

EVALUASI KUALITAS TES PSIKOLOGI KEPERIBADIAN I

Muhammad Wahyu Kuncoro

Fakultas Psikologi Universitas Mercu Buana Yogyakarta
(email : wahyu_umby@yahoo.co.id)

Abstract

This study aims to determine how well the items of the exam in the personality psychology test. *In this study, the statistical methods were used to identify the item difficulty, which is a measure of the proportion of examinees who responded to an item correctly, and the item discrimination, which is a measure of how well the item discriminates between examinees. An additional analysis that is the distractor analysis. The distractor analysis provides a measure of how well each of the incorrect options contributes to the quality of a multiple choice item.*

By using The IteMan Program showed that the 60 analyzed items were well enough and can be used. A total of 46 items has an Easy and Medium of The item difficulty index. In addition there are about 44 items with good and very good of the item discrimination index and about 57 points aitem already possess the characteristics of a good. The reliability coefficient alpha of this personality test is 0.898, so it is considered to have good reliability.

Pendahuluan

Masalah pendidikan merupakan suatu masalah yang sangat penting, karena melalui pendidikan kita akan mendapatkan insan-insan yang berkualitas bagi pembangunan bangsa. Tingkat pendidikan suatu negara akan menentukan kemajuan suatu bangsa.

Stiggins (1994) mengemukakan bahwa pendidikan sangat berperan dalam membentuk masa depan individu dan masyarakat, baik dalam bidang ilmu pengetahuan, bidang seni, bidang olah raga maupun bidang lain.

Kualitas pendidikan ditunjukkan oleh peningkatan kualitas proses dan hasil pendidikan itu sendiri. Proses pendidikan yang berkualitas akan menghasilkan tenaga terampil dan ahli yang sangat dibutuhkan

dalam pembangunan yang terus berkembang.

Prestasi belajar sebagai salah satu tolok ukur peningkatan mutu pendidikan, termasuk pendidikan tinggi. Para pengelola pendidikan telah melakukan berbagai usaha untuk memperbaiki dan meningkatkan prestasi belajar mahasiswa. Usaha-usaha tersebut antara lain : melalui perbaikan kurikulum, penyetaraan kualitas pengajar, penambahan fasilitas pelajaran dan lain-lain (Rustaman, 2003).

Menurut Azwar (2002) prestasi belajar merupakan indikator utama dari proses belajar, sebagaimana dari nilai yang diperoleh. Demikian juga Masrun dan Martaniah (1976) menyatakan bahwa prestasi belajar dipakai sebagai ukuran untuk mengetahui hasil kegiatan belajar,

yaitu sejauhmana siswa dapat menguasai bahan pelajaran yang telah diajarkan.

Evaluasi merupakan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja atau produktivitas suatu lembaga dalam melaksanakan programnya Mardapi (2008). Oleh karena itu, evaluasi merupakan salah satu subsistem yang penting dalam sistem pendidikan. Dalam Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, evaluasi diatur dalam Bab XVI Pasal 57,58, dan 59. Pelaksanaan evaluasi bertujuan untuk mengukur dan mengendalikan mutu pendidikan.

Bentuk tes yang digunakan diantaranya berupa tes tertulis (*paper and pencil test*). Tes tertulis merupakan teknik penilaian yang seringkali digunakan untuk menilai prestasi belajar siswa. Melalui tes prestasi belajar, dapat diperoleh informasi yang dapat menggambarkan kemampuan siswa (Stiggins, 1994)). Oleh karena itu, pengelolaan ujian dan mutu bahan ujian yang digunakan perlu mendapat perhatian agar hasil tes dapat mencerminkan kemampuan siswa yang sebenarnya.

Secara sederhana Allen & Yen (1979) menyebut tes sebagai perangkat untuk memperoleh sampel suatu perilaku individu. Ahli pengukuran yang lain, Djaali (2006) menyatakan tes adalah suatu cara atau alat untuk mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang harus dikerjakan oleh siswa atau sekelompok siswa sehingga

menghasilkan nilai tentang tingkah laku atau prestasi siswa sebagai peserta didik.

Pada saat ini peneliti adalah dosen pengampu mata kuliah psikologi kepribadian, sebagai salah satu mata kuliah wajib dalam kurikulum program studi psikologi. Selama ini peneliti telah menggunakan tes pilihan ganda sebagai bahan ujian mid dan akhir pada mata kuliah psikologi kepribadian.

Butir-butir soal yang disusun lebih mendasarkan pada pengungkapan kompetensi mahasiswa berdasarkan Satuan Acara Perkuliahan (SAP) yang telah ditetapkan dalam RPKPS.

Sehubungan dengan penyelenggaraan evaluasi dalam pendidikan, maka pengampu mata kuliah (dosen) bertanggungjawab pada kualitas perangkat tes yang digunakan dalam evaluasi, termasuk di dalamnya perangkat tes psikologi kepribadian 1.

