

Perbandingan Metode *Concordance* Berdasarkan Teori Tes Klasik

Oleh : Heri Retnawati & Kana Hidayati
FMIPA Universitas Negeri Yogyakarta

ABSTRACT

This research is aimed at revealing the accuracy of concordance method in the classical test theory using parallel method, linear parallel method and equipercentil method, applying the concordance and making the conversion tables as result of concordance.

The object of this study was the regional science examination tests at the JHSs in Yogyakarta city in 2006/2007 academic year. This study used tests and testees' responses to the test. The analysis has done using 3 steps, they were estimated the characteristic items of tests, equated the tests and compare the methods of concordance, and then made conversion table to score testees' achievement.

The result indicated that equating without common item (concordance) from science score of SMPN 1 Paliyan Gunungkidul 2006/2007 to science score of SMPN 10 Yogyakarta using the linear method satisfied the equation $p_{YK}(x) = 1.22722x + 5.0004$, and the almost similar result is obtained using the equipercentile method, using linear method, obtained RMSE 9.333964, while with the method equipersentil, obtained RMSE 9.326277, indicating the equipercentil method is little more accurate compared with linear method, and the conversion tables as result of concordance in this research is enclosed.

Keyword : characteristic items, tests, concordance

A. Pendahuluan

Pada Ujian Akhir Nasional (UAN) tahun ajaran 2004/2005, pemerintah pusat menyiapkan 3 perangkat tes UAN (Matematika, Bahasa Indonesia dan Bahasa Inggris), sedangkan sekolah menyiapkan 7 mata pelajaran yang lain. Karena tidak semua sekolah mampu menyiapkan sendiri ujian akhir ini, pelaksanaan di DI Yogyakarta dikembangkan oleh Musyawarah Kerja Kepala Sekolah (MKKS), yang masing-masing mempunyai wilayah kerja di tiap kabupaten/kota. Ketujuh ujian akhir dengan perangkat tes yang dibuat oleh MKKS ini dinamai dengan Ujian Akhir Sekolah (UAS).

Ujian akhir tingkat menengah pertama dilaksanakan dengan tujuan yang sama, yakni mengukur kompetensi siswa di akhir pembelajaran, dan dikembangkan dari standar kompetensi pokok yang sama. Terkait dengan hal ini, tes-tes yang dibuat diharapkan mengukur konstruk yang sama, dan memenuhi syarat tes yang baik, ditinjau dari karakteristik butir-butir penyusunnya. Mengingat perangkat ujian di DI Yogyakarta dibuat oleh MKKS yang berbeda-beda, butir-butir penyusunnya mempunyai tingkat kesulitan yang berbeda-beda pula. Terkait

dengan hal ini, perlu dilakukan menghubungkan asesmen (*linking assesment*) untuk dapat terjadi adanya pembakuan tingkat kemampuan (*ability*) yang diperoleh siswa, dan juga mengetahui pembakuan pencapaian pembelajaran di sekolah-sekolah.

Hubungan (*linking*) antar tes dapat dikelompokkan menjadi 3, yakni penyetaraan (*equiting*), *concordance*, dan prediksi (*prediction*) (Kollen dan Brennan, 2004). Yang membedakan ketiga hubungan-hubungan ini adalah konstruk tes dan distribusinya. Jika tes-tes tersebut secara statistik dan konseptual dapat saling menggantikan, maka hubungan dapat diketahui dengan penyetaraan (*equiting*), jika sama distribusinya (mengukur konstruk yang sama) dengan *concordance*, dan jika kondisi untuk penyetaraan dan *concordance* tidak terpenuhi, digunakan prediksi skor harapan.

Ada 2 pendekatan yang dapat ditempuh untuk melakukan *concordance* pada 2 perangkat tes, dengan pendekatan teori tes klasik dan teori respons butir. Berdasarkan teori tes klasik, ada 3 metode yang dapat digunakan, yakni metode linear, metode linear sejajar, dan metode equipercentile (Yin, Brennan dan Kolen, 2004). Dengan menggunakan pendekatan teori respons butir, ada 4 metode yang dapat ditempuh, yakni metode regresi, metode rerata dan sigma, metode rerata tegar dan sigma, dan metode kurva karakteristik (Hambleton, Swaminathan, dan Rogers, 1991). Dalam penelitian ini hanya akan diteliti mengenai *concordance* pada 2 perangkat tes dengan pendekatan teori tes klasik

B. Tujuan dan Manfaat Penelitian

Tujuan penelitian ini untuk mengetahui urutan ketepatan metode *concordance* dengan dengan pendekatan teori tes klasik (metode linear sejajar dan metode *equipersentil*). Selain itu, penelitian ini juga bertujuan bagaimana penerapan *concordance* pada pada perangkat tes IPA UAN SMP/MTs yang disusun oleh MKKS di lima kabupaten/kota di DI Yogyakarta dengan menyusun tabel konversi yang dapat digunakan untuk mengukur keberhasilan pembelajaran IPA SMP/MTs di wilayah DI Yogyakarta.

Secara teoretis, hasil penelitian ini diharapkan dapat memberikan sumbangan teori kepada pengetahuan di bidang pengukuran dan pengujian, khususnya metode *concordance* dalam situasi/kasus yang tertentu. Secara praktis, hasil penelitian ini diharapkan dapat memberikan sumbangan kepada lembaga pengukuran dan pengujian di Indonesia, dalam

memilih metode *concordance* yang akurat, untuk memperoleh pembakuan kompetensi peserta tes (siswa) dan pembakuan kualitas pembelajaran di sekolah-sekolah.

