

PERBEDAAN KETIDAKWAJARAN SKOR BERDASARKAN TEORI RESPON BUTIR DITINJAU DARI MODEL PENSKORAN JAWABAN SISWA SMP SE KOTA PALANGKA RAYA

(Difference of Unfair Score Based on the Item Response Theory (IRT) Referred to the Model of Students' Answers Scoring of Junior High School Sthroughout the City of Palangka Raya)

SUGIHARTO

Program Studi Pendidikan Matematika Fakultas Keguruan dan dan Ilmu Pendidikan Universitas Palangka Raya
Kampus UPR : Jl. Yos Sudarso Palangka Raya, Kalimantan Tengah 73111

e-mail : giharto64@yahoo.com

ABSTRACT

This study aims to know about the difference of unfair score based on item response theory referred to the model of students' answers scoring of Junior High Schools throughout the city of Palangka Raya. The samples were taken from 17 state and private Junior High Schools throughout the city of Palangka Raya. The instrument was in the form of a multiple choice test with three options that included math materials of Junior High School in the seventh grade students in second semester of 2015/2016. Furthermore, the students' answers were corrected by using scoring models, namely punishment score and correct score. Before doing the test to take the data, the data had to be validated first, both in the content and empirical data. From the 40 test items, they were obtained 30 valid items. To obtain the proportion of fair score, it was used the estimation using BILOG-MG program. Furthermore, the data were analyzed with the different proportions (Z) of two groups.

The results of the data analysis showed that Z_{count} is -2.806, while Z_{table} is -1.65 so Z_{count} rank outside the receipt area of H_0 . It shows that the students with the scoring models of punishment score encompass the score that more than the scoring models of correct score with a significant difference in the proportion. It can be concluded that the students who were corrected by using scoring models of punishment score have a fair index better than the fair index of the students were corrected by using scoring models of correct score.

Keywords: the unfair score, scoring models

ABSTRAK

Penelitian ini bertujuan untuk mengetahui perbedaan ketidakwajaran skor berdasarkan teori respon butir ditinjau dari model penskoran siswa SMP se Kota Palangka Raya. Sampel diambil dari 17 SMP negeri dan swasta yang ada di Kota Palangka Raya. Instrumen berupa tes pilihan ganda dengan tiga pilihan yang mencakup materi matematika SMP Kelas VII semester genap tahun 2015/2016. Selanjutnya dikoreksi dengan model penskoran *punishment score* dan *correct score*. Sebelum tes digunakan untuk mengambil data terlebih dahulu dilakukan validasi, baik isi maupun empiris. Dari 40 butir soal tes diperoleh 30 butir soal yang valid. Untuk memperoleh data proporsi skor wajar digunakan estimasi menggunakan program BILOG-MG. Selanjutnya data dianalisis dengan perbedaan proporsi (Z) dua kelompok.

Hasil analisis data menunjukkan bahwa Z_{hitung} sebesar -2,806, sedangkan Z_{tabel} adalah -1,65 sehingga Z_{hitung} terletak di luar daerah penerimaan H_0 . Hal ini menunjukkan bahwa pada siswa dengan model penskoran *punishment score* menjaring skor lebih banyak dibandingkan dengan model penskoran *correct score* dengan perbedaan proporsi yang signifikan. Dapat disimpulkan bahwa siswa yang dikoreksi dengan model penskoran *punishment score* mempunyai indeks kewajaran yang lebih baik dibandingkan dengan indeks kewajaran siswa yang dikoreksi dengan model penskoran *correct score*.

Kata kunci : ketidakwajaran skor, model penskoran

PENDAHULUAN

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 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 ketidakwaajaran skor. Menurut Naga (2001: 43) 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. Hal senada juga diungkap oleh Hulin, Drasgow, and Parsons (1983: 110).

Menurut Nitko (1996: 91 – 94) ketidakwaajaran 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. Sebagai akibatnya menurut Wiersma & Jurs (1990: 340) siswa semacam ini akan mendapatkan skor yang tidak tepat, yakni tidak sesuai dengan kemampuan mereka semestinya.

Apabila terjadi ketidakwaajaran skor dan hal tersebut sulit dihindari, maka perlu dideteksi ketidakwaajaran 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.

Menurut Crocker & Algina (1986: 399 – 401) 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*. Dalam penelitian ini yang digunakan hanya *correct score (CS)* dan *punishment score (PS)*. 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 ketidakwaajaran skor.