Untuk menjamin kualitas tes psikologi kepribadian diperlukan pengembangan bank soal. Bank soal yang biasa dikenal pendidik didefinisikan sebagai kumpulan dari butir-butir tes. Namun bank soal tidak hanya mengacu pada sekumpulan soal-soal saja. Bank soal mengacu pada proses pengumpulan soal-soal, pemantauan dan penyimpanannya dengan informasi yang terkait sehingga mempermudah pengambilannya untuk merakit soal-soal (Thorndike, 1982).

Permasalahan yang diajukan adalah bahwa sampai saat ini peneliti dan sekaligus pengampu mata kuliah psikologi kepribadian 1 belum memiliki bank soal tersebut.

Terkait dengan permasalahan di atas, maka tujuan penelitian ini adalah menguji kualitas tes kepribadian 1 melalui analisis butir soal yang meliputi indeks kesukaran (p), daya beda (d), dan distribusi respons.

Manfaat dari hasil studi ini adalah memberikan informasi empiris guna melakukan revisi terhadap aitem bilamana diperlukan dan guna meningkatkan kualitas tes yang pada gilirannya akan meningkatkan validitas hasil pengukuran tes psikologi kepribadian 1 dan sekaligus untuk menyusun bank soal

Evaluasi Kualitas Tes Psikologi Kepribadian 1

Tes (test) merupakan suatu alat penilaian dalam bentuk tulisan untuk mencatat atau mengamati prestasi siswa yang sejalan dengan target penilaian (Jacobs & Chase, 1992). Jawaban yang diharapkan dalam tes menurut Mardapi (2008) dapat secara tertulis, lisan, atau perbuatan. Menurut Zainul dan Nasution (2001) tes didefinisikan sebagai pertanyaan atau tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang suatu atribut pendidikan atau suatu atribut psikologis tertentu.

Setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar. Dengan demikian apabila suatu tugas atau pertanyaan menuntut harus dikerjakan oleh seseorang, tetapi tidak ada jawaban atau cara pengerjaan yang benar dan salah maka tugas atau pertanyaan tersebut bukanlah tes. Tes merupakan salah satu upaya pengukuran terencana yang digunakan oleh guru untuk mencoba menciptakan kesempatan bagi siswa dalam memperlihatkan prestasi mereka yang berkaitan dengan tujuan yang telah ditentukan (Gronlund, 1981).

Tes terdiri atas sejumlah soal yang harus dikerjakan siswa. Setiap soal dalam tes menghadapkan siswa pada suatu tugas dan menyediakan kondisi bagi siswa untuk menanggapi tugas atau soal tersebut. Tes menurut Arikunto dan Jabar (2004) merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dengan menggunakan cara atau aturan yang telah ditentukan. Dalam hal ini harus dibedakan pengertian antara tes, testing, testee, tester.

Testing adalah saat pada waktu tes tersebut dilaksanakan (saat pengambilan tes). Sementara itu Ebel (1972) menyatakan bahwa testing menunjukkan proses pelaksanaan tes. Testee adalah responden yang mengerjakan tes. Mereka inilah yang akan dinilai atau diukur kemampuannya. Sedangkan Tester adalah seseorang yang

diserahi tugas untuk melaksanakan pengambilan tes kepada responden.

Dewasa ini tes masih merupakan alat evaluasi yang umum digunakan untuk mengukur keberhasilan siswa dalam mencapai tujuan pendidikan dan pengajaran (Subekti & Firman, 1989). Tes yang baik harus memenuhi beberapa kriteria berdasarkan karakteristik butir soal antara lain meliputi indeks kesukaran (p), daya beda (d), dan distribusi respons.

Tes Psikologi Kepribadian 1 adalah sebuah tes prestasi belajar untuk mendapatkan data, yang merupakan informasi untuk melihat seberapa banyak pengetahuan yang telah dimiliki dan dikuasai oleh mahasiswa sebagai akibat dari pendidikan dan pelatihan yang diperoleh di dalam perkuliahan selama setengah atau satu semester.

Berdasarkan uraian tersebut di atas, maka evaluasi kualitas tes psikologi kepribadian 1 adalah upaya peninjauan kualitas butir-butir soal yang mengukur keberhasilan mahasiswa dalam menguasai materi-materi pengetahuan mata kuliah psikologi kepribadian 1, sehingga diperoleh informasi tentang indeks kesukaran (p), daya beda (d), dan distribusi respons setiap butir soal tersebut.

Tes, pengukuran, asesmen dan evaluasi

Rustaman (2003) mengungkapkan bahwa asesmen lebih ditekankan pada penilaian proses. Sementara itu evaluasi lebih ditekankan pada hasil belajar. Apabila

dilihat dari keberpihakannya, menurut Stiggins (1993) asesmen lebih berpihak kepada kepentingan siswa. Siswa dalam hal ini menggunakan hasil asesmen untuk merefleksikan kekuatan, kelemahan, dan perbaikan belajar.

