

## PENGARUH JUMLAH ALTERNATIF JAWABAN DAN TEKNIK PENSKORAN TERHADAP RELIABILITAS TES

**Yoga Budi Bhakti**

Dosen Program Studi Pendidikan Matematika Universitas Indraprasta PGRI  
E-mail : [yogabudi.bhakti@yahoo.co.id](mailto:yogabudi.bhakti@yahoo.co.id)

**Abstrak :** Penelitian ini bertujuan untuk mengetahui pengaruh jumlah alternatif jawaban dan teknik penskoran terhadap reliabilitas tes IPA Terpadu. Bentuk tes ini meliputi tes bentuk pilihan ganda tiga pilihan jawaban dan tes bentuk pilihan ganda empat pilihan jawaban dengan teknik penskoran penalti dan teknik penskoran kompensasi. Populasi dalam penelitian ini adalah siswa SMP Negeri 57, SMP Negeri 58, SMP Negeri 67 dan SMP Negeri 145 Jakarta Selatan. Metode yang digunakan adalah metode eksperimen, sedangkan desain yang digunakan adalah *treatment by level 2 x 2*. Hipotesis penelitian diuji dengan menggunakan analisis varians (ANOVA) 2 jalan. Hasil penelitian ini menyimpulkan bahwa: (1) Reliabilitas tes bentuk pilihan ganda dengan tiga pilihan jawaban lebih tinggi dibandingkan tes bentuk pilihan ganda empat pilihan jawaban, (2) Reliabilitas tes dengan teknik penskoran penalti lebih tinggi dibandingkan reliabilitas dengan teknik penskoran kompensasi. (3) Tidak terjadi interaksi antara jumlah alternatif jawaban dengan teknik penskoran. Hasil dari penelitian ini dapat dijadikan pertimbangan bagi guru untuk memberikan tes dengan menggunakan bentuk tes pilihan ganda 3 pilihan dan 4 pilihan jawaban dengan teknik penskoran penalti.

*Kata kunci: reliabilitas, teknik penskoran dan alternatif jawaban*

**Kata kunci : reliabilitas, teknik skoring, pilihan jawaban.**

### PENDAHULUAN

Tes bentuk pilihan ganda ini telah menjadi model untuk mengukur hasil belajar baik yang bersifat formatif maupun sumatif. Penggunaan tes bentuk ini tidak hanya berhenti pada penilaian lingkup kelas dan institusi, akan tetapi juga pada lingkup yang lebih luas. Tes bentuk pilihan ganda digunakan secara luas karena tes ini mempunyai beberapa keunggulan dibandingkan tes bentuk lainnya. Salah satu bentuk keunggulan tes bentuk pilihan ganda yaitu dapat mencakup banyak pokok bahasan dan mudah penyekorannya.

Teknik penilaian dalam pilihan ganda ada tiga jenis teknik penilaian yaitu teknik penilaian *correct score*, teknik penilaian penalti dan teknik penilaian kompensasi. Secara umum dalam tes yang dilaksanakan di tingkat sekolah Dasar dan Menengah untuk penskorannya dalam tes bentuk pilihan ganda menggunakan teknik penilaian *correct score*, sehingga siswa dapat menerka jawaban tanpa ada rasa khawatir adanya pengurangan nilai jika jawabannya salah.

Jumlah alternatif jawaban pada bentuk tes pilihan ganda hanya didasarkan pada tingkatan jenjang pendidikan saja, misalnya untuk Sekolah Dasar

menggunakan tiga pilihan jawaban, Untuk Sekolah Menengah Pertama menggunakan empat pilihan jawaban dan untuk Sekolah Menengah Atas menggunakan lima pilihan jawaban. Fungsi dari alternatif jawaban ini sebagai pengecoh atau distraktor. Tes objektif pilihan ganda dengan tiga alternatif jawaban, maka tingkat penerkaannya adalah 0,33 sedangkan untuk butir soal dengan empat alternatif jawaban, maka tingkat penerkaannya 0,25 kemudian untuk butir soal dengan lima alternatif jawaban maka tingkat penerkaannya adalah 0,20.

Reliabilitas tes dapat dipengaruhi oleh teknik penilaian dan jumlah alternatif jawaban. Semakin objektif dalam memberikan skor, semakin besar keajegannya. Jika dikaitkan dengan reliabilitas, objektifitas memberi tekanan pada ketetapan sistem pemberian skor sedangkan reliabilitas memberikan penekanan pada hasil. Jelas bahwa keduanya saling mempengaruhi atau saling ketergantungan, sistem pemberian skor mempunyai dampak terhadap ketetapan hasil. Begitu juga dengan Jumlah alternatif jawaban semakin banyak maka semakin besar daya pembedanya serta semakin banyak pengecohnya maka semakin tinggi reliabilitas tes tersebut.

## **TINJAUAN PUSTAKA**

### **Reliabilitas**

Nama lain dari reliabilitas adalah keterpercayaan, keterandalan, keajegan atau kestabilan. Sedangkan yang dimaksud dengan reliabilitas adalah sebagai berikut: Menurut Djaali dan Muljono (2004) reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya. Sukardi (2008) mengatakan, Reliabilitas adalah karakter lain dari hasil evaluasi. Reliabilitas juga dapat diartikan sama dengan konsistensi atau keajegan. Suatu instrumen evaluasi, dikatakan memiliki reliabilitas tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak di ukur. Menurut Surapranata (2004) mengatakan, Reliabilitas atau keajegan suatu skor adalah hal yang sangat penting dalam menentukan apakah tes telah menyajikan pengukuran yang baik.

