

**ANALISIS HUBUNGAN ANTARA VALIDITAS BUTIR, RELIABILITAS,
TINGKAT KESUKARAN, DAN DAYA PEMBEDA SOAL UJIAN
SEMESTER GENAP BIDANG STUDI BIOLOGI KELAS XI
SMA/MA NEGERI DI KOTA PADANG
TAHUN PELAJARAN 2010/2011**

SKRIPSI

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar
Sarjana Pendidikan*



Oleh:

**RAHMA NOVIYANTI
NIM.01838**

**PROGRAM STUDI PENDIDIKAN BILOGI
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 Ujian
Semester Genap Bidang Studi Biologi Kelas XI SMA/MA
Negeri di Kota Padang Tahun Pelajaran 2010/2011

Nama : Rahma Noviyanti

NIM/ TM : 01838 / 2008

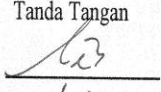
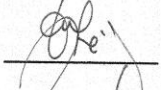
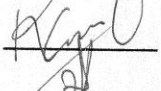
Program Studi : Pendidikan Biologi

Jurusan : Biologi

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 16 April 2012

Tim Penguji

	Nama	Tanda Tangan
1. Ketua	: Drs. Anizam Zein, M.Si.	1. 
2. Sekretaris	: Muhyiatul Fadilah, S.Si., M.Pd.	2. 
3. Anggota	: Drs. H. Rusdi Adnan.	3. 
4. Anggota	: Drs. Mades Fifendy, M.Biomed.	4. 
5. Anggota	: Dra. Helendra, M.S.	5. 

ABSTRAK

Rahma Noviyanti : Analisis Hubungan antara Validitas Butir, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas XI SMA/MA Negeri di Kota Padang Tahun Pelajaran 2010/2011.

Sebuah tes dikatakan baik sebagai alat ukur bila memenuhi persyaratan tes, seperti memiliki validitas butir, reliabilitas, obyektivitas, praktikabilitas, dan ekonomis. Dengan demikian tes sebagai alat ukur hasil belajar hendaknya sesuai dengan standar tersebut. Penelitian ini bertujuan untuk mendapatkan informasi mengenai hubungan antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda dari soal ujian semester genap bidang studi biologi kelas XI SMA/MA Negeri di Kota Padang Tahun Pelajaran 2010/2011.

Penelitian ini merupakan penelitian deskriptif. Populasi penelitian ini adalah soal beserta lembar jawaban ujian semester genap bidang studi biologi kelas XI yang berjumlah 1 set soal SMA dan 1 set soal MA Negeri di Kota Padang tahun pelajaran 2010/2011 yang terdiri dari 40 butir pada soal MA dan 35 butir pada soal SMA. Pada penelitian ini seluruh anggota populasi dijadikan sampel (*total sampling*). Data diolah dengan analisis butir soal menggunakan program SPSS 16.

Dari hasil analisis data diperoleh bahwa 1) Tingkat kesukaran tidak berkorelasi dengan daya pembeda (66,67%), 2) Tingkat kesukaran berkorelasi dengan validitas butir (66,67%), 3) Daya Pembeda berkorelasi dengan validitas butir (100%), 4) Nilai r reliabilitas berbeda antara nilai r semua butir dengan nilai r butir yang valid saja. Selanjutnya dapat disimpulkan bahwa 1) Tidak terdapat hubungan antara tingkat kesukaran dengan daya pembeda pada taraf signifikansi 1%, 2) Terdapat hubungan searah antara tingkat kesukaran dan daya pembeda dengan validitas butir pada taraf signifikansi 1%, 3) Hubungan antara reliabilitas dengan validitas dideskripsikan oleh reliabilitas tes yang memiliki butir-butir yang valid lebih tinggi daripada tes yang tidak semua butirnya valid.

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 Soal Ujian Semester Genap Bidang Studi Biologi Kelas XI SMA/MA Negeri di Kota Padang Tahun Pelajaran 2010/2011". 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. Bapak Drs. Anizam Zein, M.Si., sebagai dosen pembimbing I dan penasehat akademis 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 Ibu Dra. Helendra, M.S., sebagai dosen penguji.
4. Bapak Ketua Jurusan Biologi FMIPA UNP.
5. Bapak Kepala SMA N 15 Padang dan Bapak Kepala MAN 1 Padang .

