi hasil reliabilitas suatu tes :

1. Hal yang berhubungan dengan tes itu sendiri, yaitu panjang dan pendek kualitas butir-butir soalnya. Makin pendek tes maka makin tinggi reliabilitas suatu tes dan sebaliknya.
2. Hal yang berhubungan dengan terdoba, tes yang dicobakan kepada bukan kelompok terpilih akan menunjukkan reliabilitas yang besar dari pada yang dicobakan pada kelompok tertentu yang diambil secara dipilih.

Reliabilitas menunjuk kepada *consistency* pengukuran. Bagaimana hasil pengukuran itu bila instrumen tetap sedang objeknya berlainan atau objeknya yang tetap, sedang instrumennya berlainan. Atau dapat terjadi instrumen dan objeknya tetap hanya waktu yang berlainan. Bila dalam pengukuran dengan kejadian demikian hasil pengukurannya tetap, maka reliabilitasnya tinggi.

Beberapa metode yang biasa digunakan untuk menentukan reliabilitas suatu tes menurut Arikunto (2008: 90) adalah sebagai berikut:

1) Metode tes paralel (*Equivalent*)

Tes paralel atau tes *equivalent* yang dalam istilah bahasa Inggris disebut *Alternate-Forms Method (Parallel Forms)* adalah dua tes yang bersifat paralel dengan mempunyai kesamaan dalam tingkat kesukaran dan susunan, tetapi butir-butir soalnya berbeda yang diberikan kepada sekelompok peserta tes dalam periode waktu yang relatif singkat.

Jika koefisiennya tinggi, maka tes tersebut sudah reliabel dan dapat digunakan sebagai alat pengetes yang handal. Kelemahan dari metode ini adalah bahwa pengetes kerjanya berat karena harus menyusun dua jenis tes. Lagi pula harus tersedia waktu yang cukup lama untuk mencoba dua kali tes.

2) Metode tes ulang (*Test Retest Method*)

Metode tes ulang dilakukan untuk menghindari penyusunan dua seri tes. Dalam menggunakan teknik atau metode ini pengetes hanya memiliki satu seri tes tetapi dicoba dua kali dengan selang waktu. Kemudian hasil tes pertama dan kedua di korelasikan. Apabila diperoleh koefisien korelasi tinggi berarti hasil tes stabil, berarti siswa yang pada tes pertama memperoleh skor tinggi, pada tes yang kedua juga memperoleh skor yang tinggi pula. Untuk tes bakat dan prestasi

belajar dianggap memiliki stabilitas bila koefisien korelasi antara 0,80 - 0,90.

Persyaratan demikian umumnya digunakan pada tes yang telah distandarisasikan. Untuk tes yang banyak mengungkap pengetahuan (ingatan) dan pemahaman, cara ini kurang mengena karena yang diuji akan masih ingat terhadap butir-butir soalnya.

3) Metode belah dua (*Split Half Method*)

Dalam menggunakan metode ini pengetes hanya menggunakan sebuah tes dan dicoba satu kali. Kemudian skor yang diperoleh dikelompokkan menjadi dua, yaitu kelompok peserta nomor ganjil dan kelompok peserta nomor genap. Bila jumlah peserta ganjil nomor terakhir tidak di ikut sertakan dalam analisis. Skor tes peserta ganjil dan nomor genap di korelasikan, koefisien korelasi yang diperoleh merupakan koefisien reliabilitas yang mengukur konsistensi internal.

Oleh karena itu disebut juga *single-test-single-real-method*. Ada dua cara membelah butiran soal ini yaitu :

- a). Membelah atas item-item genap dan item-item ganjil dan selanjutnya disebut belahan ganjil dan belahan genap.
- b) Membelah atas item-item awal dan item-item akhir yaitu separuh jumlah pada nomor-nomor awal dan separuh nomor-nomor akhir yang selanjutnya disebut belahan akhir.

