

Vol. 2 No. 1

**RE-ANALISIS TINGKAT KEBAIKAN ITEM TES INTELIGENSI:  
ADVANCED PROGRESIVE MATRICES**

yayasunarya@upi.edu

**Yaya Sunarya  
Universitas Pendidikan Indonesia**

**ABSTRAK**

Selain dari aspek budaya, sumber bias atau kesalahan menafsirkan hasil suatu asesmen adalah dari tes itu sendiri (validitas, reliabilitas, daya pembeda dan pengecoh yang tidak berfungsi). Oleh karena itu tes yang standar (*standardized test*) adalah tes yang sudah dianalisis berbagai aspeknya dari waktu ke waktu, sehingga unsur kejegan, ketepatan, daya bedanya teruji. Hasil penelitian menunjukkan bahwa APM dilihat dari tingkat kesukarannya memiliki prosentase soal yang hampir seimbang antara soal yang mudah, sedang, dan sukar secara berimbang, setiap soal memiliki indeks homogenitas yang baik, daya pembeda setiap soal secara umum baik, dan pilihan-pilihan jawabannya telah berfungsi secara baik pula (tidak ada yang tidak pernah tidak dipilih).

*Kata kunci: tingkat kesukaran, daya beda, dan pengecoh.*

**A. LATAR BELAKANG**

Beberapa ahli asesmen menjelaskan, bahwa tujuan akhir dari suatu pelaksanaan asesmen atau penilaian (apakah asesmen psikologis atau pun asesmen hasil belajar) adalah pengambilan keputusan tentang individu (Murphy, 1998), tentang proses dan hasil belajar (Nitko, 1996), atau tentang aspek yang diukur, yang sering disebut dengan istilah *atribut psikologis* (Crocker, 1986) dengan tes sebagai alat ukurnya (Wright, 2011). Dengan hasil tes itulah seseorang (guru, konselor, psikolog atau helper) menginterpretasikan kualitas perilaku atau atribut psikologis seseorang dalam bentuk "label". Label-label ini lah yang nantinya menjadi ciri (trait) seseorang.

Menurut Nitko (1996: 3 – 4) dalam bidang pendidikan, beberapa keputusan yang dapat dilakukan/diambil oleh guru dan orang lain yang terkait, dapat dikelompokkan pada kegiatan sebelum, selama dan sesudah pembelajaran dilakukan. Sebelum pembelajaran guru memutuskan materi apa yang harus dipelajari, kegiatan apa yang harus dilakukan; selama belajar, mengarahkan aktivitas siswa dan guru untuk mencapai tujuan, dan sesudah pembelajaran guru menentukan siapa yang perlu pengayaan, siapa yang masuk percepatan (akselerasi), dan siapa yang perlu dibimbing.

Dewasa ini hasil tes (terutama hasil ujian akhir) di sekolah memegang peranan sangat penting. Di setiap akhir tahun sering menjadi masalah yang cukup pelik. Pentingnya hasil ujian nasional ini terkait dengan kelulusan dan lanjutan studi. Banyak orang tua yang berusaha "menggoda" para guru untuk melakukan kecurangan atau

*Jurnal Pendidikan Pascasarjana Magister PAI*

**Vol. 2 No. 1**

pemalsuan. Ini menunjukkan pengambilan keputusan dari hasil tes ini sangat menentukan.

Dilain pihak hasil tes itu tidak hanya terkait dengan prestasi hasil belajar, tetapi terkait dengan mengidentifikasi kemampuan dasar umum (inteligensi) dan kemampuan khusus (bakat) peserta didik. Pemerintah sangat menghargai peserta didik yang memiliki kecerdasan istimewa dan bakat istimewa, anak-anak ini dipersiapkan untuk mengikuti program khusus. Oleh karena itu psikotes menjadi sangat perlu.

Dengan kemampuan umum dan kemampuan khusus ini, diyakini bahwa anak mampu belajar lebih cepat dan lebih baik. Oleh karena itu pemerintah melakukan program percepatan (akselerasi) dengan program CIBI (cerdas istimewa dan berbakat istimewa). Beberapa sekolah menyelenggarakan kelas unggulan percepatan dengan sebutan Kelas Akselerasi (AKSEL). Salah satu persyaratannya anak yang mengikuti program ini harus memiliki IQ minimal 130 di skala Weshcler. Salah satu instrumen yang sering digunakan untuk ini adalah tes *Advanced Progressive Matrices* (APM), tes yang diyakini memiliki kemampuan untuk memilih anak yang berkemampuan khusus tersebut. Agar keputusan yang diambil tepat, maka tes ini perlu dianalisis dan diuji keampuhannya (khususnya daya beda item) di setiap saat (Subino, 1987). Selain itu, tes APM ini sesuai dengan namanya, perlu memiliki tingkat kesukaran progresif.

Dalam penelitian ini inteligensi dipandang sebagai kemampuan dasar yang dimiliki oleh seseorang, yang ditunjukkan dengan *efisiensi kerja intelektual*, yang akan menentukan sukses atau tidaknya seseorang dalam belajar. Kemampuan dasar tersebut dinyatakan dengan skor IQ yang didapat dari hasil tes inteligensi ini. Yang dimaksud dengan *efisiensi kerja intelektual* seseorang adalah kesanggupan total untuk melakukan pengamatan dan berfikir jernih. Kemampuan berfikir efisien dan jernih itu akan muncul dan terukur bila seseorang mengerjakan soal, khususnya tes APM, dalam waktu yang dibatasi. Seseorang dapat dikatakan berfikir efisien apabila kegiatan tersebut dilakukan secara mudah, cepat, dan tepat (Subino, 1984: 9).

APM adalah tipe tes yang kedua dari tes yang dikembangkan Raven. Tes ini mengukur kinerja intelektual dari mereka yang memiliki inteligensi di atas rata-rata; dan tes ini juga mampu membedakan secara tajam antara mereka yang tergolong memiliki inteligensi unggul dari yang lainnya. Tes ini terdiri dua set; Set I terdiri dari 12 soal waktu 5 menit dan tes II adalah 36 soal dengan waktu pengerjaan 40 menit. Set I diberikan kepada testi dengan maksud untuk menjelaskan prinsip-prinsip kerjanya, dan kemudian dilanjutkan ke set II. Menurut Subino (1984: 97), jika dianalisis soal-soal

**Vol. 2 No. 1**

pada set II meliputi persoalan-persoalan yang mampu menjadi alat pengukur semua operasi integral maupun analitis yang ada pada proses berpikir tinggi. Itulah sebabnya APM sangat berguna untuk mendapatkan pertelaah tentang laju kecepatan dan keberhasilan belajar yang mungkin dicapai seseorang didalam suatu bidang studi.

Dalam beberapa literature dinyatakan bahwa skor tes inteligensi selalu ditafsirkan berdasarkan perbandingannya dengan skor-skor pada kelompok anak seusianya. Oleh karena itu norma yang digunakan selalu mencantumkan kelompok usia tersebut. Disisi lain, berdasarkan hasil uji coba juga ditemukan bahwa skor total tes inteligensi (khususnya APM), pada anak-anak sampai remaja akhir (mahasiswa) skor total naik, sedangkan pada orang dewasa cenderung menurun berbarengan dengan bertambahnya umur. Hal tersebut dapat ditelaah dari beberapa hasil uji coba berikut.