6. Bapak Drs. Erman Siswadi, M.Si., guru IPA Biologi kelas XI SMA N 15 Padang dan Ibu Neneng Maryamah, S.Pd, guru IPA Biologi kelas XI MAN 1 Padang.
7. 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 sebaik 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
C. Batasan Masalah.....	6
D. Rumusan Masalah.....	7
E. Pertanyaan Penelitian.....	7
F. Tujuan Penelitian.....	8
G. Manfaat Penelitian.....	8
BAB II KERANGKA TEORITIS	
A. Kajian Teori.....	9
1. Pengertian Evaluasi.....	9
2. Pengertian Tes.....	10
3. Fungsi Tes.....	11
4. Bentuk-bentuk Tes.....	11
5. Ciri-ciri Tes yang Baik.....	13
6. Teknik Penganalisisan <i>Item</i> Tes.....	19
7. Hubungan antara Validitas Butir, Reliabilitas, Tingkat	

	Kesukaran, dan Daya Pembeda.....	22
	B. Kerangka Konseptual.....	24
BAB III	METODE PENELITIAN	
	A. Jenis Penelitian.....	25
	B. Definisi Operasional.....	25
	C. Populasi dan Sampel.....	26
	D. Variabel dan Data.....	27
	E. Prosedur Penelitian.....	28
	F. Teknik Analisis Data.....	29
BAB IV	HASIL PENELITIAN	
	A. Hasil Penelitian.....	31
	1. Deskripsi Data.....	31
	2. Analisis Data.....	32
	C. Pembahasan.....	33
BAB V	PENUTUP	
	A. Kesimpulan.....	44
	B. Saran.....	44
	DAFTAR PUSTAKA	
	LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Kriteria tingkat kesukaran.....	20
2. Kriteria daya pembeda.....	21
3. Distribusi populasi penelitian.....	26
4. Jumlah butir soal yang dianalisis pada masing-masing set.....	31
5. Nilai koefisien reliabilitas soal.....	33
6. Variasi Nilai tingkat kesukaran dan daya pembeda (SMA (B)).....	34
7. Distribusi indeks tingkat kesukaran dan koefisien validitas butir pada soal MA.....	37
8. Distribusi indeks daya pembeda dan validitas butir (MA).....	40
9. Perbandingan nilai r semua butir dengan butir yang valid.....	42

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Distribusi Skor Ujian MAN Padang.....	47
2. Data Faktor Analisis Soal MAN Padang.....	48
3. Analisis Korelasi Soal MAN Padang.....	49
4. Interpretasi Analisis Data MAN Padang.....	50
5. Distribusi Skor Ujian SMA Padang Paket A.....	54
6. Data Faktor Analisis Soal SMA Padang Paket A.....	55
7. Analisis Korelasi Soal SMA Padang Paket A.....	56
8. Interpretasi Analisis Data SMA Padang Paket A.....	57
9. Distribusi Skor Ujian SMA Padang Paket B.....	61
10. Data Faktor Analisis Soal SMA Padang Paket B.....	62
11. Analisis Korelasi Soal SMA Padang Paket B.....	63
12. Interpretasi Analisis Data SMA Padang Paket B.....	64
13. Soal Ujian Semester Genap MAN Kelas XI Bidang Studi Biologi.....	67
14. Kunci Jawaban Soal Ujian Semester Genap MAN Kelas XI Bidang Studi Biologi.....	71
15. Contoh Lembar Jawaban MAN Padang.....	72
16. Soal Ujian Semester Genap SMAN Kelas XI Bidang Studi Biologi Paket A.....	73
17. Kunci Jawaban Soal Ujian Semester Genap SMAN Kelas XI Bidang Studi Biologi Paket A.....	77
18. Contoh Lembar Jawaban SMAN Padang Paket A.....	78

19. Soal Ujian Semester Genap SMAN Kelas XI Bidang Studi Biologi Paket B.....	79
20. Kunci Jawaban Soal Ujian Semester Genap SMAN Kelas XI Bidang Studi Biologi Paket B.....	83
21. Contoh Lembar Jawaban SMAN Padang Paket B.....	84
22. Surat Izin Penelitian dari Dekan FMIPA UNP.....	85
23. Surat Izin Penelitian dari Kantor Dinas Pendidikan Kota Padang.....	87
24. Surat Izin Penelitian dari Kantor Kementrian Agama Kota Padang...	88
25. Surat Keterangan Selesai Penelitian dari SMAN 15 Padang.....	89
26. Surat Keterangan Selesai Penelitian dari MAN 1 Padang.....	90

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Sistem pendidikan yang ada di negara kita, Indonesia, sering kali mengalami perubahan dan tidak seimbang. Pertukaran kurikulum yang terlalu cepat membuat guru dan siswa menjadi kebingungan. Belum lagi jumlah mata pelajaran yang terlalu banyak dan tidak berfokus pada sasaran yang tepat. Selain itu, kurangnya tindakan evaluasi membuat dunia pendidikan tidak mengalami perbaikan.

Perbaikan pada sistem pendidikan harus dimulai dari hal yang paling penting yaitu lingkungan sekolah. Pada lingkungan sekolah inilah terjadi proses pembelajaran. Dalam kegiatan proses pembelajaran, terdapat berbagai unsur yang menjadi bagian pembelajaran tersebut, mulai dari tujuan, pelaksanaan hingga pencapaian hasil belajar. Untuk mengetahui pencapaian hasil belajar perlu diadakan sebuah 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 dan perbaikan pelaksanaan pengajaran oleh guru. Joesmani (1998: 19), berpendapat bahwa evaluasi adalah proses menentukan sampai berapa jauh kemampuan yang dapat dicapai siswa dalam proses pembelajaran.