4) Metode Kuder Richardson 20 dan Kuder Richardson 21

Yaitu dengan menggunakan rumus yang dikemukakan oleh dua orang ahli "*measurement*" yang bernama Kuder dan Richardson.

Koefesien korelasinya terkenal dengan KR 20 dan KR 21. Menggunakan rumus KR 20 cenderung memberikan hasil yang lebih tinggi, tetapi pekerjaannya lebih rumit.

5) Metoda alpha

Sedangkan untuk tes yang berbentuk uraian (*essay*) digunakan metoda Alpha (*Cronbac's Alpha*). Menilai soal bentuk uraian tidak dapat dilakukan seperti tes berbentuk objektif. Suatu butir soal uraian menghendaki gradualisasi penilaian. Contohnya butir soal nomor 1 penilaian terendah 0 dan tertinggi 8, tetatapi butir soal nomor 2 nilai teringgi hanya 5 dan butir soal nomor 3 sampai 10 dan sebagainya.

Untuk keperluan mencari reliabilitas soal keseluruhan perlu juga dilakukan analisis butir soal seperti halnya soal bentuk objektif. Skor untuk masing-masing butir soal dicantumkan pada kolom item menurut apa adanya.

Dalam penelitian ini digunakan rumus Kuder Richardson 20 (KR 20). Karena untuk tes berbentuk objektif dengan pilihan ganda (*multiple choice*), para ahli lebih banyak menggunakan rumus Kuder Richardson 20 (KR 20).

Reliabilitas juga menggunakan prinsip korelasi. Untuk itu, penentuan reliabilitas instrumen juga mengacu pada pemaknaan nilai r (koefisien korelasi) melalui perbandingan terhadap nilai r tabel dan kategori bertingkat (sangat tinggi, tinggi, sedang, cukup, kurang).

Besarnya reliabilitas menurut Arikunto (2008: 83) adalah sebagai berikut:

Tabel 1. Kriteria reliabilitas soal

Reliabilitas	Kategori
0,80 – 1,00	Sangat tinggi
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah

(Sumber: Arikunto, 2008:210)

Reliabilitas yang digunakan adalah $> 0,59$.

c. Tingkat Kesukaran Soal (TK)

Menurut Sudijono (2008: 370) “Bermutu atau tidaknya butir-butir item tes hasil belajar pertama-tama dapat diketahui dari drajat kesukaran atau taraf kesulitan yang dimiliki oleh masing-masing butir item tersebut”. Sejalan dengan itu Sudjana (2004: 135) menyatakan bahwa “Tingkat kesukaran soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dilihat dari sudut guru sebagai pembuat soal”.

Persoalan yang penting dalam melakukan analisis tingkat kesukaran soal adalah penentuan proporsi dan kriteria soal yang termasuk mudah,

sedang, dan sukar”. Sudijono (2008: 370) juga menyatakan “Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir item yang baik, apabila butir-butir item tersebut tidak terlalu sukar dan tidak pula terlalu mudah dengan kata lain derajat kesukaran item itu adalah sedang atau cukup”. Selain itu Rafi’i (1990: 120) menyatakan “Derajat kesukaran menunjukkan berapa orang peserta yang mampu menjawab benar dan gagal terhadap suatu item ”. Salah satu butir soal yang baik bahwa ia tidak terlalu sukar dan tidak terlalu mudah untuk kelompok tertentu yang akan dites. Menurut Arikunto (2008: 212) “ Soal yang baik itu mempunyai tingkat kesukaran antara 0,30-0,70”. Sejalan dengan itu Nurkancana (1986: 36) mengemukakan bahwa “ Item yang baik adalah item yang mempunyai derajat kesukaran tertentu yaitu yang bergerak antara 25% sampai 75%”.

