

## PENGARUH JUMLAH ALTERNATIF JAWABAN TES OBYEKTIF BENTUK PILIHAN GANDA TERHADAP RELIABILITAS, TINGKAT KESUKARAN DAN DAYA PEMBEDA

Idrus Alwi

Program Studi Pendidikan Matematika, Fakultas Teknik, Matematika & Ilmu Pengetahuan Alam  
Universitas Indraprasta PGRI

**Abstrak.** The purpose of this research was to examine the effect of the alternative numbers in a multiple choice test on reliability, level of difficulty and discriminatory power. The random technique was implemented to take the sample which consisted of the third year students of MTs N 7 Jakarta. The sample was 150 student. The data were collected through achievement test of mathematics. The one way analisis of variance was used to analyze the data. The result indicated that : (1) the alternative number in a multiple choice test had a significant effect on reliability of the test, (2) The alternative number in a multiple choice test had no significant effect on the level of difficulty, (3) but the alternative number in a multiple choice test had a significant effect on the discriminatory power.

Keywords: alternative numbers, multiple choice, reliability

### PENDAHULUAN

Peranan tes tertulis sebagai alat ukur untuk mendapatkan informasi dari berbagai aspek pendidikan sangat penting. Hal yang paling pokok di dalam penilaian pendidikan adalah pengukuran hasil belajar atau prestasi belajar. Morrison mengemukakan bahwa suatu keputusan membutuhkan informasi yang akurat, relevan serta yang dapat dipercaya. Keputusan yang lebih sempurna bila alat ukur yang digunakan untuk menapatkan informasi tersebut objektif dan tidak bias adalah tes objektif bentuk pilihan ganda.

Sampai sekian jauh penggunaan tes objektif bentuk pilihan ganda pada pelaksanaan ujian nasional, senantiasa ada kelompok yang mengkritik yang pada akhirnya menimbulkan pro dan kontra terhadap ujian nasional tersebut. Menurut Habibi, bahwa jawaban benar pada tes bentuk pilihan ganda merupakan faktor kebetulan karena menebak, lagi pula tes bentuk pilihan ganda hanya mengukur rekognisi saja.

Walaupun adanya kritikan-kritikan terhadap tes objektif bentuk pilihan ganda, namun peranan tes objektif bentuk pilihan ganda di Indonesia berkembang semakin luas. Di samping kelemahan tes objektif

bentuk pilihan ganda ada keunggulan dari tes bentuk pilihan ganda.

Kualitas tes objektif bentuk pilihan ganda tergantung oleh kemampuan pembuat item. Menurut Suryabrata, penyusunan tes objektif bentuk pilihan ganda memerlukan ketrampilan dan kemampuan professional, kesungguhan dan daya kreatif yang tinggi sehingga dihasilkan tes yang benar-benar objektif dan baik.

Ditinjau dari segi kualitatif kelemahan item yang kurang baik terletak pada materi, konstruksi dan bahasa yang umumnya menimbulkan kesalahan konsep. Ditinjau dari segi konstruksi, pemilihan atau penentuan alternative jawaban dapat menyebabkan item-item tidak memenuhi persyaratan sebagai item yang baik, karena untuk mendapatkan pengecoh yang relevan dan bermakna sangat terbatas.

Penelitian ini ingin melihat berapa jumlah alternative jawaban untuk tes objektif bentuk pilihan ganda, sebagai alat ukur untuk mendapatkan informasi yang objektif dan tidak bias, di samping itu penelitian ini bermaksud untuk mengetahui pengaruh banyaknya alternative jawaban tes objektif bentuk pilihan ganda terhadap reliabilitas tes, tingkat kesukaran soal dan daya pembeda soal.

## TINJAUAN PUSTAKA

### 1. Penilaian, pengukuran, dan pengambilan keputusan

Dalam lingkungan pendidikan di sekolah, informasi tentang keberhasilan guru menyajikan bahan pelajaran serta sejauhmana peserta didik telah menyerap materi yang diajarkan dapat diperoleh melalui penilaian dan pengukuran. Karena itu, penilaian dan pengukuran merupakan bagian yang penting di dalam proses belajar mengajar. Salah satu alat ukur yang digunakan untuk memperoleh informasi itu adalah dalam bentuk tes prestasi belajar berupa tes tertulis (*paper and pencil test*).

Peranan tes tertulis sebagai alat ukur untuk mendapatkan informasi dari berbagai aspek pendidikan sangatlah penting. Tes tertulis sangat diperlukan dalam proses belajar mengajar di kelas oleh semua guru. Karena itu tes tertulis sangat berperan dalam mengevaluasi proses belajar mengajar di kelas. Seperti diketahui bahwa hasil tes dapat digunakan untuk keperluan membuat keputusan atau kebijakan-kebijakan yang penting baik di tingkat kelas maupun nasional.