Evaluasi terdiri dari pengukuran dan penilaian. Kurniawan (2009: 3) menyatakan bahwa pengukuran (*measurement*) merupakan proses pemberian

angka atau usaha memperoleh deskripsi numerik sejauh mana peserta didik telah mencapai suatu tingkatan. Penilaian (*assessment*) merupakan dasar untuk memperoleh balikan dalam memperbaiki proses pembelajaran dan sistem pembelajaran secara keseluruhan yang dilakukan secara berkesinambungan, baik terhadap proses itu sendiri maupun terhadap hasil yang akan dicapai. Penilaian proses dapat dilakukan melalui pengamatan terhadap kegiatan siswa dalam belajar dengan menggunakan panduan pengamatan, sedangkan penilaian hasil belajar dapat dilakukan dengan teknik-teknik tertentu, baik teknik tes maupun non tes.

Tayibnapis (2008: 4) menyatakan bahwa evaluasi memiliki dua fungsi yaitu fungsi formatif dan fungsi sumatif. Fungsi formatif mencakup perbaikan dan pengembangan kegiatan yang sedang berjalan, sedangkan fungsi sumatif mencakup pertanggungjawaban, keterangan, seleksi atau lanjutan. Jadi, evaluasi hendaknya dapat membantu pengembangan, implementasi, kebutuhan suatu program, perbaikan program, pertanggungjawaban, seleksi, motivasi, menambah pengetahuan dan dukungan dari mereka yang terlibat.

Salah satu cara melakukan evaluasi adalah dengan soal ujian, baik itu soal objektif maupun essay. Soal yang dominan digunakan di sekolah adalah soal objektif, terutama pilihan ganda. Soal pilihan ganda dapat mewakili lebih banyak materi, ekonomis dan efisien, serta memudahkan penilaian oleh guru.

Soal yang digunakan untuk mengevaluasi hasil belajar siswa ini hendaknya sesuai dengan standar yang telah ditentukan. Menurut Arikunto (2008: 57-58), “Sebuah tes yang dapat dikatakan baik sebagai alat pengukur

harus memenuhi persyaratan tes, yaitu memiliki validitas, reliabilitas, objektivitas, praktikabilitas, dan ekonomis”. Soal sebagai alat ukur hasil belajar siswa hendaknya sesuai dengan standar tersebut. Begitu juga halnya dengan soal-soal yang digunakan guru se-Kota Padang.

Berdasarkan pengamatan awal dan wawancara dengan beberapa guru SMA/MA Negeri di Kota Padang pada bulan Juni dan Juli tahun 2011, diketahui ternyata sedikit dari guru tersebut yang melakukan analisis soal untuk mengetahui kualitas soal ujian. Seyogyanya, soal ujian merupakan instrumen bagi seorang guru untuk mengetahui hasil belajar siswanya. Jika hasil belajar siswa rendah berarti seorang guru harus mampu menganalisis bagian mana dari rencana pembelajarannya yang masih memerlukan perbaikan, termasuk instrumen penilaian. Menurut Arifin (2009: 286), “rencana pelaksanaan pembelajaran merupakan bagian integral dari tugas guru sebagai pelaksana dan pengelola pembelajaran”.

Menurut Gronlund (1985) dalam Arifin (2009: 91), ada beberapa langkah yang harus ditempuh dalam perencanaan penilaian hasil belajar yaitu merumuskan tujuan penilaian, mengidentifikasi kompetensi dan hasil belajar, menyusun kisi-kisi atau *blueprint*, mengembangkan *draft* instrumen, uji coba dan analisis instrumen, revisi, dan merakit instrumen baru. Akan tetapi, seringkali seorang guru tidak melakukan langkah uji coba, analisis instrumen, revisi, dan merakit instrumen baru. Hal ini terjadi karena kurangnya panduan dan keinginan guru untuk melakukan analisis soal.

Analisis soal dilakukan untuk mengetahui berfungsi atau tidaknya suatu soal. Menurut Surapranata (2005: 1), "Analisis soal dilakukan melalui dua cara, yaitu analisis kualitatif (validitas logis) dan analisis kuantitatif (validitas empiris). Analisis kualitatif dapat dilakukan sebelum soal diujikan, sedangkan analisis kuantitatif dilakukan sesudah soal diujikan pada sampel yang representatif".

Surapranata (2005: 1) mengatakan bahwa analisis kualitatif disebut juga validitas logis yang dimaksudkan untuk menganalisis soal dari segi teknis, isi, dan editorial. Analisis teknis berkaitan dengan prinsip pengukuran dan teknik penulisan soal. Analisis isi berkaitan dengan kelayakan pengetahuan yang ditanyakan. Analisis editorial berkaitan dengan penggunaan bahasa Indonesia yang baik dan benar menurut EYD.

Analisis kuantitatif adalah penelaahan butir soal berdasarkan pada karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal yang dimaksud meliputi parameter soal tingkat kesukaran, daya pembeda, validitas butir dan reliabilitas. Analisis kuantitatif ini dilakukan setelah soal diujikan (Kurniawan, 2009: 30).

