



DOKUMENTASI  
UNIVERSITAS TERBUKA

LAPORAN PENELITIAN

KESETARAAN SOAL UJIAN SECARA KUALITATIF  
MATAKULIAH METODE PERAMALAN (STAT4330)  
MASA UJIAN 872, 881, DAN 882

Oleh:

Ir. Kristanti Ambar Puspitasari, MEd.

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS TERBUKA  
November, 1990

i.

B0162



LEMBAR IDENTITAS DAN PENGESAHAN LAPORAN AKHIR HASIL PENELITIAN

1. a. Judul Penelitian : Kesetaraan Soal Ujian Secara Kualitatif  
Matakuliah Metode Peramalan (STAT4330)  
Masa Ujian 872, 881, dan 882
- b. Macam Penelitian : Deskriptif
- c. Kategori Penelitian : I (Magang)
2. Peneliti  
a. Nama : Ir. Kristanti Ambar Puspitasari, MEd.  
b. N I P : 131 601 345  
c. Pangkat/Jabatan/Gol. : Penata Muda/Asisten Ahli (III/a)
3. Lokasi Penelitian : Universitas Terbuka
4. Jangka Waktu Penelitian : 3 bulan
5. Biaya yang disetujui : Rp. 350.000,00 (tiga ratus lima puluh  
ribu rupiah)

Jakarta, November 1990

Menyetujui,  
Pembimbing

Peneliti

  
Dr. Rachmat Wiradinata, MA  
NIP. 130 427 529

  
Ir. Kristanti Ambar Puspitasari, MEd.  
NIP. 131 601 345

Mengetahui,  
Dekan FMIPA-UT

Dra. P. Moerwani, MSc.  
NIP. 130 047 606

## KATA PENGANTAR

Puji syukur kepada Tuhan Yang Mahakasih atas tersusunnya laporan penelitian ini.

Atas terlaksananya kegiatan penelitian dan tersusunnya laporan penelitian ini, peneliti mengucapkan terima kasih kepada :

1. Prof. Dr. Setijadi, MA sebagai Rektor UT yang telah memberikan kesempatan kepada seluruh staf edukatif UT untuk melakukan penelitian mandiri.
2. Dra. Patimah Moerwani, MSc. sebagai Dekan FMIPA-UT yang telah memberikan izin untuk dilaksanakannya penelitian ini.
3. Dr. Aria Djalil selaku Kepala Puslitabmas UT yang memberikan kesempatan untuk dilaksanakannya penelitian ini.
4. Dr. Rachmat Wiradinata, MA sebagai Kepala Pusat Pengolahan Pengujian yang bertindak selaku pembimbing dalam kegiatan penelitian dan penulisan laporan penelitian ini.
5. Dr. Asmawi Zainul yang telah memberikan konsultasi dalam kegiatan penyusunan laporan penelitian ini.

Mudah-mudahan hasil penelitian ini dapat berguna bagi Universitas Terbuka (UT), khususnya bagi Pimpinan Fakultas dan Pusat Pengolahan Pengujian, serta bagi semua pihak yang ikut membaca tulisan ini.

Penyusun

## ABSTRAK

Di Universitas Terbuka (UT), pengujian merupakan kegiatan pendidikan yang terpenting karena merupakan satu-satunya alat untuk mengukur mutu mahasiswa dan mutu pengajarannya. Dengan asumsi bahwa mahasiswa UT mempunyai kemampuan belajar mandiri yang setara, maka perlu diusahakan agar naskah ujiannya setara pada setiap masa ujian.

Penelitian ini bertujuan untuk mengetahui kesetaraan perangkat soal ujian secara kualitatif untuk matakuliah Metode Peramalan (STAT4330) dan untuk mengetahui hasil pengukuran masing-masing perangkat soal terhadap mahasiswa UT. Perangkat ujian yang dipilih adalah perangkat ujian yang diujikan pada masa ujian 872, 881, dan 882.

Hasil penelitian menunjukkan bahwa : (1) penyusunan soal ujian untuk ketiga perangkat soal tidak mengacu pada kisi-kisi soal ujian standar baik ditinjau dari materi modul yang dicakup (nomor BMP yang diwakili), jumlah soal yang mewakili tiap BMP yang bersangkutan, prosentase jenjang kemampuan yang diukur, dan jumlah total soal; (2) ditinjau dari ruang lingkup materi pelajaran yang dicakup (nomor BMP yang diujikan materinya), ketiga perangkat soal tidak menunjukkan kesetaraan materi ujian satu sama lain; (3) perangkat soal 882 adalah perangkat soal yang terbaik ditinjau dari hasil analisis soalnya; dan (4) ketiga perangkat soal terlalu sukar bagi mahasiswa UT yang menempuh ujian pada masa ujian yang bersangkutan.

Berdasarkan hasil penelitian tersebut, peneliti memberikan saran agar : (1) kisi-kisi soal ujian perlu ditinjau kembali dan diperbaiki; (2) penyusunan perangkat soal ujian harus mengacu pada kisi-kisi standar; (3) ketiga perangkat soal perlu ditelaah untuk mengetahui apakah ada kesalahan soal ujian; (4) pemilihan soal ujian dan penentuan jumlah soal hendaknya mempertimbangkan waktu ujian yang disediakan oleh UT (90 menit); dan (5) BMP Metode Peramalan perlu ditinjau kembali untuk mengetahui apakah materi ujian berdasarkan kisi-kisi cukup jelas dibahas dalam modul yang bersangkutan.

## DAFTAR ISI

	halaman
HALAMAN JUDUL	i
KATA PENGANTAR	ii
ABSTRAK	iii
DAFTAR ISI	iv
DAFTAR TABEL	v
BAB I. PENDAHULUAN	1
1. Latar Belakang	1
2. Tujuan Penelitian	3
3. Manfaat Penelitian	3
4. Pertanyaan Penelitian	3
BAB II. TINJAUAN PUSTAKA	5
BAB III. METODOLOGI PENELITIAN	16
1. Populasi dan Sampel Penelitian	16
2. Prosedur Penelitian	16
BAB IV. HASIL PENELITIAN DAN PEMBAHASAN	19
1. Hasil Penelitian	19
2. Pembahasan	27
BAB V. KESIMPULAN DAN SARAN	36

## DAFTAR TABEL

	halaman
Tabel 1. Kisi-kisi Soal Universitas Terbuka	7
Tabel 2. Kriteria Penilaian di Universitas Terbuka	15
Tabel 3. Kisi-kisi Soal Ujian Metode Peramalan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Terbuka	19
Tabel 4. Komposisi Soal Ujian Matakuliah Metode Peramalan Menurut Kisi-kisi Soal Ujian Universitas Terbuka	21
Tabel 5. Komposisi Soal Metode Peramalan Menurut Tingkat Kesukaran dan Daya Beda Soal	24
Tabel 6. Statistik Hasil Analisis Soal Matakuliah Metode Peramalan	25
Tabel 7. Distribusi Nilai Mahasiswa UT Matakuliah Metode Peramalan (STAT4330) Berdasarkan Kategori Penilaian Standar Masa Ujian 872, 881, dan 882	26
Tabel 8. Distribusi Nilai Mahasiswa UT Matakuliah Metode Peramalan (STAT4330) Berdasarkan Kategori Penilaian III Masa Ujian 872, 881, dan 882	27

## BAB I

### PENDAHULUAN

#### 1. Latar Belakang

Testing atau pengujian mempunyai fungsi yang penting dalam dunia pendidikan. Menurut Ebel dan Fresbie (1986), fungsi utama dari testing adalah untuk mengukur keberhasilan belajar siswa dan untuk memberikan masukan bagi evaluasi kemajuan kegiatan belajar-mengajar. Nilai hasil testing akan memberikan gambaran evaluasi terhadap keberhasilan belajar siswa. Nilai tersebut dapat digunakan sebagai dasar penentuan apakah seorang siswa dapat melanjutkan pendidikannya ke jenjang yang lebih tinggi, dan bagi mereka yang telah bekerja bahkan dapat menentukan status kepegawaiannya di masa depan. Nilai yang diperoleh siswa juga merupakan masukan bagi guru atau institusi pendidikan untuk mengevaluasi cara mengajar ataupun program belajarnya.