Dalam proses pendidikan dan pengajaran setiap saat akan selalu ada situasi yang memerlukan pengambilan keputusan. Setiap orang yang terlibat dalam proses pendidikan pada suatu ketika akan harus mengambil suatu bentuk keputusan pendidikan, yaitu keputusan-keputusan yang menyangkut berbagai hal dalam pendidikan sebagai suatu sistem. Hal tersebut akan tampak lebih jelas dalam sistem pendidikan formal, baik di perguruan tinggi ataupun di tingkat-tingkat pendidikan menengah dan dasar.

Di antara keputusan-keputusan pendidikan itu, dapat berupa keputusan didaktik yang diperlukan guna memenuhi kebutuhan pengajaran seperti misalnya keputusan yang menyangkut ketepatan kurikulum yang berlaku, keputusan mengenai bimbingan penyuluhan guna memberikan bimbingan dalam penjurusan dan penentuan karir. Keputusan pendidikan dapat pula berupa keputusan administratif guna memenuhi kebutuhan administrasi seperti misalnya keputusan mengenai nilai yang

hendak diberikan pada subjek atau keputusan mengenai kelulusan.

Apapun bentuknya, agar suatu keputusan pendidikan menjadi bermanfaat haruslah didasari oleh informasi-informasi yang tepat, akurat, dan reliable yang berkaitan dengan permasalahannya. Sebagai contoh, keputusan untuk memberikan nilai tertentu terhadap hasil ujian siswa tidaklah dapat dilakukan semata-mata berdasarkan pertimbangan subjektif guru dikarenakan pertimbangan subjektif tidak akurat dalam penilaian prestasi belajar siswa. Untuk memberikan bimbingan dan pengarahan kepada siswa mengenai jurusan yang akan diambilnya dalam suatu program studi, tidaklah dapat disandarkan hanya pada keinginan dan minat siswa itu saja akan tetapi haruslah pula didasari oleh informasi lain mengenai bakat dan kemampuannya.

Dapat dikatakan bahwa informasi yang paling besar sumbangannya dalam kelayakan suatu keputusan pendidikan yang berkaitan dengan siswa umumnya diperoleh dari kegiatan pengukuran dan penilaian pendidikan atau secara lebih spesifik diperoleh dari tes.

Banyak sekali keputusan pendidikan yang diambil berdasarkan hasil tes. Sebagai contoh antara lain adalah pemberian nilai suatu mata pelajaran, penentuan lulus tidaknya seorang siswa, perlu tidaknya penyelenggaraan remedial dan pengayaan, penempatan siswa pada suatu program dan lain sebagainya. Berbagai macam keputusan pendidikan itu menempatkan tes dalam beberapa fungsi yaitu fungsi penempatan (*placement*), fungsi formatif, fungsi diagnostik, dan fungsi sumatif (Silverius, 1991)

Antara tes, pengukuran dan penilaian di bidang pengujian sering digunakan secara rancu oleh banyak guru dan pemakai tes. Kadang-kadang istilah tes, pengukuran dan penilaian dapat dipertukarkan (Mehrens & Lehmann, 1972). Sementara orang memang lebih cenderung mengartikan ketiga kata tersebut sebagai satu pengertian yang sama sehingga dalam memakainya hanya tergantung dari kata mana yang siap untuk diucapkannya. Akan tetapi sementara orang yang lain membedakan ketiga istilah tersebut.

Tes dalam pengertian terbatas mengandung arti gambaran dari suatu kumpulan pertanyaan standart yang akan dijawab. Tes juga diartikan sebagai himpunan pertanyaan yang harus dijawab, atau pernyataan-pernyataan yang harus dipilih/ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang dites (*testee*) dengan tujuan untuk mengukur suatu aspek (prilaku) tertentu dari orang yang dites. Tes disusun berdasarkan kriteria-kriteria tertentu dan diberikan kepada siswa untuk menjawabnya, sehingga prilaku siswa yang abstrak dapat terjaring.

Batasan seperti tersebut di atas tentu masih terlalu sederhana karena pada kenyataannya tidak semua kumpulan pertanyaan cukup berharga untuk dinamakan tes. Banyak syarat-syarat kualitas yang harus dipenuhi oleh rangkaian pertanyaan atau tugas itu agar dapat disebut tes.

Anne Anastasi (1988) dalam bukunya *Psychological Testing* mengatakan bahwa tes pada dasarnya merupakan suatu pengukuran yang objektif dan standar terhadap sampel prilaku. Frederick G. Brown (1976) mengatakan bahwa tes adalah prosedur yang sistematis guna mengukur sampel perilaku seseorang. Tampaknya Brown menganggap bahwa ciri sistematis itu telah mencakup pengertian objektif, standar, dan syarat-syarat kualitas lainnya.

Dalam berbagai macam batasan mengenai tes dapatlah ditarik beberapa kesimpulan pengertian yaitu tes adalah pengukuran untuk mengambil sampel prilaku secara sistematis, objektif dan berdasarkan standar tertentu.

Arti tes hampir sama dengan pengukuran hanya pengukuran mengandung arti yang lebih luas. Tes merupakan bagian dari pengukuran, karena itu untuk mengukur karakter dapat pula dilakukan dengan cara lain.