Pada pengujian reliabilitas, dari anak-anak menuju dewasa angka reliabilitas semakin naik. Misal berdasarkan hasil uji coba Foulds (Subino, 1984: 98) secara *re-test* diperoleh angka reliabilitas 0,76 untuk anak usia 10,5 tahun, 0,86 untuk kelompok usia 12,5 tahun, dan 0,91 untuk kelompok mahasiswa dan dewasa.

Mengacu pada hasil uji coba yang dilakukan oleh Subino (1984:182), dengan 36 soal set II terhadap 981 orang siswa, dapat disimpulkan: *Pertama*: Indeks keterandalan (reliabilitas) dengan pendekatan/rumus KR20 antara 0,81 sampai 0,85. Reliabilitas atau keterandalan ini menunjukkan angka keterandalan yang tinggi (mantap). *Kedua*: Soal-soal APM terbangun dari soal-soal yang mudah sampai sukar, dan APM lebih peka bagi mereka yang berinteligensi unggul daripada yang berinteligensi asor. *Ketiga*: Butir soal APM tidak tersusun secara progresif sempurna (tidak berurutan dari yang termudah ke sukar), ada beberapa item yang lebih awal memiliki daya pembeda yang lebih baik.

Angka reliabilitas uji coba ulang dengan *re-test* yang dilakukan oleh Yaya Sunarya (2015) dengan tiga kali tes ulang dengan menggunakan perhitungan korelasi produk moment dari Pearson, didapat angka seperti pada tabel berikut.

**Tabel 1**  
**REKAP HASIL UJI COBA APM SECARA RE-TEST**

<b>Tahun Re-test</b>	<b>N</b>	<b>Rerata 1</b>	<b>Rerata 2</b>	<b>SD 1</b>	<b>SD 2</b>	<b>r-Spearman</b>
<b>I: 2008 – 2010</b>	230	21,735	22,813	4,883	5,200	0,756*
<b>II: 2009 – 2011</b>	225	20,925	22,305	4,947	5,092	0,725*
<b>III: 2010 – 2012</b>	221	19,873	21,896	5,073	5,406	0,732*

**Vol. 2 No. 1**

\* Signifikan pada  $p < 0,01$

Data di atas menunjukkan bahwa reliabilitas yang diperoleh dengan cara tes ulang berada pada kisaran antara 0,73 sampai 0,76; yang berarti bahwa APM ini memiliki keajegan/konsistensi (reliabilitas) yang cukup tinggi. Yang mengandung arti bahwa secara ulang skor yang diperoleh setiap testee masih akan berada pada fluktuasi yang wajar (tidak keluar dari standar galat baku pengukuran).

Menurut para ahli pengukuran dan asesmen, instrumen tes yang baik paling tidak harus memiliki validitas, reliabilitas, daya pembeda, dan tingkat kesukaran (Crocker, 1986; Murphy, 1998; Drummond, 2010; Wright, 2011).

**A. Validitas**

Validitas tes adalah tingkat keabsahan atau ketepatan suatu tes dalam mengukur apa yang ingin diukur. Tes yang valid adalah tes benar-benar mengukur apa yang hendak diukur. Tes matematika kelas dua SMP, hendaknya benar-benar mengukur hasil belajar matematika siswa SMP kelas dua; bukan siswa SMP kelas tiga atau siswa SD kelas enam; serta bukan pula mengukur hasil belajar dalam bidang studi lainnya. Tes yang disusun untuk mengukur hasil belajar mata pelajaran Bahasa Indonesia pada kelas tertentu, *hendaknya tidak menyimpang* sehingga mengukur hasil belajar matematika, IPS, atau Bahasa Indonesia pada tingkat/kelas lainnya.

Terdapat beberapa macam validitas tes hasil belajar dan cara mengetahui / menghitung koefisien validitas tes. Hal tersebut adalah:

**1. Validitas Permukaan (*face validity*)**

Tingkat validitas permukaan dapat diketahui dengan melakukan analisis rasional (semata-mata dengan berdasarkan pertimbangan logis, bukan pada hitungan angka-angka empirik). Berbagai hal berikut merupakan hal yang perlu atau aspek berikut ini perlu dianalisis/diperiksa kualitasnya.

- a. Apakah bahasa dan susunan kalimat (redaksi) tiap butir soal cukup jelas dan sesuai dengan kemampuan siswa ?
- b. Apakah isi jawaban yang diminta tidak membingungkan ?
- c. Apakah cara menjawab sudah dipahami siswa ?
- d. Jangan sampai siswa tahu isi jawabannya tetapi tidak tahu bagaimana cara menjawab soal bersangkutan.

**Vol. 2 No. 1**

- e. Apakah tes itu telah disusun berdasar kaidah/prinsip penulisan butir soal? Tes yang tidak mengikuti kaidah penulisan butir soal akan tampak semrawut sehingga membingungkan siswa.

Setiap tes paling sedikit harus diperiksa melalui analisis validitas permukaan. Walaupun analisis ini tergolong analisis yang paling lemah, namun lebih baik daripada tidak ada analisis sama sekali. Tentu saja akan lebih baik bila suatu tes dianalisis lebih lanjut dengan teknik analisis faktor (Crocker, 1986).

**2. Validitas isi (content validity)**

Tingkat validitas isi juga dapat diketahui dengan analisis rasional (Subino, 1987). Hal ini pada prinsipnya dilakukan pemeriksaan terhadap tiap butir soal, apakah soal tersebut sudah sesuai dengan indikator atau hasil belajar, atau pokok bahasan yang akan diteskan. Cara yang lazim ialah dengan cara mencocokkan tiap butir soal dengan kisi-kisi yang disusun berdasar silabus atau GBPP yang berlaku.

Pengujian validitas isi dilakukan dengan untuk menjawab pertanyaan berikut.

- a. Apakah keseluruhan tes (soal) telah sesuai dengan kisi-kisi yang diacu ?

Kisi-kisi adalah suatu bagan yang dapat berupa matrik yang menggambarkan penyebaran soal-soal sesuai dengan aspek atau pokok bahasan yang hendak diukur, tingkat kesukaran yang diharapkan, dan jenis soal akan digunakan. Kisi-kisi itu harus disusun sedemikian rupa sehingga mencakup seluruh materi/bahan pelajaran yang akan diteskan. (Salah satu fungsi kisi-kisi adalah untuk melihat penyebaran soal).

*Tingkat kesesuaian seluruh butir soal (dalam satu tes) dengan kisi-kisi atau dengan bahan yang akan diteskan menunjukkan tingkat validitas isi.*

- b. Apakah terdapat butir soal yang menyimpang, atau menuntut jawaban di luar hal atau bahan pelajaran yang bersangkutan ?

Penyimpangan yang tidak kentara perlu dihilangkan. Semakin banyak soal yang menyimpang, maka akan semakin rendah tingkat validitas isi tes tersebut. Untuk melakukan analisis validitas isi diperlukan adanya kisi-kisi tes yang disusun sebelum soal-soal ditulis.