Untuk mengetahui atau pendekteksian terhadap ketidakwaajaran skor dapat digunakan teori klasik atau dengan teori respon butir. Menurut Naga (2013: 391 – 397) bahwa indeks ketidakwaajaran dapat dideteksi melalui pendekatan teori tes, yaitu teori tes klasik (*classsical test theory*) atau dapat juga dideteksi menggunakan teori test modern (*Item Response Theory*). Teori tes klasik meliputi indeks Ghiselli,

Jacob, Donlon dan Fisher, dan metode SHL. Sedangkan pendeteksian dengan teori tes modern meliputi indeks Gauss, residu bakuan kuadrat dan model logaritmik. Dalam penelitian ini yang digunakan adalah Teori tes modern.

Masalah dalam penelitian ini adalah untuk siswa berkemampuan tinggi, apakah yang dikoreksi dengan model penskoran PS memiliki skor wajar lebih tinggi dibandingkan dengan yang dikoreksi dengan model CS?

Menurut Hulin (1983: 112) ketidakwajaran skor secara individu disebut dengan *inappropriatenes*. Pendeteksian ketidakwajaran skor dapat dicari melalui indeks yang dikenal sebagai indeks kewajaran. Indeks kewajaran diestimasi dengan fungsi kebolehdjian maksimum. Ditambahkan oleh Hulin (1983: 112) bahwa Fungsi kebolehdjian 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 kebolehdjian akan mencapai nilai maksimum juga terletak diantara 0 dan 1. Namun jika diberikan logaritma kepada fungsi kebolehdjian yang terletak diantara 0 dan 1 tersebut maka nilai logaritma ini akan menjadi negatif. Kebolehdjian maksimum akan menghasilkan nilai logaritma yang paling tinggi atau mendekati 1, demikian sebaliknya. Dengan demikian, semakin negatif nilai logaritma dari fungsi kebolehdjian, maka makin tidak wajar skor peserta yang bersangkutan.

Indeks kesukaran butir dan daya beda butir dalam teori respon butir dikenal sebagai parameter butir. Indeks kesukaran butir

dinotasikan sebagai “b” dan indeks daya beda butir dinotasikan sebagai “a”. Selain kedua parameter terdapat parameter “c” yang disebut sebagai parameter probabilitas menjawab betul dengan cara menebak jawaban suatu butir. Ketiga parameter tersebut berpengaruh terhadap probabilitas menjawab betul seorang peserta tes pada butir tertentu.

Pada model logistik, ada tiga model yang dikenal yaitu model logistik satu parameter, model logistik dua parameter, dan model logistik tiga parameter. Pada penelitian ini pada awalnya direncanakan menggunakan model logistik tiga parameter, karena jika hanya dua parameter maka hanya melibatkan parameter indeks kesukaran butir dan daya beda butir. Sedangkan kalau tiga parameter di samping indeks kesukaran dan daya beda butir, maka juga melibatkan parameter tebakan. Namun karena keterbatasan peneliti dalam hal pengumpulan data, maka digunakan model logistik satu parameter, yang hanya melibatkan satu parameter yaitu taraf sukar butir.

Model logistik satu parameter (L1P) melibatkan satu parameter, bentuk karakteristik butir L1P (Hambleton, Swaminathan, dan Rogers, 1991: 12) adalah:

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

Dalam mana:

- $P_i(\theta)$ = Probabilitas peserta tes dengan kemampuan θ menjawab butir ke- i dengan benar.
- e = Bilangan eksponen yang nilainya mendekati 2,718
- θ = Tingkat kemampuan subjek
- b_i = Parameter indeks kesukaran butir ke- i
- n = Banyaknya butir dalam tes

Bagi siswa berkemampuan tinggi tidak ada masalah dengan cara penskoran. Dikoreksi dengan model penskoran apapun, mereka tetap akan memilih jawaban yang dianggap betul dulu. Jika ada soal yang dianggap sulit, siswa tersebut tidak akan menjawabnya sampai benar-benar menemukan jawaban yang dianggap paling betul. Berkaitan dengan banyaknya pilihan jawaban, semakin sedikit pilihan jawaban, maka siswa berkemampuan tinggi akan semakin besar peluang menjawab soal dengan betul. Tidak berbeda dengan kelompok siswa berkemampuan tinggi yang dihadapkan pada soal dengan empat pilihan. Siswa dengan kemampuan tinggi akan berpikir beberapa kali dan berhati-hati untuk menebak jawaban jika ragu-ragu terhadap jawaban suatu butir. Dengan demikian dapat diduga bahwa untuk siswa berkemampuan tinggi yang dikoreksi dengan model penskoran PS menghasilkan skor yang lebih wajar dibandingkan penerapan model penskoran CS atau proporsi skor yang dikoreksi dengan model penskoran PS lebih tinggi jika dibandingkan dengan model penskoran CS.