C. Tinjauan Pustaka

Teori tes klasik atau disebut teori skor murni klasik (Allen dan Yen, 1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Jika dituliskan dengan pernyataan matematis, maka kalimat tersebut menjadi

$$X = T + E \dots\dots\dots (III.1)$$

dengan :

X : skor amatan,

T : skor sebenarnya,

E : skor kesalahan pengukuran (*error score*).

Kesalahan pengukuran yang dimaksudkan dalam teori ini merupakan kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara teoretis dari skor amatan yang diperoleh dengan skor amatan yang diharapkan. Kesalahan pengukuran yang sistematis dianggap bukan merupakan kesalahan pengukuran. Validitas dan reliabilitas pada perangkat tes digunakan untuk menentukan kualitas tes. Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran dan daya pembeda.

Skor-skor pada asesmen pendidikan dapat disetarakan secara statistik, dari satu unit asesmen ke unit asesmen yang lain, atau keduanya dapat dinyatakan dalam sebuah skala skor yang biasa. Cara ini disebut dengan menghubungkan dua tes (*linking*). Pada artikel ini, istilah *linking* merujuk pada sebuah hubungan antar skor dari dua tes. Seringkali dua tes yang dikaitkan ini mengukur konstruk yang sama, namun untuk kepentingan tertentu, mengaitkan dua tes yang berbeda konstraknya.

Istilah *concordance* merujuk pada mengaitkan skor-skor pada asesmen yang mengukur konstruk yang sama dan skor-skor ini dihubungkan untuk suatu kepentingan. Sebagai contoh, calon mahasiswa dapat menempuh tes SAT I atau tes ACT . Sebuah perguruan tinggi mewajibkan siswa memiliki skor komposit 1200 untuk Verbal dan Matematika. Jika perguruan

tinggi menerima siswa yang menempuh ACT, perguruan tinggi perlu mengetahui skor komposit ACT yang sebanding dengan skor minimum SAT I.

Flanagan (Dorans, 2004) menggunakan istilah *comparability* untuk menunjukkan tes-tes yang terskalakan yang memiliki distribusi skor yang sama dalam populasi nyata peserta-peserta tes. Lebih jauh, tes-tes yang skor-skoranya dapat dibandingkan jika dalam pemakaian, tes-tes tersebut dapat secara komplet saling dipertukarkan (*interchangeable*) dalam penggunaannya. Keadaan saling dipertukarkan ini terjadi jika tes-tes tersebut memiliki akurasi yang sama dan didesain sebagai bentuk-bentuk yang ekuivalen (*equivalent forms*). Dia menekankan pada skor bentuk multipel, tes yang dapat saling dipertukarkan jika terkonstruksi sama dan skala skor harus digunakan untuk menyatakan skor-skoranya. Mengenai hubungan *comparability* dengan reliabilitas, Flanagan menyatakan bahwa untuk menetapkan skor-skor yang dapat dibandingkan, distribusi skor sebenarnya (*true score*) seharusnya sama pada pengukuran-pengukuran dari tes-tes tersebut. Jika reliabilitas pengukuran sama untuk kedua tes, hasil yang sama akan diperoleh jika distribusi dari nilai yang ada diperbandingkan. Pada tes yang berbeda, tidak perlu memberikan skor-skor yang terbandingkan. Hal ini disebabkan karena regresi pada kedua tes tidak simetri, dan ini mengakibatkan kekurangcukupan pada interpretasi.

Untuk menghubungkan skor-skor tes bentuk multiple yang memiliki bentuk-bentuk lain yang dibangun dengan spesifikasi yang sama, Angoff (Dorans, 2004) menggunakan istilah penyetaraan (*equiting*). Sedangkan istilah yang digunakan untuk menyatakan hubungan antara dua tes yang memiliki konstruk yang sama tetapi berbeda dalam tingkat kesulitan dan reliabilitas, digunakan istilah kalibrasi (*calibration*) dan *comparability* digunakan untuk menyatakan hubungan tes-tes yang secara konstruk berbeda. *Comparability* merupakan keadaan ketika skor dari tes-tes yang berbeda diskalakan untuk memiliki sebuah distribusi biasa untuk menilai kekuatan dan kelemahan peserta tes terhadap suatu grup yang dijadikan acuan. Ketika peserta tes memilih untuk menempuh suatu tes, tingkat kemampuannya akan berbeda jika menempuh tes yang lain. Skor-skor tes yang dibandingkan menjadi tidak unik lagi. Ketidak unikan ini disebabkan oleh karena pada kenyataannya alat-alat ukur memiliki fungsi yang berbeda, sehingga tidak ada konversi tabel tunggal yang dapat diterapkan untuk semua grup. Menurut Angoff, kegunaan tabel skor yang dapat dibandingkan tergantung jawaban dari dua pertanyaan bagaimanakah kesamaan dari tes-tes untuk yang skornya dapat diperbandingkan dikembangkan

dan bagaimanakah kesesuaian dari grup pada tabel skor yang dapat diperbandingkan didasarkan pada satu pertimbangan person atau grup tujuan tabel tersebut digunakan.