**3. Validitas kriteria (criterion validity)**

Untuk menganalisis jenis validitas ini dapat diketahui dengan secara empirik, yakni menghitung koefisien korelasi antara tes bersangkutan dengan tes lain yang

**Vol. 2 No. 1**

sudah baku sebagai kriterianya. Tes lain yang dapat digunakan sebagai kriteria adalah tes yang sudah dianggap valid; atau juga dapat digunakan nilai mata pelajaran yang sama yang didapat dengan menggunakan tes lain yang dipandang cukup obyektif. Sebagai contoh, skor tes Bahasa Inggris buatan guru dikorelasikan dengan skor tes Bahasa Inggris yang telah dibakukan. Skor tes Matematika akhir tahun dikorelasikan dengan nilai rata-rata Matematika selama satu tahun.

Dengan rumus korelasi *Pearson's Product Moment* dan menggunakan kalkulator (komputer), perhitungan validitas kriteria tersebut tidak terlalu sulit. Kesulitan utama dalam menentukan validitas kriteria ialah mencari skor tes yang lain yang akan dijadikan kriteria. Bila kriterianya buruk atau tidak valid, maka validitas tes yang diperoleh akan percuma saja.

**4. Validitas ramalan (predictive validity)**

Validitas ini menunjukkan sejauh mana skor tes yang bersangkutan dapat digunakan untuk meramalkan keberhasilan siswa di masa mendatang dalam bidang tertentu. Cara menghitungnya sama seperti validitas kriteria, dalam hal ini skor tes dikorelasikan dengan keberhasilan siswa di masa datang. Misalnya antara nilai UN (Ujian Nasional) di SMP, dengan prestasi belajar di SMA dalam mata pelajaran yang sama.

Suatu tes yang baik biasanya memiliki angka validitas 0,50 atau lebih; tentu saja angka itu makin tinggi makin baik. Suatu tes dengan angka validitas kurang dari 0,50 belum tentu buruk. Mungkin kriterianya yang buruk atau keliru menentukan kriteria.

**B. MENGESTIMASI RELIABILITAS TES**

Reliabilitas tes adalah tingkat keajegan (konsistensi) suatu tes, yakni sejauh mana suatu tes dapat dipercaya untuk menghasilkan skor yang ajeg atau konsisten (tidak berubah-ubah). *Tes yang reliabel atau yang dapat dipercaya adalah tes yang menghasilkan skor secara ajeg, relatif tida berubah, walaupun di teskan pada situasi dan waktu yang berbeda beda.* Sebaliknya, tes yang tidak reliabel seperti karet untuk mengukur panjang, hasil pengukuran dengan karet dapat berubah-ubah (tidak ajeg, tidak konsisten).

Para ahli menyebut prosedur mengestimasi reliabilitas berbeda-beda. Anastasi (1988: 116) menyebutkan ada tiga macam cara untuk menaksir reliabilitas tes: *test-retest reliability*, *alternate form reliability*, dan *split-half reliability*. Murphy dan Davidshofer

*Jurnal Pendidikan Pascasarjana Magister PAI*

**Vol. 2 No. 1**

(1998: 115) membagi kedalam empat cara, yaitu: *test-retest methods*, *alternate form methods*, *split-half methods*, dan *internal consistency methods*. Drummon dan Jones (2010: 87) ada empat cara yaitu *test-retest*, *alternate form*, *internal consistency*, dan *interrater reliability*. Sementara Crocker dan Algina (1986: 132-135) membaginya menurut cara pengadministrasiannya menjadi dua jenis, yaitu pengadministrasian dua tes (*alternate form*, *test-retest*), dan pengadministrasian satu tes (*split-half method*, dan *analisis covariances*). Dari empat ahli tersebut pada prinsipnya memiliki kesamaan, yaitu diperolehnya indeks koefisien korelasi antara dua kelompok skor tes. Cara-cara tersebut dapat dijelaskan sebagai berikut.

**1. Test-retest method (metoda tes ulang).**

Satu tes (yakni tes yang akan dihitung reliabilitasnya), diteskan terhadap kelompok siswa tertentu dua kali dengan jangka waktu tertentu (misalnya satu semester atau satu catur wulan). Skor hasil pengetesan pertama dikorelasikan dengan skor hasil pengetesan kedua. Koefisien korelasi yang diperoleh menunjukkan koefisien reliabilitas tes tersebut.

**2. Paralel test method (metoda tes paralel)**

Cara ini mengharuskan adanya dua tes yang paralel, yakni dua tes yang disusun dengan tujuan yang sama (hanya sedikit perbedaan redaksi, isi atau susunan kalimatnya) dengan menggunakan format yang berbeda; jadi terdapat format A dan format B. Dua tes tersebut diadministrasikan pada satu kelompok siswa dengan perbedaan waktu beberapa hari saja. Skor dari kedua macam tes tersebut dikorelasikan dengan teknik yang sama seperti pada metode test- retest. Koefisien korelasi yang diperoleh menunjukkan tingkat reliabilitas tes.

**3. Split-half method (metode belah dua)**

Cara ini paling mudah dan seyogyanya diterapkan oleh para guru pada semua tes yang diberikan kepada siswanya. Tidak perlu mengulangi pelaksanaan tes atau menyusun tes yang paralel. Cukup satu tes dan diadministrasikan satu kali kepada sekelompok siswa (minimal 30 siswa).

Pada saat penyekoran, tes dibelah menjadi dua sehingga tiap siswa memperoleh dua macam skor, yakni skor yang diperoleh dari soal-soal bernomor ganjil dan skor dari soal-soal bernomor genap. Skor total diperoleh dengan

**Vol. 2 No. 1**

menjumlah skor ganjil dan genap. Selanjutnya skor-ganjil dikorelasikan dengan skor-genap, hasilnya adalah koefisien korelasi  $r_{gg}$ , atau koefisien korelasi ganjil-genap. Karena tes dibelah jadi dua, maka koefisien korelasi ganjil-genap tersebut dikoreksi sehingga menjadi koefisien reliabilitas.

**5. Interrater Reliability.**

Sebuah tes diberikan kepada sejumlah subjek, kemudian subjek dinilai oleh beberapa orang dengan format tertentu (dua atau tiga orang, yang mirip sebagai juri). Hasil penilaian terhadap subjek itu dikorelasikan. Hasil korelasi itu merupakan indeks reliabilitas.

**C. ANALISIS BUTIR SOAL**

Baik buruknya tes tergantung pada butir-butir soal yang ada di dalamnya. Oleh sebab itu untuk mendapatkan tes yang baik perlu dipilih butir-butir yang baik. Butir yang buruk harus dibuang, yang kurang baik perlu direvisi. Untuk mengetahui kualitas tiap butir soal perlu analisis satu persatu. Analisis meliputi perhitungan daya pembeda, tingkat kesukaran, homogenitas tes serta analisis distraktor/pengecoh pada tes pilihan ganda.

Analisis butir soal adalah segala upaya mengetahui kualitas (baik buruknya) semua butir soal dalam suatu tes dengan cara menghitung daya pembeda, tingkat kesukaran, tingkat homogenitas, dan fungsi pengecoh dalam setiap butir soal. (Crocker & Algina, 1986: 311; Izard, 1977; Subino, 1987; Sukardi, 2009).

Daya pembeda menunjukkan sejauh mana tiap butir soal mampu membedakan antara siswa yang menguasai bahan dengan siswa yang tidak menguasai bahan. Butir soal yang daya pembedanya rendah, tidak ada manfaatnya, malahan dapat merugikan siswa yang belajar sungguh- sungguh (Subino, 1987; Suryabrata, 1999).

