

LAPORAN PENELITIAN
EVALUASI TENTANG KEKUATAN DAN KELEMAHAN
TIPE BUTIR SOAL FISIP-UT YANG DIUJIKAN
PADA MASA UJIAN 87.1 S/D 89.2

DOKUMENTASI
UNIVERSITAS TERBUKA

OLEH:
D A R M A N T O
131 602 651

DEPARTEMEN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS TERBUKA
1990

00160

LEMBAR IDENTITAS DAN PENGESAHAN LAPORAN AKHIR
HASIL PENELITIAN

- 1.a. Judul Penelitian : Evaluasi Tentang Kekuatan dan Kelemahan
Tipe Butir Soal FISIP-UT Yang Diujikan
Pada Masa Ujian 87.1 s/d 89.2
- b. Macam Penelitian : Deskriptif
- c. Kategori Penelitian: IV
-

2. Peneliti

- a. Nama : Drs. Darmanto M.Ed
- b. NIP : 131 602 651
- c. Jenis Kelamin : Laki-laki
- d. Pangkat/Golongan : Penata Muda/III/a
- e. Jabatan : Asisten Ahli
- f. Unit kerja : Pusat Pengolahan Pengujian
-

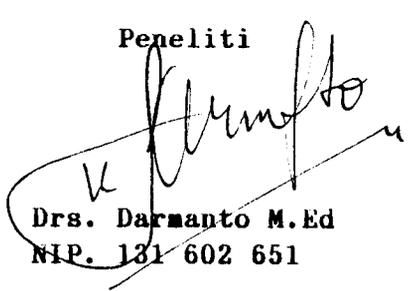
3. Pembimbing : DR. Asmawi Zainul M.Ed
4. Lokasi Penelitian : Universitas Terbuka
5. Jangka Waktu : Januari s/d April 1990
6. Biaya yang diperlukan : Rp. 350.000 (Tiga ratus lima puluh
ribu rupiah)

Jakarta, April '90

Menyetujui
Pembimbing Penelitian


DR. Asmawi Zainul M.Ed
NIP. 130 227 817

Peneliti


Drs. Darmanto M.Ed
NIP. 131 602 651

KATA PENGANTAR

Syukur alhamdulillah bahwa pada akhirnya laporan penelitian ini dapat diselesaikan. Laporan ini dibuat sebagai hasil penelitian dalam rangka tugas penelitian Universitas Terbuka.

Dalam penulisan laporan ini penulis banyak mendapatkan bimbingan dan pengarahan dari Bapak DR. Asmawi Zainul sebagai pembimbing penelitian. Pada kesempatan ini penulis mengucapkan terimakasih kepada beliau yang telah meluangkan waktu dalam mengarahkan penulis selama membuat penelitian sampai selesainya laporan ini.

Ucapan terimakasih juga penulis ucapkan kepada Ka.Puslit, Bapak DR. Aria Djalil yang telah memberikan kesempatan sehingga penulis dapat melaksanakan kegiatan penelitian.

Kepada Bapak DR. Rachmat Wiradinata sebagai pimpinan pada Pusat Pengolahan Pengujian, penulis juga mengucapkan terimakasih atas ijin yang diberikan sehingga terlaksananya penelitian ini.

Tidak lupa pula penulis sampaikan ucapan terimakasih kepada teman-teman di Pusat Pengujian dan Unit Komputer yang telah membantu dalam pengumpulan dan pengolahan data.

Kepada semua tersebut di atas penulis mendoakan semoga Allah Swt. membalas kebaikan yang telah diberikan kepada penulis.

Jakarta, April 1990

ABSTRAK

Tes sebagai alat ukur pendidikan digunakan untuk mengumpulkan informasi dalam rangka pengambilan keputusan. Untuk memperoleh informasi yang tepat dan relevan, maka alat pengumpul informasinyaapun haruslah memenuhi persyaratan antara lain mencakup validitas dan reliabilitas dari alat ukur.

Universitas Terbuka telah mengembangkan 5 tipe butir soal untuk mengukur kemampuan belajar siswa. Penelitian tentang tipe butir soal yang diujikan oleh FISIP-UT belum banyak dilaksanakan. Tujuan penelitian ini adalah untuk memperoleh gambaran yang nyata tentang tingkat kesukaran butir soal dan daya beda soal yang diujikan FISIP-UT.

Yang menjadi populasi penelitian ini adalah seluruh butir soal yang diujikan di FISIP-UT mulai masa ujian 1987 sampai dengan 1989. FISIP hanya menggunakan 3 tipe butir soal saja yaitu:

- a. Pilihan ganda biasa dengan empat option dengan kode A
- b. Analisis hubungan antar hal (sebab-akibat) dengan kode B
- c. Pilihan ganda kompleks dengan kode D.

Sampel penelitian ini adalah satu mata kuliah (ADNE4212) yang dipilih secara sengaja (purposive sampling).

Hasil penelitian menunjukkan bahwa rata-rata nilai mahasiswa yang menjawab benar untuk tipe A, B, dan D adalah kurang dari 50%.

Tipe soal B lebih sulit dibanding tipe A dan D, terbukti hanya ada satu masa ujian pada tipe B yang persentase yang

menjawab benar relatif cukup besar.

Berdasarkan distribusi tingkat kesukaran soal maka tingkat kesukaran soal untuk tipe A, B dan D termasuk dalam kategori sedang dan sukar.

Tipe soal A, B dan D memiliki daya beda dengan kategori kurang baik yang cukup besar.

Jika tingkat kesukaran soal dihubungkan dengan daya beda soal maka soal-soal cenderung mengelompok ke dalam tingkat kesukaran dengan kategori sedang dan sukar dengan daya beda soal yang kurang baik dan sedang yang relatif besar.

Dari hasil penelitian tersebut maka perlu dilakukan revisi soal untuk tipe A, B dan D berhubung rata-rata jumlah mahasiswa yang menjawab benar untuk ke tiga tipe soal tersebut kurang dari 50%.



DAFTAR ISI

	halaman
KATA PENGANTAR.....	i
ABSTRAK.....	ii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	v
BAB I. PENDAHULUAN.....	1
a. Pengantar masalah.....	1
b. Perumusan masalah.....	2
c. Tujuan penelitian.....	3
d. Manfaat penelitian.....	3
e. Definisi operasional.....	4
BAB II. TINJAUAN KEPUSTAKAAN.....	7
BAB III. METODOLOGI.....	19
a. Populasi dan sampel penelitian.....	19
b. Instrumen penelitian.....	19
c. Prosedur penelitian.....	20
d. Sistematika laporan.....	21
e. Metode analisis.....	21
BAB IV. HASIL DAN PEMBAHASAN.....	22
BAB V. KESIMPULAN DAN SARAN.....	30
a. Kesimpulan.....	30
b. Saran.....	31
DAFTAR PUSTAKA	

DAFTAR TABEL

	halaman
Tabel 1. Persentase Mahasiswa Yang Menjawab Benar Masing-masing Tipe Butir Soal Mata Kuliah ADNE 4212.....	22
Tabel 2. Distribusi Tingkat Kesukaran Setiap Soal Mata Kuliah ADNE 4212.....	24
Tabel 3. Distribusi Daya Beda Soal Mata Kuliah ADNE4212.....	26
Tabel 4. Tabulasi Silang Tingkat Kesukaran Dengan Daya Beda Soal Mata Kuliah ADNE4212.....	28

UNIVERSITAS TERBUKA

BAB I

PENDAHULUAN

A. PENGANTAR MASALAH

Universitas Terbuka (UT) telah memasuki tahun ke enam sejak didirikan pada tanggal 4 September 1984. Selama periode tersebut telah banyak kegiatan yang dilakukan dalam rangka meningkatkan kualitas pendidikan mahasiswanya. Dengan demikian dapat diharapkan UT makin dekat dengan tujuan yang ingin dicapainya yaitu agar mahasiswa UT dapat menjadi sarjana yang berkualitas serta dapat diandalkan kemampuannya.

