

TINGKAT KESUKARAN, DAYABEDA, KONSISTENSI INTERNAL BUTIR DITINJAU PENGUJIAN SEREMPAK-PARSIAL DAN KONTEN UJI PADA TES SELEKSI MASUK IKIP MATARAM TAHUN 2011/2012

Jumailiyah

Dosen FIP IKIP Mataram

E-mai: -

ABSTRAK: Tes seleksi masuk perguruan tinggi sebagai instrumen yang penting dalam rangka pengambilan keputusan yaitu memilih calon yang cocok, membutuhkan tes berkualitas dari segi tingkat kesukaran, dayabeda, dan konsistensi internal butir. Sampel peserta tes jurusan IPA sebanyak 1056, tes terdiri dari 60 butir, masing-masing 15 butir pada tes Matematika, Biologi, Kimia dan tes Fisika. Nilai butir dihitung dengan perangkat lunak ITEMAN dan pengujian data untuk penarikan kesimpulan dengan bantuan SPSS ver.15. Pengujian perbedaan dayabeda dan konsistensi internal yang dianalisis secara serempak dan secara parsial diuji dengan teknik uji T, sedangkan perbedaan tingkat kesukaran butir, dayabeda, dan konsistensi internal pada tes Biologi, Matematika, Kimia, dan tes Fisika diuji dengan analisis varians satu jalur "ANOVA" Disimpulkan: (1) Dayabeda butir tes uji parsial lebih tinggi daripada dayabeda uji serempak. (2).Konsistensi internal butir uji parsial lebih tinggi daripada konsistensi internal butir uji serempak.(3). Tidak ada perbedaan tingkat kesukaran butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika (4). Tidak ada perbedaan dayabeda butir soal pada keempat kelompok tes atau dayabeda butir soal pada tes Biologi, tes Matematika, tes Kimia, dan tes Fisika (5). Ada perbedaan konsistensi internal butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika.

Kata kunci: *Tingkat kesukaran, Dayabeda, Konsistensi internal butir, Uji serempak, Uji parsial.*

PENDAHULUAN

Fungsi pengukuran pendidikan seleksi dalam pengukuran pendidikan bertujuan untuk memilih peserta yang sesuai dengan kriteria tertentu yang di telah ditetapkan sebelumnya, demikian pula tes seleksi masuk perguruan tinggi merupakan masalah penting karena berkaitan dengan isu-isu pokok dalam penyelenggaraan pendidikan seperti pemerataan pendidikan, mutu lulusan, efisiensi penyelenggaraan pendidikan. Jika calon mahasiswa yang terjaring lulus seleksi yaitu mereka yang berkemampuan rendah maka dalam proses kuliah nanti tidak dapat menyelesaikan studinya dengan cepat, bahkan mungkin mengalami drop out. Masalah pemerataan pendidikan tampak bahwa calon mahasiswa diberikan kesempatan untuk mengikuti tes tanpa batas wilayah dengan tidak mempersoalkan asal SMA, jurusan di SMA. Jurusan ilmu pengetahuan alam (IPA) mempunyai hak yang sama dengan jurusan ilmu pengetahuan sosial dan atau bahasa dalam menentukan pilihannya di perguruan tinggi kelanjutan. Kenyataan pendidikan terkait dengan jurusan, peminatan dan kemampuan khusus siswa "terabaikan" sebagaimana ditunjukkan oleh kebebasan siswa untuk memilih bidang studi kelanjutan. Terabaikannya linieritas jurusan di SMA/MA kemungkinan akan menjadi sumber masalah

yang timbul dalam tes seleksi masuk perguruan tinggi, terutama pada kualitas tes.

Dalam teori pengukuran psikologi dan pendidikan ada dua pendekatan yaitu teori pengukuran klasik dan teori pengukuran modern. Khusus dalam teori klasik, tes yang bertujuan untuk memilih "bibit unggul" diperlukan tes berkualitas, seperti tampak pada butir-butir soal dengan tingkat kesukaran, daya pembeda, dan validitas butir atau konsistensi internal butir yang memadai. Naga mengemukakan "tiga ciri butir berupa (1) taraf kesukaran butir, (2) daya pembeda butir, dan (3) tingkat konsistensi butir. Untuk mengetahui ciri butir dikenal sebagai analisis butir, perhitungan mencakup proporsi atau probabilitas, analisis angka indeks, serta ada pula angka yang berbentuk koefisien (Naga, 1992: 50).