Sementara itu evaluasi menurut Rustaman (2003) lebih berpihak kepada kepentingan evaluator. Yulaelawati (2004) mengungkapkan bahwa terdapat perbedaan antara evaluasi dengan asesmen. Evaluasi (evaluation) merupakan penilaian program pendidikan secara menyeluruh. Evaluasi pendidikan lebih bersifat makro, meluas, dan menyeluruh. Evaluasi program menelaah komponen-komponen yang saling berkaitan tentang perencanaan, pelaksanaan, dan pemantauan. Sementara itu asesmen merupakan penilaian dalam scope yang lebih sempit (lebih mikro) bila dibandingkan dengan evaluasi. Seperti dikemukakan oleh Kumano (2001) asesmen hanya menyangkut kompetensi siswa dan perbaikan program pembelajaran.

Harlen (1982) mengungkapkan perbedaan antara asesmen dan evaluasi dalam hal metode. Evaluasi dinyatakan menggunakan kriteria dan metode yang bervariasi. Asesmen dalam hal ini hanya merupakan salah satu dari metode yang dipilih untuk evaluasi tersebut. Selain dari itu, subyek untuk asesmen hanya siswa, sementara itu subyek evaluasi lebih luas dan beragam seperti siswa, guru, materi, organisasi, dll.

Yulaelawati (2004) menekankan kembali bahwa scope asesmen hanya mencakup kompetensi lulusan dan perbaikan cara belajar siswa. Jadi hubungannya lebih pada peserta didik. Ruang lingkup evaluasi yang lebih luas ditunjukkan dengan cakupannya yang meliputi isi atau substansi, proses pelaksanaan program pendidikan, kompetensi lulusan, pengadaan dan peningkatan tenaga kependidikan, manajemen pendidikan, sarana dan prasarana, dan pembiayaan.

Pengukuran, Tes, dan evaluasi dalam pendidikan berperan dalam seleksi, penempatan, diagnosa, remedial, umpan balik, memotivasi dan membimbing. Baik tes maupun pengukuran keduanya terkait dan menjadi bagian istilah evaluasi. Meski begitu, terdapat perbedaan makna antara mengukur dan mengevaluasi. Mengukur adalah membandingkan sesuatu dengan satu ukuran tertentu. Dengan demikian pengukuran bersifat kuantitatif. Sementara itu evaluasi adalah pengambilan suatu keputusan terhadap sesuatu dengan ukuran baik-buruk. Dengan demikian pengambilan keputusan tersebut lebih bersifat kualitatif (Arikunto,2003; Zainul & Nasution, 2001).

Setiap butir pertanyaan atau tugas dalam tes harus selalu direncanakan dan mempunyai jawaban atau ketentuan yang dianggap benar (Jacobs & Chase, 1992). Sementara itu tugas ataupun pertanyaan dalam kegiatan pengukuran (measurement)

tidak selalu memiliki jawaban atau cara pengerjaan yang benar atau salah karena measurement dapat dilakukan melalui alat ukur non-tes. Maka tugas atau pertanyaan tersebut bukanlah tes. Selain dari itu, tes mengharuskan subyek untuk menjawab atau mengerjakan tugas, sementara itu pengukuran (measurement) tidak selalu menuntut jawaban atau pengerjaan tugas.

Menurut Kumano (2001) mengungkapkan bahwa meskipun terdapat perbedaan makna/pengertian, asesmen dan evaluasi memiliki hubungan. Hubungan antara asesmen dan evaluasi tersebut digambarkan sebagai berikut, “ Evaluasi adalah untuk mengevaluasi data yang diperoleh melalui pengukuran. Pengukuran adalah proses pengumpulan data yang menunjukkan perkembangan dari proses belajar.

Menurut Zainul & Nasution (2001) Hubungan antara tes, pengukuran, dan evaluasi adalah sebagai berikut. Evaluasi belajar baru dapat dilakukan dengan baik dan benar apabila menggunakan informasi yang diperoleh melalui pengukuran yang menggunakan tes sebagai alat ukurnya. Akan tetapi tentu saja tes hanya merupakan salah satu alat ukur yang dapat digunakan karena informasi tentang hasil belajar tersebut dapat pula diperoleh tidak melalui tes, misalnya menggunakan alat ukur non tes seperti observasi, skala rating, dan lain-lain.

Zainul dan Nasution (2001) menyatakan bahwa guru mengukur

berbagai kemampuan siswa. Apabila guru melangkah lebih jauh dalam menginterpretasikan skor sebagai hasil pengukuran tersebut dengan menggunakan standar tertentu untuk menentukan nilai atas dasar pertimbangan tertentu, maka kegiatan guru tersebut telah melangkah lebih jauh menjadi evaluasi.