Dari dua jenis analisis yang telah dijabarkan di atas, analisis kuantitatif menjadi fokus dalam penelitian ini. Analisis kuantitatif dilakukan dengan menelaah data yang dihasilkan langsung oleh *testee* setelah instrumen tersebut dipakaikan.

Namun, analisis kuantitatif ini jarang dilakukan oleh guru. Fakta dilapangan menunjukkan terdapat beberapa hal yang menyebabkan guru tidak

melakukan analisis soal, antara lain 1) guru merasa terbebani oleh proses analisis soal sehingga enggan melakukan dan cenderung mengabaikan, 2) guru tidak menyadari bias ketidakvalid-an hasil evaluasi hasil belajar yang ditimbulkan oleh soal tes yang tidak baik, 3) guru merasa yakin dengan kualitas soal tes yang dibuat sehingga merasa tidak perlu melakukan penelaahan lebih lanjut. Padahal, soal tes yang tidak baik dapat mencerminkan gambaran hasil belajar yang keliru.

Selain beberapa masalah yang telah disebutkan mengenai analisis soal, masih terdapat permasalahan lain yang ditemukan oleh peneliti, yaitu penentuan tindakan lanjutan terhadap soal-soal yang telah diujikan. Secara umum, tindakan yang bisa diambil ada tiga, yaitu memakai butir soal, memperbaiki, atau tidak memakai butir soal tersebut. Namun, disinilah letak kesulitannya. Sampai saat ini peneliti belum menemukan sebuah acuan yang pasti dalam menentukan syarat sebuah soal dapat dipakai, diperbaiki, atau tidak dipakai dengan mempertimbangkan faktor validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal secara bersamaan.

Untuk itu, peneliti merasa perlu menemukan hubungan dan mendeskripsikan bentuk hubungan yang terdapat antara keempat faktor dalam fungsinya sebagai penentu kualitas soal yang bisa dipakai. Dengan diperolehnya bentuk hubungan tersebut, maka pada tindakan lanjutnya bentuk hubungan tersebut dapat dipakai sebagai kriteria penentuan keterpakaian butir soal. Penelitian ini memfokuskan pada penemuan hubungan dan bentuk hubungan antara faktor analisis.

Berdasarkan hal tersebut peneliti telah melakukan penelitian dengan judul *“Analisis Hubungan antara Validitas Butir, Reliabilitas, Tingkat Kesukaran, dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas XI SMA/MA Negeri di Kota Padang Tahun Pelajaran 2010/2011”*

B. Identifikasi Masalah

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

1. Masih sedikit guru yang melakukan analisis kuantitatif soal ujian.
2. Guru tidak menyadari bias ketidakvalidan hasil evaluasi hasil belajar yang ditimbulkan oleh soal yang tidak baik.
3. Guru tidak melakukan penelaahan lebih lanjut terhadap kualitas soal ujian.
4. Belum jelasnya hubungan antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal.
5. Belum adanya acuan yang tepat dalam menentukan syarat sebuah soal dapat dipakai, diperbaiki, atau tidak dipakai dengan memperhatikan dua 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 ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011.

D. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan masalah dalam penelitian ini adalah “Bagaimana hubungan yang terdapat antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011?”

E. Pertanyaan Penelitian

Pertanyaan dalam penelitian ini adalah sebagai berikut:

1. Apakah terdapat hubungan antara validitas butir dengan reliabilitas soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011?
2. Apakah terdapat hubungan antara validitas butir dengan tingkat kesukaran soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011?
3. Apakah terdapat hubungan antara validitas butir dengan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011?
4. Apakah terdapat hubungan antara tingkat kesukaran dengan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011?
5. Bagaimana bentuk hubungan (korelasi) antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran

2010/2011?

F. Tujuan penelitian

Tujuan dari penelitian ini adalah untuk menganalisis hubungan yang terdapat antara validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA/MA Negeri di Kota Padang tahun pelajaran 2010/2011.

G. Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi guru, dalam menentukan dan memperoleh butir soal yang berkualitas untuk digunakan dalam evaluasi hasil belajar.
2. Bagi mahasiswa peserta kuliah evaluasi, agar memiliki tambahan panduan dalam menganalisis butir soal.
3. Bagi peneliti lain, untuk membantu dalam mengambil keputusan memilih butir soal yang akan digunakan, diperbaiki, atau tidak digunakan, terutama yang berkaitan dengan penggunaan soal.

BAB II

KERANGKA TEORITIS

A. Kajian Teori

1. Pengertian evaluasi

Evaluasi merupakan suatu tindakan atau proses untuk menentukan nilai dari sesuatu. Sejalan dengan hal tersebut, Sudijono (2009: 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.

Sementara itu Ralph Tyler dalam Arikunto (2008: 3) mengemukakan, “Evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagian mana tujuan pendidikan yang sudah tercapai”. Sejalan dengan itu Brink dalam Darsono (2002: 1) menyebutkan, “Evaluasi merupakan proses pengumpulan informasi dan memanfaatkannya sebagai penimbang dalam pengambilan keputusan”.