Tingkat kesukaran atau indeks kesukaran soal dapat dilihat dari kemampuan siswa dalam menjawab soal tersebut. Soal akan dikelompokkan menjadi soal mudah, sedang dan sulit. Dalam hal ini Nurkancana dan Sumartana (1986: 130) menjelaskan :

Suatu tes tidak boleh terlalu mudah dan tidak boleh terlalu sukar. Sebuah item yang terlalu mudah sehingga dijawab dengan benar oleh semua anak bukanlah merupakan item yang baik. Begitu pula item yang begitu sukar sehingga tidak dapat dijawab oleh semua anak juga merupakan item yang tidak baik. Jadi item yang baik adalah item yang mempunyai derajat kesukaran tertentu.

Dari penjelasan di atas dapat diketahui kalau item yang baik adalah item yang tidak terlalu mudah dan tidak pula terlalu sukar. Item yang terlalu mudah sehingga dapat dijawab dengan benar oleh sebagian besar peserta tes bukanlah item yang baik. Begitu pula item yang terlalu sukar, sehingga tidak dapat dijawab oleh peserta tes bukanlah merupakan item yang baik pula.

Item tes yang terlalu mudah tidak merangsang peserta untuk mempertinggi usaha memecahkannya. Sebaliknya item yang terlalu sukar menyebabkan peserta tes menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena berada di luar jangkauan kemampuannya.

Untuk mengukur tingkat kesukaran digunakan rumus sebagai berikut :

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

Dimana:

P = Tingkat kesukaran

B = Jumlah siswa yang menjawab soal dengan betul

JS = Jumlah seluruh siswa peserta tes (Arikunto, 2008:210).

Menurut ketentuannya, rentang nilai indeks kesukaran di klasifikasikan

sebagai berikut:

Tabel 2. Kriteria indeks kesukaran

Tingkat Kesukaran	Kategori Soal
0,00 – 0,30	Sukar
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

(Sumber: Arikunto, 2008:210)

Jadi dari tabel di atas dapat diketahui bahwa tingkat kesukaran yang baik adalah berkisar pada rentang 0,31- 0,70.

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya untuk keperluan ujian semester digunakan butir soal yang memiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi/sukar, dan untuk keperluan diagnostik biasanya digunakan butir soal yang memiliki tingkat kesukaran rendah/mudah.

Tingkat kesukaran butir soal juga dapat digunakan untuk memprediksi alat ukur itu sendiri (soal) dan kemampuan peserta didik dalam memahami materi yang diajarkan guru. Misalnya satu butir soal termasuk kategori mudah, maka prediksi terhadap informasi ini adalah seperti berikut:

- 1) Pengecoh butir soal itu tidak berfungsi.
- 2) Sebagian besar siswa menjawab benar butir soal itu; artinya bahwa sebagian besar siswa telah memahami materi yang ditanyakan.

Bila suatu butir soal termasuk kategori sukar, maka prediksi terhadap informasi ini adalah seperti berikut:

- 1) Butir soal itu "mungkin" salah kunci jawaban.
- 2) Butir soal itu mempunyai 2 atau lebih jawaban yang benar.
- 3) Materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, sehingga kompetensi minimum yang harus dikuasai siswa belum tercapai
- 4) Materi yang diukur tidak cocok ditanyakan dengan menggunakan bentuk soal yang diberikan (misalnya meringkas cerita atau

mengarang ditanyakan dalam bentuk pilihan ganda).

5) Pernyataan atau kalimat soal terlalu kompleks dan panjang.

d. Daya Pembeda

Daya pembeda soal maksudnya adalah kemampuan soal membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Jadi tes dilakukan dengan tujuan agar dapat mengetahui peserta tes yang berkemampuan tinggi dengan yang berkemampuan rendah. Hal ini dikemukakan oleh Daryanto (2005:183) sebagai berikut: “Daya pembeda item adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang berkemampuan rendah”.

Manfaat daya pembeda butir soal adalah seperti berikut ini.