Salah satu bidang yang menjadi prioritas utama dalam pengembangan UT adalah bidang pengujian. Pengujian ini menjadi sangat penting, karena ia merupakan satu-satunya alat bagi UT untuk mengetahui kemampuan mahasiswa dalam menyerap ilmu dan pengetahuan selama proses belajar dalam lingkungan UT.

Pada dasarnya UT berbeda dengan universitas biasa (tatap muka)lainnya. Perbedaan tersebut dapat dilihat misalnya universitas konvensional lebih menitik beratkan pada segi tatap muka, di mana frekuensi pertemuan antara dosen dengan mahasiswa lebih memegang peranan yang penting, buku menjadi referensi dan faktor pendukung dalam proses belajar mengajar. UT lebih menitik beratkan pada sistim belajar jarak jauh (SBJJ) di mana frekuensi tatap muka antara dosen (tutor) dengan mahasiswa hanya merupakan faktor pendukung. Proses belajar memprioritaskan kerja dengan belajar mandiri dengan pemakaian modul secara lebih sistimatis sebagai sarana belajar utama disamping bahan dan media belajar

yang lain seperti audio kaset serta program kuliah di TVRI. Modul UT didesain sedemikian rupa sehingga diharapkan mampu menggantikan fungsi dosen (walaupun tidak sepenuhnya) dalam proses belajar mengajar. Dengan demikian sistem belajar di UT agak berbeda dengan universitas biasa lainnya terutama dalam pengadministrasian bahan belajarnya.

Salah satu alat untuk mengukur kemampuan siswa dalam menyerap ilmu dan pengetahuan yang telah dipelajarinya selama periode tertentu disebut ujian (tes). Soal-soal yang diujikan dikakit dalam bentuk naskah yang terdiri dari beberapa soal (item test), misalnya 30, 45 atau 60 soal. Ujian yang dilakukan pada umumnya berupa tes objektif yaitu mahasiswa harus memilih jawaban yang dianggap benar dari beberapa kemungkinan pilihan yang disediakan, dan sebagian kecil menggunakan bentuk tes uraian (essay) yaitu mahasiswa menjawab pertanyaan atau mengerjakan tugas yang ada dengan menguraikan jawaban secara sistematis.

Sampai saat ini belum ada hasil penelitian tentang tipe butir soal yang diujikan di UT. Walaupun ada mungkin hasilnya masih kurang mencakup seluruh permasalahan yang ada dan tidak sama dengan kondisi UT. Oleh karena itu studi tentang kekuatan dan kelemahan macam butir soal yang diujikan di FISIP-UT perlu dilaksanakan.

B. PERUMUSAN MASALAH

Secara umum permasalahan penelitian ini dapat dirumuskan sebagai berikut: Sampai seberapa jauh kekuatan dan kelemahan tipe butir soal yang diujikan di FISIP-UT? Permasalahan penelitian

secara umum ini dapat diuraikan lebih rinci menjadi beberapa pertanyaan penelitian sebagai berikut:

- a. Berapa besarkah persentase jumlah mahasiswa yang menjawab benar masing-masing tipe soal?
- b. Bagaimanakah gambaran kesukaran butir soal masing-masing tipe soal?
- c. Bagaimanakah gambaran daya beda masing-masing tipe soal?
- d. Bagaimanakah keterkaitan antara tingkat kesukaran butir soal dengan daya beda soal untuk masing-masing tipe soal.

C. TUJUAN PENELITIAN

Penelitian ini secara umum bertujuan untuk memperoleh gambaran nyata mengenai:

- a. tingkat kesukaran butir soal
- b. daya beda butir soal

Secara lebih rinci tujuan penelitian ini adalah untuk menjawab pertanyaan penelitian yang telah dirumuskan pada bagian B yaitu perumusan masalah di atas.

D. MANFAAT PENELITIAN

Secara umum hasil penelitian ini diharapkan akan dapat memberikan masukan yang berarti kepada para pimpinan UT yang berkaitan dengan peningkatan dan pengembangan butir soal ujian di UT. Secara lebih khusus penelitian ini akan memberikan informasi dan pengetahuan yang dibutuhkan tentang kekuatan dan kelemahan macam-macam butir soal FISIP-UT. Hasil penelitian ini akan dapat juga dimanfaatkan bagi pengembangan dan peningkatan butir-butir soal yang diujikan kepada mahasiswa untuk mendukung kualitas

belajar mahasiswa UT.

E. DEFINISI OPERASIONAL

Pada penelitian ini ada beberapa istilah yang perlu didefinisikan secara operasional.

a. Kekuatan dan kelemahan tipe butir soal

Kekuatan dan kelemahan tipe butir soal dilihat dengan tingkat kesukaran dan daya beda soal. Kekuatan tipe butir soal ditentukan oleh kategori tingkat kesukaran soal yang sedang dan sukar serta daya beda soal yang sedang, baik, serta baik sekali. Kelemahan tipe butir soal ditentukan oleh kategori tingkat kesukaran soal yang mudah serta daya beda soal yang kurang baik.

b. Indeks kesukaran soal (dengan simbol P)

Tingkat kesukaran soal ditentukan oleh bagian dari seluruh kelompok siswa yang menjawab soal tersebut dengan benar dan disebut indeks kesukaran soal (item difficulty index) dan dinyatakan dengan simbol "p". Cara menghitung indeks kesukaran soal (p) ialah:

$$p = \frac{b}{T}$$

di mana p : indeks kesukaran soal

b : jumlah mahasiswa yang menjawab suatu soal dengan benar

T : jumlah mahasiswa yang mengerjakan soal tersebut

Nilai p berkisar antara 0,00 s/d 1,00. Suatu soal dikatakan mudah bila p mendekati 1,00 dan sebaliknya suatu soal disebut sukar,

bila nilai p mendekati 0,00. Indeks kesukaran soal hanya menggambarkan proporsi mahasiswa yang menjawab soal dengan benar, dan tidak ada hubungannya dengan proses mental yang digunakan dalam menjawab soal tersebut. Suatu soal yang dijawab benar atau salah oleh seluruh peserta tidak memberikan informasi apa-apa tentang perbedaan kemampuan siswa.

Derajat kesukaran tersebut dinyatakan dalam 3 kategori yaitu:

- sukar ($<0,39$)
- sedang ($0,40 - 0,79$)
- mudah ($>0,80$)

c. Daya beda soal (dengan simbol r)

Daya pembeda soal atau indeks diskriminasi soal ialah kemampuan soal tersebut dalam membedakan siswa-siswa yang memiliki skor tinggi, dan siswa-siswa yang memiliki skor rendah dalam aspek yang diukur oleh soal tersebut.