Tes Pilihan Ganda

Untuk mengukur seberapa jauh tujuan-tujuan pengajaran telah tercapai, dapat dilakukan dengan evaluasi, dalam hal ini evaluasi hasil belajar. Alat ukur untuk mengevaluasi hasil belajar tersebut di gunakan tes. Tes adalah cara (yang dapat dipergunakan) atau prosedur yang (yang perlu ditempuh) dalam rangka pengukuran dan penilaian di bidang pendidikan.

Menurut Suharsimi (2003) tes yang baik harus mempunyai syarat-syarat sebagai berikut: 1) Harus efisien (Parsimony) 2) Harus baku (Standardize) 3) Mempunyai norma 4) Objektif 5) Valid (Sahih) 6) Reliabel (Andal) Oleh sebab itu untuk memperoleh tes yang baik, tes tersebut harus di ujicobakan terlebih dahulu dan hasilnya dianalisis sehingga memenuhi syarat-syarat tersebut di atas.

Salah satu bentuk tes hasil belajar adalah Tes Pilihan Ganda. Tes pilihan ganda adalah bentuk tes obyektif yang mempunyai ciri utama kunci jawaban jelas dan pasti sehingga hasilnya dapat diskor secara obyektif.

Seperti yang dikatakan Suharsimi (2003) bahwa soal pilihan ganda terdiri dari pernyataan dan pertanyaan yang harus dijawab oleh siswa atau melengkapi dengan memilih salah satu dari beberapa alternatif yang tersedia. Satu di antaranya adalah yang paling benar, lainnya disebut pengecoh

Menurut Muhajir (dalam Chabib, 2001) mengatakan bahwa pengertian Tes Pilihan Ganda merupakan tes obyektif dimana masing-masing item disediakan lebih dari dua kemungkinan jawaban, dan hanya satu dari pilihan-pilihan tersebut yang benar atau yang paling benar.

Sedangkan keunggulan tes pilihan ganda menurut Azwar (2005) sebagai berikut : (1) Kompherhensif, karena dalam waktu tes yang singkat dapat memuat lebih banyak item. (2) Pemeriksaan jawaban dan pemberian skornya mudah dan cepat (3) Penggunaan lembar jawaban menjadikan tes efisien dan hemat bahan. (4) Kualitas item dapat dianalisis secara empirik (5) Objektifitasnya tinggi. (6) Umumnya memiliki reabilitas yang memuaskan.

Disamping keunggulan tes pilihan ganda mempunyai kelemahan sebagai berikut : (1) Pembuatannya sulit dan memakan banyak waktu dan tenaga (2) Tidak mudah ditulis untuk mengungkapkan tingkat kompetensi tinggi. (3) Ada kemungkinan jawaban benar semata-mata karena tebakan.

Langkah – langkah Menganalisis Tes

a. Menghitung Indeks kesukaran

Hasil tes setelah diperiksa di beri skor untuk jawaban benar 1 dan untuk jawaban salah 0. Skor yang diperoleh diurut dari skor yang paling tinggi ke skor yang paling rendah serta di bagi 2 menjadi kelompok atas dan kelompok bawah.

Hitung Indeks Kesukaran Butir (IKB) / p dengan formula:

$$IKB = T/R \times 100\%$$

Ket :

IKB atau p = Indeks Kesukaran Item, R = jumlah responden yang menjawab benar, dan

T = jumlah responden seluruhnya.

Indeks Kesukaran Butir (IKB) dapat bernilai 0,00-1,00.

Biasanya kategori Indeks Kesukaran Butir adalah sebagai berikut :

0,00 - 0,20 adalah *sangat sukar*,

0,20 - 0,40 *sukar*,

0,40 - 0,60 *sedang*,

0,60 - 0,80 *mudah*, dan

0,80-1,00 *sangat mudah*.

Biasanya butir yang ditoleransi sebagai tes standar adalah yang memiliki IKB = 0,30-0,70.

Bila ada butir soal yang hampir tidak ada peserta tes yang menjawab benar maka butir soal tersebut dikatakan butir yang sukar, dan sebaliknya bila hampir semua peserta tes menjawab benar maka butir tersebut dikatakan mudah. Dari hasil perhitungan indeks kesukaran maka kemungkinan tidak semua soal dapat terambil. Soal yang mempunyai indeks kesukaran sedang yang dapat di ambil.