Demikian pula yang dikatakan oleh Wiersma dan Jurs (1990) bahwa reliabilitas pengukuran adalah hasil tetap (*consistency*), ketetapan dalam mengukur instrumen (tes yang digunakan untuk mengukur). Anastasi (1988) menyatakan bahwa reliabilitas adalah kestabilan skor yang diperoleh dari orang yang sama ketika di uji ulang dengan tes yang sama pada situasi yang berbeda. Atau dengan kata lain, suatu tes dikatakan reliabel jika terhadap hasil yang dicapai pada tes, kemudian diadakan pengujian ulang atau *parallel* pada subyek yang sama, akan menunjukkan hasil yang sama. Kubiszyn dan Borich (2007) mengatakan bahwa istilah reliabilitas sering disamakan dengan *consistency*, *stability* yang pada prinsipnya menunjukkan sejauh mana pengukuran itu dapat memberikan hasil yang relatif tidak berbeda bila dilakukan pengukuran kembali terhadap subyek yang sama. Popham (1995) menyatakan reliabilitas tes menunjukkan stabilitas skor yang diperoleh peserta tes apabila diberikan tes dengan waktu yang berbeda.

Azwar (2006) mengatakan bahwa reliabilitas merupakan penerjemahan dari kata *reliability* yang mempunyai asal kata *rely* dan *ability*. Pengukuran yang memiliki reliabilitas tinggi disebut sebagai pengukuran yang reliabel (*reliable*). Walaupun reliabilitas mempunyai berbagai nama lain seperti keterpercayaan,

keterandalan, keajegan, kestabilan, konsistensi dan sebagainya, namun ide pokok yang terkandung dalam konsep reliabilitas adalah sejauhmana hasil suatu pengukuran dapat di percaya. Menurut sumber Mc Millan (2008), reliabilitas adalah ketetapan hasil pengukuran. Sebuah pengukuran dianggap reliabel jika seseorang mendapat tes yang sama walaupun diberikan dua kali dan hasilnya sama. Cangelosi (1990) mengatakan bahwa supaya absah, tes tidak saja harus relevan tetapi juga reliabel. Sebuah tes dikatakan reliabel apabila dapat memberikan hasil yang ajeg dan tak bertentangan.

Implikasinya hasil pengukuran dapat dipercaya, apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama diperoleh hasil yang relatif sama, selama aspek dalam diri subyek yang diukur memang belum berubah. Jadi reliabilitas menunjukkan sejauh mana hasil pengukuran tetap konsisten bila dilakukan pengukuran yang berulang-ulang terhadap gejala yang sama dengan alat ukur yang sama. Reliabilitas dapat juga dikatakan sebagai tingkat konsistensi atau kemandapan hasil dari dua buah pengukuran terhadap hal yang sama. Hasil pengukuran diharapkan akan sama apabila pengukuran itu diulangi. Reliabilitas suatu tes juga didefinisikan sebagai taraf sejauhmana tes itu sama dengan dirinya atau reliabilitas suatu tes adalah keajegan suatu tes.

Reliabilitas menunjukkan hasil yang diperoleh dengan suatu tes, bukan tes itu sendiri, jadi suatu tes memberikan ukuran reliabilitas yang berbeda-beda tergantung pada kelompok peserta tes dan situasi dimana tes tersebut digunakan. Koefisien atau indeks reliabilitas merupakan korelasi antara dua pengukuran yang sama yang digunakan pada waktu yang berbeda, antara dua alat ukur yang setara, antara dua bentuk tes yang *parallel* dengan waktu yang berbeda dan antara bagian dari alat ukur yang berbeda. Reliabilitas yang baik tergantung pada tujuan atau kegunaan tes serta untuk pengambilan suatu keputusan secara individu, koefisien reliabilitas harus tinggi.

Koefisien reliabilitas cenderung menjadi lebih tinggi apabila individu itu terdiri atas berbagai tingkatan yang berbeda-beda. Apabila suatu tes dirasakan terlampau sukar atau suatu kelompok peserta tes sehingga para peserta tes banyak bekerja dengan menebak, maka derajat akurasi reliabilitas mejadi lebih rendah. Demikian pula apabila suatu tes terlalu mudah untuk suatu kelompok peserta tes, maka tes ini menjadi tidak efektif karena tidak mampu membedakan kecakapan atau kemampuan di antara peserta tes.

Secara umum apabila banyaknya butir soal suatu tes makin besar, maka koefisien reliabilitas hasil tes semakin tinggi, karena makin banyak butir soal berarti faktor menebak semakin kecil. Koefisien reliabilitas ditentukan oleh posisi skor individu dari dua kali tes. Semakin tetap (ajeg) posisi skor yang diperoleh setiap individu maka makin besar koefisien reliabilitas. Hal ini disebabkan makin sukar butir soal dalam suatu kerangka tes, akan makin besar variansi skor yang diperoleh kelompok. Dengan besarnya variansi skor ini akan mengakibatkan makin tingginya reliabilitas tes itu. Sebaliknya semakin kecil variansi skor akan mengakibatkan variansi skor akan mengakibatkan makin rendahnya reliabilitas tes. Jenjang susunan butir soal dari yang mudah ke yang sukar jika tidak dipenuhi akan mempengaruhi reliabilitas tes.

### **Jumlah Alternatif Jawaban**

Soal tes bentuk pilihan ganda terdiri atas soal yang menggambarkan masalah dan serangkaian pilihan atau alternatif yang masing-masing menyatakan jawaban-jawaban yang mungkin untuk soal tersebut. Butir alternatif jawaban mencakup jawaban yang benar (kunci jawaban) dan beberapa jawaban yang salah, yang disebut pengecoh (*distraktor*). Alternatif jawaban itu jumlahnya berkisar antara tiga sampai lima buah. Fungsi distraktor ialah membelokkan perhatian siswa yang kurang pasti sikapnya tentang jawaban yang benar.