Pengukuran adalah suatu kegiatan untuk mendapatkan informasi dalam bentuk kuantitatif. Informasi yang diperoleh dari hasil pengukuran dapat menetapkan posisi suatu obyek pada skala tertentu. Data yang diperoleh dari suatu pengukuran selalu berupa angka-angka yang mewakili tingkat tertentu seperti tingkat kemampuan, tingkat prestasi, tingkat kegemaran terhadap suatu hal.

Sebagian ahli psikometri membatasi tes sebagai suatu prosedur khusus yang merupakan bagian dari pengukuran secara keseluruhan. Tyler (1971) mengatakan bahwa pengukuran adalah pemberian angka berdasarkan aturan-aturan tertentu. Jadi pemberian angka seperti dilakukan dalam tes memang merupakan suatu bentuk pengukuran.

Ciri pokok pengukuran adalah adanya proses perbandingan. Mengukur adalah membandingkan atribut yang hendak diukur dengan alat ukurnya secara deskriptif. Deskriptif artinya menyatakan hasil ukur secara kuantitatif hanya dengan satuan atau besaran ukurnya saja tanpa memberikan penilaian kualitatif. Dalam mengukur panjang sebuah meja, misalnya, hasilnya dinyatakan dalam meter atau sentimeter. Angka sentimeter atau meter itu merupakan hasil pengukuran deskriptif yang tidak diikuti oleh pernyataan apakah sekian sentimeter itu adalah panjang atau pendek, karena pendek dan panjang merupakan hasil evaluasi bukan hasil pengukuran.

Karena tes merupakan alat pengukuran maka istilah pengesanan kerap kali menggantikan istilah pengukuran, dan sebaliknya. Dalam hal ini yang terpenting adalah mengetahui di mana penggunaan kedua istilah itu dapat dipertukarkan atau saling menggantikan dan kapan kedua istilah tersebut harus dibedakan agar tidak menimbulkan salah pengertian.

Sedangkan penilaian dilaksanakan ketika pengukuran sudah dilakukan. Penilaian bagi seorang siswa yang pandai tidak dapat dengan mudah dibedakan dari siswa lainnya, yang hanya dengan melihat siswa-siswa tersebut pada proses belajar mengajar. Penampilan seorang siswa di dalam kelas hanya merupakan penunjang untuk keberhasilan dalam menjawab suatu tes.

Guru dapat menentukan seorang siswa lebih pandai dari yang lain melalui pancaran kepandaian siswa tersebut dalam mengerjakan tes. Gejala kepandaiannya di dalam mengerjakan tes dapat diukur. Hasil pengukuran tersebut diberi penilaian. Penilaian yang diperoleh dari pengukuran dapat merupakan keberhasilan yang baik atau buruk. Penilaian adalah suatu proses untuk

menggambarkan, melukiskan dan mendapatkan informasi yang digunakan secara penuh untuk mengambil alternatif keputusan. Dalam pendidikan, penilaian pada umumnya menyangkut masalah keberhasilan seseorang sebagai peserta didik. Penilaian juga berguna sebagai pertanggungjawaban terhadap kegiatan yang telah dilaksanakan.

Dari pengertian-pengertian di atas, jelaslah bahwa antara istilah tes, pengukuran dan penilaian mempunyai tingkat yang berbeda. Penilaian memerlukan data yang baik dan bersumber pada pengukuran. Pengukuran dapat dilakukan dengan beberapa macam cara dan salah satu diantaranya adalah tes.

**2. Kualitas Tes**

Soal yang bermutu adalah soal yang dapat memenuhi persyaratan mutu soal, yaitu: (1) valid, (2) reliabel, (3) memiliki taraf kesukaran yang baik, (4) mempunyai daya pembeda yang baik, dan (5) pengecohnya berfungsi. Mutu soal dapat diketahui dari hasil analisis soal (Wiersma dan Jurs, 1990).

Validitas soal adalah kecocokan atau ketepatan suatu tes dalam mengukur sesuatu yang hendak diukur (Gronlund, 1990). Ada tiga bentuk validitas soal, yaitu: (1) validitas isi (*content validity*), (2) validitas kriteria (*criterion-related validation*), dan (3) validitas konstruk (*construct validation*) (Crocker dan Algina, 1986).

Validitas isi melihat kesesuaian dan keterwakilan butir soal dengan keseluruhan materi yang diberikan dalam proses belajar mengajar. Langkah-langkah yang dilakukan untuk mendapatkan validitas isi suatu tes di antaranya: (1) mendefinisikan kemampuan yang hendak diukur secara jelas, (2) menyeleksi soal bersama pakar (*expert review*) yang jumlahnya menurut Tessmer (1995) cukup satu atau dua orang saja. Validitas tes diketahui melalui rumus Point Biserial (*r-bis*), yaitu:

$$P_{pbis}(i) = \frac{\mu_+ - \mu_x}{\sigma_x} \sqrt{\frac{p}{q}}$$

Dimana :

- $P_{pbis}$  = koefisien korelasi biserial antara skor butir soal nomor ke i dengan skor total;
- $\mu_+$  = rata-rata skor total responden yang menjawab benar butir ke i;
- $\mu_x$  = rata-rata skor total semua responden;
- $\sigma_x$  = standar deviasi;
- $p$  = populasi jawaban benar untuk butir soal nomor ke-i;
- $q$  = proporsi jawaban salah untuk butir soal nomor ke i (1-p).