Dengan demikian, evaluasi mengandung tiga unsur, yaitu pengumpulan informasi, penimbangan dengan suatu kriteria, dan pengambilan keputusan. Untuk mengambil sebuah keputusan, haruslah sesuai dengan tujuan evaluasi tersebut diadakan. Secara sistematis evaluasi harus dilakukan tahap demi tahap, yaitu pertama adalah pengukuran, tahap berikutnya adalah penilaian, dan akhirnya

pengambilan keputusan sampai sejauh mana penguasaan materi pelajaran oleh siswa. Salah satu teknik yang bisa dipakai untuk mengevaluasi hasil belajar siswa adalah dengan mengadakan 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 Nurkencana (1986: 24), “Tes adalah suatu cara mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang dikerjakan oleh anak atau sekelompok anak sehingga menghasilkan suatu nilai tentang tingkah laku prestasi anak tersebut, yang dapat dibandingkan dengan nilai yang dicapai oleh anak-anak lain atau dengan nilai standar yang ditetapkan”.

Sudijono (2009: 66) mengemukakan bahwa *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 eksperimentor, 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).

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 (2009: 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.

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 secara tertulis juga.
- b. Tes lisan (*Nonpencil and Paper Test*), yakni tes dimana *tester* didalam mengajukan butiran-butiran pertanyaan atau

soalnya dilakukan secara lisan dan *testee* memberikan jawaban dengan secara 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*). *Statement* 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.

5. Ciri-ciri tes yang baik

Baik buruknya suatu tes dapat ditinjau dari beberapa segi, yaitu: validitas, reliabilitas, daya pembeda 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”. Berdasarkan pernyataan tersebut dapat dijabarkan maksud dari istilah-istilah yang ada di dalam pernyataan tersebut, yaitu:

a. Validitas

Validitas mengacu pada kevalidan dari suatu tes. Menurut Arikunto (2008: 59), “Sebuah tes dapat dikatakan valid apabila tes itu dapat tepat mengukur apa yang hendak diukur”. Sejalan dengan itu, Purwanto (1994: 137) menyatakan: “Suatu teknik evaluasi dikatakan mempunyai validitas yang tinggi (disebut valid) jika teknik evaluasi atau tes itu dapat mengukur apa yang sebenarnya akan diukur”.

Menurut Sudjana (2004: 12), “Validitas berkenaan dengan ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai”. Selanjutnya menurut Arikunto (2008: 67-68), terdapat empat macam validitas, yaitu:

- 1) Validitas isi (*content validity*)
Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan.
- 2) Validitas konstruksi (*construct validity*)
Sebuah tes dikatakan memiliki validitas konstruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam Tujuan Instruksional Khusus (TIK).
- 3) Validitas “ada sekarang” (*concurrent validity*)
Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman.
- 4) Validitas prediksi (*predictive validity*)
Sebuah tes dikatakan memiliki validitas prediksi atau validitas ramalan apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa yang akan datang.

Analisis validitas dapat dilakukan secara kualitatif dan kuantitatif. Dalam hal ini, analisis validitas yang dibahas adalah analisis validitas secara kuantitatif, yang dikenal juga dengan validitas empiris/ validitas item/ validitas butir.

Validitas butir merupakan ketepatan butir soal pengukur konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai, yang diindikasikan oleh harga tertentu koefisien korelasi (r). Analisis validitas butir dilakukan dengan menggunakan formula statistik, yang menghasilkan koefisien yang berkisar dari -1 sampai 1. Prinsip yang mendasari validitas butir adalah korelasi antara dua

variabel, yaitu korelasi antara skor setiap butir soal (1, 2, 3, ..., n) dengan skor total (variabel bebas berkorelasi dengan variabel terikat).

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

Rumus :

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

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

N = jumlah *testee*
X = skor masing-masing butir
 $\sum X$ = jumlah skor masing-masing butir soal
 $\sum X^2$ = jumlah kuadrat masing-masing X
Y = skor masing-masing *testee*
 $\sum Y$ = jumlah masing-masing *testee*
 $\sum Y^2$ = jumlah kuadrat masing-masing Y
 $\sum XY$ = jumlah perkalian masing-masing X dan Y untuk setiap butir soal

(Arikunto, 2008: 72)

Koefesien 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 statistik). 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 $r_{xy} < r_{tabel}$,

berarti X tidak memiliki korelasi signifikan dengan Y, sehingga soal dinyatakan tidak valid.

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

- 1) Antara 0,800 – 1,000: sangat tinggi
- 2) Antara 0,600 – 0,800: tinggi
- 3) Antara 0,400 – 0,600: cukup
- 4) Antara 0,200 – 0,400: rendah
- 5) Antara 0,000 – 0,200: sangat rendah

(Surapranata, 2005: 59)

b. Reliabilitas

Reliabilitas berhubungan dengan masalah kepercayaan. Menurut Arikunto (2008: 86), “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 tes.