Menurut Popham (1995) faktor terpenting yang menentukan jumlah pilihan dalam soal pilihan ganda ialah jumlah distraktor yang tepat yang biasa di buat. Memasukkan distraktor yang tidak masuk akal atau tidak parallel dengan isi pilihan-pilihan lain akan menyebabkan soal yang sangat membingungkan. Namun harus diperhatikan pula bahwa pengecoh-pengecoh yang akan ditambahkan itu sama baiknya dengan pengecoh yang telah ada. Pertimbangan mengenai beberapa pengecoh digunakan adalah pertimbangan ekonomis mengenai waktu yang terbatas, waktu yang diperlukan oleh pengikut tes untuk membaca akan berpengaruh terhadap banyaknya soal yang dapat dijawab. Tambahan pengecoh berarti tambahan waktu untuk membaca, jadi semakin banyak jumlah pengecoh semakin kecil jumlah soal yang dapat di cakup dalam waktu tertentu. Jika pengecoh yang di tambahkan itu tidak benar-benar meningkatkan mutu soal, maka akibatnya akan merugikan reliabilitas tes.

Mahrens dan Lehmann (1991) menyatakan, hal lain yang mungkin dipertimbangkan dalam menentukan banyaknya pengecoh adalah (1) usia peserta tes, dan (2) materi yang diujikan. Peserta tes muda usia, misalnya murid sekolah dasar, tidak dapat diharapkan cukup mampu menghadapi soal-soal dengan lima atau lebih kemungkinan jawaban. Nitko (1996) menyatakan bahwa lebih banyak alternatif jawaban, maka soal tersebut lebih cenderung membedakan kelompok yang mempunyai kemampuan yang diharapkan dengan kelompok yang tidak.

Tes objektif bentuk pilihan ganda dengan tiga alternatif jawaban memiliki tingkat kesukaran lebih tinggi dibandingkan dengan yang memiliki empat atau lima alternatif jawaban. Tes objektif pilihan ganda dengan tiga alternatif jawaban, maka tingkat penerkaannya adalah 0,33 sedangkan untuk butir soal dengan empat Alternatif jawaban, maka tingkat penerkaannya 0,25 kemudian untuk butir soal dengan lima alternatif jawaban maka tingkat penerkaannya adalah 0,20.

Bila dilihat dari segi efisiensi dan dari reliabilitas tes menurut Frederic Lord (1999) Jumlah optional pilihan dalam tes pilihan ganda adalah tiga. Ia mengatakan bahwa jika penambahan jumlah pilihan tidak menambah waktu dan biaya, dari pertimbangan umum nampaknya bahwa makin banyak pilihan makin baik. Alasan matematis tentang hal ini diberikan Tversk yang juga menyimpulkan bahwa tiga pilihan jawaban tersebut adalah optimal.

### **Teknik Penskoran**

Pada hakikatnya penskoran adalah suatu proses perubahan jawaban instrumen menjadi angka-angka yang merupakan nilai kuantitatif dari suatu jawaban terhadap butir dalam instrumen. Jadi penskoran merupakan kuantifikasi terhadap jawaban instrumen. Dengan memberikan skor dapat diperoleh deskripsi tentang nilai atau harga suatu variabel untuk masing-masing unit analisis dalam

penelitian. Selanjutnya dapat dilakukan analisis kuantitatif dalam kaitannya dengan variabel lain dalam penelitian.

Menurut Djaali dan Mulyono (2004) setelah memberi skor kepada siswa maka selanjutnya adalah menentukan skor akhir berdasarkan skor mentah atau dengan kata lain menterjemahkan skor mentah ke dalam skor akhir atau nilai yang akan menentukan kelulusan. Dalam memberi nilai ada dua cara yang biasa dilakukan, yaitu :

1. Pemberian nilai berdasarkan acuan patokan atau standar mutlak
2. Pemberian nilai berdasarkan acuan kelompok atau acuan norma.

Walaupun kedua cara ini tidak hanya berlaku pada pemberian nilai, tetapi juga ikut menentukan cara penyusunan dan pengembangan tes.

Menurut Crocker dan Algina (1986) Dalam tes pilihan ganda penskoran dapat dilakukan dengan berbagai cara penskoran. Tiga metode yang paling banyak di kenal pada penskoran tes pilihan ganda adalah (1) *Numbering right scoring*,  $S=R$  (2) *Conventional correction for guessing formula*, dan (3) *Omit-corrected formula scoring*, dengan S adalah skor tes, W jumlah butir yang dijawab salah, U jumlah butir yang tidak dijawab dan C adalah jumlah pilihan jawaban.

#### **Model penskoran punishment score**

Model penskoran punishment merupakan model koreksi terhadap guessing. Model ini didasarkan pada dua asumsi, *pertama*, semua tebakan bersifat buta, masing-masing pilihan memiliki peluang yang sama untuk dipilih. Asumsi ini menolak kemungkinan bahwa seseorang peserta tes melakukan tebakan menggunakan teknik partial information yang menyebabkan peluang menjawab benar menjadi lebih besar dari  $1/k$ , dengan k jumlah pilihan jawaban. Bila model ini diterapkan pada kondisi ini maka skor terkoreksi yang diperoleh menjadi underestimate. *Kedua*, setiap jawaban yang salah diperoleh melalui hasil tebakan. Seorang peserta tes dianggap tidak pernah merespon berdasarkan kesalahan informasi atau kesalahan menuliskan jawaban pada lembar jawaban. Asumsi ini mengabaikan kemungkinan menjawab salah yang disebabkan kesalahan memahami soal.