Mislevy (1992) dan Linn (1993) (dalam Dorans, 2004) mengembangkan kerangka kerja untuk menghubungkan skor tes-tes meliputi empat tipe hubungan statistik, yakni penyetaraan (*equiting*), kaliberasi (*calibration*), moderasi statistik (*statistical moderation*) dan prediksi (*projection/prediction*). Seperti halnya Angoff, Mislevy/Linn menggunakan istilah *equiting* untuk menghubungkan skor-skor yang bentuk-bentuknya berbeda pada asesmen-asesmen. Kaliberasi digunakan untuk menghubungkan skor-skor tes yang mengukur konstruk yang sama tetapi berbeda dalam reliabilitas atau tingkat kesulitannya. Proyeksi dan moderasi statistik digunakan untuk menghubungkan tes-tes yang mengukur konstruk yang berbeda, menggunakan metode regresi. Pada istilah moderasi statistika, yang istilah ini juga digunakan oleh Kevees (1988), skor setiap tes dihubungkan dengan variabel ketiga yang disebut sebagai moderator. Grup-grup yang tesnya akan dihubungkan menempuh tes lain, yang merupakan variabel moderator. Mislevy dan Linn tidak membahas lebih lanjut hal-hal yang membedakan situasi yang mana yang mengukur konstruk yang sama dari situasi yang mengukur konstruk yang berbeda.

Freuer menggunakan istilah dan definisi yang sama dari istilahnya dan konsisten, seperti halnya yang dinyatakan oleh Mislevy dan Linn, yakni penyetaraan, kaliberasi, moderasi dan proyeksi (Dorans, 2004). Freuer menambahkan, ada 5 faktor-faktor untuk dipertimbangkan tentang skor-skor yang akan dihubungkan, yakni :

- a. kesamaan isi, tingkat kesulitan dan format butir.
- b. dapat diperbandingkannya kesalahan pengukuran yang terkait dengan skor-skor,
- c. kondisi administrasi tes,
- d. kegunaan dibuatnya tes dan konsekuensinya,
- e. akurasi dan stabilitas dari penyetaraan, termasuk stabilitas atas subgrup dan peserta ujiannya.

Dorans (2000, 2004) membedakan antara penghubungan skor-skor yang mengukur konstruk yang berbeda, dari skor-skor yang mengukur konstruk-konstruk yang sama. Penghubungan skor-skor pada tes-tes yang mengukur konstruk yang sama disebut dengan concordance. Konstruk yang sama ini diindikasikan dengan kesamaan isi, skor-skornya berkorelasi tinggi, dan hubungan antar peserta tes berbeda sedikit. Jika ada dua tes-tes yang

akan dihubungkan tidak memenuhi syarat concordance, maka untuk menghubungkannya dapat digunakan metode regresi.

Kesamaan konstruk berperan penting, dalam menentukan derajat menghubungkan skor-skor tes. Ada 3 derajat hubungan skor-skor tes-tes. Jika tes-tes tersebut secara statistik dan konseptual dapat saling menggantikan, maka hubungan dapat diketahui dengan penyetaraan (*equating*), jika sama distribusinya (mengukur konstruk yang sama) dengan *concordance*, dan jika kondisi untuk penyetaraan dan concordance tidak terpenuhi, digunakan prediksi skor harapan.

Menurut Hambleton, Swaminathan, dan Rogers (1991 : 123), ada dua jenis metode menghubungkan tes-tes, yakni metode linear dan metode equipersentil. Sedangkan Kollen dan Brennan (2004), selain kedua hal di atas, ada satu metode lagi, yakni metode linear sejajar.

Agar terjadi kekonsistenan, Kollen dan Brennan (1995) menggunakan istilah dan notasi, bentuk X didesinisikan sebagai tes yang baru, bentuk Y merupakan tes yang lama, x dan y merupakan skor observasi pada X dan Y. Notasi $l_Y(x)$, $pl_Y(x)$, dan $eq_Y(x)$ menyatakan fungsi statistik yang digunakan untuk mentransformasikan skor atas X ke skala skor Y berturut-turut menggunakan metode linear, metode paralel linear, dan metode equipersentil. Misalkan k merupakan indikator untuk grup atau sub populasi, dan misalkan K merupakan nomor total dari grup ($k=1,2,\dots,K$). Sebagai contoh, jika fokusnya jender, $K=2$ dan $k = 1,2$. Jika k tidak dinyatakan, maka persamaan dapat diterapkan untuk keseluruhan populasi.

Dengan menggunakan metode linear, persamaan transformasi untuk keseluruhan populasi dinyatakan dengan :

$$l_Y(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu(X)] + \mu(Y) \dots\dots\dots(III.6)$$

Persamaan transformasi untuk subgrup k dinyatakan dengan

$$l_{Yk}(x) = \frac{\sigma_k(Y)}{\sigma_k(X)} [x - \mu_k(X)] + \mu_k(Y) \dots\dots\dots(III.7)$$

Dengan metode linear sejajar, persamaan transformasi untuk subgrup k dinyatakan dengan

$$pl_{Yk}(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu_k(X)] + \mu_k(Y) \dots\dots\dots(III.8)$$

Adapun perbedaannya dengan metode linear, pada metode linear sejajar untuk keseluruhan populasi berlaku

$$\frac{\sigma(Y)}{\sigma(X)} = \frac{\sigma_k(Y)}{\sigma_k(X)} \dots\dots\dots(III.9)$$

Dengan metode equipersentil, perbedaan isi dan tingkat kesulitan antar tes-tes dideskripsikan dengan transformasi nonlinear. Transformasi didefinisikan dengan menghubungkan skor pada suatu tes X dengan tes Y dengan persamaan :

$$eq_Y(x) = G^{-1}[F(x)] \dots\dots\dots(III.10)$$

dengan F merupakan fungsi distribusi kumulatif dari X, G merupakan fungsi distribusi kumulatif dari Y, dan G^{-1} merupakan invers dari G.