Di Universitas Terbuka (UT), pengujian merupakan kegiatan pendidikan yang terpenting karena merupakan satu-satunya alat untuk mengukur mutu mahasiswa dan mutu program pengajarannya. Proses belajar mengajar di UT sepenuhnya dilaksanakan melalui modul yang dirancang khusus untuk keperluan belajar mandiri. Dalam hal ini, diasumsikan bahwa mahasiswa UT yang tersebar di seluruh Indonesia dari tahun ke tahun mempunyai kemampuan belajar mandiri yang setara (equivalent). Modul-modul untuk setiap matakuliah ditulis oleh para pakar yang umumnya terdiri dari para dosen universitas negeri terkemuka di Indonesia, misalnya

Universitas Indonesia, Universitas Gadjah Mada, Institut Teknologi Bandung, Universitas Diponegoro, dan sebagainya (Universitas Terbuka, 1989). Dilihat dari segi kualitas para penulisnya, maka kualitas modul-modul UT nampaknya sangat baik. Hal ini juga dibuktikan dengan banyaknya universitas negeri lain dan universitas swasta yang menggunakan modul UT sebagai bahan perkuliahan.

Sumber bahan ujian UT adalah modul matakuliah yang bersangkutan. Naskah ujian disusun berdasarkan Tujuan Instruksional Umum (TIU) dan Tujuan Instruksional Khusus (TIK), seperti yang dicantumkan dalam setiap modul (Universitas Terbuka, 1986). Soal-soal ujian dirakit berdasarkan kisi-kisi ujian yang direncanakan dengan baik. Kisi-kisi ujian mencakup materi modul yang dianggap penting untuk diujikan. Dengan demikian, diharapkan soal-soal yang disusun berdasarkan kisi-kisi tersebut dapat mencerminkan materi modul.

Untuk mengukur keberhasilan belajar mahasiswa, perlu diusahakan agar naskah ujiannya setara pada setiap masa ujian. Sampai saat ini belum diketahui dengan pasti sejauh mana kesetaraan naskah ujian tersebut untuk setiap masa ujian yang berbeda. Dalam hal ini kesetaraan naskah ujian ditinjau dari segi validitas dan reliabilitas butir soalnya.

Dengan diketahuinya komposisi dan kesetaraan masing-masing naskah ujian, maka dapat ditentukan komposisi soal yang dianggap ideal sebagai alat ukur keberhasilan belajar mahasiswa.

## 2. TUJUAN PENELITIAN

Penelitian ini bertujuan untuk :

1. menelaah kesetaraan perangkat soal ujian secara kualitatif untuk matakuliah Metode Peramalan (STAT4330), masa ujian 872, 881 dan 882
2. menelaah hasil pengukuran masing-masing perangkat soal terhadap mahasiswa UT

## 3. MANFAAT PENELITIAN

Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai:

1. pencerminan perangkat soal ujian terhadap materi modul
2. kesetaraan materi perangkat soal untuk ketiga masa ujian
3. komposisi tingkat kesukaran dan daya beda soal, serta reliabilitas perangkat soal antara ketiga masa ujian
4. hasil pengukuran perangkat soal 872,881 dan 882 terhadap mahasiswa

Gambaran-gambaran tersebut diharapkan dapat memberikan sumbangan pemikiran bagi pimpinan fakultas untuk menyusun kebijaksanaan tentang penyusunan butir-butir soal dalam naskah ujian. Hasil penelitian juga dapat dimanfaatkan untuk bahan penyempurnaan dan revisi butir soal maupun revisi naskah ujian.

## 4. PERTANYAAN PENELITIAN

Untuk mencapai tujuan penelitian, disusun pertanyaan-pertanyaan penelitian sebagai berikut:

1. Bagaimana komposisi soal ujian matakuliah Metode Peramalan menurut kisi-kisi?
2. Bagaimana komposisi soal ujian matakuliah Metode Peramalan pada masa ujian 872,881 dan 882 menurut kisi-kisi soal ujian?
3. Apakah terdapat kesetaraan perangkat soal ujian matakuliah Metode Peramalan antara masa ujian 872, 881 dan 882?
4. Bagaimana komposisi soal ujian menurut hasil analisis butir soal?
5. Bagaimana hasil pengukuran masing-masing perangkat soal terhadap mahasiswa UT?

## BAB II

### TINJAUAN PUSTAKA

Pada dasarnya seperangkat tes hanya merupakan sebagian atau sampel dari materi pelajaran karena tidak mungkin menyusun suatu tes yang meliputi seluruh materi mata pelajaran. Agar materi ujian dapat mencerminkan materi pelajaran, maka tes atau ujian harus direncanakan secara matang, dan disusun berdasarkan kisi-kisi soal ujian. Mutu suatu perangkat ujian tergantung kepada seberapa baik materi pelajaran terwakili oleh materi ujian (content validity). Pencerminan materi pelajaran dalam perangkat ujian dapat dinilai dari ruang lingkup dan penekanan materi pelajaran yang tercakup dalam kisi-kisi soal yang bersangkutan.

Menurut Sudirman dkk. (1987), kisi-kisi adalah peta distribusi soal dari berbagai topik atau pokok bahasan. Kisi-kisi soal (table of specifications) juga disebut test blueprint, test grid, atau content validity chart (Mehrens dan Lehmann, 1984: p. 67). Dalam menyusun kisi-kisi harus diperhatikan hal-hal sebagai berikut:

- a. pokok bahasan yang akan diujikan
- b. jenjang kemampuan
- c. persentase tiap kisi-kisi mengenai pokok bahasan dan jenjang kemampuan
- d. bentuk dan jumlah soal
- e. waktu yang digunakan

(Sudirman dkk., 1987).

yaitu:

- A = melengkapi empat (4) pilihan
- B = analisis hubungan antar hal
- C = analisis kasus
- D = melengkapi pilihan berganda kompleks
- F = pemakaian diagram, gambar dan grafik

5. jumlah soal setiap BMP
6. persentase soal setiap BMP
7. jumlah soal total
8. persentase soal setiap jenjang kemampuan

Menurut Panduan Penulisan Soal Universitas Terbuka (1986), persentase jumlah soal untuk setiap jenjang kemampuan adalah:

- 20 % soal C<sub>1</sub> dan C<sub>2</sub>
- 40 % soal C<sub>3</sub>
- 30 % soal C<sub>4</sub> dan C<sub>5</sub>.
- 10 % soal C<sub>6</sub>

Kisi-kisi soal UT tersebut seperti terlihat pada Tabel 1.

Tabel 1. Kisi-kisi Soal Universitas Terbuka

No BMP	Jenjang Kemampuan Macam Soal Pokok Bahasan dan TIK	Jenjang Kemampuan				Jumlah Soal	
		C1,2 ABCDE	C3 ABCDE	C4,5 ABCDE	C6 ABCDE		%
1.	.....	( )	( )	( )	( )	( )	( )
2.	.....						
3.	.....						
4.	.....						
5.	.....						
6.	.....						
Persentase		20%	40%	30%	10%	100%	100%

Perangkat tes yang baik adalah yang memenuhi persyaratan sebagai alat ukur keberhasilan belajar siswa. Salah satu

persyaratan yang harus dipenuhi oleh perangkat tes adalah bahwa alat ukur tersebut harus valid dan reliabel.

Yang dimaksudkan dengan validitas tes ialah sejauh mana tes tersebut dapat mengukur apa yang seharusnya diukur (Mehrens dan Lehmann, 1984). Validitas tes ada tiga jenis, yaitu:

1. content validity atau validitas isi; mempersoalkan sejauh mana sampel soal yang membentuk perangkat tes mewakili materi pelajaran sehingga skor yang diperoleh sebagai hasil tes tersebut dapat diinterpretasikan sebagai gambaran keberhasilan belajar siswa.
2. criterion-related validity; menunjukkan hubungan antara skor tes dengan suatu kriteria tertentu. Misalnya, hubungan antara skor tes bakat terhadap Indeks Prestasi (IP) mahasiswa.
3. construct validity atau validitas konsep; menyatakan sejauh mana hasil pengukuran dianggap dapat mencerminkan suatu rekaan (construct) dalam teori psikologi. Misalnya, suatu tes dianggap valid apabila tes tersebut dapat mengungkapkan perbedaan-perbedaan potensi akademik antara individu yang satu dengan individu yang lain.