Pembandingan adalah r-tabel pada alpha 0,05. Bila skor koefisien korelasi butir soal dengan skor total lebih besar dari r-tabel, maka butir soal tersebut dianggap valid (Crocker & Algina, 1986).

Taraf kesukaran butir soal (*item difficulty*) adalah persentase siswa yang menjawab benar suatu butir soal yang diukur melalui skor butir soal yang dihasilkan oleh jawaban sejumlah peserta ujian (Linn & Gronlund, 1995). Makin banyak peserta ujian yang berhasil dengan benar mengerjakan suatu butir soal, makin tidak sukar atau makin mudah butir soal tersebut (McBeath, 1992). Butir soal yang baik adalah butir soal yang tidak terlalu sukar dan tidak terlalu mudah, yaitu  $P = 0,5$ . Tuckman dalam Naga (1992) menganjurkan bahwa sebaiknya taraf kesukaran butir soal itu berada antara  $0,33 \leq P \leq 0,67$ . Butir soal dengan nilai  $P = < 0,33$  adalah butir soal yang sukar, sedangkan butir soal dengan nilai  $P = > 0,67$  adalah butir soal yang mudah. Hal ini menunjukkan bahwa semakin mendekati angka 1 suatu butir soal, semakin mudah soal tersebut, sebaliknya semakin mendekati angka 0 suatu butir soal, semakin sulit soal tersebut.

Indek daya pembeda adalah suatu indek yang menyatakan seberapa efisien suatu butir soal dapat membedakan peserta kelompok tinggi dengan peserta kelompok rendah (Gregory, 2000). Daya pembeda butir soal dapat dilihat melalui rumus  $D = P_u - P_l$  (Crocker dan Algina, 1996), dimana adalah proporsi kelompok atas yang menjawab butir soal dengan benar dan adalah proporsi kelompok bawah yang menjawab butir soal secara benar. Sedangkan  $D$  adalah nilai

indek diskriminasi pada rentangan antara -1,00 - 1,00 dengan ketentuan:  $D \geq 0,40$  berarti cukup memuaskan;  $0,30 \leq D \leq 0,39$  berarti sedikit atau tanpa revisi;  $0,20 \leq D \leq 0,29$  berarti perbatasan atau perlu direvisi;  $D \leq 0,19$  berarti soalnya dibuang.

Sehubungan dengan batas penerimaan butir soal dari aspek daya pembeda, Crocker dan Algina (1996) mengemukakan bahwa suatu butir soal dapat dipertahankan apabila memiliki indek daya pembeda  $D \leq 0,30$  (Linn dan Gronlund, 1995). Pengelompokan nilai butir soal di atas selanjutnya dijadikan sebagai tolok ukur membuat keputusan apakah soal yang dibuat guru sudah memenuhi target yang ditetapkan atau belum.

Syarat tes lainnya adalah reliabilitas. Reliabilitas artinya dapat dipercaya atau dapat diandalkan. Suatu alat ukur yang baik adalah alat pengukur yang mempunyai reliabilitas yang tinggi, artinya setiap kali alat pengukur digunakan untuk mengukur hal yang sama, maka hasil pengukurannya tetap (Nasoetion, 1993:103). Reliabilitas suatu tes pada hakikatnya menguji keajegan pertanyaan tes yang didalamnya berupa seperangkat butir soal apabila diberikan berulang kali pada objek yang sama. Suatu tes dikatakan reliabel apabila beberapa kali pengujian menunjukkan hasil yang relative sama (Arikunto, 2001: 86 ). Yang sering ditangkap kurang tepat bagi pembaca adalah

adanya pendapat bahwa “ajeg” atau “tetap” diartikan sebagai “sama”. Di sini ajeg atau tetap tidak selalu sama, tetapi mengikuti perubahan secara ajeg. Tentu saja tidak dituntut semuanya tetap. Besarnya ketetapan itulah yang menunjukkan tingginya reliabilitas suatu tes. Sehubungan dengan reliabilitas ini, menyatakan bahwa persyaratan bagi suatu tes, yaitu validitas dan reliabilitas ini penting. Dalam hal ini validitas lebih penting, dan reliabilitas ini perlu, karena menyokong terbentuknya validitas. Sebuah tes mungkin reliabel tetapi tidak valid. Sebaliknya, sebuah tes yang valid biasanya reliabel.

Menurut Arikunto (2001: 90), untuk melakukan analisis reliabilitas suatu tes dapat digunakan beberapa metode yaitu: metode bentuk parallel (*equivalent*), metode

tes ulang (*test-retest-method*), dan metode belah dua (*split-half-method*). Rumus yang digunakan untuk mencari reliabilitas dan banyak digunakan orang ada dua rumus yaitu rumus KR-20 dan rumus KR-21. Reliabilitas suatu tes dapat tinggi dapat pula rendah karena ada faktor-faktor yang mempengaruhi besarnya reliabilitas. Faktor-faktor yang mempengaruhi besarnya reliabilitas adalah: luas tidaknya sampling yang diambil, perbedaan bakat dan kemampuan murid yang dites serta suasana dan kondisi testing.