Kesalahan pengukuran standar (*Standard Error*, SE) digunakan untuk mengetahui keakuratan metode-metode penyetaraan. Menurut Kolen dan Brennan (1995), kesalahan penyetaraan pada suatu kasus dalam hal ini *concordance* dinyatakan sebagai selisih antara hasil *concordance* sebenarnya dibandingkan dengan nilai harapannya. Karena kedua tes memiliki skala yang sama, maka nilai harapan merupakan nilai asli hasil pengamatan. Pada desain kelompok ekivalen, kesalahan penyetaraan suatu kasus ditentukan dengan persamaan :

$$\hat{eq}_y(x_i) - E\left[\hat{eq}_y(x_i)\right] \dots\dots\dots(III.11)$$

Selanjutnya kesalahan *concordance* pada suatu data dinyatakan sebagai *Root Mean Square of Error (RMSE)*, yang dinyatakan dengan

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \left[\hat{eq}_y(x_i) - E\left[\hat{eq}_y(x_i)\right] \right]^2}{n}} \dots\dots\dots(III.11)$$

Untuk menentukan keakuratan metode, kriteria metode diindikasikan dengan RMSE yang kecil. Metode *concordance* tes C_1 dikatakan lebih akurat daripada metode *concordance* C_2 jika $RMSE(C_1) < RMSE(C_2)$.

D. Metode Penelitian

Pada penelitian ini, *concordance* dilakukan pada perangkat tes Ujian Akhir Sekolah untuk mata pelajaran IPA di Sekolah Menengah Pertama/Madrasah Tsanawiyah di DI Yogyakarta. Dipilihnya mata pelajaran IPA ini terkait bahwa IPA merupakan ilmu yang mendukung perkembangan teknologi. Selain itu, seiring dengan era otonomi daerah, tes-tes ini dibuat oleh badan daerah yakni Musyawarah Kerja Kepala Sekolah (MKKS) tiap kabupaten. Data utama dalam penelitian ini adalah tes Ujian Akhir Sekolah untuk mata pelajaran IPA di Sekolah Menengah Pertama/Madrasah Tsanawiyah di DI Yogyakarta untuk kemudian diketahui penyetaraannya. Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi yaitu dengan mengutip respons siswa peserta tes Ujian Akhir Sekolah untuk mata pelajaran IPA di Sekolah Menengah Pertama/Madrasah Tsanawiyah di DI Yogyakarta yang pada kasus ini diwakili oleh SMPN 10 Yogyakarta dan SMPN 1 Paliyan Gunungkidul.

Analisis data dilakukan melalui tiga tahap. Tahap pertama dilakukan untuk mengetahui karakteristik butir berdasarkan teori tes klasik. Kegiatan ini dilakukan dengan bantuan computer program Bilog 3.1. Tahap selanjutnya pada butir-butir soal yang baik, dilakukan *concordance* berdasarkan 3 metode penyetaraan berdasarkan teori tes klasik yakni metode linear, paralel linear dan metode equipersentil. Selanjutnya, keakuratan metode *concordance* diperbandingkan dengan melihat kesalahan standar penyetaraan. Metode dengan kesalahan standar terkecil merupakan metode yang paling akurat. Pada tahap ketiga, dibuat skala konversi untuk penyekoran untuk tiap-tiap kabupaten, yang dapat digunakan untuk membandingkan kualitas pembelajaran IPA SMP/MTs di wilayah DI Yogyakarta.

E. Hasil Penelitian dan Pembahasan

1. Hasil Penelitian

a. Karakteristik Perangkat Tes IPA berdasarkan teori tes Klasik

1). Karakteristik Perangkat Tes IPA SMPN 10 Yogyakarta

Karakteristik perangkat tes IPA Ujian Akhir Sekolah 2006/2007 dilakukan dengan bantuan program ITEMAN. Hasil analisis program *ITEMAN* berupa file statistik yang terdiri dari statistik butir soal dan statistik tes/skala. Disamping itu menghasilkan file skor yang menampilkan skor yang diperoleh tiap-tiap peserta tes. Hasil analisis untuk perangkat tes IPA Ujian Akhir Sekolah 2006/2007 disajikan sebagai berikut.

a). Tingkat kesukaran.

Tingkat kesukaran (*difficulty level*, p) pada analisis butir soal secara klasik merupakan skala rata-rata peserta tes yang menjawab benar suatu butir soal. Dalam analisis skala yang digunakan adalah proporsi menjawab benar. Adapun besaran tingkat kesukaran berkisar 0 sampai dengan 1. Tingkat kesukaran suatu butir soal pada penelitian ini dapat dikategorikan menjadi tiga, yakni mudah ($p > 0.70$), sedang ($0,30 \leq p \leq 0,70$), dan sukar ($p < 0.30$).