METODOLOGI

Penelitian ini dilaksanakan pada 4 SMP negeri maupun swasta di kota Palangka Raya. sampel diambil secara acak diperoleh 250 siswa. Pengumpulan data dilaksanakan pada akhir semester genap tahun pelajaran 2015/2016. Adapun tahapannya sebagai berikut: 1) kajian teori, 2) pembuatan instrumen, 3) panel instrumen, 4) ujicoba instrumen, 5) analisis hasil ujicoba, 6) pengambilan data, 7) analisis data, dan 8) penulisan laporan penelitian hasil penelitian, dan 9) penyusunan disertasi.

Penelitian ini merupakan *quasi eksperimen* dengan menggunakan model perbandingan proporsi. Variabel dalam penelitian ini ada dua yaitu variabel bebas dan variabel terikat. Variabel bebas dalam penelitian ini adalah model penskoran. Model penskoran yang digunakan terdiri atas dua model yakni model penskoran *Correct Score (CS)* dan *Punishment Score (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*.

Analisis yang dilakukan adalah dengan melihat perbandingan proporsi skor wajar yang diperoleh berdasarkan indeks kewajaran I_{gz} yang diperoleh antara siswa yang dikoreksi dengan model penskoran *correct score* dan *punishment score*. Analisis dilakukan terhadap kelompok siswa berkemampuan tinggi setelah siswa diurutkan berdasarkan estimasi kemampuan siswa yang diperoleh. Kemudian dilihat perbandingan proporsinya dengan batas indeks kewajaran $I_{gz} = 0$.

Model penskoran *Correct Score (CS)*, hasil jawaban peserta dikoreksi jika benar diberi skor satu dan jika salah atau tidak dijawab diberi skor nol, selanjutnya menjumlahkan jawaban benar pada setiap butir. Model penskoran dengan hukuman atau *Punishment Score (PS)* yaitu menjumlahkan jawaban benar dan

mengurangnya dengan hasil bagi banyak butir yang dijawab salah dengan banyak pilihan di kurangi satu.

Data penelitian mengenai nilai indeks kewajaran (I_{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. Caranya dengan melihat pemenuhan syarat model. Persyaratan model dilihat dari hasil program Bilog pada saat melakukan estimasi parameter butir dan kemampuan. 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 (Excel). Setelah persyaratan model dan persyaratan analisis dipenuhi, maka dilakukan pengujian hipotesis penelitian melalui pengujian hipotesis statistik yang dirumuskan. Analisis data dilakukan dengan analisis perbedaan proporsi dengan statistik Z berikut: (Walpole, 1992: 321)

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

Hipotesis yang diuji sebagai berikut:

$$H_0: \pi_{PS} \leq \pi_{CS}$$

$$H_a: \pi_{PS} > \pi_{CS}$$

π_{PS} = indeks skor wajar siswa yang dikoreksi dengan model penskoran *Punishment Score*.

π_{CS} = indeks skor wajar siswa yang dikoreksi dengan model penskoran *Correct Score*

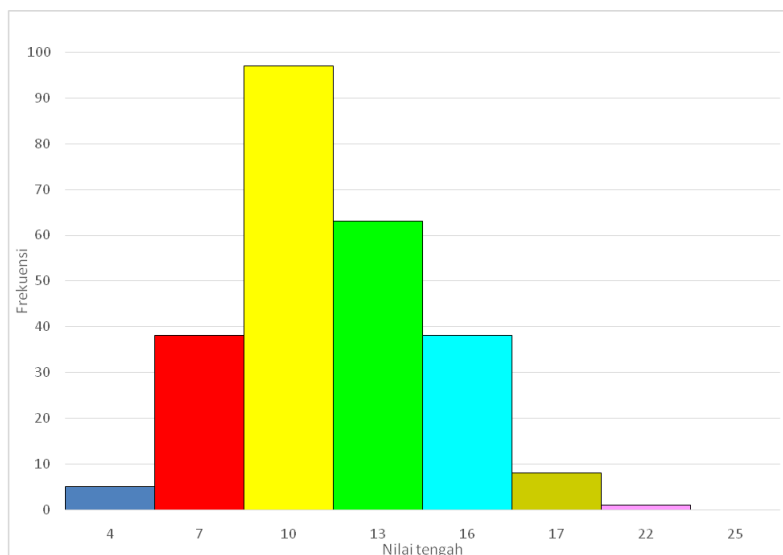
HASIL DAN PEMBAHASAN

Kelompok skor yang diperoleh dari model penskoran *Correct Score* (CS), yaitu cara penskoran dengan menjumlahkan seluruh jawaban siswa yang betul. Terdapat 250 lembar jawaban yang masing-masing terdiri dari 30 butir soal. Butir soal yang dijawab betul diberi skor 1 sedangkan yang salah 0, dan yang tidak dijawab juga diberi skor 0. Jadi dengan cara penskoran seperti ini skor maksimal yang diperoleh siswa yaitu 30 dan yang terendah adalah 0. Skor tertinggi adalah 21 dan terendah adalah 3 dengan rentang sebesar 18. Diperoleh juga rata-rata sebesar 11,448 simpangan baku 3,318. Distribusi frekuensi dan histogram skor model penskoran *correct score* disajikan dalam tabel 1 dan gambar 1.