Jika jumlah soal ganjil maka tidak mungkin dengan belah dua tetapi harus dengan rumus yang lain, yaitu yaitu rumus KR-20 dan rumus KR-21. Rumus KR-20 digunakan untuk menghitung reliabilitas *speedy test*, yaitu tes kecepatan. Sedangkan rumus KR-21 lebih tepat digunakan untuk power test yang memberi kesempatan kepada siswa untuk dapat menampilkan kemampuannya secara maksimal. Prosedur menghitung reliabilitas dilakukan dengan menghubungkan setiap butir dalam satu tes dengan butir-butir lainnya dalam tes itu sendiri secara keseluruhan. Sehingga untuk menganalisis reliabilitas soal tes dalam penelitian ini digunakan rumus Kuder Richardson (KR-20).

### 3. Bentuk Soal Pilihan Ganda.

Pengukuran secara tertulis dilakukan dengan tes tertulis (*paper and pencil test*). Tes tertulis merupakan kumpulan soal-soal yang diberikan kepada siswa dalam bentuk tulisan. Dalam menjawab soal, siswa tidak selalu harus merespon dalam bentuk jawaban, tetapi juga dapat dilakukan dalam bentuk lain seperti member tanda, mewarnai, menggambar dan sejenisnya. Tes tertulis merupakan teknik pengukuran yang banyak digunakan dalam menilai pencapaian kompetensi mata pelajaran sebagai hasil belajar.

Soal tes tertulis dapat diklasifikasikan menjadi dua yaitu soal dengan memilih jawaban yang sudah disediakan (bentuk soal pilihan ganda, benar-salah) dan soal dengan memberikan jawaban secara tertulis (bentuk soal isian, jawaban singkat dan uraian).

Dilihat dari bentuk soalnya, tes tertulis dapat dikelompokkan menjadi tes tertulis

objektif seperti pilihan ganda dan isian, dan tes tertulis non-objektif seperti bentuk soal uraian non-objektif.

Soal pilihan ganda merupakan bentuk soal yang jawabannya dapat dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Kontruksinya terdiri dari pokok soal dan pilihan jawaban. Pilihan jawaban terdiri atas kunci dan pengecoh. Kunci jawaban harus merupakan jawaban benar atau paling benar sedangkan pengecoh merupakan jawaban tidak benar, namun daya jebaknya harus berfungsi, artinya siswa memungkinkan memilihnya jika tidak menguasai materinya.

Soal pilihan ganda dapat diskor dengan mudah, cepat, dan memiliki objektivitas yang tinggi, mengukur berbagai tingkatan kognitif, serta dapat mencakup ruang lingkup materi yang luas dalam suatu tes. Bentuk ini sangat tepat digunakan untuk ujian berskala besar yang hasilnya harus segera diumumkan, seperti ujian nasional, ujian akhir sekolah, dan ujian seleksi pegawai negeri.

Hanya saja, untuk menyusun soal pilihan ganda yang bermutu perlu waktu lama dan biaya cukup besar, disamping itu, penulis soal akan kesulitan membuat pengecoh yang homogen dan berfungsi, terdapat peluang untuk menebak kunci jawaban, dan peserta mudah mencotek kunci jawaban. Secara umum, setiap soal pilihan ganda terdiri dari pokok soal (*stem*) dan pilihan jawaban (*option*). Pilihan jawaban terdiri atas kunci jawaban dan pengecoh (*distractor*).

Dalam penyusunan soal tes tertulis, penulis soal harus memperhatikan kaidah-kaidah penulisan soal dilihat dari segi materi, konstruksi, maupun bahasa. Selain itu soal yang dibuat hendaknya menuntut penalaran yang tinggi.

Hal ini dapat dilakukan antara lain dengan cara : 1) mengidentifikasi materi yang dapat mengukur perilaku pemahaman, penerapan, analisis, sintesis, atau evaluasi. Perilaku ingatan juga diperlukan namun kedudukannya adalah sebagai langkah awal sebelum siswa dapat mengukur perilaku yang disebutkan di atas, 2) membiasakan menulis soal yang mengukur kemampuan berfikir kritis dan mengukur keterampilan

pemecahan masalah; dan 3) menyajikan dasar pertanyaan (stimulus) pada setiap pertanyaan, misalnya dalam bentuk ilustrasi/bahan bacaan seperti kasus, contoh, tabel dan sebagainya.

#### 4. Jumlah Alternatif Jawaban

Dalam setiap tes objektif selalu digunakan alternative jawaban yang mengandung dua unsure sekaligus, yaitu jawaban tepat dan jawaban yang salah sebagai pengecoh (*distractor*). Tujuan pemakaian pengecoh ini adalah mengecohkan mereka yang kurang mampu (tidak tahu) untuk dapat membedakan dengan yang mampu. Oleh karena itu pengecoh yang baik adalah yang dapat dihindari oleh anak-anak yang pandai dan terpilih oleh anak-anak yang kurang pandai, jangan sampai terjadi sebaliknya.