Model punishment score merupakan model penskoran yang memperhitungkan jawaban salah yang direspon oleh peserta tes dengan jalan memberi hukuman dalam bentuk pengurangan skor dengan menggunakan rumus yang diajukan oleh Brown dan Guiford (1983) dalam Crocker dan Algina:

$$X_c = R - \frac{W}{k - 1}$$

Dengan  $X_c$  skor koreksi,  $W$  = jumlah jawaban salah dan k = jumlah alternatif pilihan jawaban. Algina menyebut penskoran ini dengan nama right minus wrongs correction. Asumsi dasar dari penggunaan rumus ini adalah jawaban salah merupakan hasil tebakan, sehingga jumlah yang salah dibagi dengan k-1 merupakan hukuman bagi peserta tes yang menjawab dengan tebakan.

#### **Model penskoran reward score**

Model penskoran ini memperhitungkan butir yang tidak diisi atau dikosongkan oleh peserta tes dengan jalan memberi hadiah dalam bentuk tambahan skor. Model penskoran ini didasarkan pertimbangan 3 kemungkinan

situasi: (1) peserta ujian mengetahui pilihan jawaban yang benar dan memilhnya, (2) peserta tidak memilih sama sekali pilihan jawaban yang disediakan dan (3) peserta ujian menebak buta dan memilih satu dari k pilihan jawaban secara acak. Berdasarkan pada model tebakn acak ini, rumus yang diterapkan untuk mempertimbangkan pengaruh tebakn untuk mengoreksi skor pada model correct score.

$$X_c = R + \frac{O}{k}$$

Dengan  $X_c$  = skor koreksi O = jumlah butir yang tidak dijawab dan k = jumlah pilihan jawabn setiap butir. Rumus penskoran ini disebut juga sebagai penskoran dengan hadiah atau reward score atau penskoran dengan kompensasi karena peserta yang tidak menjawab diberi kompensasi untuk menyamakan dengan peserta yang menebak betul butir tersebut.

Alasan rasional terhadap kedua rumus penskoran adalah rumus tersebut dapat meningkatkan reliabilitas dan validitas karena skor yang dikoreksi merupakan skor estimasi peserta tes yang lebih baik dalam pengukuran laten trait daripada skor yang diperoleh tanpa koreksi. Sebagai ilustrasi penggunaan rumus penskoran untuk tiga peserta tes yang menjawab 20 butir dengan empat pilihan dan tiga pilihan.

#### Empat Pilihan Jawaban

**Tabel 1. Simulasi Perhitungan Teknik Penskoran Kompensasi dan Penalti pada 4 Pilihan Jawaban**

Peserta	Jumlah benar	Tidak jawab	Salah	$X_c = R + \frac{O}{k}$	$X_c = R - \frac{W}{k-1}$
A	14	0	6	$X_c = 14 + \frac{0}{4} = 14$	$X_c = 14 - \frac{6}{4-1} =$
B	14	6	0	$X_c = 14 + \frac{6}{4} = 15.5$	$X_c = 14 - \frac{0}{4-1} =$
C	14	3	3	$X_c = 14 + \frac{3}{4} = 14.75$	$X_c = 14 - \frac{3}{4-1} =$

**Tiga Pilihan Jawaban**

**Tabel 2. Simulasi Perhitungan Teknik Penskoran Kompensasi dan Penalti pada 3 Pilihan Jawaban**

Peserta	Jumlah benar	Tidak jawab	Salah	$X_c = R + \frac{O}{k}$	$X_c = R - \frac{W}{k - 1}$
A	14	0	6	$X_c = 14 + \frac{0}{3} = 14$	$X_c = 14 - \frac{6}{3 - 1} = 11$
B	14	6	0	$X_c = 14 + \frac{6}{3} = 16$	$X_c = 14 - \frac{0}{3 - 1} = 14$
C	14	3	3	$X_c = 14 + \frac{3}{3} = 15$	$X_c = 14 - \frac{3}{3 - 1} = 12.5$

Dari tabel tersebut terlihat masing-masing peserta tes memiliki skor yang sama yakni 14 jika digunakan correct score. Jika digunakan formula *Reward Score* (RS) maupun *Punishment Score* (PS) akan mendapatkan skor yang berbeda. Pada rumus RS yang menerapkan tambahan skor untuk butir yang tidak dijawab oleh peserta sehingga skor meningkat untuk peserta yang sedikit melakukan kesalahan. Sedangkan rumus PS menerapkan sanksi kepada peserta yang menjawab salah sehingga semakin banyak skor salah makin banyak pengurangan. Jika dilihat kedua rumus baik RS ataupun PS, skor ketiga peserta berbeda tetapi peringkat skor mereka tidak berbeda baik untuk soal empat pilihan ataupun dengan tiga pilihan, ini menunjukkan bahwa kedua rumus tersebut dapat diterapkan secara bersamaan.

**METODOLOGI**

Penelitian ini termasuk ke dalam penelitian eksperimen dengan *treatment by level 2 x 2*.