Tingkat kesukaran menunjukkan apakah butir soal tergolong sukar, sedang atau mudah. Tes yang baik memuat kira-kira 25% soal mudah, 50% sedang dan 25% sukar. Butir soal yang terlalu sukar sehingga hampir tidak terjawab oleh semua siswa atau terlalu mudah sehingga dapat dijawab oleh hampir semua siswa, sebaiknya dibuang karena tidak bermanfaat (Izard, 1977).

Tingkat homogenitas soal menunjukkan apakah tiap butir soal mengukur aspek/pokok bahasan yang sama, atau sejauh mana tiap butir soal menyumbang

**Vol. 2 No. 1**

skor total tiap siswa. Butir soal yang homogen adalah yang menunjang skor total. Sebaliknya, butir soal yang tidak seiring dengan skor-total dikatakan tidak homogen, dan lebih baik dibuang atau direvisi (Karno To, 2003).

Pada tes pilihan ganda, tiap butir soal menggunakan beberapa pengecoh (distraktor/penyesat/option). Tiap pengecoh hendaknya bermanfaat atau berfungsi, yakni ada sejumlah siswa yang memilihnya. Pengecoh yang tidak dipilih sama sekali oleh siswa berarti tidak berfungsi mengecohkan siswa, sebaliknya pengecoh yang dipilih oleh hampir semua siswa berarti terlalu mirip dengan jawaban yang benar (Karno To, 2003).

**D. LANGKAH-LANGKAH ANALISIS BUTIR SOAL**

Butir soal tes pilihan ganda jumlahnya cukup besar, biasanya antara 50 - 100 butir, bahkan ada yang sampai 200 butir dengan ragam soal yang berbeda-beda. Untuk keperluan analisis, lembar jawaban siswa merupakan dokumen utama yang harus ada. Analisis lengkap meliputi semua hal, sedang analisis singkat hanya meliputi: reliabilitas belah-dua, daya pembeda atau tingkat kesukaran. Langkah-langkah analisis butir soal adalah sebagai berikut (Karno To, 2003; Izard, 1977).

***1. Memberi skor pada lembar jawaban.***

- a. Berilah tanda silang pada lembar jawaban, mana butir soal yang dijawab benar dan mana yang salah. Yang benar diberi skor satu, yang salah diberi nol. Untuk pemberian nilai, boleh saja jawaban benar diskor 4 dan jawaban salah didenda 1.
- b. Skor tiap lembar jawaban (tiap siswa) dijumlahkan, dengan 3 macam skor: (1) jumlah skor soal bernomor ganjil, (2) jumlah skor soal bernomor genap, dan (3) skor total.
- c. Jumlah skor ganjil dan genap digunakan untuk menghitung reliabilitas. Lihat teknik analisis reliabilitas belah-dua. Sedang skor total digunakan untuk mengurutkan dan membuat kelompok Atas - Bawah (kelompok Unggul - Asor)

***2. Menghitung daya pembeda***

- a. Berdasar skor total, susunlah nama atau nomor siswa dari tertinggi hingga terendah. Ambil 27% siswa yang skor-totalnya tinggi atau 27 % Kelompok Atas, dan 27% yang rendah (Kelompok Bawah).

**Vol. 2 No. 1**

- b. Buatlah tabel, khusus untuk siswa kelompok Atas dan kelompok Bawah. Jumlah kolom dalam tabel minimal sama dengan jumlah butir soal, sehingga memuat seluruh jawaban siswa.
- c. Tanda 1 artinya jawaban betul dan 0 artinya jawaban salah. Tabel ini digunakan untuk menghitung daya pembeda maupun tingkat kesukaran butir soal.
- d. Hitung jumlah jawaban yang benar (bertanda 1), baik pada Kelompok Atas maupun pada Kelompok Bawah.
- e. Daya pembeda dihitung dengan rumus:

$$DP = \frac{B_A - B_B}{N_A} \times 100\% \quad (\text{Subino, 1987; Karno To, 2003})$$

Catatan:

DP = indeks daya pembeda butir soal tertentu (satu butir)

$B_A$  = jumlah jawaban benar pada Kelompok Atas

$B_B$  = jumlah jawaban benar pada Kelompok Bawah

$N_A$  = jumlah siswa pada salah satu kelompok A atau B.

Kriteria daya pembeda sebagai berikut:

Negatif - 9% = sangat buruk, harus dibuang

10% - 19% = buruk, sebaiknya dibuang

20% - 29% = agak baik, kemungkinan perlu direvisi

30% - 49% = baik

50% ke atas = sangat baik

Pada prinsipnya, daya pembeda dihitung berdasar selisih jawaban benar pada Kelompok Atas dan Kelompok Bawah, dibagi dengan jumlah siswa pada salah satu kelompok tersebut. Dikalikan 100% agar diperoleh angka bulat (bukan pecahan, tetapi persen). Masih ada beberapa teknik dan rumus menghitung daya pembeda, namun cara di atas paling sederhana sehingga cocok untuk para guru.

### 3. Menghitung tingkat kesukaran

Tabel skor yang digunakan disini sama dengan tabel skor untuk menghitung daya pembeda, tetapi menggunakan rumus:

$$TK = \frac{B_A + B_B}{N_A + N_B} \times 100\% \quad (\text{Subino, 1987; Karno To, 2003})$$

TK = indeks tingkat kesukaran butir soal tertentu (satu butir)

**Vol. 2 No. 1**

$B_A$  = jumlah siswa yang menjawab benar pada kelompok A

$B_B$  = jumlah siswa yang menjawab banar pada kelompok B

$N_A$  = jumlah siswa pada kelompok A (atas/unggul)

$N_B$  = jumlah siswa pada kelompok B (bawah/asor)

Makin besar harga TK, makin mudah butir soal tersebut, sehingga dapat juga disebut “tingkat kemudahan”. Kriteria tingkat kesukaran (tingkat kemudahan) sebagai berikut:

0% - 15% = sangat sukar, sebaiknya dibuang.

16% - 30% = sukar

31% - 70% = sedang

71% - 85% = mudah

86% -100% = sangat mudah, sebaiknya dibuang.

Tingkat kesukaran tiap butir soal lebih baik bila dihitung berdasar jawaban seluruh siswa yang ikut tes (bukan hanya kelompok unggul dan asor yang berjumlah 54%). Rumus yang digunakan adalah:

$$TK = \frac{nB}{N} \times 100\% \quad (\text{Karno To, 2003})$$

TK = indeks tingkat kesukaran butir soal tertentu (satu butir soal)

$nB$  = jumlah siswa yang menjawab benar pada butir itu

$N$  = jumlah seluruh siswa yang mengikuti tes

**4. Menghitung homogenitas butir soal**

Homogen tidaknya butir soal diketahui dengan menghitung koefisien korelasi antara skor tiap butir soal dengan skor total. Diperlukan perhitungan korelasi sebanyak butir soal dalam tes . Skor tiap butir soal adalah 1 atau 0, sedang skor total tiap siswa cukup bervariasi. Teknik korelasi yang digunakan boleh dengan Pearson's Product Moment, atau teknik Korelasi Point Biserial.