IKIP Mataram salah satu perguruan tinggi swasta di Nusa Tenggara Barat yang menyeleksi calon mahasiswa baru untuk menentukan mana diantara peserta tes yang memadai untuk dapat diterima sebagai mahasiswa. Khusus untuk calon mahasiswa yang akan memasuki jurusan-jurusan atau program studi pendidikan: Biologi, Matematika, Kimia, Fisika dan Pendidikan Olahraga dan Kesehatan diberikan soal-soal tes IPA. Selanjutnya akan timbul pertanyaan yang membutuhkan kajian lebih lanjut yang terkait

dengan kualitas tes dalam tes seleksi mahasiswa baru yaitu tingkat kesukaran, dayabeda dan konsistensi butir soal. (1). Apakah ada perbedaan ciri-ciri butir jika dianalisis secara terpisah atau secara serempak? Pertanyaan ini mengacu pada banyak butir yang dianalisis. (2). Apakah ada perbedaan ciri butir pada tes Biologi, Matematika, Kimia dan Fisika?

KERANGKA TEORI

Tes sebagai alat seleksi untuk memilih calon yang berkualitas, sebagai alat untuk pengambilan keputusan, apakah peserta tes termasuk potensial atau tidak, berkelayakan atau tidak. Dalam kajian ini ciri butir dimaksudkan sebagaimana dikemukakan di atas yaitu: taraf kesukaran butir, daya pembeda butir, dan tingkat konsistensi butir.

Tingkat kesukaran butir mengandung dua pengertian yang berlawanan "suatu persentase peserta ujian yang menjawab salah, semakin besar indeks kesulitan menunjukkan item sulit. Pengertian indeks kesulitan menunjukkan persentase kelompok yang menjawab item dengan benar, dengan demikian semakin besar nilai indeks semakin menunjukkan butir yang mudah". (Robert L Ebel & David, 1991: 228). Pengertian pertama mengacu pada tingkat kesulitan sedangkan pengertian kedua mengacu pada tingkat kemudahan. Dalam penelitian ini analisis butir dilakukan dengan bantuan komputer yang menunjukkan persentase peserta menjawab benar untuk setiap butir yang diperhitungkan sehingga tingkat kesulitan mengacu pada pengertian tingkat kemudahan.

Daya beda butir menunjukkan kemampuan butir untuk menunjukkan perbedaan peserta tes yang berhasil dengan peserta tes yang gagal. (Aiken Lewis R, 2006 : 69) "daya pembeda butir adalah kemampuan butir soal dalam membedakan peserta yang berkemampuan tinggi dengan peserta yang berkemampuan rendah. Makin tinggi daya pembeda butir makin besar perbedaan skor yang dihasilkan oleh kelompok berkemampuan tinggi dengan kelompok yang berkemampuan rendah". Dari pengertian daya pembeda ini, analisis dilakukan melalui pengelompokan peserta berhasil dengan peserta yang gagal.

Konsistensi internal mencerminkan reliabilitas butir tes, analisis dilakukan melalui prosedur ujian dilancarkan satu kali, setiap butir dikorelasikan dengan skor komposit. Pada analisis butir, jika butir tes terdiri dari 60 butir ujian, akan diperoleh 60 koefisien korelasi. Dengan demikian konsistensi internal butir

diketahui dengan mengkorelasikan skor tiap butir dengan skor total ujian. Azwar Saifuddin mengemukakan bahwa "konsistensi internal prosedurnya hanya memerlukan satu kali pengenaan sebuah tes kepada sekelompok individu sebagai subjek, karena itu pendekatan ini mempunyai nilai praktis dan efisiensi yang tinggi". (Azwar Saifuddin, 2004 : 63)

Pengukuran pendidikan yang berbentuk tes, setiap peserta akan menjawab tes secara keseluruhan tes (total) butir ujian. Jika tes mengukur dimensi yang sama, maka ini berarti bahwa N butir ujian itu mengukur hal yang sama dari para peserta ujian. Karena N butir ujian mengukur hal yang sama dari para peserta ujian, maka skor setiap butir pada peserta yang sama harus kira-kira sama pula. Dengan kata lain, kalau satu peserta kurang mampu menjawab satu butir ujian, maka seharusnya ia kurang mampu menjawab N butir ujian. Dan kalau peserta lain cukup mampu menjawab N butir ujian, maka seharusnya ia cukup mampu menjawab N butir ujian, (Naga,S Dali, 1992 : 74).