Kelemahan utama indeks kesukaran soal seperti ini ialah bahwa antara indeks kesukaran soal dan taraf kesukaran soal mempunyai hubungan yang berlawanan arah, artinya makin tinggi indeks kesukarannya, maka makin rendahlah taraf kesukarannya. Dalam hal pengukuran yang bertujuan untuk membedakan subyek yang satu dengan yang lainnya dalam hal kompetensi mereka mengenai sesuatu mata pengetahuan, kebanyakan ahli berpendapat bahwa tes yang terbaik adalah tes yang terdiri dari soal-soal yang mempunyai taraf kesukaran sedang dan rentang distribusi kesukarannya yang kecil.

b. Menghitung daya beda

Menurut Suryabrata (2000) daya pembeda soal diukur dari kesesuaian soal itu dengan keseluruhan tes dalam membedakan antara mereka yang tinggi kemampuannya dan mereka yang rendah kemampuannya dalam hal yang diukur oleh tes yang bersangkutan. Teknik yang banyak digunakan untuk mengukur daya pembeda itu adalah korelasi antara skor pada soal tertentu dengan skor total. Rumus korelasi

biserial yang paling banyak digunakan adalah :

$$r_{bis} = \frac{\bar{X}_b - \bar{X}_s}{S_t} \times \frac{p(1-p)}{y}$$

Dengan keterangan :

\bar{X}_b : rata –rata skor kriteria subyek yang memilih jawaban benar

\bar{X}_s : rata-rata skor kriteria subyek yang menjawab salah

S_t : simpangan baku skor kriteria semua subyek

p : proporsi subyek yang menjawab benar terhadap semua subyek

y : ordinat dalam kurve normal yang membagi menjadi p dan 1-p

Suatu butir soal harus dapat membedakan kelompok yang pandai dengan kelompok yang lemah dalam hal ini kelompok atas dan kelompok bawah. Klasifikasi daya beda adalah sebagai berikut: 1) daya beda ≤ 0 (negatif), 2) 0,00-0,20 jelek 3) 0,21-0,20 cukup, 4) 0,41-0,71 baik, 0,71-1,00 baik sekali. Soal-soal dengan klafisifisi daya beda jelek dan negatif di buang, yang di ambil klasifikasi cukup, baik, dan baik sekali.

Dalam analisis ini daya beda dianggap memuaskan bila mencapai angka 0,25. Angka ini lebih tinggi dibanding rekomendasi Thorndike sebesar 0,20

(2005) dan rekomendasi ahli lain (Crocker & Algina, 1986) dan masih jauh lebih tinggi daripada yang disarankan oleh Kehoe yaitu 0,15 (Kehoe, 1997).

c. Analisis distraktor (pengecoh)

Analisis distraktor di perlukan hanya untuk pembuat soal. Selain menghitung indeks kesukaran dan daya beda dalam analisis butir juga perlu di ketahui apakah distraktor atau pengecoh yang di sediakan tepat atau tidak benar. Apakah semua pilihan yang disediakan dipilih semua karena dianggap betul, jawaban terkumpul pada pilihan tertentu atau pilihan yang sama sekali tidak ada pemilihnya.

Indikator lain mengenai efektivitas distraktor ditampakkan oleh koefisien r-pointbiserial bagi masing-masing distraktor. Suatu distraktor yang efektif adalah yang memiliki koefisien r_{pbis} negatif. Semakin besar harga negatif r_{pbis} menunjukkan bahwa fungsi distraktor semakin efektif sedangkan r_{pbis} yang berada di sekitar nol berarti distraktor tidak berfungsi sebagaimana mestinya. Pada aitem-aitem yang sulit, yaitu yang persentase subjek menjawab benar sangat kecil, interpretasi efektivitas distraktor tidak dapat semata-mata disandarkan pada angka statistik r_{pbis} namun harus disertai dengan pertimbangan mengenai distribusi peluang subjek yang menjawab salah pada aitem yang bersangkutan.

d. Blue Print Tes Psikologi Kepribadian
1

Gay (1987) menyatakan bahwa validitas isi (*content validity*) adalah derajat pengukuran yang mencerminkan domain isi yang diharapkan. Validitas isi penting untuk tes hasil belajar (*achievement test*). Suatu skor kurang bahkan tidak mencerminkan hasil belajar siswa apabila instrumen tidak mampu mengukur secara komprehensif apa yang telah dipelajari oleh siswa.

Prosedur yang hendak ditempuh agar suatu tes hasil belajar mampu mencerminkan domain isi secara komprehensif adalah dengan menyusun kisi-kisi tes.

METODE PENELITIAN

a. Variabel-variabel Penelitian

Variabel dalam analisis aitem adalah skor aitem tes psikologi kepribadian 1 yang meliputi akhir semester. Aitem – aitem berupa soal pilihan ganda dengan 4 pilihan jawaban yaitu a, b, c, dan d dengan satu kunci jawaban. Skor aitem merupakan skor dikotomi, yaitu 1 untuk jawaban yang benar dan 0 untuk jawaban yang salah.

b. Metode Pengumpulan Data

Data penelitian berupa skor aitem dan skor tes psikologi kepribadian 1 dari mahasiswa Program Studi Psikologi UMBY yang mengikuti ujian akhir semester mata kuliah Psikologi Kepribadian1 Semester Genap T.A. 2011/2012.