Dengan asumsi bahwa mahasiswa UT mempunyai kemampuan akademik yang merata, maka tes yang diujikan dari masa ujian ke masa ujian harus setara, baik dalam hal validitas isi atau materi ujian, tingkat kesukaran dan daya beda soalnya. Dengan demikian tes tersebut dapat diandalkan (reliabel).

Yang dimaksud dengan reliabilitas atau keandalan tes

adalah tingkat atau derajat ketetapan (konsistensi) tes yang bersangkutan (Sudirman, 1987). Suatu tes dapat dikatakan andal atau reliabel jika selalu memberikan hasil yang sama apabila diujikan pada kelompok mahasiswa yang sama pada waktu atau kesempatan yang berbeda.

Menurut Ebel dan Fresbie (1986), reliabilitas tes perlu diketahui oleh penyusun tes dan peserta tes yang bersangkutan. Mahasiswa dapat minta bukti bahwa tes yang digunakan untuk mengukur keberhasilan belajar mereka tidak hanya mewakili materi pelajaran tetapi juga mempunyai kualitas yang akan menghasilkan skor tes yang reliabel. Keinginan mahasiswa untuk mengetahui reliabilitas tes tersebut sebenarnya wajar, karena nilai matakuliah yang akan diperoleh sangat tergantung dari skor yang diperoleh dari tes.

Di UT, agar tes suatu matakuliah valid dan memberikan skor yang reliabel maka materi tes tersebut harus setara dari masa ujian ke masa ujian. Oleh karena itu ruang lingkup dan penekanan materi pelajaran yang diujikan seharusnya setara. Hal ini dapat dicapai apabila tes yang disusun hanya mengacu pada suatu kisi-kisi standar.

Penyusun tes perlu mengetahui tingkat reliabilitas tes yang disusunnya. Reliabilitas tes yang kurang baik dapat diperbaiki dengan cara memperbaiki tehnik penulisan soalnya, dengan catatan materi soal tes tersebut sudah valid. Validitas materi tes hanya dapat dinilai oleh para ahli materi untuk matakuliah yang bersangkutan. Reliabilitas tes

juga dapat diperbaiki dengan menambahkan sejumlah soal pada tes. Untuk UT, hal ini sulit dilaksanakan karena waktu ujian dibatasi 90 menit untuk setiap matakuliah, sudah termasuk waktu yang diperlukan untuk mengisi LJU (Lembar Jawaban Ujian) dan Daftar Hadir Peserta Ujian. Jumlah soal ujian UT bervariasi antara satu matakuliah dengan matakuliah yang lain sesuai dengan tingkat kesukaran soal dan waktu yang diperlukan untuk mengerjakan setiap soal.

Reliabilitas tes dihitung dengan jalan menganalisis item tes (item analysis). Analisis item adalah proses menganalisis jawaban mahasiswa terhadap setiap item untuk menilai kualitas soal yang membentuk suatu tes (Mehrens dan Lehmann, 1984). Analisis item dilakukan secara statistik dan bertujuan untuk mengukur sejauh mana keefektifan tes yang disusun, yaitu untuk membuktikan secara empirik apakah butir soal benar-benar merupakan butir soal yang baik.

UT menggunakan formula Kuder-Richardson 20 (KR-20) untuk menghitung reliabilitas tes, dengan rumus:

$$r = \frac{n}{n-1} \left[ 1 - \frac{\sum pq}{s^2} \right]$$

$n$  = jumlah soal dalam tes

$p$  = proporsi mahasiswa yang menjawab benar (tingkat kesukaran soal). Misal : untuk soal nomor 1, 6 dari 30 peserta ujian menjawab benar, maka  $p$  untuk soal nomor 1 adalah  $6/30 = 0.20$

$q$  = proporsi mahasiswa yang menjawab salah ( $q = 1 - p$ ); jika  $p = 0.20$ ,  $q = 0.80$ )

$pq$  = variansi sebuah soal yang diskor secara dikotomi (benar atau salah)

$\sum$  = tanda jumlah yang menunjukkan bahwa  $pq$  dihitung untuk seluruh soal

$s^2$  = variansi tes

Di samping reliabilitas tes, informasi yang dibutuhkan untuk mengetahui butir soal yang baik dan yang tidak baik dari suatu tes adalah tingkat kesukaran soal (Item Difficulty Index), Indeks Daya Beda Soal (Item Discrimination Index), dan Indeks Kesalahan Pengukuran (Standard Error of Measurement = SEM).

Reliabilitas tes dipengaruhi oleh tingkat kesukaran soal dan tingkat kesukaran tes. Tes yang butir-butir soalnya terlalu mudah atau terlalu sukar akan menyebabkan variabilitas skornya kecil sehingga variabilitasnya menjadi rendah (Mehrens dan Lehmann, 1984). Item-item yang sangat mudah sehingga dapat dijawab benar oleh hampir seluruh testee akan mempunyai pengaruh yang kecil atau tidak akan berpengaruh terhadap reliabilitas tes. Item-item yang sangat sukar sehingga hampir seluruh testee tidak dapat menjawab benar akan mengurangi reliabilitas tes apabila sebagian besar testee menebak jawaban secara random. Tes dianggap reliabel apabila harga koefisien reliabilitasnya tinggi dan indeks kesalahan pengukurannya rendah.

Tingkat kesukaran soal ( $p$ ) adalah persentase siswa dalam sekelompok peserta tes yang menjawab soal dengan benar (Ebel dan Fresbie, 1986). Jadi, semakin tinggi harga tingkat kesukaran, soal akan semakin mudah. Sebagai contoh, apabila sebuah soal yang diujikan pada sekelompok siswa dapat dijawab benar oleh 70% siswa tersebut maka tingkat kesukaran soalnya adalah 0.70. Tingkat kesukaran soal tergantung pada materi soal dan kepada kemampuan kelompok siswa tersebut dalam

menjawab soal yang bersangkutan (Nugraheni, 1989).

Menurut Fernandes (1984), butir soal yang baik adalah item yang mempunyai tingkat kesukaran antara 0.25 dan 0.75, dan yang terbaik adalah item yang mempunyai tingkat kesukaran 0,50. Sedangkan Ebel dan Fresbie (1986) berpendapat bahwa tes yang baik harus menyertakan beberapa item yang sulit untuk menguji siswa yang lebih pandai dan beberapa item yang mudah yang dapat dijawab oleh siswa yang kurang pandai. Butir soal yang mudah dimaksudkan untuk menambah skor setiap testee. Penambahan butir soal yang mudah akan memberikan semangat kepada testee yang secara psikologis dibutuhkan untuk dapat mengerjakan ujian dengan baik. Peringkat skor siswa tidak akan berubah karena siswa yang lebih pandai akan mendapat tambahan skor dari butir soal yang sulit. Butir soal yang mudah dan sulit tersebut diperlukan untuk menghitung indeks daya beda soal.

Indeks daya beda soal ( $r$ ) menunjukkan kemampuan soal untuk membedakan siswa yang pandai dari siswa yang kurang pandai (Fernandes, 1984). Harga indeks daya beda soal ( $r$ ) berkisar antara -1.00 dan +1.00, yaitu menunjukkan koefisien korelasi antara skor soal dengan skor tes. Harga  $r$  dipengaruhi oleh tingkat kesukaran soal. Semakin sulit atau semakin mudah soal akan menyebabkan harga  $r$  rendah. Apabila harga korelasinya positif berarti siswa yang lebih pandai cenderung menjawab soal dengan benar dan siswa yang kurang pandai cenderung menjawab salah. Semakin tinggi harga indeks daya beda soal berarti soal tersebut semakin dapat membedakan antara siswa yang pandai dari siswa yang kurang pandai.

