

**Indeks Ketidakwaian Skor Berdasarkan Teori Respon
Butir Ditinjau dari Model Penskoran**
Oleh: *Sugiharto*¹

Abstrak

Penelitian ini bertujuan untuk mengetahui apakah pada kelompok siswa yang mengerjakan soal empat pilihan, proporsi skor wajar siswa yang dikoreksi dengan model penskoran PS lebih besar dibandingkan siswa yang dikoreksi dengan model penskoran CS.

Penelitian ini dilaksanakan pada 5 SMP negeri maupun swasta di kota Palangka Raya. Pengumpulan data dilaksanakan pada akhir semester genap tahun pelajaran 2016/2017. Penelitian ini merupakan eksperimen dengan menggunakan model perbandingan proporsi. Variabel bebas dalam penelitian ini adalah model penskoran. Model penskoran yang digunakan terdiri atas dua model yakni model penskoran CS dan PS. Variabel terikatnya adalah proporsi skor wajar berdasarkan indeks kewajaran skor yang dihitung berdasarkan teori respon butir dengan menggunakan logistik satu parameter. Rancangan penelitian yang digunakan adalah penelitian komparatif, yaitu membandingkan proporsi berdasarkan indeks kewajaran yang diperoleh. Dalam melakukan perbandingan dianalisis secara terpisah terhadap dua kelompok, yaitu siswa dengan pengoreksian model correct score dan punishment score.

Berdasarkan hasil analisis data dapat disimpulkan bahwa siswa yang dikoreksi dengan model penskoran punishment score mempunyai indeks kewajaran yang tidak lebih baik dibandingkan dengan indeks kewajaran siswa pada model penskoran correct score.

Kata kunci: ketidakwaian skor, correct score, punishment score

Salah satu komponen penting dalam suatu pembelajaran adalah evaluasi. Karena dengan evaluasi, selain dapat mengetahui keberhasilan indikator, dapat pula mengetahui sejauh mana pengajar mampu mengelola pembelajaran. Evaluasi merupakan salah satu lingkup yang memerlukan standarisasi dalam meningkatkan mutu pendidikan. Sebagai instrumen utama dalam evaluasi adalah tes. Tes merupakan suatu upaya untuk melakukan pengukuran terhadap tingkat pencapaian atau hasil belajar siswa. Hasil belajar tersebut

¹ *Sugiharto* adalah staf pengajar di FKIP UPR

seharusnya betul-betul menggambarkan kemampuan siswa yang semestinya. Dengan demikian hasil tersebut dapat membedakan siswa yang sudah memenuhi dan yang belum memenuhi standar yang sudah ditentukan pada standar kompetensi lulusan.

Namun pada kenyataannya informasi yang diperoleh melalui tes belum tentu bisa menjangkau sampai dimensi atau besaran yang hendak diukur oleh tes itu sendiri. Dapat terjadi bahwa hasil tes terkontaminasi dengan besaran selain yang diukur oleh tes bersangkutan. Skor yang diperoleh dari pengujian demikian merupakan skor tidak benar atau timpang. Jika sumber dari ketimpangan skor adalah siswa maka disebut ketidakwaian skor. Hal tersebut dapat terjadi jika siswa yang berkemampuan tinggi salah dalam menjawab soal mudah dan siswa yang berkemampuan rendah benar dalam menjawab soal yang sulit, siswa tidak banyak menjawab soal yang mudah, atau seorang siswa menjawab secara acak keseluruhan tes (Naga, 2001: 43). Hal senada juga diungkap oleh Hulin, Drasgow, and Parsons (1983: 110).

Hal tersebut dapat juga terjadi pada siswa yang menempatkan jawaban tidak pada butir yang semestinya. Misalkan siswa menjawab butir nomor 15 tetapi yang disilang nomor 16, atau siswa menjawab butir 16 yang disilang butir nomor 17, dan seterusnya. Karena waktu sudah habis, siswa tidak sempat mengecek jawabannya lagi dan akhirnya skor yang diperoleh rendah. Sebaliknya dapat terjadi siswa memperoleh skor lebih tinggi daripada kemampuan yang semestinya (*spuriously high*). Hal demikian terjadi kemungkinan siswa mencontoh atau mendapatkan jawaban dari siswa lain. Siswa akan mendapatkan skor yang lebih tinggi dari skor yang seharusnya sesuai dengan kemampuannya.