Efektivitas pengecoh-pengecoh yang ada pada suatu item dianalisis dari distribusi jawaban terhadap item yang bersangkutan pada setiap alternative yang disediakan. Efektivitas pengecoh diperiksa untuk melihat apakah semua pengecoh atau semua pilihan jawaban yang bukan kunci telah berfungsi sebagaimana mestinya, yaitu apakah pengecoh-pengecoh tersebut telah dipilih oleh lebih banyak (atau semua) siswa kelompok rendah sedangkan siswa dari kelompok tinggi hanya sedikit (atau tidak ada) yang memilihnya. Pengecoh yang baik paling tidak harus terpilih oleh sedikitnya 2 % (Toha, 1996).

Pusat Penilaian Pendidikan Balitbang Departemen Pendidikan Nasional telah menentukan jumlah alternative jawaban pada soal pilihan ganda. Untuk soal tingkat sekolah dasar (SD) dan sekolah menengah pertama (SMP) jumlah alternative jawaban soal pilihan ganda adalah 4 (empat), sedangkan untuk soal tingkat sekolah menengah atas (SMA) jumlah alternative jawaban yang ditentukan adalah berjumlah 5 (lima).

#### HIPOTESIS PENELITIAN

Berdasarkan hasil kajian pustaka di atas dapat dirumuskan hipotesis sebagai berikut :

1. Tes objektif bentuk pilihan ganda dengan lima alternative jawaban

memiliki indeks reliabilitas, data pembeda dan tingkat kesukaran lebih tinggi bila dibandingkan dengan tes objektif bentuk pilihan ganda yang memiliki empat atau tiga alternative jawaban.

2. Tes objektif bentuk pilihan ganda dengan empat alternative jawaban memiliki indeks reliabilitas, data pembeda dan tingkat kesukaran lebih tinggi bila dibandingkan dengan tes objektif bentuk pilihan ganda yang memiliki tiga alternative jawaban.

**METODE PENELITIAN**

Penelitian menggunakan metode eksperimen. Untuk membuktikan hipotesis, sampel yang digunakan berasal dari satu populasi yang diambil secara random sebanyak 150 siswa-siswi kelas III MTsN 7 Jakarta. Respon peserta didik dalam eksperimen dianalisis dengan program Iteman (Item Analyzis) versi 3.00. Data yang terkumpul berupa indeks kesukaran item, dan daya pembeda item, selanjutnya analisis varians satu jalan dengan program SPSS digunakan untuk mengetahui perbedaan mean antar ketiga variable terhadap tingkat kesukaran dan daya pembeda. Untuk membandingkan koefisien reliabilitas tes dari ketiga variable digunakan transformasi Z Fisher’s.

**HASIL PENELITIAN**

**1. Reliabilitas**

Berdasarkan hasil analisis dengan menggunakan program Iteman versi 3.00 diperoleh alpha untuk peringkat tes dengan 5

alternatif jawaban 0,829, perangkat tes dengan 4 alternatif jawaban 0,707 dan perangkat tes dengan 3 alternatif jawaban 0,591. Ditinjau dari aspek reliabilitas, apabila dijadikan patokan adalah pendapat Renners (1960) yang menyatakan bahwa koefisien reliabilitas 0,5 dapat digunakan untuk tujuan penelitian. Menurut Nunnaly (1972), Kaplan dan Saccuzo (1989) koefisien reliabilitas 0,7 sampai 0,8 cukup tinggi untuk suatu penelitian dasar. Maka perangkat-perangkat tes yang digunakan ini memiliki alpha yang cukup baik digunakan sebagai alat ukur dalam penelitian ini.

Selanjutnya berdasarkan hasil perhitungan secara manual dengan menggunakan rumus transformasi Fisher’s diperoleh:  $Z_1= 1,414$  ;  $Z_2= 2.318$ ; dan  $Z_3= 0,905$  (Minium, 1970), nilai z kritis pada taraf signifikansi 0,05 diperoleh sebesar 1,96. Dengan demikian hanya  $Z_2$  yang memperoleh angka lebih besar dari nilai z kritis. Maka dapat dikatakan bahwa: (1) Tidak terdapat perbedaan indeks reliabilitas yang signifikan antara perangkat tes dengan lima alternatif jawaban dan perangkat tes dengan empat alternatif jawaban dalam taraf signifikansi 0,05. (2) Tidak terdapat perbedaan indeks reliabilitas yang signifikan antara perangkat tes dengan empat alternative jawaban dan perangkat tes dengan tiga alternative jawaban dalam taraf signifikan si 0,05. (3) Ada perbedaan indeks reliabilitas yang signifikan antara perangkat tes dengan lima alternative jawaban dan perangkat tes dengan tiga alternative jawaban dalam taraf signifikansi 0,05.