Harga korelasi 0 (nol) menunjukkan bahwa soal tersebut tidak dapat membedakan siswa yang pandai dari yang kurang pandai. Hal ini dapat disebabkan oleh (1) soal terlalu sukar atau terlalu mudah sehingga seluruh testee tidak dapat menjawab benar atau seluruh siswa dapat menjawab benar, dan (2) soal tidak jelas atau mengandung dua pengertian (ambiguous) (Mehrens dan Lehmann, 1984). Apabila harga  $r$  negatif, berarti siswa yang kurang pandai cenderung menjawab benar. Oleh karena itu, untuk soal pilihan berganda diharapkan  $r$  jawaban soal positif dan pengecohnya (distractor) negatif sehingga hanya siswa yang kurang pandailah yang menjawab salah. Lebih lanjut Mehrens dan Lehmann menyatakan bahwa jika separuh jumlah testee dianggap pandai dan yang separuh kurang pandai maka sebagian besar item kemungkinan mempunyai harga  $r = 0.20$ . Dengan asumsi bahwa separuh jumlah peserta tes kurang pandai maka item yang baik dalam tes yang bersangkutan mempunyai indeks daya beda soal sebesar atau lebih besar dari 0.20. Item-item yang harga  $r$ -nya kurang dari 0.20 harus direvisi.

Hasil analisis soal dipengaruhi oleh kemampuan peserta tes, jumlah testee, dan cara penulisan soalnya (Mehrens dan Lehmann, 1984). Oleh karena itu dalam menginterpretasikan hasil analisis penyusun tes harus selalu menilai kembali materi soalnya. Apabila soal tersebut dimaksudkan untuk mengukur aspek pengetahuan tertentu untuk suatu tujuan pelajaran yang tertentu maka soal tersebut harus tetap dipertahankan meskipun hasil analisisnya kurang baik. Lebih

lanjut Mehrens dan Lehmann menyatakan bahwa tingkat kesukaran soal yang diperoleh dari hasil analisis dapat dipengaruhi oleh hasil menebak jawaban dan letak kunci soal diantara alternatif jawaban yang lain.

Hasil analisis bersifat tentatif (sementara) karena perangkat soal yang sama dapat memberikan hasil analisis yang berbeda apabila diujikan kepada kelompok mahasiswa yang berbeda. Dengan demikian tingkat kesukaran dan daya beda soal ujian sangat dipengaruhi oleh skor yang diperoleh mahasiswa dari tes yang bersangkutan. Nilai suatu tes biasanya merupakan jumlah skor soal-soal yang membentuk tes. Untuk tes obyektif, skor mewakili jumlah soal yang dijawab benar. Nilai (grade) --sering disebut nilai akhir-- merupakan interpretasi dari nilai mentah atau skor (score) yang diperoleh dari tes. Nilai tersebut digunakan untuk melaporkan keberhasilan belajar siswa; apakah ia lulus dengan memuaskan, cukup baik, kurang atau tidak lulus. Jadi, nilai merupakan gambaran keberhasilan belajar siswa yang dapat dilaporkan kepada siswa sendiri, kepada orang tua, kepada institusi pendidikan yang lebih tinggi ataupun kepada calon pemberi kerja (Ebel dan Fresbie, 1986). Bagi siswa, nilai dapat memberikan motivasi belajar yang lebih tinggi. Menurut Nugraheni (1989), kriteria penilaian yang diterapkan di UT seperti tercantum pada Tabel 2. Apabila nilai yang diperoleh mahasiswa berdasarkan kategori standar terlalu rendah maka fakultas menentukan kategori penilaian yang akan diterapkan terhadap kelompok mahasiswa tersebut. Pemilihan kategori penilaian yang akan diterapkan untuk setiap matakuliah dan setiap masa ujian

sepenuhnya tergantung pada kebijaksanaan fakultas.

Tabel 2. Kriteria Penilaian di Universitas Terbuka

Nilai	Standar	Kategori Penilaian		
		I	II	III
A	80 - 100	75 - 100	70 - 100	65 - 100
B	70 - < 80	65 - < 75	60 - < 70	55 - < 65
C	55 - < 70	50 - < 65	45 - < 60	40 - < 55
D	40 - < 55	35 - < 50	30 - < 45	25 - < 40
E	< 40	< 35	< 30	< 25

Sebaran nilai mahasiswa menggambarkan kemampuan mahasiswa dalam setiap matakuliah. Karena kegiatan belajar mengajar di UT dilaksanakan secara mandiri maka sebaran nilai tersebut dapat memberikan gambaran mengenai (1) kemampuan belajar mandiri mahasiswa UT dan (2) kualitas modul UT sebagai bahan belajar mandiri. Dengan mengetahui sebaran nilai mahasiswa, hasil analisis item dan kesetaraan materi soal UT dari masa ujian ke masa ujian akan dapat diketahui gambaran kualitas materi ujian UT.

### BAB III METODOLOGI PENELITIAN

Pada Bab ini diuraikan tentang metode yang digunakan dalam pelaksanaan penelitian. Bab ini terdiri atas 2 sub bab, yaitu Sub Bab Populasi dan Sampel Penelitian dan Sub Bab Prosedur Penelitian.

#### 1. Populasi dan Sampel Penelitian

Populasi penelitian ini adalah seluruh perangkat soal ujian UT untuk setiap matakuliah yang telah diujikan. Sebagai sampel dipilih perangkat soal untuk matakuliah Metode Peramalan (STAT4330) yang diujikan pada masa ujian 872, 881, dan 882. Pemilihan matakuliah Metode Peramalan didasarkan kepada keperluan (purposive sampling) karena matakuliah ini sedang dialami materinya oleh peneliti.

#### 2. Prosedur Penelitian

Data yang dibutuhkan untuk penelitian ini berupa (1) kisi-kisi standar untuk matakuliah Metode Peramalan, (2) perangkat soal matakuliah ini untuk masa ujian 872, 881, dan 882, (3) hasil analisis soal matakuliah ini untuk ketiga masa ujian, dan (4) sebaran nilai mahasiswa yang mengikuti ujian pada ketiga masa ujian tersebut.

Kisi-kisi soal ujian standar dan perangkat soal diperoleh dari Fakultas Matematika dan Ilmu Pengetahuan Alam UT (FMIPA-UT). Sedangkan hasil analisis soal dan sebaran nilai mahasiswa diperoleh dari komputer mini UT.

Kegiatan penelitian dilaksanakan dalam empat tahap. Tahap pertama adalah melakukan penelaahan terhadap ketiga perangkat soal secara kualitatif. Komposisi soal ujian untuk masing-masing masa ujian dibandingkan dengan kisi-kisi berdasarkan kriteria :

- nomor modul
- jenjang kemampuan menurut analisis penulis soal atau analisis peneliti jika penulis soal tidak menyebutkannya
- jumlah soal per modul
- jumlah soal total

Dalam hal ini peneliti memilih Nomor BMP sebagai kriteria untuk penetapan ruang lingkup dan materi modul, dan tidak menggunakan TIK ataupun Pokok Bahasan. Keputusan ini diambil untuk memudahkan peneliti dalam mengklasifikasikan soal ke dalam form kisi-kisi karena penyusun tes tidak menyebutkan TIK ataupun Pokok Bahasan pada setiap soal.

Ketiga perangkat soal ujian dibandingkan satu sama lain berdasarkan kriteria tersebut. Kemudian kesetaraan soal ujian ketiga perangkat soal dianalisis secara kualitatif, yang didasarkan pada ruang lingkup dan penekanan materi modul yang diujikan.

Tahap kedua, dilakukan penelaahan ketiga perangkat soal berdasarkan hasil analisis butir soal. Pertama-tama jawaban mahasiswa pada ketiga masa ujian dianalisis dengan menggunakan fasilitas program Item Analysis yang tersedia di komputer mini UT. Dari hasil analisis soal yang diperoleh, setiap perangkat soal kemudian diklasifikasikan menurut:

a) derajat atau tingkat kesukaran soal (p-value) untuk soal-soal yang harga indeks daya beda ( $r$ ) nya di atas harga minimum, dengan kriteria sebagai berikut

- terlalu mudah (  $p > 0,90$  )
- mudah (  $0,71 < p < 0,90$  )
- sedang (  $0,41 < p < 0,70$  )
- sukar (  $0,20 < p < 0,40$  )
- terlalu sukar (  $p < 0,20$  )

Nomor soal untuk soal-soal yang tersisa, yaitu soal-soal yang r nya tidak memenuhi syarat dicatat.

b) daya beda soal ( $r$ ) untuk soal-soal dengan p-value yang memenuhi syarat (  $0,20 < p < 0,90$  ), dengan kriteria :

- tidak memenuhi syarat (  $r < r_{\text{minimum}}$  )
- kurang baik (  $r_{\text{min.}} < r < 0,20$  )
- sedang (  $0,20 < r < 0,30$  )
- baik (  $0,30 < r < 0,40$  )
- baik sekali (  $r > 0,40$  )

Soal-soal dengan p-value yang tidak memenuhi syarat dicatat nomornya.

c) reliabilitas perangkat soal dengan rumus KR-20 dan indeks kesalahan pengukuran (SEM) masing-masing perangkat soal

Kemudian ketiga perangkat soal dibandingkan berdasarkan klasifikasi tersebut.