Butir soal yang homogen, koefisien korelasinya sama atau di atas batas signifikansi (batas kritis korelasi). Butir soal yang tidak/kurang homogen koefisien korelasinya negatif atau lebih kecil dari batas signifikansi. Butir soal tersebut mungkin mengukur aspek lain di luar bahan yang diajarkan (soal tidak sesuai dengan tujuan pengajaran), maka sebaiknya direvisi atau dibuang.

**Vol. 2 No. 1****5. Analisis distraktor/pengecoh.**

Pada tes pilihan ganda ada beberapa option/alternatif jawaban yang sengaja dimasukkan sebagai pengecoh (distraktor). Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta yang menjawab salah dalam soal tersebut, tidak berkumpul pada satu pilihan. Sebaliknya, butir soal yang buruk pengecohnya akan dipilih secara tidak merata, berkumpul di salah satu pilihan. Demikian juga butir soal yang baik, jawaban pada pilihan yang benar, tidak akan terkumpul (lebih banyak) pada salah satu jawaban yang salah. Jika terjadi demikian, pilihan jawaban tersebut akan merupakan pilihan jawaban (distraktor) “yang menyesatkan” (distraktor tersebut seperti pilihan jawaban benar, tetapi salah).

Pengecoh dianggap baik bila jumlah siswa yang memilih pengecoh itu sama atau mendekati jumlah ideal. Jumlah pemilih ideal adalah jumlah yang memilih salah dibagi dengan banyaknya pilihan salah tersebut.

Indeks pengecoh dihitung dengan rumus:

$$IPc = \frac{nPc}{(N - nB) / (Alt - 1)} \times 100\%$$

IPc = Indeks Pengecoh/Distraktor

nPc = Jumlah siswa yang memilih pengecoh itu

N = Jumlah seluruh subyek yang ikut tes

nB = Jumlah subyek yang menjawab benar pada butir soal itu

Alt = Banyak alternatif jawaban/option (3, 4, atau 5)

Dengan kriteria suatu soal dikatakan:

Sangat baik            IPc = 76% - 125%

Baik                     Ipc = 51% - 75% atau 126% - 150%

Kurang baik            Ipc = 26% - 50% atau 151% - 175%

Buruk                    IPc = 0% - 25%    atau 176% - 200%

Sangat buruk            IPc = lebih dari 200%

Bila semua peserta menjawab benar pada butir soal tertentu (semua sesuai kunci), maka IPc = 0, artinya semua pengecoh tidak berfungsi (semua pengecoh buruk).

**Vol. 2 No. 1**

Untuk analisis pengecoh perlu dibuat tabel khusus agar setiap butir soal diketahui berapa siswa yang menjawab a, berapa yang menjawab b, berapa yang menjawab c, dan seterusnya. Dengan menggunakan komputer pengolahan ini hanya memerlukan waktu beberapa detik saja.

Pada penelitian ini pengujian validitas dan reliabilitas tidak dilakukan karena sudah dilakukan pada penelitian tahun sebelumnya (tahun 2015). Analisis instrumen hanya dilakukan pada aspek tingkat kesukaran, daya pembeda, homogenitas item dan analisis pengecoh.

**E. METODOLOGI**

Penelitian ini merupakan penelitian deskriptif dengan pendekatan *expost-facto*, melibatkan responden sebanyak 5400 peserta dari 20 sekolah. Pengumpulan data dilakukan dengan teknik studi dokumentasi, di Labortorium Psikologi Pendidikan dan Bimbingan FIP UPI. Instrumen yang dianalisis adalah tes inteligensi *Advanced Progressive Matrices* (APM), yang dikembangkan oleh John C. Raven, dengan dasar teorinya adalah teori inteligensi dari Spearman (yaitu teori “g” factor). Tes ini terdiri dari dua set, set I terdiri dari 12 soal (sebagai soal latihan), dan set II sebanyak 36 soal, model pilihan ganda dengan masing-masing soal 8 pilihan (dikotomus). Yang dianalisis adalah set II. Teknik yang dilakukan adalah analisis item dengan aspek analisis (1) tingkat kesukaran, (2) daya pembeda, (3) homogenitas item, dan (3) analisis distraktor.

**F. HASIL PENELITIAN****1. Tingkat Kesukaran.**

Hasil perhitungan dengan cara menghitung proporsi antara siswa yang menjawab benar dengan seluruh peserta tes diperoleh indeks tingkat kesukaran sebagai berikut.

**Tabel 2****HASIL PERHITUNGAN TINGKAT KESUKARAN**

No Item	Indeks	Tafsiran	No Item	Indeks	Tafsiran	No Item	Indeks	Tafsiran
1	85.40	Mudah	13	52.48	Sedang	25	35.85	Sedang
2	84.04	Mudah	14	67.74	Sedang	26	35.75	Sedang
3	84.90	Mudah	15	53.51	Sedang	27	25.65	Sukar

**Vol. 2 No. 1**

4	82.48	Mudah	16	64.39	Sedang	28	18.07	Sukar
5	80.88	Mudah	17	65.10	Sedang	29	14.75	Sukar
6	85.16	Mudah	18	49.45	Sedang	30	29.60	Sukar
7	73.60	Mudah	19	68.35	Sedang	31	22.12	Sukar
8	67.05	Sedang	20	59.42	Sedang	32	15.74	Sukar
9	88.06	Mudah*)	21	45.49	Sedang	33	33.85	Sedang
10	71.71	Mudah	22	35.20	Sedang	34	18.24	Sukar
11	76.98	Mudah	23	44.79	Sedang	35	24.11	Sukar
12	74.94	Mudah	24	26.68	Sukar	36	7.077	Sukar*)

Dari hasil perhitungan di atas dengan menggunakan tiga kategori, didapat 11 soal tergolong kategori mudah (31%), 15 sedang (42%), dan 10 soal termasuk kategori sukar (27%). Jika dibuat dalam lima kategori, dengan menambahkan kategori Sangat Mudah dan Sangat Sukar, maka didapat 1 soal Sangat Mudah, 10 Mudah, 15 Sedang, 9 Sukar, dan 1 soal Sangat Sukar. Kriteria soal yang ideal, perbandingan antara soal Mudah: Sedang: dan Sukar adalah 1: 2 : 1 atau 25% : 50% : 25% hampir terpenuhi; kriteria soal yang termasuk Sedang masih lebih banyak dibanding dengan soal Mudah dan Sukar. Jika dilihat dari ketentuan soal yang progresif, dimana susunan soal harus dari Mudah, Sedang dan Sukar terpenuhi. Namun jika dilihat dari urutan indeks kesulitan, dimana soal harus dimulai dari soal yang sangat mudah, kemudian mudah, sedang, sukar dan sangat sukar, tidak terpenuhi secara ideal (tidak progresif betul). Maka direkomendasikan jika ingin memenuhi kriteria soal progresif secara ideal, perlu ada perubahan urutan soal berdasarkan susunan tingkat kesukaran tersebut. Hasil penelitian ini sama seperti yang dilakukan Subino (1984).

## 2. Daya Pembeda (DP)

Untuk melihat daya pembeda ini dilakukan dengan dua pendekatan, yaitu (1) menggunakan pendekatan kelompok unggul – asor. Berdasarkan pendekatan yang pertama (unggul – asor) dilakukan prosedur: (a) soal diurutkan berdasarkan skor jawaban tertinggi sampai terendah; (b) diambil 25% sampai 33% untuk masing-masing kelompok unggul dan asor (Izard, 1977). Untuk keperluan ini diambil 30% dari peserta tes untuk masing-masing kelompok unggul dan asor (Crocker, 1986), diperoleh hasil perhitungan sebagai berikut.