Tabel.1 data Statistik dengan rumus Fisher’s

Nomor	Variabel	Alpha	Standar Error	Z’
1	$r_1$	0.829	0,22	1,188
2	$r_2$	0,707	0,22	0,877
3	$r_3$	0,591	0,22	0,678

**2. Tingkat Kesukaran**

Indeks kesukaran rata-rata item untuk ketiga perangkat tes adalah sebagai berikut : (1) perangkat tes dengan lima alternative jawaban indeks kesukaran rata-rata 0,484, (2) perangkat tes dengan empat alternative

jawaban, indeks kesukaran rata-rata 0,525, (3) perangkat tes dengan tiga alternative jawaban, indeks kesukaran rata-rata= 0,591. Menurut Suryabrata (1987), tes yang baik ayes yang terdiri dari butir-butir soal yang mempunyai taraf kesukaran sedang. Adapun

yang dimaksud dengan taraf kesukaran sedang adalah perangkat soal yang apabila dijawab oleh orang yang mempunyai kemampuan sedang, harus dapat menjawab dengan benar separoh dari semua butir soal dalam perangkat, dengan mengabaikan factor berhasil secara kebetulan. Dengan menggunakan dasar penentuan tingkat kesukaran di bawah 0,30 termasuk soal yang sukar, 0,30 sampai 0,70 termasuk soal yang sedang dan diatas 0,70 termasuk soal yang

mudah, maka ketiga perangkat tes merupakan perangkat yang cukup baik.

Indeks kesukaran yang diperoleh dengan program analisa Iteman versi 3.0 untuk masing-masing perangkat tes, pengolahan dilanjutkan dengan analisis varians satu jalan untuk memperoleh harga F. Data rangkuman analisis varians satu jalan untuk indeks kesukaran disajikan di bawah ini :

Tabel.2 Rangkuman Anava 1 jalan Tingkat Kesukaran.

Sumber	JK	db	RK	F	p
Antar A	0,2816	2	0,1408	2,395	0,0952
Galat	7.7626	132	0,0588	-	-
Total	8,0442	134			

Berdasarkan hasil analisis varian dengan program SPSS diperoleh  $F = 2,395$  dan  $p = 0,0952$ , ini berarti tidak ada perbedaan yang signifikan pada taraf signifikansi 0,05 antara rata-rata tingkat kesukaran dari ketiga perangkat tes tersebut.

Menurut Popham (1981), Fernandes (1984) bahwa jumlah alternative jawaban dapat mempengaruhi tingkat kesukaran rata-rata. Popham (1981) menyatakan bahwa tingkat kesukaran rata-rata untuk tiga alternative jawaban adalah 0,67, untuk empat alternative adalah 0,63 dan untuk lima alternative jawaban adalah 0,60. Hopkins dan Stanley (1987) menyatakan bahwa, makin banyak alternative jawaban pada soal bentuk pilihan ganda makin sukar soal tersebut. Crocker (1986) menyatakan bahwa indeks kesukaran 0,50 akan menghasilkan korelasi antara item dan skor totalnya terbesar (maksimum) dan untuk tes pilihan ganda dengan lima alternative jawaban tingkat kesukaran yang baik adalah 0,60, untuk empat alternative jawaban 0,62 dan untuk tiga alternative jawaban 0,67. Bilangan ini diperoleh dari  $(0,50 + 0,50/m)$ . 0,50 adalah peluang menjawab benar berdasarkan pengetahuannya ditambah dengan peluang jawaban menerka  $(0,50/m)$  adalah banyaknya alternative jawaban.

### 3. Daya Pembeda

Daya pembeda yang diperoleh dengan program analisis Iteman versi 3,0 untuk

ketiga perangkat tes adalah sebagai berikut : (1) perangkat tes dengan lima alternative jawaban, rerata daya beda = 0,426, (2) perangkat tes dengan empat alternative jawaban, rerata daya beda = 0,355, (3) perangkat tes dengan tiga alternative jawaban, rerata daya beda = 0,295.

Menurut Ebel (1979) bahwa lebih banyak alternative jawaban dapat menyebabkan daya pembeda item lebih tinggi. Korelasi biserial yang dapat diterima minimal 0,30 (Pakpahan, 1990). Nunnally (1972) daya pembeda di atas 0,30 maka item tersebut termasuk kategori dapat membedakan secara maksimal kelompok yang berkemampuan tinggi dan kelompok yang berkemampuan rendah.

Daya pembeda ditinjau dari korelasi biserial rata-rata, apabila besarnya indek dijadikan patokan, maka menurut Nunnally (1972), Popham (1981) dan Nitko (1983) bahwa korelasi di atas 0,30 merupakan butir soal yang baik. Maka ketiga perangkat tes yang digunakan dalam penelitian ini memiliki daya pembeda yang cukup baik.

Korelasi biserial yang diperoleh dengan program Iteman versi 3 untuk masing-masing perangkat tes, pengolahan dilanjutkan dengan analisis varian satu jalan untuk memperoleh harga F dan kemudian dilakukan uji t. Data rangkuman analisis varians satu jalan untuk daya pembeda disajikan dalam table berikut:

Tabel.3 Rangkuman Anava 1 jalan untuk daya pembeda.