Tahap ketiga, menelaah hasil pengukuran masing-masing perangkat soal terhadap mahasiswa yang bersangkutan yang didasarkan pada sebaran atau distribusi skor yang diperoleh mahasiswa untuk masa ujian yang bersangkutan.

Tahap keempat, membahas hubungan antara kesetaraan soal dengan hasil ujian mahasiswa. Hasil analisis butir soal ditelaah dengan jalan mengkaitkannya dengan kesetaraan soal antara ketiga masa ujian. Kemudian hasil penelaahan tersebut dikaitkan dengan hasil pengukuran terhadap mahasiswa yang telah dilakukan pada tahap ketiga penelitian ini.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 1. Hasil Penelitian

1.1. Komposisi soal ujian Metode Peramalan menurut kisi-kisi ujian standard. Komposisi soal untuk masing-masing perangkat soal yang diujikan pada masa ujian 872, 881 dan 882 diperoleh dengan jalan memasukkan setiap nomor soal ke dalam form kisi-kisi, berdasarkan kriteria:

- nomor BMP
- jenjang kemampuan
- jumlah soal per BMP
- jumlah soal total

Ruang lingkup dan penekanan materi modul yang harus diujikan berdasarkan kisi-kisi soal ujian standar dapat dilihat pada Tabel 3.

**TABEL 3. KISI-KISI SOAL UJIAN METODE PERAMALAN  
FAKULTAS MATEMATIKA DAN ILMU PENETAHUAAAN ALAM UNIVERSITAS TERBUKA**

No BMP	Jenjang Kemampuan Macam Soal Pokok Bahasan	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	Jml. Soal	Σ
		ABCDE	ABCDE	ABCDE		
I.	PELHDARULUAN	1			1	1
	1. Arti dan Kegunaan Peramalan					
	2. Peramalan Kuantitatif dan Kualitatif					
	3. Dasar-dasar Peramalan Kuantitatif					
II.	METODE ANALISIS REGRESI	2			2	3
	1. Metode Regresi dalam Analisis Runtun Waktu					
III.	METODE RATA-RATA BERGERAK		3	1	4	7
	1. Rata-rata Bergerak Sederhana untuk Pola Stasioner					
	2. Rata-rata Bergerak untuk Pola Trend Linier					

NO BAP	Jenjang Kemampuan	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	Jml. Soal	%
	Macam Soal Pokok Bahasan	ABCDE	ABCDE	ABCDE		
IV.	METODE PENGHALUSAN EKSPONENSIAL	2	3	1	6	10
	1. Penghalusan Eksponensial untuk Pola Stasioner					
	2. Penghalusan Eksponensial untuk Pola Trend Linier					
	3. Penghalusan Eksponensial Orde-N					
V.	METODE PENGHALUSAN LANGSUNG	1	3	2	6	10
	1. Kuadrat Terkecil Tersusut dan Penghalusan Langsung					
	2. Penerapan pada Model Polinom dan Model Transenden					
VI.	BEBERAPA MODEL UNTUK DATA MUSIMAN	1	3	5	9	15
	1. Model-model Trigonometri					
	2. Model Multiplikatif dan Model Aditif					
VII.	INTERVAL PREDIKSI	1	5	6	12	20
	1. Variansi Galat Ramalan					
	2. Interval Prediksi					
VIII.	PENAKSIRAN MEAN DAN VARIANSI GALAT RAMALAN	1	5	7	13	22
	1. Penaksiran Mean dan Variansi dalam Peramalan Periodik					
	2. Penaksiran Variansi dalam Peramalan T - Periode di depan					
IX.	PENGAWASAN KUALITAS PERAMALAN	1	3	3	7	12
	1. Penelusuran Jejak					
	2. Pemeriksaan Data Pengamatan					
JUMLAH SOAL		12	24	24	60	100
PERSENTASE		20	40	40	100	

1.2. Komposisi soal ujian Metode Peramalan pada masa ujian 872, 881 dan 882. Komposisi soal untuk ketiga masa ujian

dibandingkan dengan komposisi soal menurut kisi-kisi standar (Tabel 4.) Di sini nampak jelas bahwa ketiga perangkat soal tidak menunjukkan kesetaraan materi tes apabila dilihat dari ruang lingkup, penekanan materi modul yang diujikan dan jenjang kemampuan yang diukur apabila dibandingkan dengan kisi-kisi soal standar.

TABEL 4. KOMPOSISI SOAL UJIAN MATAKULIAAH METODE PERAMAALAN MENURUT KISI-KISI SOAL UJIAN UNIVERSITAS TERBUKA

No BMP	STANDARD			872				881				882					
	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	%	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	%	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	%	C <sub>1,2</sub>	C <sub>3</sub>	C <sub>4,5,6</sub>	%	
	ABCDE	ABCDE	ABCDE		ABCDE	ABCDE	ABCDE		ABCDE	ABCDE	ABCDE		ABCDE	ABCDE	ABCDE		
I.	1			1	3			12									
II.	2			3	2			8									
III.	2	1	1	7	3	2		20	1	1		6,67					
IV.	2	3	1	10	2	1		12	3	3		20	7	1			27
V.	1	3	2	10		1		4	1	1		6,67	1	1			6,5
VI.	1	3	5	15		2		8	1	1		6,67	1	1			6,5
VII.	1	5	6	20			1	4	1	5	6	40	1	2	4		23
VIII.	1	5	7	22		1	3	16									10
IX.	1	3	3	12	1	2	1	16	1	2	3	20	2	3	3		27
JS	12	24	24		11	9	5		8	13	9		12	8	10		100
%	20	40	40	100	44	36	20	100	26,67	43,33	30	100	40	26,67	33,33		100

**Keterangan**

MU : Masa Ujian  
 BPM : Buku Materi Pokok  
 JS : Jumlah Soal

1.3. Kesetaraan perangkat soal ujian Metode Peramalan antara masa ujian 872, 881 dan 882. Ditinjau dari ruang lingkup materi pelajaran yang dicakup (nomor BMP yang diujikan materinya), ketiga perangkat soal tidak menunjukkan kesetaraan materi ujian satu sama lain.

Perangkat soal 872 terdiri atas 25 soal yang mewakili semua nomor BMP. Penekanan materi tes paling banyak dari modul 3 (20%), modul 8 (16%) dan modul 9 (16%). Persentase jenjang kemampuan yang diukur adalah 44% ( $C_{1,2}$ ), 36% ( $C_3$ ), dan sebesar 20% untuk  $C_{4,5,6}$ .

Pada masa ujian 881, materi modul yang dicakup dalam tes hanya meliputi enam BMP dari sembilan BMP. BMP yang materinya tidak diujikan dalam tes adalah BMP 1, 2 dan 8. Penekanan materi tes paling banyak berasal dari BMP 7 (40 %) serta BMP 4 dan 9 (masing-masing 20%). Persentase aspek pengetahuan yang dicakup adalah 26,67% ( $C_{1,2}$ ), 43,33% ( $C_3$ ) dan 30% ( $C_{4,5,6}$ ).

Pada masa ujian 882, materi modul yang dicakup dalam tes juga hanya terdiri atas 6 BMP. Penekanan materi modul paling banyak dari BMP 4 dan 9 (masing-masing 27 %) dan dari BMP 7 (23 %). Persentase materi modul ditinjau dari aspek pengetahuan yang diukur adalah 40 % ( $C_{1,2}$ ), 26,67 % ( $C_3$ ), dan 33,33 % ( $C_{4,5,6}$ ).

Ditinjau dari jumlah soal total, ketiga perangkat soal tidak mengacu pada kisi-kisi standar. Jumlah soal ujian untuk matakuliah ini menurut kisi-kisi adalah 60 soal. Pada masa ujian 872 jumlah soal hanya 25 butir soal, sedangkan pada masa ujian 881 dan 882 jumlah soal ujiannya masing-masing 30 butir soal.