**Tabel 3**  
**HASIL PERHIRUNGAN DAYA PEMBEDA ITEM**

No Soal	Kel. Unggul	Kel. Asor	Indeks DP	Tafsiran	No Soal	Kel. Unggul	Kel. Asor	Indeks DP	Tafsiran
1	1447	1010	29.1	Cukup	19	1299	645	43.6	Baik
2	1469	856	40.9	Baik	20	1266	477	52.6	Sgt.baik
3	1469	909	37.3	Baik	21	1186	252	62.3	Sgt.baik
4	1450	876	38.3	Baik	22	1029	418	40.7	Baik
5	1450	832	41.2	Baik	23	1062	236	55.1	Sgt.baik
6	1458	1000	30.5	Baik	24	754	180	38.3	Baik
7	1355	806	36.6	Baik	25	943	178	51.0	Sgt.baik
8	1304	609	46.3	Baik	26	860	239	41.4	Baik
9	1484	1023	30.7	Baik	27	758	146	40.8	Baik
10	1456	501	63.7	Sgt.baik	28	516	126	26.0	Cukup
11	1460	647	54.2	Sgt.baik	29	414	116	19.9	Cukup
12	1437	588	56.6	Sgt.baik	30	803	192	40.7	Baik
13	1169	427	49.5	Sgt.baik	31	668	109	37.3	Baik
14	1396	509	59.1	Sgt.baik	32	422	115	20.5	Cukup
15	1214	387	55.1	Sgt.baik	33	842	221	41.4	Baik
16	1393	419	64.9	Sgt.baik	34	566	103	30.9	Baik
17	1320	553	51.1	Sgt.baik	35	695	142	36.9	Baik
18	1168	285	58.9	Sgt.baik	36	171	71	6.7	Kurang

Dari tabel di atas dapat dilihat bahwa daya pembeda (DP) setiap soal termasuk dalam kategori cukup ke atas kecuali soal nomor 36 daya pembedanya kurang. Jika dirinci terdapat 1 soal dengan daya pembeda kurang, 4 soal cukup baik, 18 soal memiliki daya pembeda baik, dan 13 soal sangat baik. Hasil penelitian ini jika dibandingkan dengan penelitian Subino (1984) memiliki kesamaan. Yaitu butir soal 29, 32 dan 36 merupakan soal yang daya bedanya paling kecil. Hasil analisis Subio (1984:239) tentang soal nomor 29 dan 36 adalah “soal ini mengukur proses intelektual yang berbeda dari soal lainnya. Kedua soal ini, mempersoalkan bangun-bangun yang tidak jelas aturannya, melebihi tuntutan intelektual yang lain dari soal lainnya. Nomor 29 *Jurnal Pendidikan Pascasarjana Magister PAI*

**Vol. 2 No. 1**

mempersoalkan letak tiga garis lurus pada bangun yang tidak jelas, dan nomor 36 mempersoalkan kedudukan titik didalam bangun yang tidak jelas juga.

**3. Homogenitas Item**

Pendekatan di atas oleh beberapa ahli dianggap kurang baik, karena hanya memperhitungkan sebagian peserta (kelompok unggul dan asor saja), sebagian lagi tidak diikutsertakan. Oleh karena itu digunakan pendekatan lain yaitu pendekatan *korelasi point biserial* (rpbis). Hasil uji ini menghasilkan indeks homogenitas item. Hasilnya adalah sebagai berikut.

**TABEL 4**  
**HASIL PERHITUNGAN HOMOGENITAS ITEM**

No Item	rpbis	Tafsiran	No Item	rpbis	Tafsiran	No Item	rpbis	Tafsiran
1	0.37	Signifikan	13	0.40	Signifikan	25	0.42	Signifikan
2	0.48	Signifikan	14	0.51	Signifikan	26	0.36	Signifikan
3	0.48	Signifikan	15	0.44	Signifikan	27	0.38	Signifikan
4	0.44	Signifikan	16	0.54	Signifikan	28	0.28	Signifikan
5	0.45	Signifikan	17	0.43	Signifikan	29	0.23	Signifikan
6	0.38	Signifikan	18	0.47	Signifikan	30	0.36	Signifikan
7	0.36	Signifikan	19	0.39	Signifikan	31	0.35	Signifikan
8	0.41	Signifikan	20	0.43	Signifikan	32	0.24	Signifikan
9	0.40	Signifikan	21	0.49	Signifikan	33	0.36	Signifikan
10	0.56	Signifikan	22	0.51	Signifikan	34	0.33	Signifikan
11	0.52	Signifikan	23	0.43	Signifikan	35	0.36	Signifikan
12	0.54	Signifikan	24	0.34	Signifikan	36	0.12	Signifikan

Berdasarkan hasil perhitungan homogenitas item dengan menggunakan korelasi point biserial dapat disimpulkan bahwa semua soal memiliki indeks homogenitas yang signifikan. Ini berarti bahwa semua soal merupakan soal-soal yang homogen, yang berarti item-item yang dikembangkan mengukur hal yang sama dan berasal dari konstruk yang sama. Soal 29, 32, 36 merupakan soal yang indeks homogenitasnya paling rendah. Hal yang sama terjadi pada penelitian Subino (1984).

#### 4. Analisis Distraktor (Pengecoh).

Hasil analisis terhadap distraktor pada seluruh soal, dengan menggunakan rumus Indeks Pengecoh (IPc) dari Karno To (2003), diperoleh data sebagai berikut.

**Tabel 5**  
**REKAP HASIL PENGUJIAN DISTRAKTOR**

No	Nomor Pilihan Jawaban								>
Item	1	2	3	4	5	6	7	8	200%
1	237.3	87.8	19.0	70.6	K	16.3	134.9	134.0	1
2	K	141.9	27.2	40.8	17.0	272.7	98.5	101.9	6
3	65.5	16.6	66.3	75.9	41.9	224.3	K	209.5	6
4	82.9	119.5	92.8	K	90.5	48.7	68.5	197.1	
5	237.2	69.0	K	56.7	60.2	30.8	192.1	54.0	1
6	K	73.3	40.2	42.0	24.1	20.6	453.3	46.5	7
7	12.3	557.3	7.9	18.8	28.6	K	21.2	53.8	7
8	K	211.3	128.8	87.8	82.5	38.2	55.0	96.4	2
9	150.2	49.7	206.5	107.1	43.1	69.6	74.0	K	3
10	67.8	96.3	46.3	K	99.5	16.8	144.9	228.5	8
11	144.3	18.8	99.3	132.4	K	84.4	198.5	22.2	2
12	25.0	34.4	41.2	138.6	225.2	K	87.0	148.5	5
13	34.6	K	60.4	14.8	377.9	103.7	65.3	43.4	
14	K	33.9	53.8	217.7	65.8	108.0	135.8	84.9	4
15	20.4	K	13.4	100.5	14.3	57.1	23.2	471.1	8
16	99.8	40.8	88.3	K	281.6	41.9	63.8	83.8	5
17	17.3	26.7	360.2	153.0	25.6	K	78.2	39.1	3
18	236.9	49.5	49.5	151.8	86.9	49.5	K	76.0	1
19	40.7	25.9	K	42.8	388.7	44.8	83.5	73.6	5
20	12.5	440.4	27.3	141.5	28.6	31.2	18.3	K	2
21	229.2	38.9	27.0	272.0	46.2	12.2	74.6	K	1, 4
22	35.4	59.1	34.2	48.4	82.6	51.9	K	388.3	8
23	123.2	46.7	116.5	54.9	111.0	K	37.3	210.4	8
24	361.8	78.5	K	28.3	65.9	44.3	55.3	65.9	1