Aitem yang diujikan berjumlah 60 butir soal, dengan durasi waktu mengerjakan 65 menit. Soal-soal dikerjakan secara individual dan bersifat tertutup. Penelitian dilaksanakan pada 23 Juli 2012.

c. Subyek Penelitian

Subyek penelitian ini adalah para mahasiswa Universitas Mercu Buana Yogyakarta yang mengikuti ujian akhir pada mata kuliah psikologi kepribadian 1 berjumlah 70 mahasiswa.

d. Teknik Analisis Data

Data yang diperoleh dalam penelitian ini, selanjutnya dianalisis secara kuantitatif (empiris) dengan bantuan program *Item and Test Analysis (ITEMAN)* Versi 3.00. Analisis ini akan menghasilkan karakteristik butir soal dan perangkat tes berupa statistik. Statistik butir tes, meliputi: (1) tingkat kesukaran, (2) daya beda, dan (3) efektivitas distraktor.

Pada saat memasukkan data/skor subyek ke dalam program ITEMAN, maka pilihan jawaban a diubah menjadi 1, b menjadi 2, c menjadi 3 dan d menjadi 4). Hal ini untuk memudahkan dalam input data.

DISKUSI

a. Deskripsi data penelitian

Berdasarkan hasil pengambilan data yang dilakukan pada subjek penelitian sebanyak 70 orang, diperoleh gambaran data sebagai berikut :

Tabel 1.

Deskripsi data
(N=70 Subyek)

Jumlah aitem (N of aitem)	60 butir
Jumlah subyek (N of examinees)	70 mahasiswa
Rerata	34,814
Varian	106,923
Standar deviasi	10,340
Minimum	15
Maksimum	59
Alpha	0,898

bahwa butir soal tersebut terlalu sukar atau terlalu mudah untuk peserta tes. Indeks ini disebut juga indeks tingkat kesukaran soal secara klasikal.

Berdasarkan hasil analisis data yang dilakukan pada subjek penelitian sebanyak 70 orang dengan 60 butir, diperoleh gambaran data sebagai berikut yang tercantum pada tabel 2.

b. Hasil hitung Indeks Kesukaran Butir

Hasil indeks kesukaran aitem dalam program ITEMAN ditunjukkan oleh **Prop. Correct** yaitu Proporsi mahasiswa yang menjawab benar butir tes. Nilai ekstrem mendekati nol atau satu menunjukkan

Tabel 2.

Indeks Kesukaran Butir (IKB)

No. soal	IKB	No. soal	IKB	No. soal	IKB
1	0.886	21	0.371	41	0.643
2	0.186	22	0.600	42	0.600
3	0.700	23	0.686	43	0.586
4	0.671	24	0.414	44	0.557
5	0.471	25	0.543	45	0.443
6	0.400	26	0.957	46	0.443
7	0.557	27	0.629	47	0.514
8	0.729	28	0.800	48	0.757
9	0.443	29	0.757	49	0.571
10	0.643	30	0.500	50	0.957
11	0.571	31	0.629	51	0.543
12	0.043	32	0.586	52	0.686
13	0.871	33	0.843	53	0.229
14	0.800	34	0.686	54	0.386
15	0.229	35	0.586	55	0.657
16	0.571	36	0.600	56	0.771
17	0.586	37	0.443	57	0.700
18	0.700	38	0.157	58	0.571
19	0.343	39	0.686	59	0.457
20	0.629	40	0.729	60	0.514

Berdasarkan Indeks Kesukaran Butir (IKB) tersebut maka didapatkan aitem dengan kategori *sangat sukar* (3 aitem), *sukar* (5 aitem), *sedang* (23 aitem), *mudah* (22 aitem), dan *sangat mudah* (7 aitem). Biasanya butir yang ditoleransi sebagai tes standar adalah yang memiliki IKB = 0,30-0,70, dalam hal ini didapatkan sebanyak 43 aitem.

Ditinjau dari tujuan pelaksanaan tes, perlu diperhatikan bahwa soal yang sangat mudah atau sangat sukar mungkin memang kurang memberikan informasi yang berguna bagi peserta tes pada umumnya, di antaranya kemungkinan karena belum berfungsinya pengecoh dengan baik, namun demikian pada soal yang terlalu mudah atau terlalu sukar ini, apabila

pengecohnya berfungsi dengan baik yakni *Prop Endorsing* memenuhi ($>0,05$) serta daya pembedanya negatif maka soal tersebut masih memenuhi untuk diterima sebagai salah satu alternatif untuk disimpan dalam bank soal.