Ketidakwaian skor dapat juga berasal dari kondisi penilaian. keadaan psikologis siswa, misalnya cemas, khawatir, takut gagal, tidak dapat menulis dengan baik, dapat menyebabkan siswa tidak berhasil menyelesaikan soal secara benar butir-butir tes (Nitko, 1996: 91). Sebagai akibatnya siswa semacam ini akan mendapatkan skor yang tidak tepat, yakni tidak sesuai dengan kemampuan mereka semestinya (Crocker and Algina, 1986: 399).

Apabila terjadi ketidakwaian skor dan hal tersebut sulit dihindari, maka perlu dideteksi ketidakwaian skor tersebut baik secara individu maupun kelompok. Hasil deteksi dapat digunakan untuk mengambil keputusan yang sesuai tentang apa yang harus dilakukan

penyelenggara tes terhadap hasil pengukuran tersebut. Ada hal lain yang dapat mempengaruhi dalam perolehan skor, yaitu model penskoran. Tidak semua siswa terutama yang berkemampuan rendah menyadari bahwa model penskoran mempengaruhi skor akhir yang diperoleh.

Ada tiga model penskoran yang dikenal pada tes pilihan ganda yaitu: 1) penskoran dengan menghitung jumlah jawaban yang benar saja (*correct score*) sering juga disebut *number right score* atau *conventional scoring*, 2) penskoran dengan memberi sanksi pada jawaban yang salah (*punishment score*) atau *rights minus wrong correction*, dan 3) penskoran dengan memberi hadiah pada butir yang tidak dijawab (*reward score*) atau *correcting row score*, kedua model terakhir ini sering juga disebut sebagai *formula scoring* (Crocker and Algina, 1986: 401). Penerapan model penskoran yang berbeda-beda dapat berdampak pada skor yang diperoleh masing-masing peserta karena peserta akan mempertimbangkan kemungkinan untuk menjawab dengan cara menebak atau tidak menjawab butir soal yang sedang dikerjakan dan pada gilirannya dapat menyebabkan terjadinya ketidakwajaran skor.

Mata pelajaran yang dianggap sangat penting adalah matematika. Menurut Cornelius ada lima dasar perlunya belajar matematika: 1) berpikir jelas, logis, dan kritis 2) menyelesaikan masalah di kehidupan sehari-hari, 3) mengenali berbagai pola hubungan dan generalisasi pengalaman, 4) fasilitas untuk mengembangkan kreativitas, dan 5) fasilitas dalam mengingatkan kesadaran siswa terhadap perkembangan budaya (Abdurrahman, 1999: 253). Ditambahkan pula oleh Cockroft (1983: 5) bahwa kenapa guru perlu mengajarkan matematika kepada siswa: 1) diterapkan pada segala segi kehidupan, 2) ketrampilan matematika diperlukan bidang studi lain, 3) fasilitas komunikasi yang kuat, singkat dan jelas, 4) menyampaikan informasi dengan berbagai cara 5) melatih kemampuan berpikir logis, teliti, dan kemampuan keruangan, dan 6) menawarkan penyelesaian terhadap usaha pemecahan masalah yang rumit. Tujuan penelitian ini adalah untuk mengetahui yang dikoreksi dengan model penskoran PS memiliki skor wajar lebih tinggi dibandingkan dengan yang dikoreksi dengan model CS.

Tujuan diberikan tes terhadap peserta tes adalah untuk menyajikan informasi tentang besaran atau dimensi yang diukur oleh tes tersebut. Jika langkah-langkah yang ditempuh dalam pelaksanaan tes berjalan dengan baik, maka akan menghasilkan informasi yang benar

tentang peserta tes. Termasuk di dalamnya memperoleh informasi tentang butir tes yang digunakan dalam pengujian tersebut.