- a) Untuk meningkatkan mutu setiap butir soal melalui data empiriknya.
Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak.
- b) Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru. Apabila suatu butir soal tidak dapat membedakan kedua kemampuan siswa itu, maka butir soal itu dapat dicurigai "kemungkinannya" seperti berikut ini:
 - 1) Kunci jawaban butir soal itu tidak tepat.
 - 2) Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar

- 3) Kompetensi yang diukur tidak jelas
- 4) Pengecoh tidak berfungsi
- 5) Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
- 6) Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya

Indeks daya pembeda setiap butir soal biasanya juga dinyatakan dalam bentuk proporsi. Semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan warga belajar/siswa yang telah memahami materi dengan warga belajar/peserta didik yang belum memahami materi. Indeks daya pembeda berkisar antara -1,00 sampai dengan +1,00. Semakin tinggi daya pembeda suatu soal, maka semakin kuat/baik soal itu. Jika daya pembeda negatif (<0) berarti lebih banyak kelompok bawah (warga belajar/peserta didik yang tidak memahami materi) menjawab benar soal dibanding dengan kelompok atas (warga belajar/peserta didik yang memahami materi yang diajarkan guru).

Untuk mengetahui daya pembeda soal bentuk pilihan ganda adalah dengan menggunakan rumus berikut ini.

$$DP = \frac{BA - BB}{\frac{1}{2}N} \quad \text{atau} \quad DP = \frac{2(BA - BB)}{N}$$

Atau

$$DP = \frac{BA - BB}{\frac{27}{100} N} \quad \text{atau} \quad DP = \frac{100(BA - BB)}{27 N}$$

DP = daya pembeda soal,

BA = jumlah jawaban benar pada kelompok atas,

BB = jumlah jawaban benar pada kelompok bawah,

N = jumlah siswa yang mengerjakan tes.

Untuk mengetahui daya pembeda soal bentuk uraian adalah dengan menggunakan rumus berikut ini.

$$DP = \frac{\text{Mean kelompok atas} - \text{Mean kelompok bawah}}{\text{Skor maksimum soal}}$$

Hasil perhitungan dengan menggunakan rumus di atas dapat menggambarkan tingkat kemampuan soal dalam membedakan antar peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum/tidak memahami materi yang diujikan. Adapun klasifikasinya adalah seperti berikut ini:

- Menurut Arikunto (2008:221)

Tabel 3. Kriteria daya pembeda soal

Daya Pembeda	Kategori Soal
0,00 – 0,20	Lemah
0,21 – 0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Baik sekali
Hasil negatif	Tidak baik

- Menurut Crocker dan Algina (1986: 315).

Tabel 4. Kriteria daya pembeda soal

Daya Pembeda	Kategori Soal
0,40 – 1,00	Diterima
0,30 – 0,39	Diterima tapi perlu diperbaiki
0,20 – 0,29	Diperbaiki
0,00 – 0,19	Tidak dipakai/dibuang
Hasil negatif	Tidak baik

e. Hubungan antara validitas butir, reliabilitas, tingkat kesukaran dan daya pembeda

Suatu tes yang baik memuat kriteria tertentu pada empat aspek yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Sejauh ini, peneliti belum menemukan hubungan antara validitas dengan reliabilitas, validitas dengan tingkat kesukaran dan validitas dengan daya pembeda.

Penulis menemukan literatur yang mengemukakan hubungan reliabilitas dengan tingkat kesukaran soal. Crocker dan Algina (1986) dalam Sumarna (2005: 47) menyatakan bahwa tinggi rendahnya koefisien reabilitas dipengaruhi oleh beberapa faktor. Faktor tersebut adalah panjang suatu tes, kecepatan, homogenitas belahan dan tingkat kesukaran soal. Tingkat kesukaran soal memegang peranan paling dominan. Semakin sukar soal, semakin tinggi reliabilitasnya, begitu juga sebaliknya. Luas-tidaknya sampling yang diambil sebagai subjek uji coba juga mempengaruhi reliabilitas suatu tes. Reliabilitas tes juga dipengaruhi oleh karakter testee dan penyelenggaraan uji coba (Arikunto, 2005: 51). Tes yang diujicobakan kepada bukan kelompok terpilih akan menunjukkan reliabilitas yang lebih besar

daripada yang diujicobakan pada kelompok tertentu yang diambil secara dipilih. Pengaruh penyelenggaraan berupa kondisi psikologis yang tercipta saat tes diselenggarakan.