Selanjutnya dalam analisis butir ciri butir dapat dipengaruhi oleh beberapa faktor antara lain: Faktor pertama, faktor kemampuan peserta tes, dalam teori klasik ciri-ciri butir tes sangat tergantung pada ciri kelompok peserta tes atau sampel uji. Jika peserta tes memiliki kemampuan yang berbeda maka nilai ciri butir itu akan berbeda atau berubah sesuai dengan perbedaan atau perubahan kemampuan peserta tes. Misalnya soal yang sama diberikan kepada peserta ujian masuk tahun 2011 cenderung diikuti oleh peserta tes yang berkemampuan tinggi, dan jika calon mahasiswa tahun 2012 cenderung berkemampuan sedang atau kurang, maka ciri butir tahun 2012 diperoleh angka-angka yang lebih tinggi daripada angka yang diperoleh pada tahun 2011. Akan tetapi jika kelompok calon mahasiswa peserta tes tahun 2011 memiliki kemampuan yang sama dengan peserta tahun 2012, maka angka-angka ciri butir akan sama pula. Walaupun ini dianggap faktor dominan dalam mempengaruhi tingkat kesukaran, daya beda dan konsistensi internal butir, akan tetapi keterbatasan penelitian yang dijadikan dasar tulisan ini yang hanya meneliti satu tes seleksi masuk yaitu tahun 2011.

Faktor kedua yaitu panjang tes dapat mempengaruhi nilai tingkat kesukaran, daya beda dan koefisien reliabilitas tes. Dalam beberapa literatur dicontohkan pada ciri butir dengan konsistensi internal butir, yang sebenarnya panjang tes akan mempengaruhi ciri butir yang lain. "Tes yang panjang cenderung menghasilkan reliabilitas lebih tinggi,

banyak butir tes yang dianalisis (panjang tes) mempengaruhi konsistensi internal butir, semakin panjang tes semakin reliabel tes semakin banyak butir yang terkandung dalam suatu tes cenderung reliabilitas semakin tinggi" (Gronlund, Norman E, Robert Linn, 1990 : 93). Adapun kriteria ujian pendek atau panjang dikemukakan (Crocker Linda & James Algina, 1986 : 146), ujian dikategorikan pendek jika tes terdiri dari 10 – 75 butir uji tes, berarti ujian yang lebih dari 75 butir termasuk ujian panjang. Kembali ke definisi tingkat kesulitan butir yaitu "indeks kesulitan menunjukkan persentase kelompok yang menjawab item dengan benar, dengan demikian semakin besar nilai indeks semakin menunjukkan butir yang mudah". (Robert L Ebel & David, 1991: 228). Tampak bahwa tingkat kesukaran butir tidak terkait berdasarkan panjang dan pendeknya tes. Setiap perhitungan tingkat kesukaran butir tes akan berdiri sendiri untuk setiap butir, tidak akan terkait dengan banyaknya butir tes yang dijawab atau banyak butir yang dianalisis.

Faktor ketiga homogenitas materi atau konten soal. Jika soal yang dijawab peserta tes terdiri dari materi soal yang sama akan berbeda dengan materi soal bervariasi. "Pengestimasian reliabilitas ujian yang baik dengan menggunakan ujian dalam satu trait, homogen dalam konten ujian" (Yen, Wendy , Mery J Allen, 1979 : 88), atau dikatakan bahwa "butir-butir dalam satu perangkat ujian berdimensi sama atau satu dimensi ukur" (Naga, Dali S, 1992 : 74). Dua pendapat ini mempersyaratkan bahwa butir soal yang homogen dalam isi, artinya butir-butir soal mengungkapkan materi soal yang sama, jika banyak materi yang diungkapkan berarti butir bersifat heterogen atau multidimensi akan mempengaruhi ciri butir. Khusus terkait dengan ciri butir ketiga yaitu konsistensi internal, konsistensi internal butir akan tinggi jika butir-butir ujian mengandung materi/konten yang sama atau disebut unidimensi. (Crocker Linda & James Algina, 1986 : 135). Satu perangkat tes yang terdiri dari berbagai materi uji dimana peserta ujian menjawab bermacam-macam materi pertanyaan yang berakibat pada tingkat kesukaran, daya beda, dan konsistensi internal tes.