Namun kenyataan di lapangan, skor yang diberikan oleh peserta tes tidak memberikan informasi yang benar tentang peserta tes. Ada kemungkinan informasi tersebut tidak menjangkau sampai kepada besaran atau dimensi yang akan diukur oleh tes tersebut. Kemungkinan lain bahwa hasil tes tersebut tercampur oleh besaran lain yang seharusnya tidak diukur dalam tes dimaksud, sehingga hasil tes atau skor menjadi bias dan akan menyajikan informasi yang kurang tepat. Apapun penyebabnya, jelas bahwa hasil yang diperoleh yang berupa skor tes adalah tidak benar atau skor yang timpang. Dengan demikian bahwa ketimpangan skor pada tes merupakan skor yang tidak memberikan informasi yang benar tentang hal-hal yang dimaksudkan dalam suatu tes.

Ketimpangan skor di atas, dalam pengukuran pendidikan harus dihindari. Namun bagaimana jika hal tersebut tidak mampu dihindari, maka hanya bisa dideteksi. Berdasarkan hasil pendeteksian tersebut dapat ditentukan keputusan apa yang diambil terhadap hasil yang dicapai dalam pengujian tersebut.

Sumber ketimpangan skor berasal dari peserta tes dan perangkat tes. Ketimpangan yang bersumber dari peserta tes, dapat terjadi secara individu maupun kelompok. Sedangkan ketimpangan skor yang bersumber dari perangkat tes, dapat terjadi dari masing-masing butir atau dari seperangkat tes. Sumber lain dari ketimpangan skor adalah bentangan skor. Dengan demikian ketimpangan pada tes dapat dipandang dari kombinasi antara letak ketimpangan dan bentangan ketimpangan, yaitu a) pada peserta tes atau pada butir tes, dan 2) secara individu atau secara kelompok. Dua pasang pernyataan tersebut dapat dibentuk empat kombinasi ketimpangan skor. Pertama, ketimpangan skor pada peserta tes secara individu. Kedua, ketimpangan skor secara kelompok. Ketiga, ketimpangan skor pada suatu butir, dan keempat, ketimpangan skor pada butir secara kelompok atau seperangkat tes.

Dari penjelasan di atas bahwa pada kombinasi pertama dinamakan sebagai ketidakwaian skor, kombinasi ketiga adalah bias butir, dan yang keempat adalah bias perangkat tes. Dalam penelitian ini yang akan dibahas adalah ketidakwaian skor secara individu, sebagaimana disebut oleh Hulin dengan *inappropriatenes* (Hulin, Drasgow, and Parsons, 1983: 110).

Penjelasan sebelumnya telah menyinggung tentang ketidakwajaran skor, yaitu seorang peserta tes dapat memperoleh skor yang tidak sesuai dengan kemampuannya. Pada hal semua butir tes sudah baik, sedangkan kelompok peserta secara keseluruhan memperoleh skor yang baik. . seorang peserta tes dapat saja memperoleh skor jauh di bawah kemampuan yang sebenarnya. Sebaliknya seorang peserta tes dapat memperoleh skor lebih tinggi dari kemampuan sebenarnya. Menurut Hullin, Drasgow, and Parsons, 1983: 608) bahwa ketidakwajaran skor dapat dilihat dari keanehan pola jawaban siswa, yaitu pola jawabannya tidak normal. Ada sejumlah jawaban benar terhadap butir soal yang sulit pada setengah perangkat tes dan ada sejumlah jawaban salah terhadap butir tes yang mudah pada setengah perangkat tes kedua. Ada juga siswa yang kreatif memberikan penafsiran yang berbeda pada butir yang mudah. Akibatnya respons butir seperti ini tidak cocok dengan teori respon butir yang mengasumsikan peluang jawaban benar merupakan fungsi dari kecerdasan peserta tes.