Sejalan dengan itu Purwanto (1994: 138), menyatakan “Keandalan (*reliability*) adalah ketetapan atau ketelitian suatu alat evaluasi. Suatu tes atau alat evaluasi dikatakan andal jika ia dapat dipercaya, konsisten, atau stabil dan produktif”. Jadi, yang dipentingkan disini ialah ketelitiannya, sejauh mana tes atau alat evaluasi tersebut dapat dipercaya kebenarannya.

Rumus yang digunakan untuk mengukur reliabilitas soal bentuk obyektif adalah rumus Kuder-Richardson (K-R 20) yang dikemukakan oleh Arikunto (2008: 100-101), rumus K-R 20 sebagai berikut:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \sum pq}{S^2} \right)$$

$$S^2 = \frac{\sum y^2}{N}$$

$$\sum y^2 = (\sum y)^2 - \left\{ \frac{(\sum y)^2}{N} \right\} / N$$

Dimana: r_{11} = reliabilitas tes secara keseluruhan
 P = proporsi jumlah subjek yang menjawab item dengan benar
 q = proporsi jumlah subjek yang menjawab item dengan salah ($q = 1 - p$)
 $\sum pq$ = jumlah hasil perkalian antara p dan q
 n = jumlah butir soal
 S = standar deviasi dari tes
 X = jumlah skor masing-masing siswa
 N = jumlah *testee*

Interprestasi terhadap koefisien reliabilitas tes (r_{11}) menurut Sudijono (2009: 209) pada umumnya digunakan patokan sebagai berikut:

1. Apabila r_{11} sama dengan atau lebih besar daripada 0,70 berarti tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan telah memiliki reliabilitas yang tinggi.
2. Apabila r_{11} lebih kecil daripada 0,70 berarti bahwa tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan belum memiliki reliabilitas yang tinggi.

c. Objektivitas

Sebuah tes dikatakan memiliki objektivitas apabila dalam pelaksanaan tes tidak ada faktor subjektif yang mempengaruhi, terutama pada sistem skoringnya.

Purwanto (1994: 141) mengemukakan bahwa:

Kualitas objektivitas suatu tes dapat dibedakan menjadi tiga tingkatan, yaitu:

- 1) *Objektivitas tinggi* ialah jika hasil-hasil tes itu menunjukkan tingkat kesamaan yang tinggi. Contohnya *tes yang sudah terstandarisasi*, hasil penskorannya sangat objektif.
- 2) *Objektivitas sedang* ialah seperti tes yang sudah distandarisasi, tetapi pandangan subjektif skor masih mungkin muncul dalam penilaian dan interpretasinya.
- 3) *Objektivitas fleksibel* ialah seperti beberapa jenis tes yang digunakan oleh LBP (Lembaga Bimbingan dan Penyuluhan) untuk keperluan *counseling*, misalnya tes yang bersifat *open-end item* (*open-end questionnaires*).

d. Praktikabilitas

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

Purwanto (1994: 141) menyatakan bahwa:

Suatu tes dikatakan mempunyai kepraktisan yang baik jika kemungkinan untuk menggunakan tes itu besar. Kriteria untuk mengukur praktis-tidaknya suatu tes dapat dilihat dari:

- 1) biaya yang diperlukan untuk menyelenggarakan tes itu,
- 2) waktu yang diperlukan untuk menyusun tes itu,
- 3) sukar-mudahnya menyusun tes itu,
- 4) sulit-tidaknya menginterpretasikan (mengolah) hasil tes itu,
- 5) lamanya waktu yang diperlukan untuk melaksanakan tes itu.

e. Ekonomis

Menurut Sudijono (2009: 97), “Suatu tes yang dikatakan bersifat ekonomis mengandung pengertian bahwa tes hasil belajar tersebut tidak memakan waktu yang panjang dan tidak memerlukan tenaga serta biaya yang banyak”.

6. Teknik penganalisisan *item* tes (butir soal)

Menurut Sudjana (2004: 135), “Analisis butir soal atau analisis *item* adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai”. Dengan melakukan analisis soal guru akan mendapatkan perbaikan terhadap soal yang akan diberikan kepada siswa sehingga soal sebagai alat ukur dapat benar-benar mengukur kemampuan siswa.

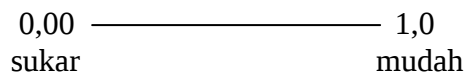
Analisis terhadap butir-butir tes dapat dilakukan dari tiga segi, yaitu: tingkat kesukaran, daya pembeda, dan fungsi distraktornya (*option pengecoh*).

a. Tingkat kesukaran soal

Bermutu atau tidaknya butir soal dapat diketahui melalui tingkat kesulitan dari tiap-tiap butir *item* tersebut. Sudijono (2009: 370) menyatakan bahwa “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”.