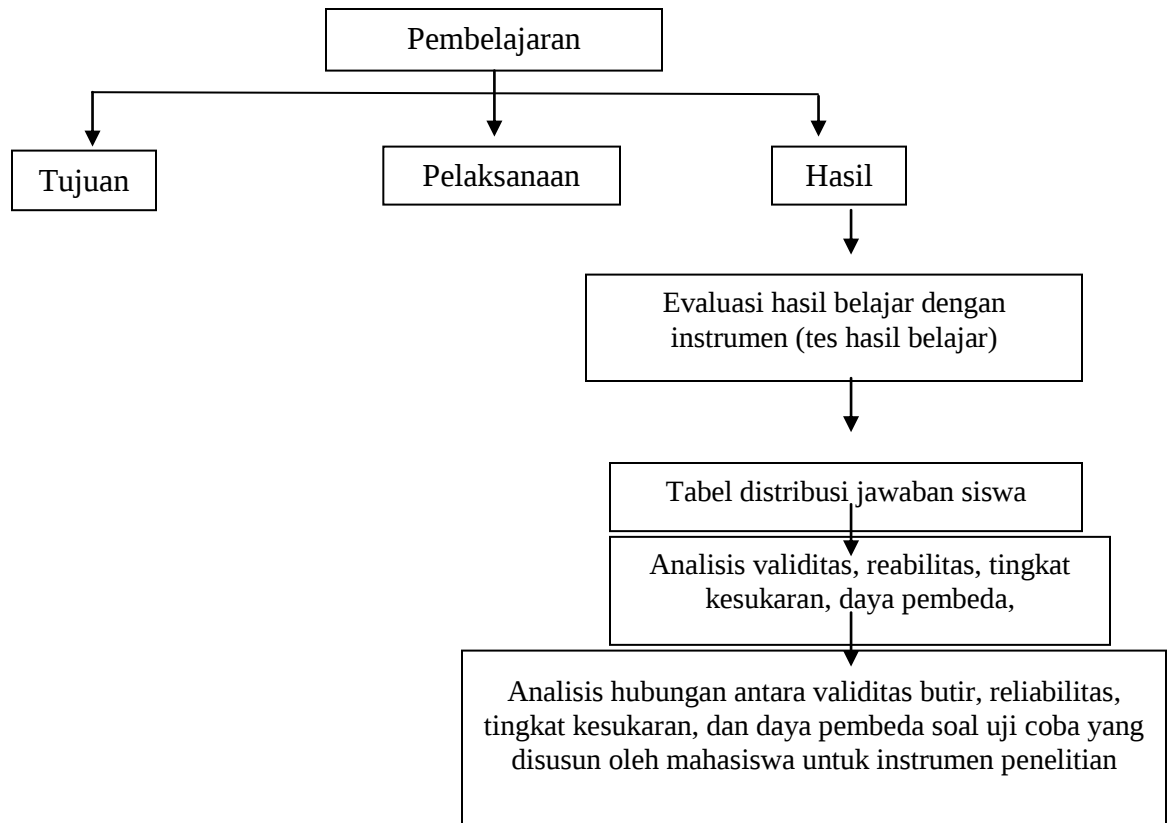
Dari penjelasan diatas, diketahui terdapat hubungan antara reliabilitas dengan tingkat kesukaran soal. Namun, belum diketahui hubungan antara reliabilitas dengan validitas, reliabilitas dengan daya pembeda.

Hubungan antara Tingkat Kesukaran Soal dan Daya Pembeda telah dicoba dideskripsikan oleh Arikunto (2005: 55) melalui telaah rumus, bahwa nilai-nilai P (tingkat kesukaran) sebesar 0,5 memungkinkan untuk mendapatkan daya pembeda yang paling tinggi. Nilai-nilai P yang dianjurkan adalah antara 0,3-0,7, namun harus diingat bahwa soal-soal tersebut tidak berarti mempunyai daya pembeda yang tinggi.