Salah satu cara menghitung indeks diskriminasi soal adalah dengan menggunakan korelasi point-biserial (Point Biserial Correlation). Rumus koefisien point-biserial adalah:

$$r_{p \text{ bis}} = \frac{M_p - M_t}{s^t} \sqrt{\frac{p}{q}}$$

$r_{p \text{ bis}}$: koefisien korelasi point-biserial

M_p : mean skor pada tes dari subjek-subjek yang memiliki jawaban benar pada butir soal

M^t : mean skor total (skor seluruh subjek)

s^t : simpangan baku (Sd) dari skor total

p : proporsi subjek yang menjawab benar terhadap butir soal

q : $1 - p$

Daya beda tersebut dinyatakan dalam 4 kategori yaitu:

- kurang baik ($<0,19$)
- sedang ($0,20 - 0,29$)
- baik ($0,30 - 0,39$)
- baik sekali (>40)

UNIVERSITAS TERBUKA

BAB II

TINJAUAN KEPUSTAKAAN

Salah satu faktor yang sangat erat kaitannya dengan pendidikan adalah tes prestasi belajar. Ada beberapa ahli yang menyamakan pengertian tes (test) dengan evaluasi (evaluation), pengukuran (measurement) maupun penilaian (assessment). Menurut Wayan Nurhancana, tes adalah "suatu cara untuk mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang harus dikerjakan oleh anak sehingga menghasilkan suatu nilai tentang tingkah laku atau prestasi anak tersebut, yang dapat dibandingkan dengan nilai-nilai yang dicapai oleh anak-anak lain atau dengan nilai standar yang ditetapkan" (Sudirman dkk, 1988, hal. 243). Mehrens dan Lehman (1973) menyatakan bahwa tes merupakan seperangkat pertanyaan yang terstandar yang harus dijawab. Dilain pihak Ebel dan Frisbie (1986) menyatakan tes sebagai suatu ukuran yang terdiri atas seperangkat pertanyaan yang masing-masing mempunyai jawaban yang benar. Dari beberapa pendapat yang telah diuraikan tersebut maka dapat dikatakan bahwa pada dasarnya tes merupakan pertanyaan atau tugas yang harus dijawab atau harus dikerjakan, dan jelas ada jawaban yang benar dan salah, di mana dari hasil tersebut akan diperoleh skor berupa angka dari karakteristik orang yang dites.

Tes bukan alat untuk mengevaluasi siswa tetapi terhadap ciri-ciri yang melekat pada siswa tersebut. Sebagai contohnya, kemampuan siswa terhadap mata pelajaran Kimia, Biologi, dan ataupun terhadap karakteristik siswa tentang ketekunan dalam

belajar, kejujuran dan sebagainya. Tes bukan alat pengganti observasi atau pengamatan yang dilakukan seorang guru di kelas. Namun tes hanyalah suatu alat untuk melengkapi metode guru dalam mengevaluasi anak didiknya agar diperoleh bukti-bukti secara lebih sistematis dan ilmiah tentang kemampuan dan prestasi belajar siswa.

Universitas Terbuka, sebagai suatu lembaga pendidikan mempunyai kewajiban merencanakan dan mengembangkan penulisan soal sebagai bagian dari fungsi pendidikan. Intisari fungsi pendidikan dapat dispesifikasikan menjadi fungsi selektif dan fungsi pengembangan. Yang dimaksud dengan fungsi selektif adalah bahwa tugas pendidikan antara lain mengidentifikasi anggota populasi yang diijinkan untuk meneruskan pendidikannya ke jenjang yang lebih tinggi. Sedangkan fungsi pengembangan adalah bahwa tugas pendidikan antara lain mengembangkan efektifitas pengalaman belajarnya. Soal tes sebagai salah satu alat pengukuran pendidikan disusun untuk mengukur sampai seberapa jauh ke dua fungsi pendidikan tersebut berhasil dicapai. Hal-hal yang perlu dipertimbangkan dalam menulis soal tes/ujian antara lain:

- a. Apakah soal yang ditulis memenuhi persyaratan sebagai alat pengukur pendidikan.
- b. Apakah format soal yang dipilih selaras dengan maksud pengukuran.
- c. Apakah jenjang kemampuan yang akan diukur terarah jelas pada perumusan pertanyaannya.
- d. Apakah derajat kesukaran soal optimal bagi populasi yang diuji sesuai dengan tingkat pendidikannya.

- e. Apakah soal dapat memisahkan golongan kemampuan yang akan diuji menjadi golongan pandai, kurang pandai, dan bodoh sehingga soal dapat mengemban tugas selektif pendidikan.
- f. Apakah soal terungkap dengan bahasa yang jelas sehingga dapat dimengerti oleh peserta ujian.

UT telah mengembangkan 5 tipe soal untuk mengukur kemampuan belajar siswa yaitu:

1. Pilihan ganda biasa dengan 4 option dengan kode A
2. Analisis hubungan antar hal (sebab akibat) dengan kode B.
3. Analisis kasus dengan kode C
4. Pilihan berganda kompleks dengan kode D.
5. Analisis diagram dengan kode E.

Soal dengan kode A terdiri dari kalimat pokok yang berupa pernyataan yang belum lengkap yang diikuti oleh 4 kemungkinan jawaban yang dapat melengkapi pernyataan tersebut. Hanya ada satu jawaban yang benar. Contoh soal dengan kode A tersebut misalnya:

Salah satu sarjana yang menyatakan perlunya studi ekologi adalah

- A. Robert Dahl
- B. Thomas R. Dye
- C. David Easton
- D. Charles Lindblom

Soal dengan kode B (analisis hubungan antar hal) terdiri atas satu kalimat pernyataan yang diikuti oleh satu kalimat alasan. Ditanyakan apakah pernyataan itu benar dan apakah alasan

itu benar. Apabila pernyataan dan alasan keduanya benar apakah ada hubungan antara pernyataan dan alasan. Apabila salah satu dari pernyataan dan alasan tidak benar, sudah tentu tidak ada hubungan sebab dan bilamana pernyataan dan alasan keduanya salah, juga tidak ada hubungan sebab. Contoh soal ini adalah:

Selama negara masih ada maka administrasi negara tetap ada

sebab

Eksistensi administrasi negara lekat dengan eksistensi negara

Soal dengan kode C (analisis kasus) merupakan simulasi keadaan nyata, seolah-olah yang diuji dihadapkan dengan keadaan yang sebenarnya. Contoh soal ini adalah:

Sebuah benda digantung dengan pegas, dengan h^0 adalah tinggi benda dari tanah. Bila sekarang benda ditarik sedikit ke bawah, kemudian dilepaskan, benda akan berayun naik turun secara harmonis dengan frekuensi ayunan per detik. Ayunan dari benda tersebut disebabkan:

- A. tarikan searah dari pegas
- B. tarikan searah dari gravitasi
- C. antar aksi antara pegas dan gaya gesekan udara
- D. antara antar pegas dan gaya gravitasi