Teknik Penskoran (B)	Jumlah Alternatif Jawaban (A)	
	3 Pilihan Jawaban (A <sub>1</sub> )	4 Pilihan Jawaban (A <sub>2</sub> )
Skor penalti (B <sub>1</sub> )	A <sub>1</sub> B <sub>1</sub> (r <sub>1</sub> , r <sub>2</sub> , ..... r <sub>30</sub> )	A <sub>2</sub> B <sub>1</sub> (r <sub>1</sub> , r <sub>2</sub> , ..... r <sub>30</sub> )
Skor kompensasi (B <sub>2</sub> )	A <sub>1</sub> B <sub>2</sub> (r <sub>1</sub> , r <sub>2</sub> , ..... r <sub>30</sub> )	A <sub>2</sub> B <sub>2</sub> (r <sub>1</sub> , r <sub>2</sub> , ..... r <sub>30</sub> )

**Gambar 1. Desain Penelitian *Treatment by Level 2 x 2***

Keterangan:

- A<sub>1</sub> = Bentuk tes pilihan ganda 3 pilihan jawaban
- A<sub>2</sub> = Bentuk tes pilihan ganda 4 pilihan jawaban
- B<sub>1</sub> = Teknik penskoran dengan skor penalti
- B<sub>2</sub> = Teknik penskoran dengan skor kompensasi
- A<sub>1</sub>B<sub>1</sub> = Reliabilitas soal pilihan ganda 3 pilihan jawaban dengan teknik penskoran skor penalti (X<sub>1</sub>)

- $A_1B_2$  = Reliabilitas soal pilihan ganda 3 pilihan jawaban dengan teknik penskoran skor kompensasi  
 $A_2B_1$  = Reliabilitas soal pilihan ganda 4 pilihan jawaban dengan teknik penskoran skor penalti ( $X_3$ )  
 $A_2B_2$  = Reliabilitas soal pilihan ganda 4 pilihan jawaban dengan teknik penskoran skor kompensasi  
 $r_1, r_2, \dots, r_{30}$  = koefisien reliabilitas dengan replikasi sebanyak 30 kali

## Variabel Penelitian

### 1. Variabel bebas

Variabel bebas terdiri dari satu variabel aktif (perlakukan) dan satu variabel atribut (variabel moderator). Variabel aktifnya adalah jumlah alternatif pilihan bentuk tes pilihan ganda yang terdiri dari bentuk tes pilihan ganda tiga pilihan jawaban dan bentuk tes pilihan ganda empat pilihan jawaban. Sedangkan variabel atributnya adalah teknik penskoran yang terdiri dari skor kompensasi dan skor penalti.

### 2. Variabel terikat

Variabel terikat dalam penelitian ini adalah reliabilitas.

## Populasi dan Sampel

### 1. Populasi Penelitian

Objek penelitian adalah instrumen tes, yaitu membandingkan dua buah tes dengan bentuk soal pilihan ganda tiga pilihan jawaban dan pilihan ganda empat pilihan jawaban, maka populasi penelitian ini adalah populasi peserta tes sebagai responden bagi pengujian kedua bentuk tes tersebut, supaya diperoleh data empirik untuk keperluan membandingkan kedua tes ditinjau dari reliabilitas.

Populasi penelitian adalah seluruh siswa SMP Negeri Jakarta Selatan tahun ajaran 2012/2013. Sedangkan populasi target dalam penelitian ini adalah seluruh siswa SMP Negeri 57, SMP Negeri 58, SMP Negeri 67 dan SMP Negeri 145 Jakarta Selatan.

### 2. Sampel Penelitian

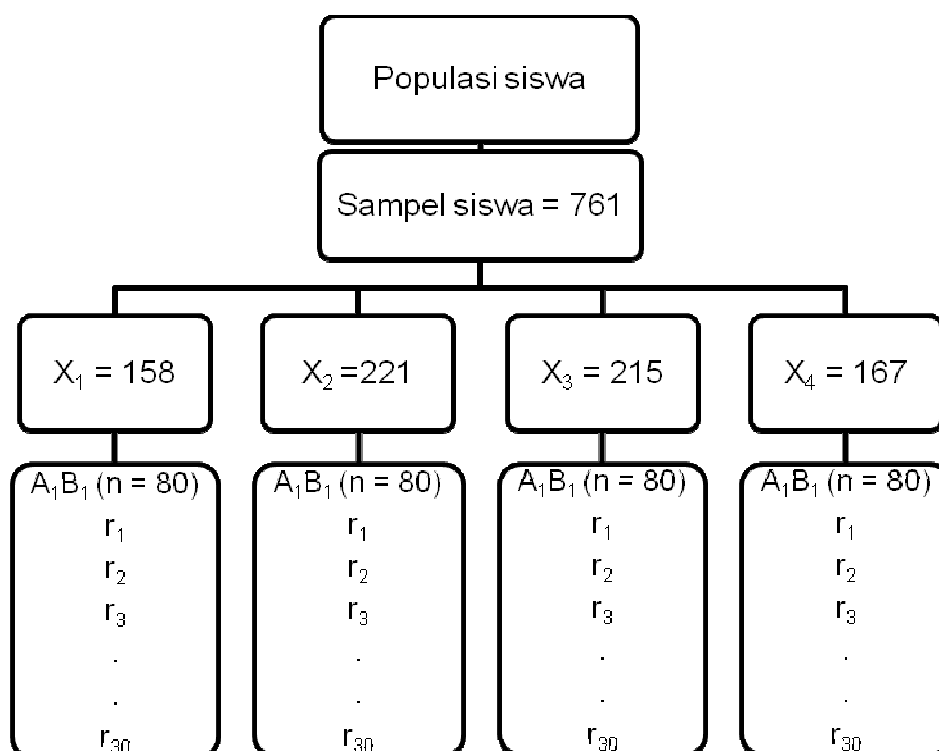
Sampel penelitian adalah siswa siswa SMP Negeri 57, SMP Negeri 58, SMP Negeri 67 dan SMP Negeri 145 Jakarta Selatan Kelas VIII semester ganjil (semester 1). Sampel diambil dengan teknik *simple random sampling*. *Simple random sampling* dilakukan pengambilan anggota sampel dari populasi dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu. Cara demikian dilakukan bila anggota populasi dianggap homogen.