Pendeteksian ketidakwajaran skor dapat dicari melalui indeks yang dikenal sebagai indeks kewajaran. Indeks kewajaran diestimasi dengan fungsi kebolehjadian maksimum. Fungsi kebolehjadian maksimum merupakan perkalian probabilitas keberhasilan dan kegagalan atau perkalian probabilitas jawaban benar dan jawaban salah di dalam menjawab butir tes bersangkutan. Probabilitas tersebut memiliki rentang skor antara 0 sampai dengan 1, sehingga perkaliannya atau fungsi kebolehjadian akan mencapai nilai maksimum juga terletak diantara 0 dan 1. Namun jika diberikan logaritma kepada fungsi kebolehjadian yang terletak diantara 0 dan 1 maka nilai logaritma ini akan menjadi negatif. Kebolehjadian maksimum akan menghasilkan nilai logaritma yang paling tinggi atau mendekati 1, demikian sebaliknya. Dengan demikian, semakin negatif nilai logaritma dari fungsi kebolehjadian, maka makin tidak wajar skor peserta yang bersangkutan.

Diantaranya ada empat metode indeks kewajaran skor yang dibicarakan di sini. Pada tiga metode pertama, metode yang satu merupakan perbaikan dari metode yang lainnya. Tiga metode indeks kewajaran itu adalah indeks kewajaran I_D , indeks kewajaran I_G , dan indeks kewajaran I_x .

Metode Penelitian

Penelitian ini dilaksanakan pada 7 SMP negeri maupun swasta di kota Palangka Raya. Pengumpulan data dilaksanakan pada akhir semester genap tahun pelajaran 2016/2017.

Rancangan penelitian yang digunakan adalah penelitian komparatif, yaitu membandingkan proporsi berdasarkan indeks kewajaran yang diperoleh. Perbandingan siswa dengan pengoreksian model *correct score* dan *punishment score*, yaitu perbandingan proporsi skor wajar yang diperoleh berdasarkan indeks kewajaran l_{gz} yang diperoleh antara siswa yang dikoreksi dengan model penskoran *correct score* dan *punishment score*. Selanjutnya dilihat perbandingan proporsinya dengan batas indeks kewajaran $l_{gz} = 0$. Proses dalam penelitian ini adalah: 1) pembuatan kisi-kisi instrumen tes, 2) pembuatan soal tes pilihan ganda sebanyak 30 butir, 3) validasi isi, 4) revisi instrumen, 5) pengambilan data, 6) uji persyaratan teori respon butir meliputi: uji unidimensi dan invariansi serta uji kecocokan model, 7) perhitungan indeks kewajaran, dan 8) membandingkan proporsi skor wajar berdasarkan indeks kewajaran l_{gz} .

Populasi dalam penelitian ini adalah seluruh siswa SMP negeri dan swasta se kota Palangka Raya sebanyak 2.233 siswa. Sampel minimal dihitung menggunakan rumus Isaac dan Michael (Sugiyono, 2012: 126) dengan taraf kesalahan 1% diperoleh 288 siswa. Namun untuk kepentingan analisis butir diambil per blok 250 siswa. Karena ada dua blok (kelompok) yaitu untuk kelompok CS dan PS, sehingga seluruh sampel menjadi 500 siswa.

Langkah-langkah pengumpulan data adalah: 1) pengelompokan siswa menjadi dua kelompok, 2) pemberian seperangkat soal matematika berbentuk pilihan ganda, 3) pengoreksian data berdasarkan kedua model penskoran PS dan CS, 4) pengestimasi parameter kemampuan masing-masing siswa dan parameter butir, 5) perhitungan indeks kewajaran skor l_{gz} siswa berdasarkan fungsi karakteristik butir yang diperoleh dari hasil estimasi menggunakan model karakteristik logistik satu parameter, dan 7) perhitungan proporsi skor wajar berdasarkan hasil estimasi kemampuan.