METODE

1. Sumber data

Hasil ujian masuk calon mahasiswa IKIP Mataram tahun 2011, khususnya peserta tes yang terdaftar program studi Ilmu Pengetahuan Alam, berjumlah 1056

orang calon mahasiswa, yang diteliti melalui lembar jawaban mereka saat mengikuti tes seleksi masuk IKIP Mataram tahun 2011. Jumlah sampel 1056 orang ini ditetapkan setelah dilakukan pembersihan data melalui peserta tes dikeluarkan sebagai sampel yaitu mereka yang menjawab minimal 20 % soal.

2. Instrumen tes.

Tes seleksi masuk IKIP Mataram tahun 2011 berbentuk tes pilihan ganda dengan lima option yaitu respon A, B, C, D, dan E. Dilengkapi dengan kunci jawaban. Terdiri dari empat bidang/materi soal yaitu Biologi, Matematika, Kimia dan Fisika, masing-masing terdiri dari 15 butir soal, dengan demikian jumlah keseluruhan tes seleksi tahun 2011 sebanyak 60 butir.

Ukur dan butir tes yang pendek, sedangkan penganalisaan secara keseluruhan yaitu 60 butir mencerminkan butir tes panjang dengan multidimensi ukur.

3. Prosedur dan Desain Penelitian

Unit analisis dalam penelitian melalui perhitungan: (1) Tingkat kesukaran butir, proporsi jawaban benar setiap butir (2) Daya beda butir ditunjukkan koefisien korelasi biserial, (3) Konsistensi internal butir yaitu koefisien korelasi dengan korelasi biserial titik (Kuder-Richardson 20) karena data yang dianalisis dikotomi yaitu korelasi skor butir dengan skor butir total. Perhitungan ketiga ciri butir ini dianalisis dengan bantuan ITEMAN versi 3.5, dari MICROCAT (tm) System, program ini menghasilkan (output): tingkat kesukaran butir yang dianalisis dari proporsi jawaban betul, daya beda butir, konsistensi internal butir.

4. Prosedur pertama secara serempak dan parsial

analisis data respons keseluruhan sampel 1056 dengan jawaban mereka pada 60 butir tes (butir total) secara serempak semua butir, analisis selanjutnya berdasarkan konten tes yaitu: tes Biologi, Matematika, Kimia dan Fisika atau tes terdiri dari empat konten dilakukan analisis secara terpisah atau parsial. Pada tahap ini, out put menghasilkan tingkat kesukaran, daya beda, dan konsistensi internal butir. Respons jawaban 60 butir ini terdiri dari respons jawaban Biologi (15) butir, Matematika (15) butir, Kimia (15) butir, Fisika (15) butir. Prosedur pertama analisis data dilakukan dengan menganalisis daya beda dan konsistensi internal butir berdasarkan pengujian serempak pada

seluruh butir soal dan berdasarkan analisis parsial. Adapun tingkat kesukaran butir tidak dianalisis lebih lanjut karena ciri butir ini akan memiliki angka tingkat kesukaran butir yang sama pada analisis serempak dan analisis parsial. Untuk memperoleh angka tingkat kesukaran butir, setiap butir berdiri sendiri dan tidak berpengaruh pada banyak butir yang dianalisis sehingga pada saat dianalisis 60 butir akan sama dengan 4×15 . Selanjutnya pertanyaan penelitian yaitu: (a). Apakah ada perbedaan dayabeda butir tes yang dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012. (b) Apakah ada perbedaan konsistensi internal butir dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012.

5. Prosedur analisis kedua

analisis data respons sampel 1056 terdiri dari respons jawaban Biologi (15) butir, Matematika (15) butir, Kimia (15) butir, Fisika (15) butir, masing-masing tes ini dianalisis menurut bidang, konten tes sehingga mencerminkan konten unidimensi dan tes pendek. Analisis tahap ini menghasilkan tingkat kesukaran, daya beda dan konsistensi internal pada masing-masing bidang uji. Dari segi konten butir tes setiap bidang tes dianalisis secara sendiri dan mencerminkan satu dimensi atau unidimensi yaitu tes Biologi, Matematika, Kimia, dan tes Fisika. Analisis tahap ini akan menjawab pertanyaan: (a) Apakah ada perbedaan tingkat kesukaran butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012, (b) Apakah ada perbedaan dayabeda butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012, (c) Apakah ada perbedaan konsistensi internal butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012.