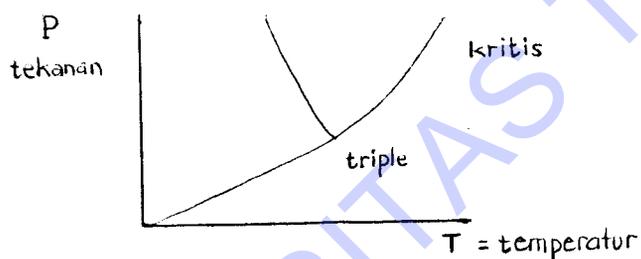
Pada soal melengkapi berganda (kode D), bentuk soal ini hampir sama dengan bentuk soal dengan kode A, yaitu satu pernyataan yang tidak lengkap yang diikuti dengan berbagai kemungkinan. Yang membedakan dengan bentuk pada kode A adalah pada melengkapi berganda ini kemungkinan yang benar dapat satu, dua atau tiga. Contoh dari soal ini adalah:

Menurut SP Siagian, organisasi adalah:

- 1). suatu bentuk persekutuan
- 2) berusaha untuk mencapai tujuan
- 3) ada unsur bawahan dan atasan

Pada soal dengan kode E (analisis diagram), masalah yang ada didalam soal dengan bentuk ini ialah: gambar, diagram, grafik, yang jenisnya adalah soal dengan bentuk kode A. Yang ditanyakan adalah kelainan, keadaan atau gejala yang terungkap di dalamnya. Misalnya:

Petunjuk: Dalam menjawab pertanyaan berikut ini, hendaknya digunakan gambar serta data yang ada di dalamnya.



Soal: Berdasarkan data tersebut dapat disimpulkan bahwa jika tekanannya berkurang maka

- A. titik lebur es dan titik didih air naik
- B. titik lebur es naik, dan titik didih air turun
- C. titik lebur es dan titik didih air turun
- D. titik lebur es dan titik didih air ke duanya turun

Tes sebagai alat ukur pendidikan digunakan untuk mengumpulkan informasi dalam rangka pengambilan keputusan. Karena pengambilan keputusan pendidikan selalu mengenai seseorang atau sekelompok anak didik, maka keputusan itu haruslah berdasarkan informasi yang tepat dan relevan agar setiap keputusan pendidikan

merupakan suatu keputusan yang tepat. Untuk memperoleh informasi yang tepat dan relevan itu, maka alat pengumpul informasinyaapun haruslah memenuhi persyaratan tertentu. Hal-hal yang ada di luar alat ukur yang mempengaruhi pengukuran pendidikan antara lain adalah peranan dan keinginan mereka yang terlibat dalam lembaga pendidikan seperti guru, murid, masyarakat, lembaga pemerintah, lembaga pendidikan dan para pengambil keputusan. Persyaratan alat ukur yang baik bila dilihat dari alat ukur itu sendiri (persyaratan intra) antara lain mencakup validitas dan reliabilitas.

Yang dimaksud dengan validitas alat ukur ialah sejauh mana alat ukur tersebut mengukur apa yang seharusnya diukur. Dengan kata lain, seberapa jauh alat pengukur tersebut memenuhi fungsinya sebagai alat ukur. Sebagai contohnya, tes matematik sebagai alat ukur dikatakan memiliki tingkat validitas yang tinggi apabila tes benar-benar mengukur pengetahuan seseorang dalam bidang matematik, bukan mengukur taraf kemampuan yang lain (misalnya kemampuan berbahasa Inggris). Tes itu valid jika tes itu dapat mengukur apa yang seharusnya diukur, tidak mengukur konsep lain yang tidak akan diukur. Selain itu dapat digunakan pula untuk memprediksi tingkah laku individu yang akan diukur. Berlakunya validitas suatu alat ukur hanya terbatas pada kelompok tertentu saja. Dengan kata lain, bila alat ukur tersebut diterapkan untuk kelompok yang lain yang berbeda kondisinya, maka alat ukur itu tidak valid lagi. Misalnya tes bahasa Inggris yang memiliki tingkat validitas yang tinggi untuk siswa SMTP akan menjadi tidak valid lagi apabila digunakan untuk siswa SMTA. Validitas

biasanya dibedakan menjadi beberapa jenis yaitu content-validity, criterion- related validity dan construct validity.

a) Content validity (validitas isi)

Yang dimaksud content validity suatu ukur ialah sejauh mana isi suatu alat ukur dianggap dapat mengukur hal-hal yang mewakili keseluruhan isi bahan pelajaran yang akan diukur. Adalah tidak mungkin mengujikan semua soal untuk mata pelajaran Biologi kelas III yang begitu banyak kepada siswa. Oleh karena itu perlu dilakukan sampling atau pemilihan contoh soal dari semua kemungkinan yang ada yang dapat diujikan kepada siswa. Untuk menentukan sampel yang masih dapat menjamin validitas isi haruslah berdasarkan pertimbangan dari para ahli bidang studi yang akan diukur dan lebih baik lagi apabila melibatkan ahli bidang pengukuran. Demikian juga diperlukan pertimbangan untuk melihat kesesuaian isi butir soal, kesesuaian jenjang kemampuan, dengan bahan yang diberikan, serta derajat kesukaran butir soal.

b). Criterion-related validity (validitas yang dikaitkan dengan kriteria lain)

Criterion-related validity ialah validitas yang ditinjau dalam hubungannya dengan alat pengukur yang lain yang dianggap sebagai kriterium. Untuk menentukan tinggi rendahnya validitas alat pengukur yang sedang dipersoalkan, dihitung korelasi antara alat pengukur yang sedang dipersoalkan dengan kriterium. Makin tinggi korelasi antara kedua alat pengukur tersebut maka makin tinggi validitas alat pengukur yang sedang dipersoalkan itu. Contohnya adalah seperti membandingkan hasil SIPENMARU dengan

kemampuan mahasiswa selama di universitas atau membandingkan nilai rapor dengan hasil EBTANAS.

Menurut saat tersedianya kriterium, criterion-related validity ini dapat dibedakan antara concurrent validity dan predictive validity. Jika kriterium itu tersedia atau terdapat dalam waktu yang bersamaan dengan alat pengukur yang sedang diteliti, maka validitasnya disebut concurrent validity (validitas sama alat). Jika kriteriumnya didapat di waktu yang akan datang, validitasnya disebut predictive validity (validitas prediktif). Di sini alat pengukur yang sedang dipersoalkan membuat prediksi (ramalan) tentang bagaimana kriteriumnya. Karena alat pengukur yang sedang dipersoalkan mengadakan prediksi, maka alat pengukur tersebut disebut prediktor. Kedua contoh di atas adalah contoh untuk criterion-related validity yaitu hasil pengukuran pada waktu SIPENMARU dikaitkan dengan kemampuan di universitas. Sekiranya kedua hasil ini mempunyai hubungan yang positif dapat disimpulkan bahwa tes SIPENMARU mempunyai daya ramal yang tinggi atau SIPENMARU mempunyai predictive validity yang tinggi. Adakalanya pengukuran yang dilakukan beberapa kali dalam jangka waktu yang hampir sama tetapi menggunakan alat pengukur yang berbeda memberikan hasil pengukuran yang tinggi korelasinya. Dalam hal ini alat pengukur yang satu mempunyai validitas yang sama saat (concurrent validity) terhadap alat pengukur yang lainnya.

c). Construct validity

Construct validity (validitas konsep) menyatakan sejauh mana hasil pengukuran dianggap dapat mencerminkan suatu rekaan

(construct) dalam teori psikologi. Misalkan mengenai teori kreativitas. Suatu alat pengukur kreatifitas akan dianggap valid jika alat pengukur tersebut dapat mengungkapkan perbedaan-perbedaan antar individu yang satu dengan individu yang lain, dari segi kreativitas mereka. Contoh yang lain, misalnya, alat pengukur tingkat keterbacaan seseorang akan dikatakan mempunyai construct validity yang tinggi sekiranya alat pengukur tersebut mampu mengungkapkan perbedaan-perbedaan ketrampilan baca antara individu yang satu dengan individu yang lain.