Data penelitian mengenai nilai indeks kewajaran (l_{gz}) dihitung berdasarkan rumus indeks kewajaran skor setelah parameter butir dan kemampuan responden diestimasi. Sedangkan tingkat kemampuan (θ) diperoleh setelah estimasi kemampuan dan parameter butir didapat melalui perhitungan dengan program Bilog. Sebelum dilakukan analisis data, terlebih dahulu dilakukan uji kecocokan model yaitu menguji apakah model logistik satu parameter cocok digunakan untuk data yang diperoleh dari lapangan. Persyaratan yang diuji antara lain persyaratan unidimensi dan invariansi baik invariansi parameter butir maupun

invariansi parameter peserta. Syarat unidimensi dilakukan dengan analisis faktor (program SPSS) dan invariansi dilakukan menggunakan korelasi Product Moment (Ecxel).

Hasil Penelitian Dan Pembahasan

Estimasi parameter dilakukan dengan menggunakan program Bilog, yakni perangkat lunak untuk mengestimasi parameter butir dan kemampuan siswa berdasarkan fungsi karakteristik butir logistik satu parameter. Dari proses estimasi tersebut diperoleh estimasi parameter butir berupa taraf sukar butir dan kemampuan siswa (θ) untuk masing-masing kelompok.

Berdasarkan hasil estimasi parameter taraf kesukaran butir (b) diperoleh harga-harga parameter mulai dari -1,475 sampai 1,253, berarti distribusi soal mulai dari sangat mudah sampai dengan sangat sukar.

Estimasi parameter kemampuan siswa digunakan program Bilog. Selanjutnya dicopy ke program Excel kemudian diurutkan dari terendah sampai tertinggi.

Indeks kewajaran skor dapat dihitung setelah semua persyaratan analisis teori respon butir diuji. Indeks kewajaran yang digunakan adalah indeks kewajaran Drasgow (ℓ_g) yang merupakan peningkatan indeks kewajaran Levine (ℓ_o) dengan langkah-langkah sebagai berikut:

- a. Hitung fungsi karakteristik butir LIP untuk masing-masing perangkat tes berdasarkan hasil estimasi parameter butir untuk masing-masing kelompok dengan rumus sebagai berikut:

$$P_i(\theta) = \frac{e^{(n-b_i)}}{1+e^{(n-b_i)}} \quad i = 1,2,3,\dots,n$$

b_i merupakan estimasi butir untuk $i = 1,2,3,\dots,30$

- b. Hitung indeks (ℓ_o) berdasarkan jawaban peserta tes untuk setiap peserta dan fungsi karakteristik butir pada langkah a dengan rumus:

$$\ell_o = \ln L(X|n) \text{ maksimum}$$

$$\text{dengan } L(X|n) = \prod_{j=1}^n P_j(\theta)^{x_j} \cdot Q_j(\theta)^{1-x_j}$$

c. Hitung indeks (ℓ_g) berdasarkan (ℓ_o) yang diperoleh dari hasil perhitungan langkah b dengan rumus: $\ell_g = e^{\ell_o/N}$ dengan N jumlah butir yang dijawab peserta tes.

d. Hitung indeks kewajaran (ℓ_{gz}) (Indeks kewajaran baku) dengan rumus: $\ell_{gz} = \frac{\ell_g - m_{\ell g}}{S_{\ell g}}$

dengan: m_{ℓ_g} = rerata ℓ_g untuk semua peserta tes

S_{ℓ_g} = simpangan baku ℓ_g untuk semua peserta tes

Fungsi karakteristik butir untuk masing-masing butir untuk tiap perangkat tes tiap kelompok dapat ditentukan berdasarkan hasil estimasi parameter butir. Berikut disajikan contoh perhitungan fungsi karakteristik untuk soal dengan tiga pilihan jawaban yang dikoreksi dengan model penskoran *correct score*. Dari hasil estimasi parameter butir satu diperoleh harga $b = 0,532$, disubstitusi ke fungsi karakteristik butir logistik satu parameter sehingga diperoleh fungsi karakteristik butir satu sebagai berikut:

$$P(\theta) = \frac{e^{(-0,532)}}{1 + e^{(-0,532)}}$$

Berdasarkan respon peserta tes dan fungsi karakteristik butir dapat dihitung indeks kewajaran skor untuk setiap peserta sesuai dengan tingkat kemampuan.