Teknik Analisis: Analisis data dilakukan dengan menghitung tingkat kesukaran butir, dayabeda butir dengan teknik korelasi biserial dan konsistensi internal korelasi point-biserial atau KR-20 dengan ketentuan-ketentuan yang diperlukan dalam analisis ITEMAN versi 3,5 maka program akan bekerja dan menghasilkan output tingkat kesukaran soal, daya beda butir atau koefisien korelasi

biserial sedangkan hasil korelasi biserial titik mencerminkan konsistensi internal butir.

Analisis data dalam rangka penarikan kesimpulan melalui teknik uji T menguji dua rata-rata hitung, rata-rata hitung butir analisis serempak dan rata-rata hitung analisis parsial, yaitu: (a). Perbedaan dayabeda butir tes yang dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012. (b) Perbedaan konsistensi internal butir dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012.

Selanjutnya dilakukan perhitungan dengan uji F (analisis varian) satu jalur, menguji perbedaan empat rata-rata hitung, yaitu rata-rata hitung tes Biologi, Matematika, Kimia, dan tes Fisika. Secara lebih rinci analisis varian satu jalur, digunakan untuk menguji: (1). Perbedaan tingkat kesukaran butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012, (2) Perbedaan dayabeda butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012, (3) Perbedaan konsistensi internal butir tes Biologi, Matematika, Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012.

Perhitungan untuk pengujian hipotesis dalam rangka pengambilan keputusan dan kesimpulan penelitian melalui bantuan program perangkat lunak SPSS versi 15.

Analisis data dalam rangka pengujian hipotesis dilakukan melakukan uji T yaitu menguji perbedaan dua rata-rata hitung, mean pada dua kelompok. Dalam penelitian ini perbedaan ciri butir pada kelompok uji serempak dan uji parsial. Selanjutnya dilakukan analisis uji F diterapkan untuk menguji rata-rata hitung pada empat kelompok. Pengujian melalui uji F dilakukan melalui uji asumsi terlebih dahulu yaitu uji homogenitas varian dengan *Levene Statistic* "analisis bertujuan untuk menguji berlaku tidaknya asumsi untuk ANOVA, yaitu apakah keempat sampel mempunyai varians yang sama ...jika probabilitas $> 0,05$ maka asumsi kesamaan varians untuk uji ANOVA sudah terpenuhi" (Singgih Santoso, 2008: 243-244).

HASIL DAN PEMBAHASAN

Sesuai dengan pembahasan di atas prosedur tahap pertama mencerminkan analisis pengujian yang membutuhkan analisis uji T karena pengujian ini menuntut signifikansi perbedaan dua rata-rata hitung.

Pertama: menguji perbedaan dayabeda butir tes yang dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012. Banyak butir tes yang dianalisis 60 butir atau $N=60$ Hasil output SPSS menunjukkan rata-rata hitung dayabeda uji parsial = 0,2417 dan rata-rata hitung dengan uji serempak = 0,1640 kedua jenis uji parsial dan serempak berkorelasi $r= 0,438$ dan signifikansi 0,000. Ini menunjukkan bahwa dayabeda butir uji parsial atau uji berdasarkan kesamaan konten memiliki dayabeda yang lebih tinggi daripada pengujian serempak, atau diuji berdasarkan konten soal bervariasi atau multidimensi. Perbedaan kedua rata-rata hitung dimaksud diperoleh nilai $T= 6,256$ dengan signifikansi 0,000. Dengan demikian secara meyakinkan bahwa dayabeda uji parsial lebih tinggi daripada dayabeda uji serempak.

Secara teoretis semakin tinggi nilai dayabeda semakin menunjukkan kemampuan butir soal membedakan peserta tes mampu dengan peserta tes tidak mampu. Dari penelitian ini diperoleh bahwa tes yang berisi konten terbatas lebih baik daripada yang kontennya banyak. Perhitungan dayabeda butir dengan bantuan ITEMAN ditunjukkan oleh angka atau koefisien korelasi biserial dilambangkan dengan D . Koefisien korelasi yang diperoleh dari hasil analisis data tampak koefisien yang tinggi kecuali pada ujian panjang multidimensi dengan rata-rata 0,1640 sementara kriteria koefisien nilai D dikemukakan (Crocker Linda & James Algina, 1986 : 315), "jika $D \geq 0,40$ fungsi butir sempurna, $0,30 \leq D \leq 0,39$ tidak perlu direvisi, $0,20 \leq D \leq 0,29$ perlu direvisi, $D \leq 0,19$ dibuang, tidak digunakan". Berdasarkan pendapat ini maka dayabeda pada kedua kelompok yang dianalisis tidak mencerminkan butir tes yang baik.