## Vol. 2 No. 1

25	136.1	250.4	13.7	83.2	34.9	30.3	K	151.3	2
26	37.9	K	76.3	25.0	157.2	239.1	25.0	139.4	6
27	37.2	58.4	94.0	202.4	90.2	173.9	K	43.9	4
28	27.8	134.6	156.2	140.4	K	100.9	45.9	94.2	
29	74.8	83.0	157.3	25.9	118.3	K	175.0	65.6	
30	111.4	134.7	89.8	85.7	K	86.7	54.5	137.2	
31	44.6	75.9	106.3	K	153.6	110.8	118.6	90.2	
32	156.3	55.6	39.1	122.6	68.6	50.9	206.9	K	
33	59.8	126.8	121.8	83.1	K	90.6	103.3	114.5	
34	K	110.5	207.2	73.4	142.5	73.6	32.9	59.9	3
35	284.7	55.8	K	85.6	16.6	68.6	169.8	18.8	1
36	207.7	K	54.2	121.1	75.0	119.3	27.6	95.1	1

K = Kunci Jawaban

Dari tabel 5 di atas tampak bahwa hampir pada setiap soal terdapat pengecoh yang tergolong *sangat jelek atau jelek*. Kriteria *sangat jelek* disini adalah pengecoh yang terlalu banyak dipilih oleh responden (melebihi 200% setelah semua responden yang salah dibagi dengan banyaknya pilihan salah/pilihan ideal, atau sebaliknya persentase terlalu kecil, dibawah 25%). Namun jika menggunakan kriteria yang lain (misalnya kriteria dari Izard, 1977) kriteria yang diberikan adalah istraktor yang jelek adalah diatraktor yang sama sekali tidak pernah dipilih oleh responden. Dengan menggunakan kriteria ini, tidak ada distraktor yang sangat jelek.

## F. PEMBAHASAN HASIL/DISKUSI

Kegunaan hasil tes adalah untuk mengambil kesimpulan atau keputusan. Keputusan tersebut akan terkait dengan orang atau testi yang di tes dan instrumen itu sendiri. Keputusan atau kesimpulan tersebut, jika terkait dengan orang akan berkenaan dengan aspek yang diases yang ada dalam diri (*inner conditions*: kualitas bakat, tingkatan inteligensi, dan lain-lain atribut psikologisnya) dan atau kompetensi hasil belajar yang diperoleh (Naglieri & Goldstein, 2009), yang berfungsi sebagai prediktor dalam pengambilan keputusan (penempatan, diagnosis, dll). Jika terkait dengan instrumen, akan menunjukkan kualitas dari instrumen (Nitko, 1996). Dengan kata lain, untuk membuat keputusan yang tepat, data yang diperoleh harus benar-benar

*Jurnal Pendidikan Pascasarjana Magister PAI*

**Vol. 2 No. 1**

mengambarkan dirinya (valid) secara konsisten (reliabel), sehingga keputusan itu bersifat adil (*fairness*) dan tidak salah atau bias (Crocker dan Algina, 1986: 377).

Mengacu pada hasil penelitian di atas, dapat disimpulkan bahwa tes APM -- dilihat dari kualitas item-itemnya (daya beda, tingkat kesukaran, homogenitas item, dan pengecoh-pengecohnya) -- telah memenuhi sebagai tes yang baik. Oleh karena itu hasil tes APM dapat dijadikan dasar pengambilan keputusan tentang kualitas inteligensi seseorang. Konsekwensinya harus dijaga dan diamankan secara profesional, jangan sampai bocor. Karena jika suatu soal atau suatu tes sudah bocor, tes tersebut tidak akan menjadi tidak akan mampu membedakan lagi, tingkat kesukaran berubah, menjadi bias dan kualitasnya berubah, (Izard :1977).

Terkait dengan bias tes, menurut Hays (2013:62-63) ada dua kategori sumber bias, yaitu (1) bias dihubungkan dengan isi tes yaitu merujuk pada "*inappropriate selection of test item or general content coverage*, dan (2) bias dihubungkan dengan proses respons yang merujuk pada "*situations when item elicit responses not intended by the test*, yang disebut *response set*. Salah satu penyebab bias yang bersumber dari tes itu sendiri adalah tingkat kesulitan soal (*item difficulty*) (Crocker dan Algina, 1986: 388). Soal yang terlalu sulit atau terlalu mudah sehingga semua orang salah atau semua orang betul, tidak memberi manfaat apa-apa terhadap pengambilan keputusan (Subino, 1987).

APM adalah tes non verbal yang bebas dari gangguan kemampuan menggunakan bahasa verbal. Ravens (2011: 8) mengatakan bahwa " ... *APM is a nonverbal assessment tool designed to measure an individual's ability to perceive and think clearly, make meaning out of confusion and formulate new concepts when faced with novel information*". Jadi tes ini disusun untuk mengukur kemampuan berpikir jernih, yang oleh Subino (1984: 9) dikatakan sebagai *efisiensi kerja intelektual*, yang akan menentukan sukses atau tidaknya seseorang dalam belajar; yang ditunjukkan dengan belajar secara mudah, cepat, dan tepat. Karena mengerjakan tes ini membutuhkan kerja intelektual secara efisien, maka orang yang asal-asalan jelas hasilnya akan jelek.

Di sisi lain tes ini dikatakan bebas pengaruh budaya, bebas pengaruh bahasa; karena soalnya berupa disain gambar, di wilayah atau daerah mana pun orang membaca atau menafsirkan gambar ya begitu, sesuai dengan bahasa dalam dirinya (Anastasi, 1988; Naglieri 2009). Namun Matsumoto (2008: 135-136) mengatakan tidak ada tes yang bebas dari pengaruh budaya (*intelligence tests were biased and did not accurately measure the mental ability of people from different cultures*). Dikatakan lebih lanjut bahwa : "... *There are ethnic group differences in measured intelligence (although the ethnic groups*

**Vol. 2 No. 1**

*scoring low on the standard tests change across time). The average score of some minority groups in the United State are 12 to 15 percentage point lower than the average for European Americans*". Hal itu tidak berarti bahwa tidak ada orang yang pandai dalam kelompok itu, namun secara umum lebih rendah.

Contoh lainnya, dari hal di atas adalah sebagian orang Indonesia susah jika diajak berfikir kritis, atau berpikir tinggi, atau menggunakan otaknya; belum apa-apa suka mengatakan "pusing". Ini adalah budaya yang dapat membuat skor dalam tes rendah, distraktor menjadi jelek, karena pengaruh cara berpikir yang kurang baik. Oleh karena itu dalam kurikulum 2013 dikatakan bahwa hal yang sangat perlu diubah dari para guru adalah "mindset" (pola pikir) yang salah.