Berdasarkan pernyataan di atas, suatu soal dikatakan baik bila soal tersebut tidak terlalu mudah dan tidak pula terlalu sulit. Soal yang terlalu mudah, tidak dapat merangsang pikiran siswa dalam memecahkan masalah, sedangkan soal yang terlalu sulit akan membuat siswa putus asa dan tidak bersemangat dalam menjawab soal karena di luar jangkauan pemikirannya.

Menurut Arikunto (2008: 207), “Bilangan yang menunjukkan sukar dan mudahnya suatu soal disebut indeks kesukaran (*difficulty indeks*). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0”.



Semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki $TK = 0,00$ artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki $TK = 1,00$ artinya bahwa semua siswa menjawab benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal.

Untuk mengetahui tingkat kesukaran soal objektif dan essay digunakan rumus yang dinyatakan oleh Arikunto (2008: 208), yaitu:

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

Dimana: P = tingkat kesukaran soal
B = banyak siswa yang menjawab soal itu dengan benar
JS = jumlah seluruh siswa peserta tes

Indeks kesukaran diklasifikasikan menurut Arikunto (2008: 210) sebagai berikut:

Tabel 1. Kriteria tingkat kesukaran

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

Sumber: Arikunto, 2008: 210

Soal yang baik adalah soal yang indeks kesukarannya sedang, yakni: 0,31 – 0,70.

b. Daya pembeda

Menurut Arikunto (2008: 211), “Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah)”. Sejalan dengan itu, Sudijono (2009: 385-386) mengemukakan bahwa:

Daya pembeda *item* adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan (mendiskriminasi) antara *testee* yang berkemampuan (pandai) tinggi, dengan *testee* yang kemampuannya rendah (bodoh) demikian rupa sehingga sebagian besar *testee* yang memiliki kemampuan tinggi untuk menjawab butir *item* tersebut lebih banyak menjawab betul, sementara *testee* yang kemampuannya rendah untuk menjawab butir *item* tersebut sebagian besar tidak dapat menjawab *item* dengan betul.

Penentuan daya pembeda soal dapat digunakan rumus yang dinyatakan oleh Arikunto (2008: 213), yaitu:

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

Dimana: D = daya pembeda soal
B_A = banyak peserta kelompok atas yang menjawab soal dengan benar
B_B = banyak peserta kelompok bawah yang menjawab soal dengan benar
J_A = banyak peserta kelompok atas
J_B = banyak peserta kelompok bawah

Tabel 2. Kriteria daya pembeda

Daya Pembeda	Kriteria
0,00 – 0,20	Jelek
0,21 -0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Baik Sekali
Negatif	Tidak baik

Sumber: Arikunto, 2008:218

Menurut Crocker dan Algina (1986) dalam Kurniawan, 2009: 37.

0,40 - 1,00	soal diterima baik
0,30 - 0,39	soal diterima tetapi perlu diperbaiki
0,20 - 0,29	soal diperbaiki
0,00 - 0,19	soal tidak dipakai/dibuang

Soal yang baik adalah soal yang memiliki daya beda $\geq 0,40$.

Semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan siswa yang telah memahami materi dengan siswa yang belum memahami materi. Indeks daya pembeda berkisar antara -1,00 sampai dengan +1,00. Semakin tinggi daya pembeda suatu soal, maka semakin baik soal itu. Jika daya pembeda negatif (< 0) berarti lebih banyak kelompok bawah (siswa yang tidak memahami materi) menjawab benar soal dibanding dengan kelompok atas (siswa yang memahami materi yang diajarkan guru).

7. 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.

Menurut Crocker dan Algina (1986) dalam Sumarna (2005: 47) menyatakan bahwa “Tinggi rendahnya koefisien reliabilitas dipengaruhi oleh beberapa faktor. Faktor tersebut adalah panjang suatu tes, kecepatan, homogenitas belahan dan tingkat kesukaran soal”. Alen dan Yen (1979)

menambahkan bahwa tingkat kesukaran soal memegang peranan paling dominan. Semakin sukar soal, semakin tinggi reliabilitasnya, begitu juga sebaliknya.

Selain itu, reliabilitas tes juga dipengaruhi oleh karakter *testee* dan penyelenggaraan uji coba (Arikunto, 2008: 89). Tes yang diujicobakan kepada bukan kelompok terpilih akan menunjukkan reliabilitas yang lebih besar daripada yang diujicobakan pada kelompok tertentu yang diambil secara dipilih. Reliabilitas juga dipengaruhi oleh kondisi psikologis *testee* yang tercipta saat tes diselenggarakan.