Berdasarkan hasil analisis, dengan kategori butir pada tabel diatas diperoleh hasil yang disajikan pada tabel2.

Tabel 1. Kategori Butir Soal Berdasarkan Indeks Tingkat Kesulitan

No	Kategori	Butir Soal
1	Mudah	1,3,5,7,12,17,18,21,23,25,26,27,28,30,34,43,44
2	Sedang	2,8,9,10,11,13,14,15,19,22,24,29,30,32,35,36,37,38,39,40,41
3	Sukar	4,6,16,20,31,33,42,45

Mencermati hasil analisis ini, dapat diperoleh bahwa hanya terdapat 17 butir soal yang mudah atau sama dengan 38%, dan 20 butir soal yang sedang atau sama dengan 44% dan terdapat 8 butir soal yang sukar atau 18%.

b). Daya Beda

Daya pembeda dikategorikan menjadi 4 yakni (1) 0,40 ke atas baik; (2) 0,39 – 0,39 baik dapat diterima tapi perlu direvisi; (3) 0,20 – 29 cukup tetapi perlu direvisi; (4) (< 19) tidak baik (dibuang atau direvisi).

Berdasarkan hasil analisis, diperoleh hasil yang disajikan pada tabel VI.4

Tabel 2. Kategori Butir Berdasarkan Indeks Daya Beda

No.	Kategori	Butir Soal
1	Baik	1,2,3,5,7,8,9,10,11,12,13,17,18,21,22,25,26,27,28,30,32,34,35,36,38,40,43,44
2	Cukup	4,14,15,19,23,24,29,32,37
3	Perlu direvisi	6,16,20,31,41,42

4	Dibuang	45
---	---------	----

Dari data tersebut diatas diperoleh besarnya daya beda, butir soal yang baik berjumlah 28 butir atau sama dengan 62 %, Butir soal kategori sedang dan tidak perlu direvisi berjumlah 10 soal atau sama dengan 22 % dan butir soal kategori perlu direvisi berjumlah 6 butir atau sama dengan 14 %, sedangkan soal kategori jelek dan harus dibuang berjumlah 1 butir soal atau 2 %.

3). Pengecoh (*Distractor*)

Berdasarkan hasil analisis tingkat keberfungsian pengecoh, maka ditemukan butir yang pengecohnya berfungsi baik sebanyak 23 atau 51 % dan soal yang pengecohnya tidak berfungsi sebanyak 22 orang atau 49 %,

2. Karakteristik Perangkat Tes IPA SMPN 1 Paliyan Gunungkidul

a). Tingkat kesukaran.

Kriteria suatu butir soal mudah, sedang dan sukar mengacu pada yang disampaikan Allen & Yen (1973). Berdasarkan hasil analisis, pada tabel diatas diperoleh hasil sebagai berikut.

Tabel 3. Kategori Butir Berdasarkan Tingkat Kesulitannya

No	kategori	Butir Soal
1	Mudah	5, 41
2	Sedang	1, 2, 3, 6, 7, 9, 10, 14, 16, 17, 18, 21, 24, 25, 28, 30, 31, 32, 35, 36, 37, 39, 40, 43, 44, 45
3	Sukar	4, 8, 11, 12, 13, 15, 19, 20, 22, 23, 26, 27, 29, 33, 34, 38, 42

Dalam analisis ini hanya terdapat 2 butir soal yang mudah atau sama dengan 4%, dan 26 butir soal yang sedang atau sama dengan 58% dan terdapat 17 butir soal yang sukar atau 38%.

b). Daya beda

Dari hasil analisis diperoleh informasi sebagai berikut :

Tabel 4. Kategori Butir Berdasarkan Indeks Daya Beda

No.	Kategori	Butir Soal
1	Baik	6, 7, 8, 13, 20, 21, 24, 31, 34, 35, 37, 44,

2	Cukup	3, 12, 16, 30, 36, 39, 40, 42, 43
3	Perlu direvisi	1, 4, 5, 9, 10, 11, 17, 18, 27, 29, 38
4	Dibuang	2, 14, 15, 19, 22, 23, 25, 26, 28, 32, 33, 41, 45

Dari data tersebut diatas diperoleh besarnya daya beda, butir soal yang baik berjumlah 12 butir atau sama dengan 27 %, butir soal kategori diterima dan tidak perlu direvisi berjumlah 9 butir atau samadengan 20 % dan butir soal kategori perlu direvisi berjumlah 11 butir tau sama dengan 24 %, sedangkan soal kategori jelek dan harus dibuang berjumlah 13 butir soal atau 29 %.

3). Pengecoh (Distractor)

Berdasarkan hasil analisis tingkat keberfungsian pengecoh, maka ditemukan soal yang pengecohnya berfungsi baik sebanyak 18 atau 40 % dan soal yang pengecohnya tidak berfungsi sebanyak 27 butir atau 60 %,

2. Concordance dengan metode Linear

Memperhatikan persamaan III.7, III.8 dan III.9 pada penelitian ini dan mensubstitusikan hasil analisis dengan SPSS, *concordance* dengan metode linear sejajar yang memenuhi persamaan

$$l_{Yk}(x) = \frac{\sigma_k(Y)}{\sigma_k(X)} [x - \mu_k(X)] + \mu_k(Y)$$

merupakan kasus dari metode linear

$$pl_{Yk}(x) = \frac{\sigma(Y)}{\sigma(X)} [x - \mu_k(X)] + \mu_k(Y)$$

disebabkan oleh

$$\frac{\sigma(Y)}{\sigma(X)} = \frac{\sigma_k(Y)}{\sigma_k(X)}. \text{ Oleh sebab itu, dalam penelitian ini kedua metode ini dirangkum dalam satu}$$

metode saja, yakni metode linear.