Penggunaan teknik korelasi untuk menentukan validitas butir soal hingga kini masih banyak dipergunakan, sebab cara menghitung teknik korelasi lebih sederhana. Yang dihitung di sini adalah indeks korelasi dengan cara menentukan tinggi rendahnya korelasi antara butir soal dengan skor total (item total correlation). Ada beberapa cara untuk menentukan validitas suatu soal. Teknik yang paling banyak dipergunakan ialah teknik korelasi point biserial (r_{pbis}) dan teknik korelasi biserial.

Ada beberapa karakteristik yang perlu diketahui terhadap suatu soal diantaranya adalah tingkat kesukaran soal dan daya pembeda soal.

a). TINGKAT KESUKARAN SOAL

Tingkat kesukaran soal ditentukan oleh bagian dari seluruh kelompok siswa yang menjawab soal tersebut dengan benar, dan disebut indeks kesukaran soal (item difficulty index) dan dinyatakan dengan huruf "p". Pada umumnya tingkat kesukaran soal didefinisikan bahwa semakin besar nilai p, makin mudah soal

tersebut.

Tingkat kesukaran dari soal-soal dalam suatu tes mempengaruhi bentuk distribusi dari skor tes secara keseluruhan. Tes yang sukar mengandung soal-soal dengan nilai $p < 0,25$, dan distribusi skor cenderung miring ke kiri (kemiringan positif), dan soal-soal dengan $p > 0,80$ menyebabkan distribusi skor cenderung miring ke kanan (kemiringan negatif).

Thorndike dan Hagen membuat kategorisasi tingkat kesukaran butir soal sebagai berikut:

persentase yang menjawab benar	kategori butir soal
< 30%	sangat sukar
(30-40)%	sukar
(41-84)%	sedang
(85-90)%	mudah
>90 %	sangat mudah

Suatu tes yang baik ialah suatu tes yang dapat membedakan dengan rentangan yang luas kemampuan dari subjek yang dites. Jadi tes sebaiknya berisi beberapa soal mudah untuk mentes subjek yang kurang pandai dan beberapa soal yang sukar untuk mentes subjek yang pandai, sehingga distribusi skornya akan merupakan distribusi yang mendekati kurva normal.

b). DAYA PEMBEDA SOAL

Daya pembeda suatu soal ialah kemampuan soal tersebut dalam membedakan siswa-siswa yang memiliki skor tinggi, dan siswa-siswa yang memiliki skor rendah dalam aspek yang diukur oleh soal tersebut.

Menurut Robert L.Ebel (Essentials of Educational Measurement, 1979):

Indeks Diskriminasi	Evaluasi soal
0,40 dan di atasnya	soal sangat baik
0,30 - 0,39	soal cukup baik
0,20 - 0,29	soal marginal, biasanya perlu disempurnakan
di bawah 0,19	soal-soal kurang baik, dibuang, atau masih dapat dipakai jika direvisi

Thorndike dan Hagen membuat kategorisasi Indeks Daya Pembeda sebagai berikut:

Indeks daya pembeda	Kategori butir soal
>40	baik
0,20 - 0,39	cukup
<0,19	lemah

Butir soal yang mempunyai indeks daya pembeda yang negatif atau nol adalah butir soal yang tidak baik, butir soal semacam ini harus dikeluarkan dari tes atau harus diperbaiki.

Perhitungan indeks diskriminasi dengan berbagai jenis cara atau teknik tidak lepas dari kesalahan (error). Semakin kecil sampel yang digunakan dalam analisis, semakin besar errornya. Suatu butir soal yang memiliki indeks diskriminasi yang besar pada suatu sampel, ada kemungkinan indeks diskriminasinya akan menurun, jika dicobakan pada sampel yang lain. Sebaiknya dikatakan : "indeks diskriminasi, untuk sampel tertentu".

Suatu alat pengukur yang baik adalah alat pengukur yang

mempunyai reliabilitas yang tinggi artinya setiap kali alat pengukur digunakan untuk mengukur hal yang sama, hasil pengukurannya tetap. Sebagai contohnya suatu meteran, kalau digunakan untuk mengukur panjang suatu meja, hasil pengukurannya tetap walaupun digunakan beberapa kali oleh berbagai individu. Kualitas butir soalpun ada yang reliabel dan ada yang tidak misalnya, satu set soal yang reliabel jika diberikan pada hari pertama, hari ke dua dan seterusnya selalu akan memberikan skor yang sama (tetap). Reliabilitas butir soal tidak dapat dipertimbangkan, hanya dapat diketahui melalui uji coba.

Karena reliabilitas dipengaruhi oleh keadaan kelompok yang di tes (heterogen atau homogen), isi dari alat tes, dan kondisi selama dilaksanakannya tes, maka tidak ada metode tunggal yang tepat, yang dapat digunakan untuk mengestimasi besarnya koefisien reliabilitas. Pada umumnya ada 3 cara yang biasanya dipergunakan dalam mengestimasi koefisien reliabilitas dari suatu tes, ialah:

1. menghitung korelasi antara skor-skor yang diperoleh dari dua hasil pengukuran dengan menggunakan alat ukur yang sama disebut metode tes ulang (test-retest method)
2. menghitung korelasi antara skor-skor yang diperoleh dari hasil pengukuran dengan menggunakan dua bentuk tes yang paralel (metode bentuk sejajar atau ekuivalen)
3. menghitung korelasi internal (intercorrealtion) antara semua komponen dari tes (metode analisis internal).

BAB III

METODOLOGI

Dalam bab ini dijelaskan tentang metode yang dipakai dalam pelaksanaan penelitian. Ada 4 bagian yang merupakan unsur utama pada bab ini: pertama, membahas populasi dan cara memilih sampel; kedua, menguraikan tentang instrumen penelitian; ketiga, menjelaskan prosedur penelitian; dan di bagian terakhir, diterangkan tentang sistematika laporan.

A. POPULASI DAN SAMPEL PENELITIAN

Yang menjadi populasi penelitian ini adalah seluruh butir soal yang diujikan di FISIP-UT mulai masa ujian 1987 sampai dengan 1989. FISIP hanya menggunakan 3 tipe butir soal saja, yaitu:

- a. Pilihan ganda biasa dengan empat option dengan kode A
- b. Analisis hubungan antar hal (sebab-akibat) dengan kode B
- c. Pilihan ganda kompleks dengan kode D

Sampel penelitian ini adalah satu mata kuliah (ADNE4212) yang dipilih secara sengaja (purposive sampling). Adapun dasar pemilihan secara sengaja tersebut adalah matakuliah yang menggunakan 3 tipe butir soal dan telah diuji sejak tahun 1987 sampai dengan 1989 terus menerus.