Kedua: menguji perbedaan konsistensi internal butir dianalisis secara serempak dengan yang dianalisis secara parsial pada tes seleksi masuk IKIP Mataram tahun 2011/2012. Banyak butir tes yang dianalisis 60 butir atau $N=60$ Hasil output SPSS menunjukkan rata-rata hitung konsistensi internal uji parsial = 0,3118 dan rata-rata hitung dengan uji serempak = 0,1993 kedua jenis uji parsial dan serempak berkorelasi $r =$

0,565 dengan signifikansi 0,000 Ini menunjukkan bahwa dayabeda butir uji parsial atau uji berdasarkan kesamaan konten memiliki konsistensi internal yang lebih tinggi daripada pengujian serempak, atau diuji berdasarkan konten soal bervariasi atau multidimensi. Perbedaan kedua rata-rata hitung dimaksud diperoleh nilai $T= 12,756$ dengan signifikansi 0,000. Dengan demikian secara meyakinkan bahwa konsistensi internal butir uji parsial lebih tinggi daripada konsistensi internal butir uji serempak. Butir tes yang diuji berisi satu macam konten memiliki reliabilitas yang lebih baik daripada tes yang berisi konten banyak atau multidimensi.

Estimasi reliabilitas atau konsistensi internal butir dalam analisis butir, ternyata dari hasil analisis data dapat diketahui bahwa estimasi reliabilitas yang paling mantap jika butir dalam suatu ujian sedikit (ujian pendek) dan dari segi konten soal homogen atau unidimensi. Semakin banyak butir akan semakin reliabel, dalam penelitian ini uji serempak dianalisis dari 60 butir, sedangkan uji parsial dianalisis dari 4x15 butir atau konten homogen dan tes pendek. Demikian juga akan tampak bahwa butir soal multidimensi atau heterogen, walaupun butir tes banyak tidak berpengaruh dalam mengestimasi reliabilitas atau konsistensi internal. Dengan demikian pendapat Gronlund tidak terdukung dalam penelitian ini. Hasil penelitian ini juga sama dengan yang telah dilakukan peneliti tahun 2007, data sampel yang digunakan adalah hasil tes masuk IKIP Mataram tahun 2005 tes terdiri dari 130 butir tes diperoleh bahwa konsistensi internal butir dipengaruhi oleh dimensi ukur bukan pada panjang tes. Secara teoretis reliabilitas tes dipengaruhi oleh panjang tes, semakin panjang tes akan semakin menunjukkan tes reliabel, akan tetapi data hasil perhitungan reliabilitas "konsistensi internal" dipengaruhi oleh homogenitas konten tes, dengan demikian membutuhkan penelitian lebih lanjut terkait dengan validitas konstruk tes.

Tahap analisis butir dengan membandingkan ciri butir berdasarkan pengelompokan konten tes. Konten tes dimaksudkan dalam penelitian ini, instrumen tes masuk IKIP Mataram tahun 2011/2012 terdiri dari tes Biologi, Matematika, Kimia dan Fisika, keempat konten tes ini masing-masing terdiri dari 15 butir tes. Perhitungan dan analisis uji hipotesis dilakukan dengan uji F (analisis varian) satu jalur untuk mengetahui perbedaan pada empat kelompok atau empat konten uji Biologi, Matematika, Kimia, dan tes

Fisika. Secara lebih rinci analisis varian satu jalur, digunakan untuk menguji perbedaan:

Ketiga: Tingkat kesukaran butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012. Jumlah responden 1056 N butir 15 pada masing-masing konten tes. Apakah tingkat kesukaran pada keempat kelompok berbeda? Pengujian hipotesis dilakukan setelah menguji kesamaan varian pada keempat kelompok sebagai persyaratan terpenuhinya asumsi yaitu dengan uji *Levence Statistic*, jika nilai *Levence Statistic* > 0,05 maka asumsi terpenuhi dan analisis varians dapat dilanjutkan. Dari hasil output SPSS diperoleh mean tingkat kesukaran butir tes Matematika, tes Biologi, tes Kimia, dan tes Fisika berturut-turut: 0.2027, 0.2100, 0.2107, 0.2140. dan mean Total 0.2093. Nilai *Levence Statistic* 0.396 df 3 dan 56 dengan tingkat signifikansi 0.756, tingkat signifikansi > 0.05 . Berarti asumsi terpenuhi dan analisis varians dapat dilanjutkan. Adapun mean tingkat kesukaran butir pada keempat kelompok diperoleh nilai $F = 0,052$ dengan tingkat signifikansi 0,0984. Dapat disimpulkan bahwa tidak ada perbedaan tingkat kesukaran butir soal pada keempat kelompok tes atau tingkat kesukaran soal pada tes Biologi, tes Matematika, tes Kimia, dan tes Fisika sama.

Beberapa kriteria penarikan kesimpulan penelitian terkait tingkat kesukaran butir soal. Butir baik adalah butir yang tidak terlalu sukar dan tidak juga terlalu mudah dalam arti tingkat kesukaran butir soal sesuai dengan tingkat kemampuan, yaitu butir dengan taraf kesukaran 0,5 atau disekitar 0,5 yaitu $0,33 \leq P \leq 0,67$ (Ebel dan Frisbie, 1991:232). Sementara Nitko mengemukakan kriteria tingkat kesukaran butir, butir sangat sukar bernilai $p > 0.25$ dan butir mudah dengan nilai $p > 80$ (Nitko, 2001:323). Untuk keperluan tertentu, butir dengan tingkat kesulitan bervariasi terkadang dibutuhkan sehingga diperlukan butir dengan berbagai tingkat kesukaran butir dimulai dari tingkat kesukaran rendah sampai tingkat kesukaran tinggi. Bahrul Hayat (Bahrul Hayat, 1997: 18) mengemukakan kualitas butir menurut kriteria tingkat kesukaran butir yaitu $p > 0.70$ soal yang mudah, $0,30 \leq p \leq 0.70$ soal sedang, dan $p < 0.30$ butir soal yang menunjukkan sukar. Dengan memperhatikan tingkat kesukaran butir yang diperoleh dari data empiris penelitian ini maka diketahui soal-soal yang diberikan kepada peserta tes masuk tergolong sulit, sehingga sulit untuk mengestimasi kemampuan peserta tes yang mendekati kenyataan.

Keempat: Dayabeda butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012. Jumlah responden 1056 N butir 15 pada masing-masing konten tes. Apakah dayabeda butir soal pada keempat kelompok berbeda? Pengujian hipotesis dilakukan setelah menguji kesamaan varian pada keempat kelompok sebagai persyaratan terpenuhinya asumsi yaitu dengan uji *Levence Statistic*, jika nilai *Levence Statistic* > 0,05 maka asumsi terpenuhi dan analisis varians dapat dilanjutkan. Dari hasil output SPSS diperoleh mean tingkat kesukaran butir tes Matematika, tes Biologi, tes Kimia, dan tes Fisika berturut-turut: 0.2180, 0.2500, 0.2520, 0.2476. dan mean Total 0.2417. Nilai *Levence Statistic* 0.354 df 3 dan 56 dengan tingkat signifikansi 0.784, tingkat signifikansi > 0.05 . Berarti asumsi terpenuhi dan analisis varians dapat dilanjutkan. Adapun mean tingkat kesukaran butir pada keempat kelompok diperoleh nilai $F = 0,515$ dengan tingkat signifikansi 0,674. Dapat disimpulkan bahwa tidak ada perbedaan dayabeda butir soal pada keempat kelompok tes atau dayabeda butir soal pada tes Biologi, tes Matematika, tes Kimia, dan tes Fisika adalah sama.