Secara rinci, penarikan sampel dan pengelompokan sampel dilakukan sebagai berikut:

- a. Jumlah responden sebanyak 761 siswa yang digunakan dikelompokkan menjadi 4 bagian yaitu,
  - 1) Kelompok 1 adalah siswa yang mengerjakan bentuk tes pilihan ganda 3 pilihan jawaban dengan skor penalti sebanyak 158 siswa.



- 2) Kelompok 2 adalah siswa yang mengerjakan bentuk tes pilihan ganda 3 pilihan jawaban dengan skor kompensasi sebanyak 221 siswa.
  - 3) Kelompok 3 adalah siswa yang mengerjakan bentuk tes pilihan ganda 4 pilihan jawaban dengan skor penalti sebanyak 215 siswa.
  - 4) Kelompok 4 adalah siswa yang mengerjakan bentuk tes pilihan ganda 4 pilihan jawaban dengan skor kompensasi sebanyak 167 siswa.
- b. Kemudian masing-masing kelompok dipilih secara acak sebanyak 80 responden untuk dihitung reliabilitasnya dengan menggunakan teknik pengambilan sampel secara acak dengan pengembalian (*sampling with replacement*). Secara rinci yaitu diambil secara acak sebanyak 80 responden dari masing-masing kelompok untuk selanjutnya dihitung reliabilitasnya. Kemudian 80 responden tersebut dikembalikan ke kelompoknya lagi untuk ditarik 80 responden secara acak untuk dihitung reliabilitasnya, demikian seterusnya dilakukan secara berulang-ulang sebanyak 30 kali sehingga didapat 30 reliabilitas untuk setiap kelompok. Jumlah total reliabilitas yang akan dianalisis adalah sebanyak 120 reliabilitas. Dengan demikian yang menjadi unit analisis pada penelitian ini adalah kelompok reliabilitas. Proses penarikan sampel disajikan pada bagan sebagai berikut:



**Gambar 2. Proses Penarikan Sampel**

Keterangan:

A<sub>1</sub>B<sub>1</sub> = Kelompok siswa mengerjakan bentuk tes pilihan ganda 3 pilihan jawaban dengan skor penalti

A<sub>2</sub>B<sub>1</sub> = Kelompok siswa mengerjakan bentuk tes pilihan ganda 4 pilihan jawaban dengan skor penalti

- $A_1B_2$  = Kelompok siswa mengerjakan bentuk tes pilihan ganda 3 pilihan jawaban dengan skor kompensasi  
 $A_2B_2$  = Kelompok siswa mengerjakan bentuk tes pilihan ganda 4 pilihan jawaban dengan skor kompensasi  
 $r_1 \dots r_{30}$  = Reliabilitas tes pada sel sebanyak 30 kali

### HASIL PENELITIAN

Secara umum penelitian ini bertujuan untuk mengetahui pengaruh bentuk tes pilihan ganda dengan teknik penskoran ditinjau dari reliabilitas. Perlakuan diterapkan pada empat sekolah yang terdapat di Jakarta Selatan yaitu, SMP Negeri 57, SMP Negeri 58, SMP Negeri 67 dan SMP Negeri 145. Setiap sekolah yang digunakan sebagai tempat penelitian diberikan *treatment* yang sama yaitu dalam setiap sekolah siswa diberikan ulangan dengan materi pertumbuhan dan perkembangan makhluk hidup yang terdiri dari 4 paket soal. Paket 1 berupa seperangkat butir dengan bentuk tes pilihan ganda tiga pilihan jawaban dengan penskoran penalti. Paket 2 berupa seperangkat butir dengan bentuk tes pilihan ganda tiga pilihan jawaban dengan penskoran kompensasi. Paket 3 berupa seperangkat butir dengan bentuk tes pilihan ganda empat pilihan jawaban dengan penskoran penalti. Paket 4 berupa seperangkat butir dengan bentuk tes pilihan ganda empat pilihan jawaban dengan penskoran kompensasi.

Pada uji persyaratan normalitas, diperoleh bahwa semua kelompok analisis berdistribusi normal, yang berarti sampel berasal dari populasi yang berdistribusi normal. Sehingga sampel telah mewakili karakteristik populasi. Sedangkan pada uji persyaratan homogenitas varians, diperoleh bahwa kelompok-kelompok analisis memiliki varians yang homogen.

Hasil penelitian ini menunjukkan bahwa secara keseluruhan terdapat perbedaan reliabilitas antara kelompok siswa yang diberikan bentuk tes pilihan ganda tiga pilihan jawaban dengan kelompok siswa yang diberikan bentuk tes pilihan ganda empat pilihan jawaban. Penerapan bentuk tes dan pemberian teknik penskoran yang bervariasi memberikan konsekuensi pada perbedaan reliabilitas. Berdasarkan hasil analisis data yang telah dilakukan, maka dapat dikemukakan beberapa hal sebagai berikut:

Ringkasan hasil perhitungan analisis data Uji ANAVA dua jalur dapat dilihat pada tabel berikut:

**Tabel 3. Ringkasan Hasil Perhitungan Anava**

SUMBER VARIAN	db	JK	RK=JK/db	Fh=RK/RDK	Ft
<b>Antar Baris (JKAb)</b>	<b>1</b>	<b>0,018</b>	<b>0.018</b>	<b>9.491</b>	<b>3.91</b>
<b>Antar Kolom (JKAk)</b>	<b>1</b>	<b>0,040</b>	<b>0.040</b>	<b>20.563</b>	<b>3.91</b>
<b>Interaksi (bxk)(JKAi)</b>	<b>1</b>	<b>0,01</b>	<b>0.01</b>	<b>5.27</b>	<b>3.91</b>
<b>Dalam (JKD)</b>	<b>116</b>	<b>0,220</b>	<b>0.0019</b>	<b>1</b>	
<b>Total Dikoreksi (JKTR)</b>	<b>119</b>	<b>0,274</b>	<b>0.0023</b>		

Berdasarkan hasil perhitungan ANAVA dua jalan pada tabel 3, maka dapat dilakukan pengujian hipotesis sebagai berikut:

#### **Perbedaan Reliabilitas Bentuk Tes Pilihan Ganda 3 Pilihan dengan Tes Pilihan Ganda 4 Pilihan**

Berdasarkan hasil analisis varians (ANAVA) pada tabel di atas dan dihitung dari hasil perhitungan *Microsoft Excel* dan perhitungan manual diperoleh  $F_{hitung}$  yang sama, diperoleh  $F_{hitung} = 9,491$  ; sedangkan  $F_{tabel}$  pada taraf signifikansi = 0,05 adalah 3,91. Jadi  $F_{hitung} > F_{tabel}$ , dengan demikian  $H_0$  ditolak. Hal ini berarti terdapat perbedaan rata-rata reliabilitas antara kelompok siswa yang mengerjakan bentuk tes pilihan ganda 3 pilihan jawaban dan bentuk tes pilihan ganda 4 pilihan jawaban.

Dengan demikian dapat disimpulkan bahwa, rata-rata reliabilitas antara kelompok siswa yang mengerjakan bentuk tes pilihan ganda 3 pilihan lebih tinggi daripada reliabilitas pada kelompok siswa yang mengerjakan bentuk tes pilihan ganda 4 pilihan.

#### **Perbedaan Reliabilitas Bentuk Tes Pilihan Ganda Teknik Penskorannya Penalti dengan Teknik Penskorannya Skor kompensasi**

Berdasarkan hasil analisis varians (ANAVA) pada tabel 3 dan dihitung dari hasil perhitungan *Microsoft Excel* dan perhitungan manual diperoleh  $F_{hitung}$  yang sama, diperoleh  $F_{hitung} = 20,563$ ; sedangkan  $F_{tabel}$  pada taraf signifikansi = 0,05 adalah 3,91. Jadi  $F_{hitung} > F_{tabel}$ , dengan demikian  $H_0$  ditolak. Hal ini berarti terdapat perbedaan rata-rata reliabilitas antara kelompok siswa yang teknik penskorannya dengan skor penalti dan skor kompensasi.

Dengan demikian dapat disimpulkan bahwa, rata-rata reliabilitas antara kelompok siswa yang teknik penskorannya dengan skor penalti lebih tinggi daripada reliabilitas pada kelompok siswa yang teknik penskorannya dengan skor kompensasi.

#### **Interaksi antara Reliabilitas Bentuk Tes Pilihan Ganda (A) dengan Teknik Penskoran (B)**

Berdasarkan hasil analisis varians (ANAVA) pada tabel 3 yang dihitung dari hasil perhitungan *Microsoft Excel* dan perhitungan manual diperoleh  $F_{hitung}$  yang sama, diperoleh  $F_{hitung} = 5,271$  ; sedangkan  $F_{tabel}$  pada taraf signifikansi = 0,05 adalah 3,91. Jadi  $F_{hitung} > F_{tabel}$ , dengan demikian  $H_0$  ditolak. Hal ini berarti tidak terjadi interaksi antara reliabilitas bentuk tes pilihan ganda dengan teknik penskoran.

Tidak terjadi interaksi antara kedua variabel tersebut, yaitu bentuk tes pilihan ganda dan teknik penskoran (variabel bebas) ditinjau dari reliabilitas (variabel terikat).

### **PEMBAHASAN**

Soal bentuk tes pilihan ganda merupakan bentuk tes yang dapat mengukur berbagai macam kemampuan, mulai dari yang sederhana sampai dengan kemampuan yang rumit. Tes bentuk pilihan ganda dapat digunakan untuk mengukur kemampuan yang secara umum tidak dapat diukur oleh bentuk tes lainnya.

Tes bentuk pilihan ganda dengan tiga pilihan jawaban memiliki reliabilitas lebih tinggi dibandingkan dengan tes bentuk pilihan ganda empat pilihan jawaban karena ditinjau dari tingkat penerkaannya jumlah 3 pilihan jawaban yaitu 0.33 lebih tinggi dibandingkan dengan jumlah 4 pilihan jawaban yaitu 0.25. Dengan semakin banyak pilihan jawaban, maka akan semakin banyak juga jumlah pengecohnya atau distraktor. Pengecoh jawaban jika tidak bekerja dengan baik, justru akan menurunkan reliabilitas tes tersebut.