B. INSTRUMEN PENELITIAN

Pada penelitian ini tidak digunakan instrumen penelitian oleh karena data yang dikumpulkan dengan menggunakan data sekunder yang diperoleh dari unit komputer.

C. PROSEDUR PENELITIAN

Data yang diperlukan dalam rangka penelitian berupa tipe butir soal untuk mata kuliah Pengantar Ilmu Politik (ADNE4212) masa ujian 87.1, 88.1, 88.2, 89.1, dan 89.2 dan hasil analisis soal untuk ke lima masa ujian tersebut.

Data tentang tipe butir soal yang berupa naskah ujian diperoleh dari Pusat Pengolahan Pengujian sedangkan analisis soal diperoleh dari komputer mini pada Unit Komputer UT.

Tahap-tahap kegiatan penelitian adalah:

Pertama, mencari naskah ujian mata kuliah Pengantar Ilmu Politik untuk masa ujian 87.1 s/d 89.2 (87.1, 88.1, 88.2, 89.1, 89.2), namun demikian berhubung file tentang data ujian masa ujian 87.2 tidak bisa diperoleh dari komputer maka masa ujian 87.2 dibatalkan. Dari naskah ujian tersebut diperoleh data tentang jumlah soal untuk tipe A, B, dan D untuk ke lima masa ujian.

Kedua, melakukan analisis soal terhadap mata kuliah yang diteliti dengan menggunakan fasilitas komputer UT. Dari analisis soal tersebut diperoleh data tentang jumlah mahasiswa sampel yang dianalisis, proporsi mahasiswa yang menjawab soal dengan benar, tingkat kesukaran soal serta daya beda soal

Ketiga, membahas persentase mahasiswa yang menjawab benar untuk setiap tipe butir soal.

Keempat, membahas tingkat kesukaran soal berdasarkan hasil analisis butir soal untuk setiap tipe butir soal

Kelima, membahas daya beda soal berdasarkan hasil analisis butir soal untuk setiap tipe butir soal.

Keenam, menelaah hubungan antara tingkat kesukaran soal dengan daya beda soal untuk setiap tipe butir soal.

D. SISTIMATIKA LAPORAN

Laporan penelitian tentang kekuatan dan kelemahan tipe butir soal FISIP yang diujikan di UT masa ujian 87.1 s/d 89.2 disusun dalam 5 Bab. Bab I berisi pendahuluan yang menguraikan tentang latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat hasil penelitian, dan definisi operasional. Bab II menjelaskan tentang studi kepustakaan. Bab ini memuat hasil-hasil penelitian yang serupa yang dilakukan oleh para peneliti lainnya dan digunakan sebagai dasar perbandingan hasil penelitian. Bab III berisi metodologi penelitian yang terdiri atas 5 bagian yaitu populasi dan sampel penelitian, instrumen penelitian, prosedur penelitian, sistematika laporan, metode analisis. Bab IV menggambarkan hasil penelitian dan hasil analisis yang telah dilakukan dalam bentuk tabel yang meliputi: deskripsi data dan pembahasan hasil analisis. Bab V memuat kesimpulan serta saran-saran yang diajukan sebagai tindak lanjut dari penelitian ini.

E. METODE ANALISIS

Untuk menganalisis butir soal yang diteliti dalam penelitian ini digunakan program Itean (Item Analysis). Program Itean menyediakan statistik analisa item untuk menentukan reliabilitas dan tingkat kesukaran dari suatu tes serta persentase jumlah mahasiswa yang menjawab benar.

BAB IV
HASIL DAN PEMBAHASAN

Pada bab ini diuraikan hasil penelitian dan pembahasan tentang kekuatan dan kelemahan tipe butir soal FISIP-UT mata kuliah Pengantar Ilmu Politik (ADNE4212). Pembahasan ini dimaksudkan untuk menjawab pertanyaan penelitian yang diuraikan pada bagian sebelumnya.

**TABEL_1: PERSENTASE MAHASISWA YANG MENJAWAB BENAR
MASING-MASING TIPE BUTIR SOAL MATAKULIAH ADNE4212**

ADNE4212						
TIPE SOAL	12 (N:1257)	14 (N:4356)	15 (N:2519)	16 (N:1164)	17 (N:1472)	- X
	%	%	%	%	%	%
A	52,53	50,43	45,28	49,25	50,79	49,65
B	45,49	50,8	43,48	39,81	48,23	45,56
D	55,2	45,73	49,48	39,03	38,62	45,61

Keterangan: 12,14,15,16,dan 17 adalah masa ujian

Ditinjau dari distribusi persentase mahasiswa yang menjawab benar untuk setiap butir soal maka persentase mahasiswa yang menjawab benar relatif cukup besar adalah 52,53% (masa ujian 12),

50,43% (masa ujian 14), 49,25% (masa ujian 16), dan 50,79% (masa ujian 17) pada tipe soal A (pilihan ganda biasa). Persentase yang cukup besar dari tipe soal B (analisis hubungan antar hal (sebab akibat)) adalah masa ujian 14 (50,8%). Tipe soal D (pilihan ganda kompleks) memiliki dua masa ujian (masa ujian 12 dan 15) di mana persentase mahasiswa yang menjawab benar relatif cukup besar yaitu 55,2% dan 49,48%.

Rata-rata nilai mahasiswa yang menjawab benar untuk keseluruhan tipe soal (tipe A, B, dan D) adalah kurang dari 50%. Dengan demikian ada kemungkinan bahwa soal-soal yang diujikan pada mata kuliah ADNE4212 cukup sulit.

Namun dari tabel tersebut terlihat bahwa tipe soal B lebih sulit dibanding tipe A dan D, terbukti pada tipe soal B hanya ada satu masa ujian (masa ujian 14) yang persentase mahasiswa yang menjawab benar relatif cukup besar (50,8%).

Dalam menjawab atau mengerjakan soal tipe B (analisis hubungan antar hal) yaitu tipe soal yang terdiri atas satu kalimat pernyataan yang diikuti oleh satu kalimat alasan, mahasiswa harus menghubungkan kedua kalimat tersebut. Pertama, mahasiswa menyelidiki apakah pernyataan itu benar dan apakah alasan itu benar. Kedua, apabila pernyataan dan alasan keduanya benar, maka mahasiswa masih harus mengecek apakah ada hubungan antara pernyataan dan alasan tersebut.

Tipe soal A adalah yang paling mudah dibanding dengan kedua tipe soal lainnya terlihat dari adanya 4 masa ujian (masa ujian 12, 14, 16, dan 17) yang persentase mahasiswa yang menjawab

benar relatif cukup besar. Pada tipe A, mahasiswa hanya memilih satu jawaban yang benar di antara 4 kemungkinan jawaban yang tersedia.