Berdasarkan paparan di atas, hubungan yang jelas dan timbal balik antara satu faktor analisis dengan tiga faktor analisis lainnya belum tersedia dan terdeskripsikan dengan baik.

B. Kerangka Konseptual

Untuk memudahkan penelitian ini, kerangka konseptual sebagai berikut:



Gambar 1. Kerangka Konseptual

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa:

1. Terdapat hubungan antara validitas butir dengan reliabilitas soal uji coba pada taraf signifikansi 1%. Reliabilitas suatu perangkat soal lebih tinggi jika soal tersebut memiliki lebih banyak butir soal yang valid.
2. Tidak terdapat hubungan antara validitas butir dengan tingkat kesukaran soal uji coba pada taraf signifikansi 1%. Dengan demikian, tingkat kesukaran dan validitas butir tidak bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama.
3. Terdapat hubungan antara validitas butir dengan daya pembeda soal uji coba pada taraf signifikansi 1%. Dengan demikian, daya pembeda dan validitas butir bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama
4. Tidak terdapat hubungan antara tingkat kesukaran dengan daya pembeda soal pada taraf signifikansi 1%. Dengan demikian, tingkat kesukaran dan daya pembeda tidak bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama.

B. Saran

1. Dalam menganalisis soal, hal yang harus menjadi pertimbangan dalam pengambilan keputusan untuk menentukan kualitas soal yang akan dipilih

adalah faktor validitas butir, reliabilitas, dan daya pembeda soal. Faktor tingkat kesukaran dapat diabaikan, karena dalam suatu paket soal hendaknya terdiri atas soal yang memiliki tingkat kesukaran bervariasi agar dapat membekas kelompok atas dan kelompok bawah.

2. Penelitian ini belum memasukkan faktor kualitas *option* sebagai salah satu aspek dalam analisis kuantitatif. Untuk itu perlu dilakukan penelitian lanjutan yang membahas semua faktor analisis secara komprehensif.
3. Perlunya penelitian yang lebih mendalam tentang hubungan faktor analisis soal secara kualitatif untuk menjadi panduan tambahan dalam memaknai hasil analisis yang dilakukan oleh calon pendidik.

DAFTAR PUSTAKA

Arikunto, Suharsimi. 2005. *Dasar Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

-----, 2008. *Evaluasi program pendidikan*. Jakarta: bumi Aksara.

Ariyoso. 2011. *Korelasi Biserial*. <http://ariyoso.wordpress.com/2011/07/07/korelasi-biserial/>, diakses tanggal 14 Maret 2012.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston, Inc.

Daryanto. 2005. *Evaluasi Pendidikan*. Jakarta: Rineka Cipta.

Masidjo, Ing.(1995). *Penilaian Pencapaian Hasil Belajar Siswa di Sekolah*. Yogyakarta :Kanisius.

Nurkancana, Wayan dkk. 1986. *Evaluasi Pendidikan*. Surabaya: Usaha Nasional.

Purwanto, Ngalm. 1994. *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT. Remaja Rosdakarya.

Rafi'i, Suryatna. 1990. *Teknik Evaluasi*. Bandung: Angkasa.

Ridwan dan Sunarto. 2007. *Statistik Pendidikan*. Bandung : alfabeta.

Sajekti, Ratna Rusli. 1988. *Tes dan pengukuran Dalam Pendidikan*. Jakarta: Depdikbud.

Slameto. 1988. *Evaluasi Pendidikan*. Jakarta: Bina Aksara.

Sudijono, Anas. 2008. *Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.

Sudjana, Nana. 2008. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT. Remaja Rosdakarya.

Surapranata, Sumarna. 2005. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Bandung: PT Remaja Rosda Karya.