Kelima: Konsistensi Internal butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika pada tes seleksi masuk IKIP Mataram tahun 2011/2012. Jumlah responden 1056 N butir 15 pada masing-masing konten tes. Apakah dayabeda butir soal pada keempat kelompok berbeda? Pengujian hipotesis dilakukan setelah menguji kesamaan varian pada keempat kelompok sebagai persyaratan terpenuhinya asumsi yaitu dengan uji *Levence Statistic*, jika nilai *Levence Statistic* > 0,05 maka asumsi terpenuhi dan analisis varians dapat dilanjutkan. Dari hasil output SPSS diperoleh mean tingkat kesukaran butir tes Matematika, tes Biologi, tes Kimia, dan tes Fisika berturut-turut: 0.2633, 0.3353, 0.3247, 0.3240. dan mean Total 0.3118. Nilai *Levence Statistic* 0.360 df 3 dan 56 dengan tingkat signifikansi 0.782, tingkat signifikansi > 0.05 . Berarti asumsi terpenuhi dan analisis varians dapat dilanjutkan. Adapun mean tingkat kesukaran butir pada keempat kelompok diperoleh nilai $F = 4,600$ dengan tingkat signifikansi 0,006. Dapat disimpulkan bahwa ada perbedaan konsistensi internal butir soal pada keempat kelompok tes atau konsistensi internal butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika adalah berbeda.

SIMPULAN

Berdasarkan analisis data di atas maka dapat disimpulkan: (1) Dayabeda butir tes uji parsial lebih tinggi daripada dayabeda uji serempak. (2). Konsistensi internal butir uji parsial lebih tinggi daripada konsistensi internal butir uji serempak. (3). Tidak ada perbedaan tingkat kesukaran butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika (4). Tidak ada perbedaan dayabeda butir soal pada keempat kelompok tes atau dayabeda butir soal pada tes Biologi, tes Matematika, tes Kimia, dan tes Fisika (5). Ada perbedaan konsistensi internal butir tes Biologi, tes Matematika, tes Kimia, dan tes Fisika.

SARAN

Saran tindak: bagi para pengembang ujian melakukan penelitian yang lebih cermat, simpulan penelitian ini tidak sejalan dengan dasar teoretis mengkaji konsistensi internal butir pada ujian panjang (uji serempak) memiliki konsistensi internal yang lebih rendah daripada ujian pendek (uji parsial) sehingga diharapkan untuk meneliti masalah yang sama dengan menggunakan instrumen dan sampel yang lain, penelitian yang dapat mengungkap validitas konstruk, dan penelitian yang dapat mengungkapkan faktor-faktor yang mempengaruhi ciri butir yang lain, bentuk distribusi kemampuan peserta ujian.

DAFTAR RUJUKAN

- Aiken Lewis R., Gary Groth-Marnat, *Psychological Testing and Assessment*, Pearson Education Group, Boston, 2006.
- Anne, Anastasi, Susana Urbina, *Psychological Testing*, jilid 1 dan 2, alih bahasa Robertus Hariono S.Iman, Jakarta: Prehalindo, 1998.
- Anthony J Nitko, *Educational Assessment of Students*, New Jersey: Merrill Prentice Hall, 2001.
- Azwar, Saifuddin, Reliabilitas dan Validitas, Pustaka Pelajar, Yogyakarta, 2003
- Cohen, Ronald Jay, Mark E. Swedlik, *Psychological Testing and Assessment, An Introduction to Tests and Measurement*, McGraw-Hill Companies, New York, 2005.
- Jumailiyah, "Pengaruh Banyak Butir dan Dimensi Ukur terhadap Konsistensi Internal: Studi Pada Tes Masuk Mahasiswa IKIP Mataram", Jurnal Kependidikan, Vol 8, No.1,
- Kerlinger, Fred N, *Foundation of Behavioral Research*, alih bahasa Asas-Asas Penelitian Behavioral, Landung, R Simatupang, H.J Koesoemanto, Gajah Mada University Press, 2003.
- Linda Crocker dan James Algina, *Introduction to Classical and Modern Test Theory*, (Fort Worth: Holt, Rinehart dan Winston, Inc, 1986)
- Naga Dali S, *Pengantar Teori Sekor pada Pengukuran Pendidikan*, Jakarta: Gunadarma, 1992.
- Norman E Gronlund dan Robert Linn, *Measurement and Evaluation in Teaching*, New York: MacMillan Publishing Company, 1990.
- Robert L.Ebel dan David A Frisbie, *Essentials of Educational Measurement*, New Jersey: Prentice Hall, 1991.
- Singgih Santoso, *Panduan Lengkap Mengusai SPSS 16*, PT Elex Media Komputindo Kelompok Gramedia, Jakarta, 2008.