TABEL 2: DISTRIBUSI TINGKAT KESUKARAN SETIAP SOAL
MATAKULIAH ADNE4212

TIPE SOAL	TINGKAT KESUKARAN (P)	12		14		15		16		17		TOTAL	
		JML. SOAL	%										
A	MUDAH	5	15,6	1	2,9	0	0	2	7,1	5	16,6	13	8,6
	SEDANG	13	40,6	22	64,7	17	60,7	19	67,9	14	46,6	85	55,9
	SUKAR	14	43,8	11	32,4	11	39,3	7	25	11	36,8	54	35,5
Jumlah Soal		32	100	34	100	28	100	28	100	30	100	152	100
B	MUDAH	0	0	0	0	0	0	0	0	0	0	0	0
	SEDANG	11	64,7	7	63,6	10	76,9	6	42,8	7	63,6	41	62,1
	SUKAR	6	35,3	4	36,4	3	23,1	8	57,2	4	36,4	25	37,9
Jumlah Soal		17	100	11	100	13	100	14	100	11	100	66	100
D	MUDAH	1	9,1	1	6,7	0	0	1	5,5	0	0	3	3,7
	SEDANG	8	72,7	8	53,3	14	73,7	8	44,4	11	61,1	49	60,5
	SUKAR	2	18,2	6	40	5	26,3	9	50,1	7	38,9	29	35,8
Jumlah Soal		11	100	15	100	19	100	18	100	18	100	81	100
Total Soal		60		60		60		60		59		299	

Tingkat kesukaran soal (p) untuk masing-masing soal A, B dan D terbagi atas kategori Sukar ($<0,39$), Sedang ($0,40 - 0,79$) dan Mudah ($>0,80$). Tabel 2 di atas menunjukkan hasil penelitian dimana total jumlah soal ujian pada tipe A lebih banyak (152), dibanding tipe B (66) dan tipe D (81).

Dari tabel di atas terlihat bahwa pada umumnya tingkat kesukaran soal untuk tipe A, B dan D yang paling dominan adalah soal dengan kategori sedang, dan sukar.

Rata-rata jumlah soal dengan kategori mudah yang paling banyak adalah tipe A (8,6%) sedang yang paling sedikit yaitu tipe B (0%). Tipe B memiliki jumlah soal kategori sedang yang paling banyak (62,1%), sedangkan yang paling sedikit yaitu tipe A (55,9%).

Jumlah soal dengan kategori sukar yang terbanyak yaitu tipe B (37,95), sedangkan yang paling sedikit yaitu tipe A (35,5%).

Pada tipe A untuk masa ujian 15 tidak terdapat soal yang mudah, sementara tipe D hanya mempunyai satu soal yang mudah untuk masa ujian 12,14 dan 16. Pada tipe B tidak ada satu soal pun yang memiliki kategori mudah.

Dari tabel di atas terlihat bahwa tingkat kesukaran soal untuk tipe A, B dan D termasuk ke dalam kategori sedang dan sukar. Apabila dikaitkan dengan tabel 1 terdapat kesesuaian dimana jumlah soal yang sedang dan sukar pada tipe soal B lebih banyak (100%) dibanding tipe A dan D sehingga jumlah mahasiswa

yang menjawab benar pada tipe B juga relatif lebih sedikit dibanding tipe A dan D.

TABEL 3: DISTRIBUSI DAYA BEDA SOAL (r)
MATAKULIAH ADNE4212

TIPE SOAL	DAYA BEDA SOAL (r)	12		14		15		16		17		TOTAL	
		JML. SOAL	%										
A	KURANG BAIK	9	28,1	9	26,5	7	25	7	25	7	23,3	39	25,6
	SEDANG	10	31,2	12	35,3	9	32,1	4	14,3	11	36,7	46	30,3
	BAIK	7	21,8	4	11,8	8	28,6	6	21,4	9	30	34	22,4
	BAIK SEKALI	6	18,9	9	26,4	4	14,3	11	39,3	3	10	33	21,7
Jumlah Soal		32	100	34	100	28	100	28	100	30	100	152	100
B	KURANG BAIK	12	70,6	3	27,3	7	53,8	4	28,6	5	45,5	36	46,9
	SEDANG	2	11,8	5	45,5	3	23,1	6	42,9	5	45,5	21	31,8
	BAIK	1	5,8	2	18,2	2	15,4	3	21,4	1	9	9	13,7
	BAIK SEKALI	2	11,8	1	9	1	7,7	1	7,1	0	0	5	7,6
Jumlah Soal		17	100	11	100	13	100	14	100	11	100	66	100
D	KURANG BAIK	1	9	3	20	6	31,6	8	44,4	8	44,4	26	32,1
	SEDANG	4	36,4	3	20	6	31,6	3	16,7	3	16,7	19	23,5
	BAIK	2	18,2	5	33,3	7	36,8	5	27,8	6	33,3	25	30,9
	BAIK SEKALI	4	36,4	4	26,7	0	0	2	11,1	1	5,6	11	13,5
Jumlah Soal		11	100	15	100	19	100	18	100	18	100	81	100
Total Soal		60		60		60		60		60		299	

Tabel 3 menggambarkan daya beda soal yang dinyatakan dalam kategori daya beda soal yang kurang baik ($<0,19$), sedang ($0,20-0,29$), baik ($0,30-0,39$) dan baik sekali ($>0,40$).

Distribusi daya beda soal yang kurang baik, sedang, baik, dan baik sekali adalah merata untuk tipe soal A, B, dan D. Namun terlihat bahwa tipe B memiliki persentase yang besar untuk soal yang kurang baik pada masa ujian 12 (70,6%), masa ujian 15 (53,8%) dan masa ujian 17 (45,5%). Tipe D memiliki soal yang kurang baik dengan persentase yang cukup besar untuk masa ujian 16 (44,4%) dan masa ujian 17 (44,4%).

Dengan demikian dapat diketahui bahwa rata-rata jumlah soal dengan kategori kurang baik yang terbanyak adalah pada tipe B (46,9%) dan yang paling sedikit adalah tipe A (25,6%).

Tipe B juga memiliki rata-rata jumlah soal dengan kategori sedang, baik dan baik sekali yang paling sedikit (53,1%) dibandingkan dengan tipe A (74,4%) dan D (67,9%).

Dikaitkan dengan tabel 1 maka terdapat kesesuaian di mana soal-soal tipe B kurang mampu membedakan mahasiswa - mahasiswa yang pandai dengan yang tidak pandai sehingga mahasiswa kesulitan dalam menjawab pertanyaan-pertanyaan pada tipe butir soal B.