Demikian pula pada butir soal yang memiliki IKB sangat sukar atau sangat mudah, dapat tetap digunakan untuk

mengetahui sebaran tingkat penguasaan materi siswa. Misalnya dalam tes yang bertujuan untuk evaluasi formatif.

c. Hasil hitung Indeks Daya Beda Butir

Berdasarkan hasil analisis data yang dilakukan pada subjek penelitian sebanyak 70 orang dengan 60 butir, diperoleh gambaran data IDB sebagai berikut yang tercantum pada tabel 3

Tabel 3.
Indeks Daya Beda (IDB)

No. soal	IDB	No. soal	IDB	No. soal	IDB
1	0.180	21	0.351	41	0.555
2	0.129	22	0.434	42	0.594
3	0.184	23	0.321	43	0.518
4	0.514	24	0.352	44	0.346
5	0.142	25	0.219	45	0.280
6	0.068	26	0.126	46	0.472
7	0.384	27	0.318	47	0.654
8	0.331	28	0.367	48	0.541
9	0.539	29	0.418	49	0.431
10	0.301	30	0.496	50	0.235
11	0.462	31	0.258	51	0.613
12	0.058	32	0.459	52	0.309
13	0.455	33	0.448	53	0.108
14	0.316	34	0.512	54	0.497
15	0.395	35	0.582	55	0.156
16	0.610	36	0.558	56	0.332
17	0.366	37	0.628	57	0.462
18	0.308	38	0.224	58	0.096
19	0.284	39	0.300	59	0.396
20	0.447	40	0.508	60	0.347

Hasil evaluasi IDB, diperoleh aitem dengan dengan kategori : jelek (10aitem), perlu perbaikan (6 aitem), bagus (18 aitem), dan bagus sekali (26 aitem).

Daya diskriminasi yang baik memang pada umumnya terdapat pada aitem yang tidak terlalu mudah dan juga tidak terlalu sukar, yaitu apabila harga p berkisar antara 0,40 sampai dengan 0,60.

Berdasarkan kategori tersebut, maka sangat perlu dilakukan peninjauan ulang terhadap butir-butir soal yang masuk dalam kategori jelek dan perlu perbaikan.

d. Hasil hitung efektivitas distraktor (pengecoh)

Indikator lain mengenai efektivitas distraktor ditampakkan oleh koefisien r -pointbiserial bagi masing-masing distraktor. Suatu distraktor yang efektif adalah yang memiliki koefisien r_{pbis} negatif. Semakin besar harga negatif r_{pbis} menunjukkan bahwa fungsi distraktor semakin efektif sedangkan r_{pbis} yang berada di sekitar nol berarti distraktor tidak berfungsi sebagaimana mestinya.

Pada aitem-aitem yang sulit, yaitu yang persentase subjek menjawab benar sangat kecil, interpretasi efektivitas distraktor tidak dapat semata-mata disandarkan pada angka statistik r_{pbis} namun harus disertai dengan pertimbangan mengenai distribusi peluang subjek yang menjawab salah pada aitem yang bersangkutan.

Ditinjau dari distribusi jawaban yaitu persentase peserta tes merespons alternatif jawaban, semua pengecoh telah berfungsi dengan baik. Dapat dilihat pada kolom *Prop Endorsing*, tampak bahwa pada masing masing alternatif jawaban sebagian besar ada yang memilih.

Berdasarkan indeks Point biserial (pada alternative statistics), diperoleh masing-masing alternatif pengecoh memiliki angka negatif dan memiliki presentasi yang baik dalam hal jumlah pemilih. Namun demikian terdapat beberapa nomor butir yang mendapatkan peringatan dari hasil ITEMAN, yaitu butir nomor : 6, 12 dan 58. Selengkapnya dapat dilihat pada tabel 4.

Tabel 4
Analisis distraktor (6, 12, 58)

No. butir	Alternatif Jawaban	Point Biserial	Porp. Endorsing	Prop. Kunci
6	1	0,196	0,30	0,40
12	4	0,102	0,20	0,043
58	1	0,247	0,186	0,571

Berdasarkan data tersebut di atas, butir soal no 6, memiliki pengecoh yang tidak efektif, yaitu alternatif 1 (atau a). pengecoh tersebut memiliki indeks Point biserial positif dan dipilih oleh 30% hampir setara dengan kunci (40%). Artinya bahwa alternatif ini justru lebih banyak dipilih oleh kelompok yang tinggi (pandai) dari pada kelompok rendah.

Demikian juga butir soal nomor 12, menunjukkan pada alternatif 4 (atau d), memiliki indeks point biserial positif dan dipilih 20% lebih banyak dari kunci jawaban (4,3%). Artinya pengecoh 4 (atau d) kurang efektif karena kelompok tinggi banyak yang memilihnya dari pada kunci jawaban.

Butir soal nomor 58 memiliki karakteristik yang mirip dengan butir soal nomor 6 pada alternatif 1 (a).

Oleh karena itu sangat perlu dilakukan penggantian alternatif jawaban pada alternatif-alternatif tersebut di atas.

e. Reliabilitas

Sebenarnya tidak terdapat suatu ukuran yang pasti mengenai berapa tinggi koefisien reliabilitas. Reliabilitas yang baik tergantung pada tujuan atau kegunaan tes. Menurut Suryabrata (2000) menyatakan bahwa kebanyakan tes-tes di bidang pendidikan pada umumnya memiliki koefisien reliabilitas minimal 0,8 untuk populasi yang sesuai.