Dengan menggunakan data pada lampiran 3, dapat diperoleh bahwa dengan metode linear, *concordance* skor siswa dari SMPN 1 Paliyan ke SMPN 10 Yogyakarta memenuhi persamaan :

$$p_{Y_k}(x) = \frac{4.97191}{4.05135} [x - 18.8744] + 28.1635$$

atau disederhanakan menjadi $p_{Y_k}(x) = 1.22722 x + 5.0004$.

Selanjutnya dapat dibuat tabel konversi sebagai berikut.

Tabel 5. Tabel Konversi dari Tes UAS SMPN 1 Paliyan ke SMPN 10 Yogyakarta

Skor	Skor Baru	Skor	Skor Baru	Skor	Skor Baru
1	6.22762	16	24.63592	31	43.04422
2	7.45484	17	25.86314	32	44.27144
3	8.68206	18	27.09036	33	45.49866
4	9.90928	19	28.31758	34	46.72588
5	11.1365	20	29.5448	35	47.9531
6	12.36372	21	30.77202	36	49.18032
7	13.59094	22	31.99924	37	50.40754
8	14.81816	23	33.22646	38	51.63476
9	16.04538	24	34.45368	39	52.86198
10	17.2726	25	35.6809	40	54.0892
11	18.49982	26	36.90812	41	55.31642
12	19.72704	27	38.13534	42	56.54364
13	20.95426	28	39.36256	43	57.77086
14	22.18148	29	40.58978	44	58.99808
15	23.4087	30	41.817	45	60.2253

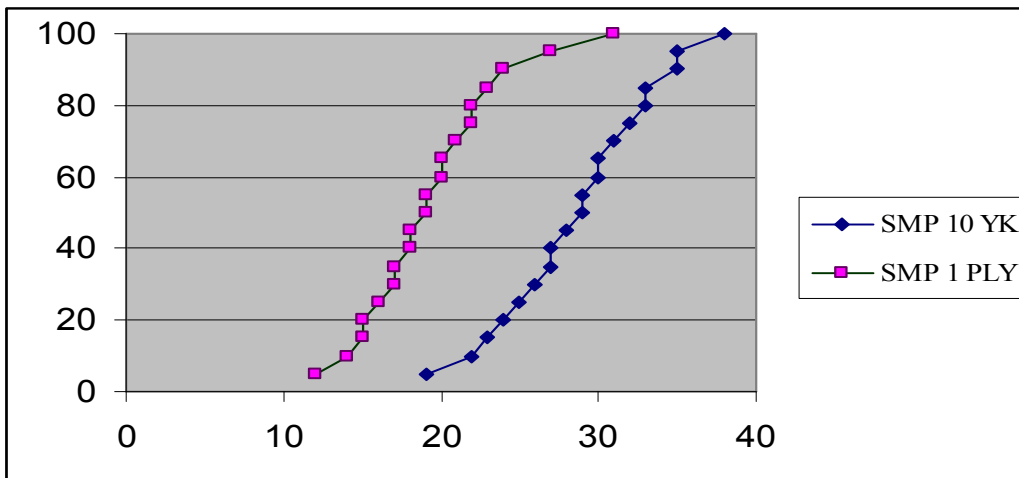
3. Concordance dengan metode Equipercetile

Dengan bantuan SPSS, data skor siswa dianalisis untuk mengetahui persentil data tersebut. Selanjutnya, hasil analisis ini disusun sebagai berikut.

Tabel 6. Persentil Data

Persentil ke-	SMPN 10 YK	SMPN 1 PALIYAN	Persentil ke-	SMPN 10 YK	SMPN 1 PALIYAN
5	19	12	55	29	19
10	22	14	60	30	20
15	23	15	65	30	20
20	24	15	70	31	21
25	25	16	75	32	22
30	26	17	80	33	22
35	27	17	85	33	23
40	27	18	90	35	24
45	28	18	95	35	27
50	29	19	100	38	31