TABEL 4. TABULASI SILANG TINGKAT KESUKARAN DENGAN DAYA BEDA SOAL
NATAKULIAH ADNE4212

TIPE SOAL	TINGKAT KESUKARAN	DAYA BEDA	12		14		15		16		17		TOTAL	
			JML SOAL	%										
A	MUDAH	KURANG BAIK	1	3,1	0	0	0	0	0	0	3	10	4	2,6
		SEDANG	2	6,2	1	2,9	0	0	0	0	2	6,7	5	3,3
		BAIK	1	3,1	0	0	0	0	1	3,6	0	0	2	1,3
		BAIK SEKALI	1	3,1	0	0	0	0	1	3,6	0	0	22	1,3
	SEDANG	KURANG BAIK	2	6,2	3	8,8	5	17,8	5	17,8	2	6,7	17	11,2
		SEDANG	4	12,5	10	29,4	4	14,3	2	7,1	5	16,7	25	16,4
		BAIK	5	15,6	3	8,8	4	14,3	4	14,3	5	16,7	21	13,8
		BAIK SEKALI	2	6,2	6	17,6	4	14,3	8	28,6	2	6,7	22	14,5
	SUKAR	KURANG BAIK	6	18,7	6	17,6	2	7,1	2	7,1	2	6,7	18	11,8
		SEDANG	4	12,5	1	2,9	5	17,8	2	7,1	4	13,3	16	10,5
		BAIK	1	3,1	1	2,9	4	14,3	1	3,6	4	13,3	11	7,2
		BAIK SEKALI	3	9,4	3	8,8	0	0	2	7,1	1	3,3	9	5,9
Jumlah Soal			32	100	34	100	28	100	28	100	30	100	152	100
B	MUDAH	KURANG BAIK	0	0	0	0	0	0	0	0	0	0	0	0
		SEDANG	0	0	0	0	0	0	0	0	0	0	0	0
		BAIK	0	0	0	0	0	0	0	0	0	0	0	0
		BAIK SEKALI	0	0	0	0	0	0	0	0	0	0	0	0
	SEDANG	KURANG BAIK	8	47,1	1	9,1	5	38,4	2	14,2	3	27,3	19	28,8
		SEDANG	0	0	4	36,4	2	15,3	3	21,4	3	27,3	12	18,2
		BAIK	1	5,9	1	9,1	2	15,3	1	7,1	1	9,1	6	9,1
		BAIK SEKALI	2	11,8	1	9,1	1	7,6	0	0	0	0	4	6,1
	SUKAR	KURANG BAIK	4	23,5	2	18,2	2	15,3	2	14,2	2	18,2	12	18,2
		SEDANG	2	11,8	1	9,1	1	7,6	3	21,4	2	18,2	9	13,6
		BAIK	0	0	1	9,1	0	0	2	14,2	0	0	3	4,5
		BAIK SEKALI	0	0	0	0	0	0	1	7,1	0	0	1	1,5
Jumlah Soal			17	100	11	100	13	100	14	100	11	100	66	100
D	MUDAH	KURANG BAIK	0	0	0	0	0	0	0	0	0	0	0	0
		SEDANG	1	9,1	0	0	0	0	1	5,5	0	0	2	2,4
		BAIK	0	0	0	0	0	0	0	0	0	0	0	0
		BAIK SEKALI	0	0	1	6,7	0	0	0	0	0	0	1	1,2
	SEDANG	KURANG BAIK	0	0	1	6,7	4	21,4	2	11,1	2	11,1	9	11,1
		SEDANG	3	27,3	2	13,3	4	21,4	1	5,5	4	22,2	14	17,3
		BAIK	1	9,1	4	26,7	6	31,6	3	16,6	4	22,2	18	22,2
		BAIK SEKALI	4	36,4	1	6,7	0	0	2	11,1	1	5,5	8	9,8
	SUKAR	KURANG BAIK	1	9,1	2	13,3	2	10,5	5	27,7	5	27,5	15	18,5
		SEDANG	0	0	1	6,7	2	10,5	2	11,1	0	0	5	6,2
		BAIK	1	9,1	1	6,7	1	5,2	2	11,1	2	11,1	7	8,6
		BAIK SEKALI	0	0	2	13,3	0	0	0	0	0	0	2	2,4
Jumlah Soal			11	100	15	100	19	100	18	100	18	100	81	100
Total Soal			60		60		60		60		59		299	

Tabel 4 di atas menggambarkan keterkaitan antara tingkat kesukaran soal dengan daya beda soal untuk mata kuliah ADNE4212.

Terlihat bahwa jika tingkat kesukaran soal dihubungkan dengan daya beda soal maka soal-soal cenderung mengelompok ke dalam kategori tingkat kesukaran yang sedang dan sukar untuk seluruh masa ujian dan tipe soal.

Jumlah soal yang termasuk dalam kategori mudah dengan daya beda soal baik dan baik sekali hampir tidak ada perbedaan antara tipe soal A dan D.

Soal dengan kategori mudah dengan daya beda soal yang kurang baik, sedang, baik dan baik sekali tidak terdapat pada tipe B.

Soal-soal yang termasuk dalam kategori soal yang sukar, daya beda soalnya lebih cenderung kurang baik dan sedang untuk seluruh tipe butir soal. Jumlah soal dengan daya beda soal yang sedang, baik dan baik sekali yang termasuk dalam kategori soal yang sukar pada tipe A adalah adalah yang paling besar (23,6%), sedangkan tipe B = 19,6% dan tipe D = 17,2%.

BAB V

KESIMPULAN DAN SARAN

A. KESIMPULAN

Jumlah soal mata kuliah Pengantar Ilmu Politik (ADNE4212) di antara tipe soal A, B dan D berbeda-beda. Dari masa ujian 12, 14, 15, 16 dan 17, jumlah soal untuk tipe A yaitu 152 soal (50,8%), tipe B sebanyak 66 soal (22,1%) dan tipe D sebanyak 81 soal (27,1%). Rata-rata jumlah mahasiswa menjawab benar yang paling sedikit adalah tipe soal B.

Tingkat kesukaran soal yang mudah tidak terdapat pada tipe soal B, namun tipe soal B memiliki jumlah soal dengan kategori sedang dan sukar yang terbanyak dibanding tipe A dan D.

Tipe soal B mempunyai jumlah soal terbanyak untuk daya beda soal dengan kategori kurang baik, namun memiliki soal dengan daya beda yang sedang, baik dan kurang baik yang paling sedikit dibanding tipe A dan D.

Untuk kategori soal yang sukar, tipe A mempunyai jumlah soal dengan daya beda sedang, baik dan baik sekali yang terbanyak bila dibandingkan dengan tipe soal B dan D. Untuk kategori soal yang mudah, tipe B tidak mempunyai daya beda soal yang kurang baik, sedang, baik dan baik sekali.

B. Saran-saran

Berdasarkan hasil penelitian tentang kekuatan dan kelemahan tipe-tipe butir soal FISIP-UT untuk mata kuliah Pengantar Ilmu Politik (ADNE4212) maka penulis mengajukan saran-saran sebagai berikut :

Walaupun hanya tipe soal B yang memiliki daya beda kurang baik yang terbanyak dibanding tipe A dan D dan tidak mempunyai soal yang mudah, namun secara keseluruhan perlu dilakukan revisi soal untuk tipe soal A, B, dan D berhubung rata-rata jumlah mahasiswa yang menjawab benar untuk ke tiga tipe soal tersebut adalah kecil yaitu kurang dari 50%.



DAFTAR PUSTAKA

1. Berk, R.A., A Guide to Criterion-Referenced Test Construction. The Johns Hopkins University Press, London, 1984.
2. Ebel, L.Robert., and Frisbie, A.David, Essentials of Educational Measurement. Prentice-Hall, Inc, New Jersey, 1986.
3. Fernandes, H.J.X., Testing and Measurement. National Education Planning, Evaluation and Curriculum Development, Jakarta, 1984.
4. Gronlund, N.E., Measurement and Evaluation in Teaching. Macmillan, Inc., New Jersey, 1985.
5. Mehrens, W.A., and Lehmann, I.J., Measurement and Evaluation Holt, Rinehart and Winston, New York, 1984.
6. Pembakuan Petunjuk Pengerjaan Soal Ujian Universitas Terbuka
7. Pengertian dan Peranan Pengukuran, Tes dan Evaluasi. Penataran Penulisan Soal Ujian II, Pusat Pengolahan Pengujian UT. Oleh: Asmawi Zainul, 9 Januari 1990.



UNIVERSITAS TERBUKA