Perhitungan indeks kewajaran ℓ_{gz} untuk siswa yang mengerjakan soal tiga pilihan diperoleh tujuh siswa mempunyai skor wajar dan sembilan siswa mempunyai skor tak wajar dari 16 siswa berkemampuan rendah. Perhitungan selanjutnya diperoleh 18 siswa mempunyai skor wajar dan 14 siswa mempunyai skor tidak wajar dari 32 siswa dengan kemampuan tinggi. Untuk siswa yang mengerjakan soal empat pilihan diperoleh 20 siswa mempunyai skor wajar dan 11 siswa mempunyai skor tidak wajar dari 31 siswa yang mempunyai kemampuan rendah. Terdapat 16 siswa mempunyai skor wajar dan 15 siswa mempunyai skor tidak wajar dari 31 siswa yang mempunyai kemampuan tinggi.

Proporsi skor wajar untuk masing-masing kelompok siswa berdasarkan tingkat kemampuan (θ) disajikan pada tabel berikut:

Tabel 4.1 Proporsi skor wajar kelompok *Correct Score*

Kemampuan (θ)	Empat pilihan	
	n	Proporsi
Rendah	20	0,645
Tinggi	16	0,516

Indeks kewajaran skor siswa yang mengerjakan soal empat pilihan dengan model *punishment score*. Sedangkan proporsi skor wajar untuk kedua model penskoran dengan empat pilihan dapat dilihat pada tabel berikut:

Tabel 4.2 Proporsi skor wajar tes empat pilihan

Kemampuan (θ)	<i>Correct score</i>		<i>Punishment score</i>	
	n	Proporsi	N	Proporsi
Rendah	20	0,645	52	0,481
Tinggi	16	0,516	13	0,565

Uji persyaratan analisis unidimensi dimaksudkan untuk mengetahui apakah setiap butir mengukur satu macam ciri dari seluruh peserta tes. Pemeriksaan persyaratan unidimensi dilakukan dengan menggunakan analisis faktor dengan melihat salah satu faktor analisis yaitu nilai eigen. Apabila nilai eigen faktor pertama mempunyai nilai yang dominan dibandingkan dengan nilai eigen faktor kedua dan seterusnya mempunyai nilai eigen hampir sama, maka dapat dikatakan bahwa syarat unidimensi terpenuhi. Perhitungan persyaratan unidimensi untuk empat kelompok dengan menggunakan analisis faktor menggunakan software SPSS versi 18.

Untuk melihat apakah karakteristik butir tetap atau tidak berubah sekalipun kelompok peserta yang menjawab butir yang sama berubah-ubah dan untuk kelompok yang sama, ciri butir-butir tersebut tetap sekalipun butir yang dijawab berubah-ubah, maka dilakukan uji invariansi parameter. Sehingga ada dua pengujian invariansi parameter yaitu invariansi parameter butir dan invariansi parameter kemampuan peserta.

Langkah-langkah yang dilakukan dalam pengujian invariansi parameter butir adalah:

- 1) mengurutkan kemampuan peserta dari yang terendah sampai tertinggi atau sebaliknya, 2)

mengambil 27% jawaban siswa baik kelompok bawah maupun kelompok atas, 3) mengestimasi parameter butir kedua kelompok tersebut, dan 4) mengkorelasikan parameter kelompok atas dengan parameter kelompok bawah. Jika hasil r_{hitung} lebih besar atau sama dengan r_{tabel} , maka terdapat invariansi butir. Berikut merupakan hasil perhitungan korelasi pada masing-masing kelompok:

Tabel 4.3. Nilai Korelasi Parameter Butir

Parameter	CS	PS	r_{tabel}
Taraf Sukar	0,521	0,536	0,361

Dari tabel di atas terlihat bahwa seluruh nilai korelasi atau r_{hitung} lebih besar dari r_{tabel} , sehingga dapat disimpulkan bahwa terdapat invariansi parameter butir untuk semua kelompok.