Reliabilitas paket butir-butir soal tes kepribadian 1 ini memiliki koefisien

reliabilitas alpha sebesar ; 0,898, sehingga dianggap memiliki reliabilitas yang baik.

KESIMPULAN DAN SARAN

Hasil analisis butir soal dengan menggunakan program iteman berdasarkan Statistik butir soal (*Prop.corret*, *Point biser*, dan *Point biser (alt statis)*) dan Statistik tes (koefisien reliabilitas alpha) menunjukkan bahwa dari 60 butir soal yang dianalisis **cukup baik dan dapat digunakan**. Sebanyak 43 aitem memiliki IKB standar yaitu berkisar 0,30-0,70, dan terdapat 44 butir soal yang memiliki IDB Bagus dan Sangat bagus.

Terkait dengan efektifitas distraktor, maka sebanyak 57 butir aitem telah memiliki karakteristik distraktor yang bagus.

Analisis butir soal dengan menggunakan program iteman sangat cepat dan tepat untuk digunakan sebagai kesimpulan analisis akhir dari suatu butir soal atau tes secara keseluruhan apakah layak digunakan atau tidak. Hal ini penting untuk peningkatan kualitas pembelajaran sebagai acuan untuk menganalisa secara cepat, tepat agar menghasilkan suatu alat evaluasi yang baik.

Penelitian ini menghasilkan suatu analisis butir soal yang murah, tepat dan cepat karena menggunakan program computer yang direkomendasikan untuk

diinterpretasikan pada soal-soal tes yang lainnya.

DAFTAR PUSTAKA

- Allen, M.J., & Yen, W.M. 1979. *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.
- Arikunto, S & Jabar. 2004. *Evaluasi Program Pendidikan*. Jakarta: Bumi Aksara
- Azwar, S. 2002. *Tes Prestasi*. Yogyakarta : Pustaka Pelajar.
- Azwar, S .2005. *Dasar-dasar psikometri (Edisi 1)*. Yogyakarta: Pustaka Pelajar.
- Chabib,M. 2001. *Teknik Evaluasi Pendidikan*, Jakarta: PT Raja Grafindo Persada.
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston, Inc.
- Djaali. 2006. *Hasil belajar evaluasi dalam evaluasi pendidikan: Konsep dan aplikasi*. Jakarta: Uhamka Press.
- Ebel, RL. (1972). *Essential of educational measurement and evaluating in education and psycology*. New York: Holt, Rine hart, and Winston. Inc.
- Gay, L. R. 1987. *Education research, Competencies for analysis and application*. Third edition. Columbus: Merrill Publishing Company.
- Gronlund, NE. (1981). *Measurement and evaluating in teaching*. New York:Macmillan Publishing Co., Inc.
- Harlen, W. 1983. *Guides to Assessment in Education Science*. London: Macmillan Education
- Jacobs & Chase. 1992. *Developing and Using test Effectively*. San Fransisco: Jossey-Bass Publishers.
- Kehoe, J. 1997. Basic item analysis for multiple-choice tests. *ERIC Digest*. <http://www.ericdigests.org/1997-1/basic.html>
- Kumano, Y. 2001. *Authentic Assessment and Portfolio Assessment-Its Theory and Practice*. Japan: Shizuoka University.
- Mardapi, D. 2008. *Teknik penyusunan instrument tes dan nontes*. Yogyakarta: Mitra Cendekia Press.
- Masrun & Martaniah. 1976. *Psikologi Pendidikan : Seri Pedagogik dan Psikologi*. Yogyakarta : Yayasan Penerbitan Fakultas Psikologi UGM.

- Rustaman, N. 2003. Asesmen Pendidikan IPA. Makalah penataran guru-guru NTT di Jurusan pendidikan Biologi.
- Stiggins, R.J. 1994. *Student-Centered Classroom Assessment*. New York : Macmillan College Publishing Company
- Subekti, R. & Firman, H. 1989. *Evaluasi Hasil Belajar dan Pengajaran Remedial*. Jakarta: UT.
- Suharsimi. 2003. *Dasar-Dasar Evaluasi Pendidikan* (edisi revisi). Jakarta. Penerbit Bumi Aksara.
- Suryabrata, S. 2000. *Pengembangan alat ukur psikologis*. Yogyakarta : Penerbit Andi.
- Thorndike, R.L. 1982. *Applied Psychometrics*. Boston : Houghton Mifflin.
- Thorndike, R.M. 2005. *Measurement and evaluation in psychology and education (7th ed)*. New Jersey: Pearson Education, Inc.
- Yulaelawati, E. 2004. *Kurikulum dan Pembelajaran*. Jakarta: Pakar Raya Jakarta.
- Zainul, A. (2001). *Alternative assessment*. Jakarta: Dirjen Dikt