Sumber	JK	db	RK	F	p
Antar A	0,2869	2	0,1434	3,4462	0,0347
Galat	5,4909	132	0,0416	-	-
Total	5,7777	134			

Berdasarkan hasil analisis varian satu jalan dengan menggunakan program SPSS diperoleh  $F = 3,4462$  dan  $p = 0,0347$  ini

berarti ada perbedaan mean daya beda yang signifikan para taraf signifikan 0,05 dari ketiga perangkat tes tersebut.

Tabel.4 Rangkuman Anava 1 jalan untuk daya pembeda

Sumber	t	p
A1 – A2	1,51	0,134
A1 – A3	2,56	0,012
A2 – A3	1,11	0,268

Menurut Ebel (1984) dan Gronlund (1985) bila distraktor yang baik digunakan maka makin banyak alternative jawaban makin tinggi daya pembedanya. Selanjutnya bila ditinjau hasil analisis perbedaan mean antara A1A2, A1A3 dan A2A3 dengan menggunakan uji t diperoleh : (1) tidak ada perbedaan mean daya pembeda yang signifikan antara A1 dan A2 pada taraf signifikansi 0,05, (2) Ada perbedaan mean daya pembeda yang signifikan antara A1 dan A3 dalam taraf signifikansi 0,05, (3) tidak ada perbedaan mean daya pembeda yang signifikan antara A2 dan A3 dalam taraf signifikansi 0,05.

**KESIMPULAN DAN SARAN**

**Kesimpulan**

Kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

- a. Semakin banyak jumlah alternative jawaban tes objektif bentuk pilihan ganda semakin tinggi reliabilitas tes.
- b. Dari uji F untuk tingkat kesukaran soal diperoleh bahwa tidak terdapat perbedaan mean yang signifikan antara ketiga perangkat tes, itu berarti jumlah alternative jawaban tes tidak mempengaruhi tingkat kesukaran soal
- c. Dari hasil uji F untuk daya pembeda soal diperoleh bahwa ada perbedaan mean yang signifikan antara ketiga perangkat tes pada taraf signifikansi 0,05 itu berarti semakin banyak alternative jawaban tes objektif bentuk

pilihan ganda semakin tinggi daya pembeda tes.

**Saran**

Berdasarkan kesimpulan diajukan saran-saran lebih lanjut dalam bidang penyusunan soal bentuk pilihan ganda untuk masa yang akan datang sebagai berikut :

- a. Untuk menyusun perangkat tes objektif pilihan ganda, diharapkan para guru dan tenaga kependidikan terkait agar mempertimbangkan tingkat realibilitas, tingkat kesukaran dan daya pembeda soal sehingga perangkat soal yang disusun memiliki kualitas yang baik.
- b. Sebelum menggunakan tes prestasi belajar dalam bentuk pilihan ganda, sebaiknya perangkat tes tersebut diujicobakan terlebih dahulu agar hasil yang diperoleh tidak merugikan siswa
- c. Penelitian ini masih perlu dilanjutkan, dengan menggunakan sampel yang lebih besar.

**DAFTAR PUSTAKA**

Anastasi, Anne. 1988. **Psychological Testing**. New York: Macmillan Publishing Company.

Azwar, Saifuddin. 1999. **Dasar-Dasar Psikometri**. Yogyakarta: Pustaka Pelajar.

Bahan Penataran Pengujian Pendidikan. 1993. Jakarta: Puslitbang Sisjian Balitbang Departemen Pendidikan dan Kebudayaan.

- Brown, Frederick G. 1976. **Principles of Educational and Psychological Testing**. New York: Holt, Rinehart & Winston.
- Crocker, Linda & James Algina. 1986. **Introduction to Classical and Modern Test Theory**. Florida: Holt, Rinehart and Winston, Inc.
- Ebel, Robert L. 1979. **Essentials of Education Measurement**. New York: Prentice-Hall, Inc.
- Fernandes, HJX. 1984. **Testing and Measurement**. Jakarta: National Educational Planning.
- Gronlund, E.N. 1985. **Measurement and Evaluation in Teaching**. New York: Mc-Millan Publihsing Company.
- Hopkins, KD & Stanley, JC. 1978. **Educational and Psychology Measurement and Evaluation and Issues**. California: Englewood Cliffs Prentice-Hall.
- Kaplan, RM & Saccuzo, DP. 1989. **Psychological Testing, Principles Application and Issues**. California: Brooks/Cole Publishing Company.
- Master, Geofferey N & John P. Keeves (ed), 1999. **Advance in Measurement in Education Reseach and Assessment**. United Kingdom: Elsevier Science Ltd.
- Nitko, Anthony J. 1983. **Educational Test and Measurement an Introduction**. New York: Harcourt Brace Javanovich, Inc.
- Popham William. J. 1994. **Classroom Assessment, What Teacher Need to Know**. Boston: Allyn and Bacon.
- Silverius, Suke. 1991. **Evaluasi Hasil Belajar Dan Umpan Balik**. Jakarta: Gramedia Widiasarana Indonesia.
- Surapranata, Sumarna. 2004. **Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes**. Bandung: Rosda.