**1.4. Komposisi soal ujian Metode Peramalan menurut hasil analisis butir soal.** Ketiga perangkat soal tidak mempunyai butir soal yang terlalu mudah maupun butir soal yang mudah

ditinjau dari tingkat kesukaran soalnya. Apabila ditinjau hanya dari tingkat kesukaran butir soalnya, perangkat soal 872 adalah yang paling mudah karena sepuluh butir soal dari 21 soal yang memenuhi syarat termasuk dalam kategori tingkat kesukaran sedang.

Menurut persyaratan tingkat kesukaran dan daya beda soal yang diterapkan di UT, maka perangkat soal pada masa ujian 872 terdiri atas 21 soal yang memenuhi persyaratan, yaitu 10 soal dengan tingkat kesukaran sedang dan 11 soal yang sukar. Dari ke-21 soal tersebut enam soal mempunyai daya beda soal sangat baik, dua soal termasuk baik, enam soal mempunyai daya beda soal sedang, dan tujuh soal kurang memenuhi syarat daya beda soalnya (lihat Tabel 5).

Pada masa ujian 881, ada 26 soal yang memenuhi persyaratan apabila ditinjau dari tingkat kesukaran dan daya beda soalnya. Dua soal mempunyai tingkat kesukaran sedang, dan sisanya termasuk soal yang sukar. Semua soal yang memenuhi syarat tersebut terdiri atas 5 soal dengan daya beda baik sekali, 4 soal termasuk baik, 6 soal dengan daya beda soal sedang, dan 11 soal kurang baik daya bedanya.

Set soal untuk masa ujian 882 terdiri atas 28 butir soal yang tingkat kesukaran dan daya beda soalnya memenuhi syarat. Tujuh soal mempunyai tingkat kesukaran sedang, dan sisanya termasuk dalam kategori soal yang sukar. Dari kedua puluh delapan soal tersebut, 20 soal mempunyai daya beda baik sekali, 4 soal termasuk baik dan 4 soal kurang baik daya beda soalnya.

Dari Tabel 6 dapat diketahui bahwa reliabilitas perangkat soal 872 adalah yang terendah di antara ketiga perangkat soal, yaitu sebesar 0,407. Sedangkan yang paling baik reliabilitasnya adalah masa ujian 882 (0,816).

**TABEL 5. KOMPOSISI SOAL METODE PERAMALAN  
MENURUT TINGKAT KESUKARAN DAN DAYA BEDA SOAL**

Tingkat Kesukaran Soal	Daya Beda Soal	872 N = 236		881 N = 251		882 N = 195	
		NS	JS	NS	JS	NS	JS
Sedang	- Baik Sekali	1,6,7,8,9,10	6	1	1	3,8,14,17	4
	- Baik	2,5	2	9	1	5,7,22	3
	- Sedang	13,14	2	-	-	-	-
	- Kurang	-	-	-	-	-	-
Sukar	- Baik Sekali	-	-	2,3,21,22	4	1,2,4,6,9,13,15, 16,18,19,20,24, 25,26,27,28	16
	- Baik	-	-	11,14,20	3	21	1
	- Sedang	11,12,20,25	4	4,10,15,19,23,24	6		
	- Kurang	3,15,18,19,21,22 23	7	5,6,7,13,16,17, 18,26,27,28,29	11	10,12,29,30	4
J U M L A H			21	26	26	28	

**Keterangan:**

NS = Nomor Soal

JS = Jumlah Soal

Apabila ditinjau dari hasil analisis soalnya, terlihat bahwa ketiga perangkat soal tidak menunjukkan kesetaraan dalam hal daya bedanya, meskipun ketiganya termasuk perangkat

soal yang sukar. Ditinjau dari daya beda soal dan KR-20, perangkat soal yang diujikan pada masa ujian 882 merupakan perangkat soal yang terbaik dibandingkan dengan perangkat soal 872 dan 881.

**TABEL 6. STATISTIK HASIL ANALISIS SOAL  
MATAKULIAH METODE PERAMALAN**

Kriteria	872 N = 236	881 N = 251	882 N = 195
KR-20	0,407	0,490	0,816
Jumlah Mahasiswa	236	251	195
Mean	8,720	9,050	10,180
SD	2,920	3,400	5,550
SEM	2,248	2,426	2,379
$r_{\text{minimum}}$	0,092	0,090	0,092

1.5. Hasil pengukuran masing-masing perangkat soal terhadap mahasiswa UT. Tabel 7 menunjukkan distribusi nilai mahasiswa untuk matakuliah ini pada ketiga masa ujian berdasarkan kategori penilaian standar.

Set soal yang diujikan pada masa ujian 872 merupakan soal yang sangat sulit bagi mahasiswa. Tidak satupun mahasiswa yang mengikuti ujian pada masa ujian tersebut mendapat nilai A maupun B. Sebanyak 65,2% peserta ujian tidak lulus ujian atau mendapat nilai E.

Pada masa ujian 881, hanya ada seorang mahasiswa yang mendapat nilai A tetapi tidak ada yang mendapat nilai B. Persentase mahasiswa yang tidak lulus paling banyak apabila dibandingkan dengan kedua masa ujian yang lain, yaitu sebesar

75,4%.

Ada 12 peserta ujian yang mendapat nilai A dan 6 orang mendapat nilai B pada masa ujian 882. Sebanyak 72,2% mahasiswa mendapat nilai E.

**TABEL 7. DISTRIBUSI NILAI MAHASISWA UT  
MATAKULIAH METODE PERAMALAN (STAT4330)  
BERDASARKAN KATEGORI PENILAIAN STANDAR  
MASA UJIAN 872,881 DAN 882**

Masa Ujian	872 N = 236		881 N = 251		882 N = 195	
	JM	%	JM	%	JM	%
A ( > 80 )	0	0	1	0,4	12	6,4
B ( 70 - 79 )	0	0	0	0	6	3,2
C ( 55 - 69 )	23	7,4	10	3,8	6	3,2
D ( 40 - 54 )	85	27,4	54	20,8	37	19,8
E ( < 40 )	202	65,2	196	75,4	135	72,2

**Keterangan:**

JM = Jumlah Mahasiswa

Dari Tabel 7 tersebut secara sepintas terlihat bahwa perangkat soal yang diujikan pada masa ujian 881 adalah soal ujian yang paling sulit dibandingkan dengan kedua perangkat soal yang lain. Ketiga perangkat soal tersebut termasuk soal yang sangat sukar karena sebagian besar peserta ujian (lebih dari 50%) tidak lulus ujian.

Oleh karena hasil ujian mahasiswa sangat tidak memuaskan, penilaian untuk ketiga masa ujian adalah berdasarkan kategori penilaian III (Tabel 8). Dari Tabel 8 tampak jelas bahwa penerapan kategori penilaian III sangat membantu mahasiswa. Dengan kategori penilaian tersebut

mahasiswa yang tidak lulus atau mendapat nilai E jauh lebih sedikit dibandingkan dengan nilai yang didapat bila didasarkan pada kategori standar. Tetapi jumlah mahasiswa yang mendapat nilai A, B dan C tetap kurang dari 50% dari jumlah seluruh peserta ujian.

**TABEL 8. DISTRIBUSI NILAI MAHASISWA UT MATAKULIAH METODE PERAMALAN (STAT4330) BERDASARKAN KATEGORI PENILAIAN III MASA UJIAN 872,881 DAN 882**

Masa Ujian	872 N = 236		881 N = 251		882 N = 195	
	JM	%	JM	%	JM	%
A ( > 65 )	7	2,26	3	1,15	18	9,18
B ( 55 - 64 )	16	5,16	8	3,06	6	3,06
C ( 40 - 54 )	85	27,42	54	20,69	37	18,88
D ( 25 - 39 )	150	48,39	114	43,68	84	42,86
E ( < 25 )	52	16,77	82	31,42	51	26,02

**Keterangan:**

JM = Jumlah Mahasiswa

## 2. Pembahasan

2.1. Komposisi soal ujian Metode Peramalan menurut kisi-kisi ujian standard. Menurut kisi-kisi standard, ruang lingkup materi soal ujian Metode Peramalan mencakup seluruh modul, yaitu mewakili kesembilan modul (BMP) yang ada. Pada kisi-kisi yang asli, kisi-kisi disusun berdasarkan TIK (Tujuan Instruksional Khusus). Namun ternyata TIK yang tertera pada kisi-kisi berlainan dengan yang tertera pada modul: ada yang tertulis pada kisi-kisi tetapi tidak tertulis pada

modul, dan tidak jelas disebutkan TIK tersebut mewakili Pokok Bahasan/Kegiatan Belajar yang mana. Oleh karena itu, peneliti memutuskan untuk memodifikasi kisi-kisi standard tersebut berdasarkan Nomor Modul (BMP).