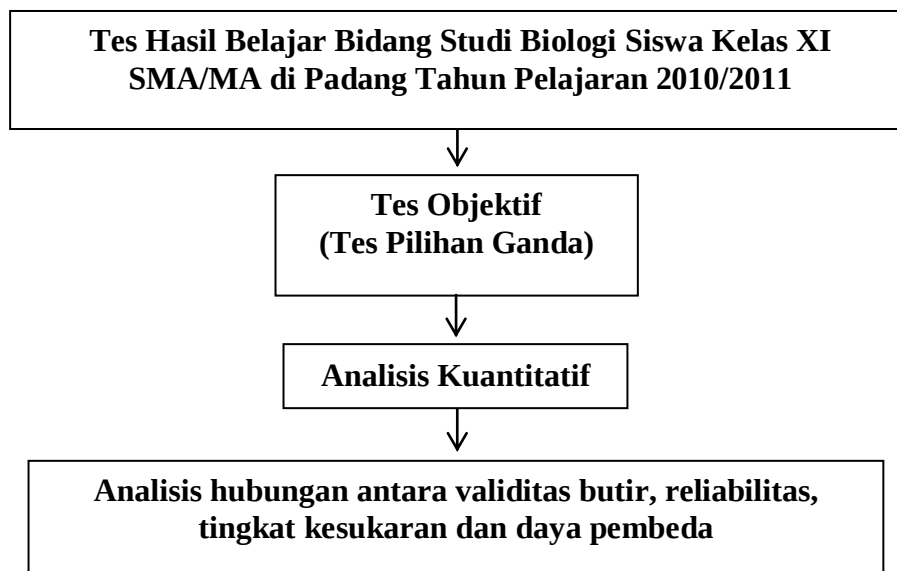
Hubungan antara tingkat kesukaran soal dan daya pembeda pernah dideskripsikan oleh Arikunto (2008: 210) melalui telaah rumus, bahwa nilai-nilai P (tingkat kesukaran) sebesar 0,5 memungkinkan untuk mendapatkan daya pembeda yang paling tinggi. Nilai-nilai P yang dianjurkan adalah antara 0,3-0,7. Namun, harus diingat bahwa soal-soal tersebut tidak berarti mempunyai daya pembeda yang tinggi.

Hubungan yang telah diketahui adalah hubungan antara tingkat kesukaran dan daya pembeda, tingkat kesukaran dan reliabilitas. Akan tetapi, belum diketahui hubungan antara tingkat kesukaran dan validitas, reliabilitas dengan validitas. Selain itu, hubungan antara daya pembeda dan tingkat kesukaran telah diketahui, tetapi belum diketahui hubungan antara daya pembeda dengan validitas, dan daya pembeda dengan reliabilitas.

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

Berdasarkan latar belakang dan kerangka teori yang telah dikemukakan, maka kerangka konseptual dari penelitian ini adalah sebagai berikut:



BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa:

1. Tingkat kesukaran sebagian besar tidak berkorelasi signifikan dengan daya pembeda pada taraf signifikansi 1%. Dengan demikian, tingkat kesukaran dan daya pembeda tidak bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama.
2. Tingkat kesukaran sebagian besar berkorelasi signifikan dengan validitas butir pada taraf signifikansi 1%. Dengan demikian, tingkat kesukaran dan validitas butir bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama.
3. Daya pembeda berkorelasi signifikan dengan validitas butir pada taraf signifikansi 1%. Dengan demikian, daya pembeda dan validitas butir bisa dijadikan acuan dalam pengambilan keputusan soal secara bersama-sama.
4. Reliabilitas suatu soal lebih tinggi jika soal tersebut memiliki *item-item* yang valid lebih banyak.

B. Saran

1. Pada penelitian ini peneliti terbatas pada parameter indeks dengan indeks pada masing-masing faktor analisis dari butir soal. Untuk itu perlu dilakukan penelitian lanjutan yang lebih lengkap parameter penelitiannya.
2. Perlunya penelitian yang lebih mendalam tentang hubungan (multivariabel) faktor analisis soal secara kuantitatif untuk menjadi panduan yang jelas dalam memaknai hasil analisis yang dilakukan oleh guru.
3. 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.

DAFTAR PUSTAKA

- Arifin, Zaenal. 2009. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arikunto, Suharsimi. 2008. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, Suharsimi. 2006. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Asdi Mahasatya.
- Ariyoso. 2011. *Korelasi Biserial*. <http://ariyoso.wordpress.com/2011/07/07/korelasi-biserial/>, diakses tanggal 14 Maret 2012.
- Darsono. 2002. *Belajar dan Pembelajaran*. Semarang: IKIP Semarang Press.
- Daryanto. 2005. *Evaluasi Pendidikan*. Jakarta: PT Rineka Cipta.
- Joesmani. 1998. *Pengukuran dan Evaluasi dalam Pengajaran*. Jakarta: Rajawali Press.
- Kurniawan, Endang dan Endah Mutaqimah. 2009. *Penilaian*. Jakarta: Depdiknas.
- Nurkancana, Wayan. 1986. *Evaluasi Pendidikan*. Surabaya: Usaha Nasional.
- Purwanto, Ngilim. 1994. *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Sudijono, Anas. 2009. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Sudjana, Nana. 2004. *Penilaian Hasil dan Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Surapranata, Sumarna. 2005. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Bandung: PT Remaja Rosda Karya.
- Tayibnapis, Farida Yusuf. 2008. *Evaluasi Program dan Instrumen Evaluasi untuk Program Pendidikan dan Penelitian*. Jakarta: Rineka Cipta.
- Thoha, Chabib. 2003. *Teknik Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.