Langkah-langkah yang dilakukan untuk menguji invariansi parameter kemampuan siswa adalah: 1) membagi dua butir tes menjadi dua bagian berdasarkan butir ganjil dan genap. Sehingga masing-masing terdiri atas 15 butir ganjil dan 15 butir genap, 2) mengestimasi parameter kemampuan pada kedua kelompok tersebut, 3) mengkorelasikan kemampuan hasil estimasi untuk kedua kelompok. Jika hasil r_{hitung} lebih besar atau sama dengan r_{tabel} , maka terdapat invariansi kemampuan siswa. Berikut merupakan hasil perhitungan korelasi pada masing-masing kelompok:

Tabel 4.4 Nilai Korelasi Parameter Kemampuan Siswa

Parameter	CS	PS	r_{tabel}
Taraf Sukar	0,567	0,662	0,123

Dari tabel di atas terlihat bahwa seluruh nilai korelasi atau r_{hitung} lebih besar dari r_{tabel} , sehingga dapat disimpulkan bahwa terdapat invariansi parameter kemampuan siswa untuk semua kelompok.

Hasil pengujian kecocokan model diperoleh harga-harga chi kuadrat hitung beserta chi kuadrat tabel. Untuk kelompok CS terdapat 17 butir yang cocok dan 13 butir yang tidak cocok. Sedangkan pada kelompok PS terdapat 16 butir yang cocok dan 14 butir yang tidak cocok.

Adapun hasil pengujian terhadap hipotesis adalah:

$$H_0: \pi_{PS4} \leq \pi_{CS4}$$

$$H_a: \pi_{PS4} > \pi_{CS4}$$

Berdasarkan hasil perhitungan diperoleh harga-harga sebagai berikut:

$$x_1 = 13 \quad x_2 = 16$$

$$n_1 = 34 \quad n_2 = 31$$

$$p_1 = \frac{x_1}{n_1} = \frac{13}{34} = 0,382 \quad p_2 = \frac{x_2}{n_2} = \frac{16}{31} = 0,516$$

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{13 + 16}{34 + 31} = \frac{29}{65} = 0,443 \quad q = 1 - 0,443 = 0,557$$

$$z = \frac{p_1 - p_2}{\sqrt{pq \left[\left(\frac{1}{n_1} \right) + \left(\frac{1}{n_2} \right) \right]}}$$

$$z = \frac{0,382 - 0,516}{\sqrt{(0,443)(0,557) \left[(0,029) + (0,032) \right]}} = -1,078$$

Ternyata Z_{hitung} sebesar -1,078, sedangkan Z_{tabel} adalah -1,65 sehingga tidak cukup alasan untuk menolak H_0 . Hal ini menunjukkan bahwa pada siswa berkemampuan tinggi untuk model penskoran *punishment score* empat pilihan jawaban menjaring skor tidak lebih banyak dibandingkan dengan model penskoran *correct score* soal empat pilihan jawaban dengan perbedaan proporsi yang signifikan. Dapat dikatakan siswa berkemampuan tinggi yang mengerjakan soal empat pilihan dikoreksi dengan model penskoran *punishment score* mempunyai indeks kewajaran yang tidak lebih baik dibandingkan dengan indeks kewajaran siswa pada model penskoran *correct score*.