Ditinjau dari jumlah soal per BMP, kisi-kisi tersebut menekankan materi modul VI, VII, dan VIII sebagai materi pelajaran yang harus dikuasai siswa (lihat Tabel 3). Jenjang kemampuan yang diukur dengan menggunakan kisi-kisi tersebut adalah 20% untuk C1,2 dan masing-masing 40% untuk C3 dan C4,5,6. Di sini terdapat penggabungan 2 atau 3 jenjang kemampuan karena dianggap soal yang bersangkutan dapat mengukur 2 atau lebih jenjang kemampuan.

**2.2. Komposisi soal ujian Metode Peramalan pada masa ujian 872, 881 dan 882.** Dalam hal topik materi pelajaran yang dicakup, tes yang diujikan pada masa ujian 872 lebih mendekati kisi-kisi standar karena materi tes mewakili semua BMP dari modul matakuliah ini. Tabel 4 menunjukkan bahwa penyusunan soal ujian yang diujikan pada masing-masing masa ujian tidak disusun mengacu pada kisi-kisi soal ujian standard baik ditinjau dari materi modul yang dicakup, jumlah soal yang mewakili setiap modul yang bersangkutan, prosentase jenjang kemampuan yang diukur maupun dari jumlah total soal pada masing-masing masa ujian.

Jumlah soal menurut kisi-kisi standard sebanyak 60 soal, sedangkan pada masa ujian 872 hanya 25 soal dan pada masa ujian 881 serta 882 jumlah soal yang diujikan hanya 30 soal.

Hal ini menunjukkan bahwa kisi-kisi soal standard dikembangkan tanpa memperhitungkan waktu ujian yang disediakan bagi mahasiswa untuk menempuh ujian yang bersangkutan. Penentuan jumlah soal yang berlainan kemungkinan didasarkan pada perbedaan tingkat kesukaran masing-masing soal yang menyusun perangkat soal. Seharusnya persentase jumlah soal per modul pada ketiga perangkat soal sama dengan kisi-kisi standard.

**2.3. Kesetaraan perangkat soal ujian Metode Peramalan antara masa ujian 872, 881 dan 882.** Kesetaraan materi tes dari segi pokok bahasan serta jumlah soal per pokok bahasan yang diujikan pada ketiga perangkat soal ternyata tidak dapat dibandingkan. Hal ini disebabkan karena kisi-kisi standar hanya menunjukkan jumlah soal dan persentase soal per BMP saja, dan tidak menekankan materi ujian pada pokok bahasan tertentu.

Ditinjau dari ruang lingkup materi pelajaran yang dicakup, ketiga perangkat soal tidak menunjukkan adanya kesetaraan materi pelajaran baik antara masing-masing perangkat soal dengan kisi-kisi standard maupun antara perangkat soal yang satu dengan yang lainnya. Perangkat soal ujian yang diujikan pada masa ujian 872 adalah yang paling mendekati materi soal ujian menurut kisi-kisi standard.

Meskipun ketiga perangkat soal tidak menunjukkan kesetaraan dengan kisi-kisi, tetapi bukan berarti bahwa perangkat soal tersebut tidak valid. Menurut peneliti, materi soal ketiga perangkat soal memang mengukur kemampuan

mahasiswa dalam matakuliah ini, tetapi penekanan materi modul yang diujikan pada ketiga masa ujian berlainan. Jadi, ketidaksetaraan perangkat soal tersebut kemungkinan hanya menunjukkan bahwa tingkat kesukaran ketiga perangkat soal berbeda-beda.

2.4. Komposisi soal ujian Metode Peramalan menurut hasil analisis butir soal. Apabila ditinjau dari hasil analisis soalnya secara statistik, ketiga perangkat soal termasuk perangkat soal yang sukar. Namun, perangkat soal 872 adalah perangkat soal yang paling mudah di antara ketiga perangkat soal yang diteliti ini. Perangkat soal 872 lebih mudah karena mungkin jumlah soalnya lebih sedikit daripada kedua perangkat soal yang lain sehingga mahasiswa lebih banyak mempunyai waktu pada saat ujian. Diujikannya materi BMP nomor awal pada masa ujian 872 seharusnya memang mencerminkan bahwa perangkat soal tersebut lebih mudah. Namun, sebagai perangkat soal yang termudah belum tentu perangkat soal ini merupakan yang terbaik karena ternyata daya beda soalnya kurang bagus.

Soal yang diujikan pada masa ujian 882 merupakan yang terbaik. Sejumlah 20 soal dari 30 soal yang diujikan pada masa ujian 882 mempunyai daya beda soal sangat baik ( $r > 0,40$ ) meskipun 16 soal di antaranya termasuk soal yang sukar (harga  $p$  antara  $0,20 - 0,40$ ). Dengan demikian, meskipun sukar sebagian besar soal 882 tersebut sangat baik dalam membedakan antara mahasiswa yang lebih pintar dari pada yang kurang pintar.

Reliabilites tes untuk perangkat soal 882 sangat baik, yaitu sebesar 0,816 (lihat Tabel 6). Hal ini menunjukkan bahwa korelasi antara butir-butir soal yang menyusun perangkat soal ini sangat tinggi. Menurut Popham (1981), internal realibilitas tes yang tinggi menunjukkan bahwa butir-butir tes yang menyusun tes tersebut sangat homogen satu sama lain.

Sementara itu perangkat soal 872 dan 881 mempunyai reliabilitas tes yang cukup rendah, yaitu masing-masing sebesar 0,407 dan 0,490 . Bila dilihat dari hasil analisis soalnya (Tabel 5), terlihat bahwa lebih dari setengah jumlah soal pada kedua masa ujian ini mempunyai daya beda soal yang termasuk sedang dan kurang. Daya beda soal yang rendah biasanya disebabkan oleh karena soal tersebut terlalu sukar ataupun terlalu mudah bagi mahasiswa (Mehrens dan Lehmann, 1984). Dalam hal ini, hasil analisis soal menunjukkan bahwa soal-soal tersebut termasuk sukar.

Hasil analisis suatu tes antar lain dipengaruhi oleh kemampuan peserta tes, jumlah peserta tes, dan proses belajar mengajar. Oleh karena itu, hasil analisis tes bersifat tentative karena dapat berubah dari waktu ke waktu. Jadi, meskipun menurut penulis soal perangkat soal yang diteliti ini mempunyai materi dan tingkat kesukaran tes yang setara, namun hasil analisis tes akan berbeda apabila tingkat kemampuan peserta tesnya berbeda. Untuk memutuskan bahwa suatu perangkat soal lebih baik dari yang lain kemampuan rata-rata peserta tes pada setiap masa ujian juga perlu dievaluasi. Apabila mahasiswa banyak yang menebak jawaban

karena soalnya terlalu sulit atau karena konstruksi tesnya tidak jelas atau karena adanya kesalahan kunci jawaban soal maka hasil analisis tes akan sulit diinterpretasikan. Dalam hal ini tampaknya banyak mahasiswa UT yang menebak jawaban, karena kurangnya pengetahuan mereka dalam materi atau pokok bahasan yang diujikan atau karena soalnya memang membutuhkan perhitungan yang teliti dan perlu waktu untuk mengerjakan. Sebagai contoh, di bawah ini dikutip soal ujian nomor 3 dan 15 yang diujikan pada masa ujian 872 (soal termasuk sukar dan daya bedanya kurang, lihat Tabel 7), sebagai berikut:

3. Sekelompok data yang terdiri dari 5 buah hasil pengamatan, menggunakan dua buah model analisis regresi. Kedua model yang memberikan galat taksiran  $e_i = y_i - \bar{y}_i$  adalah sebagai berikut:

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
Model I	1	1	1	1	1
Model II	2	2	2	2	2

maka:

- A. kedua model tidak baik
- B. model I lebih baik daripada model II
- C. model II lebih baik daripada model I
- D. kedua model tidak dapat dibandingkan

Sebetulnya pertanyaannya cukup sederhana dan seharusnya mahasiswa dapat menjawabnya dengan mudah karena materi soal termasuk konsep dasar yang harus dipahami oleh mahasiswa mengenai matakuliah Metode peramalan. Materi soal tersebut dijelaskan dalam modul awal, yaitu BMP 1. Dengan demikian jelas mahasiswa kurang belajar dan hanya menebak jawaban secara random.