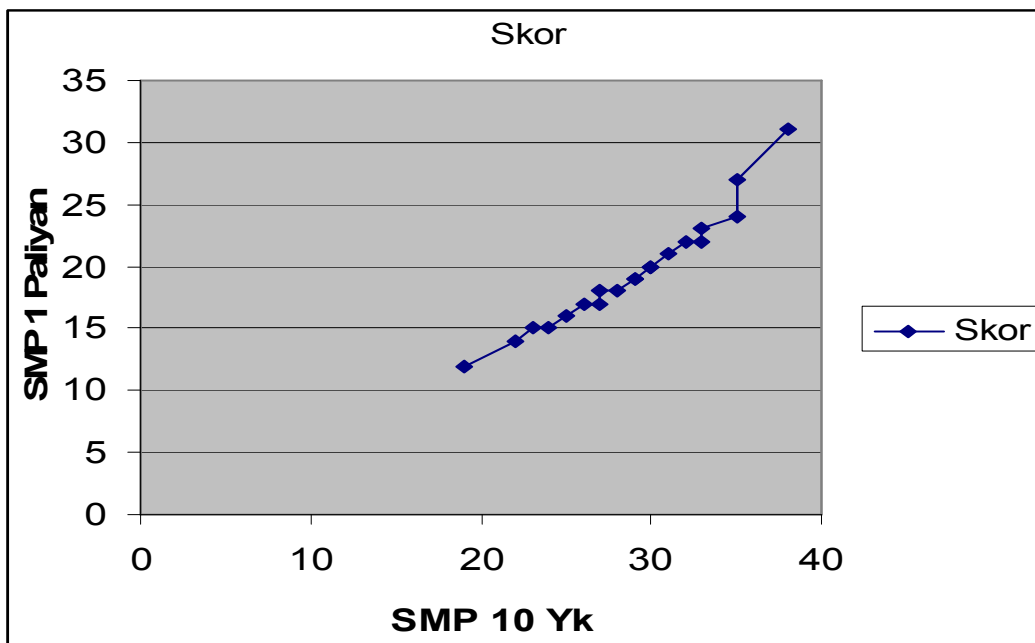
Jika hasil ini disajikan dalam bentuk grafik, maka akan nampak seperti pada gambar berikut.



Gambar 1. Grafik Persentil Data

Hasil ini menunjukkan bahwa pada persentil yang sama, skor perolehan siswa di SMPN 10 Yogyakarta lebih tinggi dibandingkan dengan skor perolehan siswa di SMPN 1 Paliyan Gunungkidul.

Untuk menyusun tabel konversi, selanjutnya dibuat grafik hubungan antara skor perolehan siswa di kedua SMPN tersebut Gambar 3.



Gambar 2. Hubungan Antara Skor Perolehan Siswa di SMPN 10 Yogyakarta dan SMPN 1 Paliyan Gunungkidul

Selanjutnya dapat disusun tabel konversi dengan metode equipersentile sebagai berikut.

Tabel 11. Tabel Konversi dengan Metode Equipersentile
dari Tes UAS SMPN 1 Paliyan Gunungkidul ke SMPN 10 Yogyakarta

Skor Siswa	Skor Siswa yang Baru	Skor Siswa	Skor Siswa yang Baru
12	19	21	31
13	20.5	22	22
14	22	23	33
15	23	24	35
16	25	25	35.6
17	26	26	36.2
18	27	27	36.8
19	29	28	37.4
20	30	29	38

4. Perbandingan Metode Concordance pada Teori Tes Klasik

Membandingkan langsung hasil concordance pada tabel konversi baik dengan menggunakan metode linear maupun dengan menggunakan metode equipersentil, dapat diperoleh bahwa perbedaan hasil dengan menggunakan kedua metode ini sangat kecil, dengan kata lain hasil konversi tidak terlalu berbeda.. Hasil ini diperkuat juga setelah menghitung RMSE. Dengan menggunakan metode linear, diperoleh RMSE sebesar 9.333964, sedangkan dengan metode equipersentil, diperoleh RMSE sebesar 9.326277 (hasil terlampir). Jika dicermati lebih lanjut, nampak bahwa dengan menggunakan metode equipersentil hasil RMSE sedikit lebih kecil. Hasil ini menunjukkan bahwa metode Equipersentil sedikit lebih cermat dibandingkan dengan dengan metode linear.

2. Pembahasan

Melihat dari sisi historisnya, perangkat tes ini dikembangkan berdasarkan kisi-kisi yang sama, seperti yang telah digariskan Departemen Pendidikan, dan kisi-kisi ini dikembangkan berdasarkan Standar Kompetensi Lulusan (SKL). Selanjutnya berdasarkan kisi-kisi, forum Musyawarah Kepala Sekolah dan Musyawarah Guru Mata Pelajaran (MGMP) tingkat kabupaten. Selanjutnya perangkat tes ini digunakan untuk 1 kabupaten/kota yang bersangkutan. Mencermati karakteristik perangkat tes UAS mata pelajaran IPA di SMPN 10 Yogyakarta dan SMPN 1 Paliyan Gunungkidul, nampak bahwa ada perbedaan yang cukup kontras dari kedua perangkat tes tersebut.

Rerata tingkat kesulitan butir pada tes UAS untuk SMPN 1 Paliyan Gunungkidul lebih rendah dibandingkan dengan tes UAS untuk SMPN 10 Yogyakarta. Hal ini menunjukkan bahwa

tes untuk Gunungkidul lebih sulit dibandingkan dengan tes di kota Yogyakarta. Reliabilitas tes di SMPN 10 Yogyakarta juga lebih baik dibandingkan dengan di SMPN 1 Paliyan. Hal ini mengakibatkan tes UAS di SMPN 10 Yogyakarta memiliki varians yang besar dalam menjelaskan kemampuan IPA siswa, dan juga mengakibatkan rerata prestasi siswa dengan indikator skor perolehan siswa di SMPN 1 Paliyan Gunungkidul lebih rendah dibandingkan dengan rerata skor perolehan siswa di SMPN 10 Yogyakarta.