Terkait dengan perhitungan distraktor di atas, mungkin banyaknya yang salah atau yang memilih option salah sebarannya menjadi tidak seimbang, disebabkan karena menebak, memilih dengan tidak sepenuh hati, atau asal memilih saja (Sunda: puraga tamba kadenda), yang oleh Karno To (2003) dikatakan memilih secara tidak bijak. Berdasarkan hasil analisis rasional, option-option (7 options dalam setiap soal) yang dibuat dalam setiap pola gambar di setiap soal APM, sudah dibuat sedemikian rupa, dibuat se-logis mungkin. Namun karena budaya "menebak" atau "asal mengisi", menyebabkan beberapa option tidak ideal.

Berdasarkan penelitian Subino (1984: 241-242) soal-soal pada APM 90,7% memuat faktor "g", yang menurut Freeman (1965) mengukur hubungan logis dari hal-hal yang dinyatakan secara *non-verbal*. Untuk mengerjakan soal-soal APM ini seseorang dituntut untuk mampu berfikir analisis-terpadu, abstrak, dan komparatif. Pola berpikir demikian merupakan unjuk perbuatan intelektual yang harus dilakukan se-efisien mungkin (dalam waktu yang terbatas). Itu sebabnya APM merupakan tes yang mengukur *efisiensi kerja intelektual*. Soal sisanya (9,3%) mengukur faktor "s" yang diperoleh dari belajar dan pengalaman. Artinya model-model soal yang mengukur faktor "s" ini, merupakan persoalan yang ditemukan oleh testi dalam belajar dan kehidupan sehari-hari.

Untuk mendapatkan informasi lengkap, faktor apa saja yang mempengaruhi pencapaian skor APM seseorang, Matsumoto dan Juang (2008) menyarankan untuk mengaitkannya dengan kualifikasi sekolah, usia, cita-cita, kondisi rumah (keluarga), jenis kelamin, tempat tinggal, status anak dalam keluarga (jumlah saudara), dan cara belajar.

**G. KESIMPULAN**

**Vol. 2 No. 1**

Mengacu pada hasil penelitian di atas, dapat disimpulkan bahwa tes APM -- dilihat dari kualitas item-itemnya (daya beda, tingkat kesukaran, homogenitas item, dan pengecoh- pengecohnya) -- telah memenuhi sebagai tes yang baik. Oleh karena itu hasil tes APM dapat dijadikan dasar pengambilan keputusan tentang kualitas inteligensi seseorang. Konsekwensinya harus dijaga dan diamankan secara profesional, jangan sampai bocor. Karena jika suatu soal atau suatu tes sudah bocor, tes tersebut tidak akan menjadi tidak akan mampu membedakan lagi, tingkat kesukaran berubah, menjadi bias dan kualitasnya berubah, (Izard :1977).

APM adalah tes non verbal yang bebas dari gangguan kemampuan menggunakan bahasa verbal. Ravens (2011: 8) mengatakan bahwa “ ... *APM is a nonverbal assessment tool designed to measure an individual's ability to perceive and think clearly, make meaning out of confusion and formulate new concepts when faced with novel information*”. Jadi tes ini disusun untuk mengukur kemampuan berpikir jernih, yang oleh Subino (1984: 9) dikatakan sebagai *efisiensi kerja intelektual*, yang akan menentukan sukses atau tidaknya seseorang dalam belajar; yang ditunjukkan dengan belajar secara mudah, cepat, dan tepat. Karena mengerjakan tes ini membutuhkan kerja intelektual secara efisien, maka orang yang asal-asalan jelas hasilnya akan jelek.

**H. DAFTAR PUSTAKA**

- Anastasi, Anne (1988), *Psychological Testing*, Sixth Edition, New York: Macmillan Publishing Company
- Crocker, Linda; dan Algina, James, (1986), *Introduction to Classical & Modern Test Theory*, New York: Holt, Rinehart, and Winston Inc.
- Drummond, Robwrt J. & Jones, Karyn D., (2010), *Assessment Procedures for Counselors and Helping Professionals*, 7th Edition, Boston: Pearson
- Fudyartanta, K. (2004), *Tes Bakat dan Penskalaan Kecerdasan*, Yogyakarta: Pustaka Pelajar
- Hays, Danica G. (2013). *Assessment in Counseling: A Guide to Use of Psychological Assessment Procedures*. Fifth edition. New York: Library of Congress Cataloging in Publication Data
- Izard, John F. (1977). *Construction and Analysis of Classroom Tests*. The Australian Council for Educational Research Limited: Hawthorn, Victoria: Australia.

**Vol. 2 No. 1**

- Matsumoto, D. dan Juang, Linda, (2008), *Cultural & Psychology*, Fouth Edition, Belmont: Thomson High Education
- Murphy, Kevin R., & Davidshofer, Charles O. (1988). *Psychological Testing: Principle and Applicaations*. Fourt Edition. New York: Macmillan Publishing Company.
- Nitko, Anthony J., (1996), *Educational Assessment of Student*, second edition, New Jersey: Merrill, an imprint of Prentice Hall
- Naglieri, Jack A. & Goldstein, Sam. (2009). *Practitioners Guide To Assessing Inteligence and Achievement*. New Jersey: John Wiley and Sons
- Raven, J.C. (2011). *Raven's Advanced Progressive Matrices: International Technical Manual*. U.K.: Pearson. Inc. All rights reserved.
- Rimm, Sylvia B. (2000). *Why Bright Kids Get Poor Grade. Mengapa Anak Pintar Memperoleh Nilai Buruk*. alih bahasa A. Mangunhardjana. Jakarta: PT. Grasindo
- Subino, Hadisubroto. (1984). *Analisis terhadap Kecocokan APM bagi Siswa-siswa Kelas I SMA Dikaitkan dengan Beberapa Faktor yang Melatarbelakangi Kehidupannya*. Disertasi. Bandung: Sekolah Pasca Sarjana IKIP Bandung. Tidakditerbitkan
- Subino, Hadisubroto. (1987). *Konstruksi dan Analisis Tes: Suatu Pengantar Kepada Teori Tes dan Pengukuran*. Jakarta: P2LPTK
- Sukardi, D. K, dan Kusmawati, N. (2009). *Analisis Tes Psikologis Teori dan Praktek (dalam Penyelenggaraan Layanan Bimbingan dan Konseling di Sekolah)*. Jakarta: RinekaCipta.
- Surapranata, S. (2006). *Analisis Validitas Reliabilitas dan Interpretasi Hasil Tes (Implementasi Kurikulum 2004)*. Bandung: RemajaRosdakarya.
- Suryabrata, S. (1999). *Pengembangan Alat Ukur Psikologis*. Jakarta: Direktorat Jenderal Pendidikan Tinggi Departemen Pendidikandan Kebudayaan.
- Surya, Moch. (1983). *Pengaruh Faktor-Faktor Non Intelektual Terhadap Gejala Berprestasi Kurang*. Jakarta: Depdikbud. DirjenDikti
- Widiawati, D. (2006). *Modul Mata Kuliah Psikodiagnostikpada Program Magister Profesi Psikologi Universitas Gadjah Mada*. Yogyakarta: Universitas Gadjah Mada
- Wright, A. Jordan, (2011), *Conducting Psychological Assessment: A Guide for Practitioners*, New Jersey: John Wiley & Sons, Inc.