Tabel 7).

Berdasarkan kategori penilaian III nilai mahasiswa memang menjadi lebih baik, namun jumlah mahasiswa yang memperoleh nilai A, B dan C tetap kurang dari separuh jumlah peserta ujian. Nilai yang diperoleh berdasarkan kategori III tersebut sebetulnya juga tidak dapat mencerminkan kemampuan mahasiswa dalam matakuliah ini. Kategori III ini harus diterapkan karena perangkat soal ujian pada ketiga masa ujian tersebut tampaknya memang terlalu sulit bagi peserta ujian. Oleh karena itu, sebagai dasar pembahasan peneliti memilih mengacu pada nilai mahasiswa yang diperoleh berdasarkan penilaian dengan kategori standar.

Ditinjau dari hasil analisis butir soalnya, perangkat soal 872 lebih baik dari pada perangkat soal 881. Meskipun pada masa ujian 881 ada seorang mahasiswa yang berhasil memperoleh nilai A (berdasarkan kategori penilaian standar), namun prosentase mahasiswa yang tidak lulus (E) lebih banyak dibandingkan dengan masa ujian 872. Mahasiswa yang mengikuti ujian pada masa ujian 881 yang memperoleh nilai D juga lebih sedikit daripada pada masa ujian 872.

Perangkat soal 882 adalah perangkat soal yang paling baik karena internal konsistensinya tinggi (0,816) dan sebanyak 24 butir soal dari 30 soal yang diujikan mempunyai daya beda soal baik dan sangat baik. Hal ini juga didukung oleh data distribusi nilai mahasiswa, yaitu sejumlah 11,8% mahasiswa dapat mencapai nilai di atas D, bahkan 6,4% di antaranya memperoleh nilai A. Namun, jumlah mahasiswa yang

tidak lulus sangat tinggi, yaitu sebesar 72,2%. Dalam hal ini penulis soal perlu meninjau kembali perangkat soal tersebut. Meskipun perangkat soal sudah memenuhi syarat baik dari segi materi maupun reliabilitas soalnya, tetapi ada kemungkinan waktu yang disediakan terlalu singkat. Oleh karena jam ujian UT sudah ditentukan selama 90 menit per matakuliah, maka perlu dipertimbangkan perlunya mengurangi jumlah soal ujian atau menyederhanakan soal sehingga mahasiswa dapat menyelesaikan soal dengan lebih cepat.

## BAB V

### KESIMPULAN DAN SARAN

#### Kesimpulan

Berdasarkan hasil-hasil penelitian yang ditemukan pada penelitian ini dapat disimpulkan bahwa:

1. Kisi-kisi soal ujian standar tidak disusun berdasarkan TIK, pokok bahasan ataupun kegiatan belajar dari setiap BMP yang diwakili oleh soal ujian. Kisi-kisi tersebut hanya menunjukkan jumlah soal setiap BMP dan jenjang kemampuan yang diukur oleh masing-masing butir soal.
2. Ruang lingkup materi soal yang harus diujikan berdasarkan kisi-kisi standar ditekankan pada BMP VI, VII, dan VIII.
3. Penyusunan soal ujian untuk masa ujian 872, 881, dan 882 tidak mengacu pada kisi-kisi soal ujian standar baik ditinjau dari materi modul yang dicakup (nomor BMP yang diwakili), jumlah soal yang mewakili setiap modul yang bersangkutan, prosentase jenjang kemampuan yang diukur maupun dari jumlah total butir soalnya.
4. Ketiga perangkat soal tidak menunjukkan kesetaraan satu sama lain, baik dari segi nomor BMP yang diwakili maupun dari persentase soal per BMP dan prosentase jenjang kemampuan yang diukur.
5. Perangkat soal 872 paling mudah dibandingkan perangkat soal 881 dan 882. Tetapi berdasarkan daya beda soal dan reliabilitas soalnya, perangkat soal 882 adalah yang

terbaik.

6. Hasil pengukuran ketiga perangkat soal menunjukkan bahwa ketiga perangkat soal terlalu sulit bagi mahasiswa.

### Saran

Berdasarkan hasil penelitian yang diperoleh peneliti mengusulkan saran-saran sebagai berikut:

1. Penulisan kisi-kisi soal ujian untuk matakuliah Metode Peramalan sebaiknya tidak perlu didasarkan pada TIK, tetapi cukup didasarkan pada pokok bahasan/kegiatan belajar. Kisi-kisi soal ujian standar yang ada sekarang perlu ditinjau kembali dengan memperjelas pokok bahasan/kegiatan belajar yang harus diujikan materi BMP-nya dan prosentase jumlah butir soalnya.
2. Penyusunan perangkat ujian hendaknya mengacu pada kisi-kisi soal standar.
3. Ketiga perangkat soal perlu ditelaah kembali untuk mengetahui apakah ada kesalahan soal ujian, kesalahan kunci soal dan untuk memastikan bahwa tingkat kesukaran soal-soal yang diujikan sesuai dengan waktu ujian 90 menit.
4. Penulisan soal hendaknya mempertimbangkan waktu ujian yang diperlukan. Oleh karena waktu yang disediakan untuk mengerjakan ujian hanya 90 menit, maka sebaiknya soal-soal yang sulit disederhanakan dengan jalan mengurangi jumlah data yang harus dihitung mahasiswa dalam menjawab soal sehingga mahasiswa dapat lebih cepat dalam

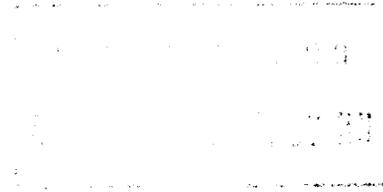
mengerjakan soal. Dalam hal ini yang diutamakan adalah mengukur kemampuan mahasiswa dalam menerapkan pengetahuan/rumus yang tepat untuk setiap soal, dan bukan mengutamakan mengukur ketelitian mahasiswa dalam menghitung.

5. Perlu dilakukan penelaahan terhadap Buku Materi Pokok Metode Peramalan sebagai bahan belajar mahasiswa. Ada kemungkinan materi modul tersebut terlalu sulit bagi mahasiswa. Misalnya, mahasiswa memerlukan lebih banyak contoh-contoh penyelesaian soal --baik untuk contoh soal maupun untuk soal-soal tes formatif--. Dalam hal ini perlu diteliti apakah materi soal yang diujikan pada setiap masa ujian cukup jelas dibahas dalam modul yang bersangkutan.



## DAFTAR PUSTAKA

- Ebel, R.L. dan Frisbie, D.A. (1986). Essentials of Educational Measurements. (4<sup>th</sup> Ed.). New Jersey : Prentice-Hall, Inc.
- Fernandes, H.J.X. (1984). Testing and Measurement. INS/78/030 - National Educational Planning, Evaluation, and Curriculum Development. Jakarta : Badan Penelitian dan Pengembangan Pendidikan dan Kebudayaan.
- Mehrens, W.A. dan Lehmann, I.J. (1984). Measurement and Evaluation in Education and Psychology. (3<sup>rd</sup> Ed.). New York : Holt-Saunders International.
- Nugraheni, E. (1987). Measuring Student Achievement At Universitas Terbuka (A Master Degree Project). Vancouver, B.C.: Simon Fraser University.
- Popham, W.J. (1981). Modern Educational Measurement. London : Prentice-Hall, Inc.
- Sudirman dkk. (1987). Ilmu Pendidikan. Bandung : Penerbit Remaja Karya CV.



UNIVERSITAS TERBUKA