Terkait dengan skor siswa SMPN 1 Paliyan yang lebih rendah, diperkuat oleh hasil penyetaraan tes paralel atau *concordance* dengan metode linear memenuhi persamaan $pl_{YK}(x) = 1.22722x + 5.0004$. Demikian pula halnya dengan metode *equipersentil*. Jadi nilai siswa SMPN 1 Paliyan akan lebih tinggi jika di bawa ke Kota Yogyakarta, dan bisa terjadi sebaliknya. Ini akan menjadi permasalahan yang pelik bagi masyarakat yang belum mengerti konsep pengukuran, sehingga konsep penyetaraan akan disebut sebagai “subsidi silang” nilai perolehan siswa.

Mencermati hasil *concordance* dengan metode linear dan metode *equipercentile*, nampak bahwa hasil *concordance* dengan metode *equipercentile* sedikit lebih cermat dibandingkan dengan metode linear. Hal ini disebabkan karena keduanya tetap menggunakan pendekatan yang hampir sama, yakni pendekatan garis lurus untuk mengestimasi skor konversi. Pada metode linear, skor konversi didekati dengan persamaan garis dengan mengestimasi gradien, sedangkan pada metode *equipercentile* pendekatan dilakukan dengan garis yang melalui titik titik percentile dari data skor perolehan siswa kedua kelompok yang mengerjakan tes-tes yang disetarakan.

F. Penutup

Berdasarkan hasil analisis, dapat disimpulkan sebagai berikut.

1. Penyetaraan tanpa butir bersama (*concordance*) skor tes UAS siswa SMPN 1 Paliyan mata pelajaran IPA tahun 2006/2007 ke skor tes UAS siswa SMPN 10 Yogyakarta dengan metode linear memenuhi persamaan $pl_{YK}(x) = 1.22722x + 5.0004$, dan diperoleh hasil yang hampir sama untuk metode *equipercentile*.
2. Dengan menggunakan metode linear, diperoleh RMSE sebesar 9.333964, sedangkan dengan metode *equipersentil*, diperoleh RMSE sebesar 9.326277, yang menunjukkan bahwa metode *equipercentil* sedikit lebih cermat dibandingkan dengan metode linear.

3. Tabel konversi hasil concordance pada penelitian ini terlampir.

Tes untuk Ujian Akhir Sekolah (UAS) yang dikembangkan oleh MKKS dan MGMP ini dikembangkan secara langsung dari kisi-kisi, kemudian langsung dirakit dalam suatu tes. Mengingat proses perakitan tanpa melalui ujicoba ini, masih dipertanyakan karakteristik masing-masing butir dalam mengestimasi kemampuan. Suatu permasalahan yang krusial, mengingat ujian yang dilaksanakan ini ujian akhir siswa di SMP yang hasilnya digunakan sebagai salah satu penentu kelulusan siswa.

Pemerintah melalui lembaga yang bersangkutan dapat mengatasi anggapan masyarakat dengan adanya anggapan "subsidi silang" skor perolehan siswa. Agar estimasi kemampuan siswa tepat, perlu adanya butir bersama (*common items*) untuk masing-masing wilayah, sehingga evaluasi untuk siswa dan mutu sekolah dapat lebih akurat dan bersifat lebih adil. Penelitian seperti ini juga perlu dilakukan di wilayah lain untuk mengetahui hubungan antara beberapa tes yang setara.

Penelitian ini menggunakan pendekatan teori tes klasik, yang mempunyai kelemahan adanya ketergantungan pada grup peserta tes (*group dependent*). Hal ini mengakibatkan, parameter butir soal tergantung pada kelompok peserta tes. Untuk lebih cermat, perlu dilaksanakan penelitian perbandingan metode concordance dengan pendekatan teori respons butir. Penelitian ini juga hanya menggunakan data real, tanpa melalui replikasi. Untuk hasil yang lebih akurat, perlu adanya penelitian simulasi yang membandingkan metode concordance dengan pendekatan teori tes klasik maupun teori respons butir.

DAFTAR KEPUSTAKAAN

- Abel, R. L. 1972. *Essential of Educational Measurement* (2rd ed). Englewood Cliffs. New Jersey: Prentice-Hall, Inc.
- Allen, M.J.& Yen, W.M. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth, Inc.
- Dorans, N.J. (2004). *Equating, Concordance and Expectation*. *Jurnal Applied Psychological Measurement*, Vol 28. No. 4, July 2004, p. 219-226

- Hanson, B.A dan Beguin, A.A. (2002) *Obtaining a common scale for IRT item parameters using separate versus concurrent estimation in the common item non equivalent groups equating design*. *Jurnal Applied Psychological Measurement*, Vol 26. No. 7, p. 24.
- Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.
- Hullin, C. L. , et al. (1983). *Item response theory : Application to psychological measurement*. Homewood, IL : Dow Jones-Irwin.
- Kerlinger, F.N. (1986). *Asas-asas penelitian behavioral* (Terjemahan L.R. Simatupang). Yogyakarta : Gajahmada University Press.
- Kolen, M.J. dan Brennan, R.L. (1995). *Test Equating : Methods and Practices*. New York : Springer.
- Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. New York : Hold, Rinehart and Wiston, Inc.
- Yinn, P.Y., Brennan, R.L., dan Kolen, M.J. (2004). *Concordance Between ACT and ITED Scores From Different Popolation*. *Jurnal Applied Psychological Measurement*, Vol 28. No. 4, July 2004, p. 219-226