Penskoran tes bentuk pilihan ganda dalam penelitian ini dilakukan dengan teknik pensoran penalti dan teknik penskoran kompensasi. Penskoran penalti untuk tes pilihan ganda memiliki reliabilitas lebih tinggi dibandingkan dengan Penskoran Kompensasi. Hal ini dikarenakan dalam teknik penskoran penalty yang diperhitungkan adalah jumlah jawaban salah sebagai denda pengurang nilainya, dengan penskoran ini peserta didik lebih berhati-hati dalam menjawab soal tes sehingga dapat meningkatkan reliabilitas tes tersebut. Penskoran penalti membuat siswa yang tidak mengetahui dan tidak memiliki pengetahuan pada soal tersebut, mereka tidak akan menjawabnya karena khawatir jawabannya akan salah dan mengurangi nilai mereka. Berbeda dengan penskoran penalti, dalam penskoran kompensasi jumlah butir soal yang tidak dijawab diperhitungkan sebagai penambah skor peserta didik. Teknik penskoran ini tidak membuat peserta didik hati-hati dalam menjawab, sehingga masih banyak yang melakukan tebakan karena mereka tidak khawatir akan adanya pengurangan nilai. Hal ini menyebabkan teknik penskoran kompensasi memiliki reliabilitas lebih rendah dibandingkan teknik penskoran penalti.

Berdasarkan penelitian jumlah tiga pilihan jawaban dan teknik penskoran penalty memiliki reliabilitas yang lebih tinggi. Dengan tiga pilihan jawaban, membuat pilihan lebih efisien dari segi waktu pengerjaan dan dalam pembuatan distraktor atau pengecoh. Sedangkan dengan teknik penskoran penalti, peserta didik lebih cenderung takut jika jawabannya salah dan akan dikenakan denda sehingga dapat membuat reliabilitas pada teknik penskoran ini tinggi.

## SIMPULAN

Berdasarkan hasil dan pembahasan penelitian, maka dapat disimpulkan bahwa Reliabilitas tes bentuk pilihan ganda dengan tiga pilihan jawaban lebih tinggi dibandingkan tes bentuk pilihan ganda empat pilihan jawaban, Reliabilitas tes dengan teknik penskoran penalti lebih tinggi dibandingkan reliabilitas dengan teknik penskoran kompensasi, Tidak terjadi interaksi antara jumlah alternatif jawaban dengan teknik penskoran.

Bila dilihat dari reliabilitas, pada semua model penskoran dan jumlah pilihan jawaban yang memiliki reliabilitas paling tinggi adalah tes bentuk pilihan ganda dengan tiga pilihan jawaban dan skor penalti.

Ditinjau dari model penskoran, pada umumnya siswa yang dikoreksi dengan menggunakan model penskoran penalti dan penskoran kompensasi menunjukkan hasil lebih akurat sesuai dengan kemampuannya dan jawaban mereka mencerminkan kemampuan mereka. Model penskoran penalty (*Punishment Score*) dan penskoran kompensasi (*Reward Score*) akan membuat siswa hati-hati dan membatasi diri tidak melakukan tebakan acak terhadap butir yang betul-betul tidak diketahuinya.

## SARAN

Pengukuran hasil belajar selama ini dilakukan di tingkat Sekolah Menengah Pertama (SMP) menggunakan tes bentuk pilihan ganda dengan jumlah empat pilihan jawaban dan penskoran skor benar. Dengan tes selama ini tidak menggambarkan kemampuan peserta didik yang sebenarnya karena banyak peserta didik melakukan tebakan dalam memilih jawaban.

Diharapkan dari penelitian yang dilakukan dapat diterapkan oleh guru di sekolah dalam penyusunan instrumen tes kepada peserta didik. Sehingga dengan tes dengan teknik penskoran penalti dapat menggambarkan kemampuan peserta didik serta dapat untuk menganalisa butir soal yang masih belum dikuasai peserta didik. Sehingga dapat dilakukan remedial atau pengayaan pada materi tersebut.

Sebagai seorang pendidik marilah kita membuat inovasi pembelajaran khususnya dalam hal penyusunan instrumen tes untuk peserta didik dengan menerapkan teknik penskoran yang dapat menggambarkan kemampuan sebenarnya dari peserta didik. Meskipun evaluasi akhir Ujian Nasional tingkat SMP masih menggunakan jumlah empat pilihan jawaban dengan penskoran skor benar.

## DAFTAR PUSTAKA

- Anastasi, Anne. 1988. *Psychological Testing*. New York : Mac Millan Publishing Company.
- Azwar, Saifuddin. 2006. *Reliabilitas dan Validitas*. Yogyakarta : Pustaka Pelajar.
- Cangelosi, James S. 1990. *Evaluating Classroom Instruction*. New York : Publishing Group.
- Crocker, L. and Algina, J. 1986. *Introduction to Classical and Modern Test Theory*. New York : Holt, Reinhart dan Wiston Inc.
- Djaali dan Pudji Muljono. 2004. *Pengukuran dalam Bidang Pendidikan*. Jakarta : Program Pascasarjana UNJ.
- Kubiszyn, Tom and Borich, Gary. 2007. *Educational Testing and Measurement: Classroom Application and Practice*. United States : John Wiley and Sons, Inc.
- Mahrens, W.A. and Lehmann, J. 1991. *Measurement and Evaluation in Education and Psychology*. New York : Brace College Publisher.
- McMillan, James H. 2008. *Assesment Essentials for Standard-Based Education*. California: Corwin Press.
- Nitko, J. Anthony. 1996. *Educational Assessment of Student*. New Jersey : Prentise-Hall, Inc.

Popham, W. James. 1995. *Classroom Assesment: What Teachers Need to Know*. California : Allyn and Bacon.

Sukardi. 2008. *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta : Bumi Aksara.

Surapranata, Sumarna. 2004. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Bandung : Remaja Rosdakarya.

Surapranata, Sumarna. 2004. *Panduan Penulisan Tes Tertulis*. Bandung : PT Remaja Rosdakarya.

Wiersma, William dan Stephen G. Jurs. 1990. *Educational Measurement and Testing*. Boston: Allyn and Boston.