Hasil perhitungan menunjukkan bahwa model penskoran yang digunakan dan banyak pilihan jawaban pada soal tes pilihan ganda dapat menghasilkan proporsi skor wajar yang berbeda atau sama sesuai dengan tingkat kemampuan siswa. Hasil pengujian proporsi menunjukkan bahwa pada kelompok siswa yang dikoreksi dengan model *correct score* pada kedua tingkat kemampuan (θ) rendah maupun tinggi, proporsi skor wajar siswa yang mengerjakan soal empat pilihan tidak berbeda secara signifikan dibandingkan dengan siswa

~~Suhibarto: Indeks Ketidakwaiban Skor Berdasarkan Teori Respon Butir ditinjau dari Model Penskoran yang mengunakan tiga pinhan. Proporsi skor waja siswa berkemampuan rendah yang~~

mengerjakan soal empat pilihan sebesar 64,5% lebih besar dibandingkan dengan siswa yang mengerjakan soal tiga pilihan, yakni sebesar 43,8%. Sedangkan proporsi skor wajar siswa berkemampuan tinggi yang mengerjakan soal empat pilihan sebesar 51,6% lebih kecil dibandingkan dengan siswa yang mengerjakan soal tiga pilihan, yakni sebesar 56,2%. Namun kedua perbedaan tersebut tidak signifikan. Bila dilihat dari tingkat kemampuan siswa, pada tingkat kemampuan tinggi untuk siswa yang mengerjakan soal tiga pilihan tidak serta merta menghasilkan skor wajar lebih banyak dari pada siswa yang mengerjakan soal empat pilihan. Demikian halnya pada tingkat kemampuan rendah untuk siswa yang mengerjakan soal empat pilihan tidak dapat dikatakan menghasilkan skor wajar lebih banyak dari siswa yang mengerjakan soal tiga pilihan. Hal tersebut sesuai dengan salah satu hasil penelitian yang dilakukan oleh Rosana dan Sukardiyono bahwa hasil perhitungan dengan teknik *person-fit statistic* dari *output Quest*, terdapat 2,78% responden memiliki skor tidak wajar.

Hasil pengujian proporsi untuk siswa berkemampuan tinggi pada model *correct score* sebesar 56,3% dan pada model *punishment score* sebesar 26,3%. Hal ini menunjukkan bahwa model penskoran memang berpengaruh terhadap banyak pilihan jawaban.

Kesimpulan

Berdasarkan hasil analisis data dapat disimpulkan: siswa berkemampuan tinggi yang dikoreksi dengan model penskoran *punishment score* mempunyai indeks kewajaran yang tidak lebih baik dibandingkan dengan indeks kewajaran siswa pada model penskoran *correct score*.

Daftar Pustaka

- Abdurrahman, M. 2009. *Pendidikan bagi anak berkesulitan belajar*. Jakarta: PT. Rineka Cipta.
- Aiken, L. R. 1997. *Psychological Testing and Measurement*. Ninth Edition. Allyn and Bacon. Needham Height, MA.
- Allen, M. J. and Yen, W. M. 1979. *Introduction to Measurement Theory*. Ci. State College: A Division of Wadsworth Inc.
- Cockroft, W.H. 1983. *Mathematics counts, report of the committee of inquiry into the teaching of mathematics id schools*. London: Her Majesty's Stationery Office.

- Crocker, L. and Algina, J. 1986 *Introduction to Classical and Modern Test Theory*. Florida: Holt, Rincardt and Winston, Inc.
- Cronbach, L. J. 1960. *Essential of Psychological Testing*. New York: Harper & Row Publisher.
- Gronlund, N. E. and Linn, R. L. 1990 *Measurement and Evaluation in Teaching 6th Edition*. New York: Macmillan Publishing Company.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. 1991. *Fundamental of Item Response Theory*. London: Sage Publication, In.
- Hulin, C.L., Drasgow, P., & Parsons, C.K. 1983 *Item response theory application to psychological measurement*. New York: Dow Jones-Irwin.
- Naga, D. S. 2012 *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Nitko, A. J. 1996. *Educational Assessment of Students*. (2nd ed). Columbus Ohio: Prentice Hall.
- Thorndike, R. M. 1982. *Applied Psychometrics*. Boston: Houghton Mifflin Co.
- Walpole, R. E. 1992. *Pengantar Staistik*. Jakarta: PT Gramedia Pustaka Utama.
- Wiersma, W. & Jurs, S.G. 1990. *Educational Measurement and Testing*. (2nd ed.). Needhem Heights, MA: Allyn and Bacon.