Penskoran dengan model *Punishment Score* (PS) atau penskoran dengan hukuman adalah penskoran dengan cara menjumlahkan skor jawaban betul dan mengurangnya dengan skor jawaban yang salah dibagi banyaknya pilihan dikurangi satu dari setiap lembar jawaban siswa. Diperoleh data skor tertinggi adalah 27 dan terendah adalah -9 dengan rentang sebesar 36. Diperoleh juga rata-rata sebesar 4,62 simpangan baku 6,53. Distribusi frekuensi dan histogram skor model penskoran *punishment score* untuk soal dengan tiga pilihan disajikan dalam tabel 2 dan gambar 2.

Tabel 1. Distribusi frekuensi skor dengan model penskoran *correct score*

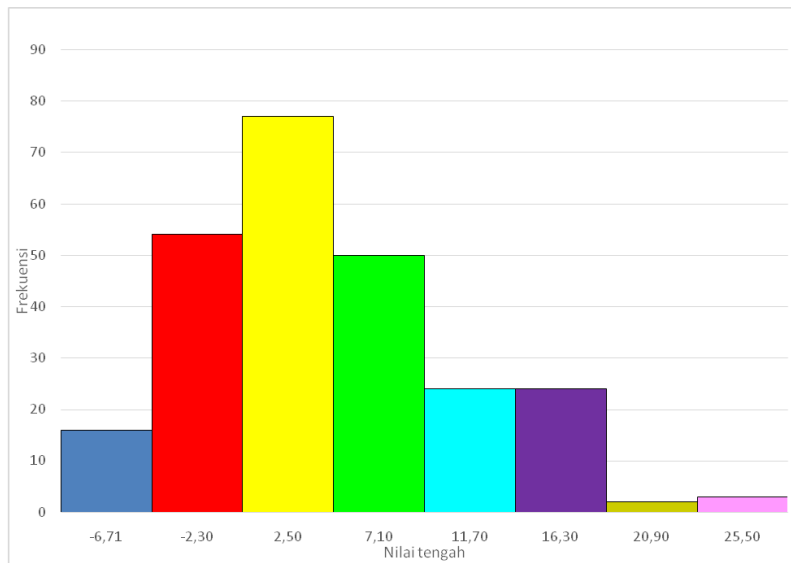
No	Interval	Nilai tengah	Frekuensi	Frek. relatif	Frek. rel kum
1.	3 - 5	4	5	2	2
2.	6 - 8	7	38	15,2	17,2
3.	9 - 11	10	97	38,8	56
4.	12 - 14	13	63	25,2	81,2
5.	15 - 17	16	38	15,2	96,4
6.	18 - 20	17	8	3,2	99,6
7.	21 - 23	22	1	0,4	100
8.	24 - 26	25	0	0	
			250	100	



Gambar 1. Histogram skor dengan model penskoran *correct score*

Tabel 2. Distribusi frekuensi skor dengan model penskoran *punishment score*

No.	Interval	Nilai tengah	Frekuensi	Frek. relatif	Frek. rel kum
1.	-9,00 - (-4,41)	-6,705	16	6,4	6,4
2.	-4,40 - 0,19)	-2,295	54	21,6	28
3.	0,20 - 4,79	2,495	77	30,8	58,8
4.	4,80 - 9,39	7,095	50	20	78,8
5.	9,40 - 13,99	11,695	24	9,6	88,4
6.	14,00 - 18,59	16,295	24	9,6	98
7.	18,60 - 23,19	20,895	2	0,8	98,8
8.	23,20 - 27,79	25,495	3	1,2	100
			250	100	



Gambar 2. Histogram skor dengan model penskoran *punishment score*

Tabel 3. Hasil estimasi parameter butir

No. Butir	CS3	No. Butir	CS3	No. Butir	PS3	No. Butir	PS3
1	0.532	16	0.406	1	0.376	16	0.278
2	-0.057	17	0.232	2	-0.121	17	0.329
3	0.244	18	0.066	3	0.087	18	-0.054
4	-0.136	19	-0.006	4	-0.077	19	0.118
5	-0.077	20	0.077	5	-0.183	20	0.385
6	0.153	21	0.220	6	0.542	21	0.183
7	0.131	22	-0.292	7	-0.032	22	-0.273
8	0.077	23	0.131	8	-0.099	23	0.260
9	-0.486	24	-0.546	9	-0.543	24	-0.397
10	-0.379	25	-0.516	10	-0.318	25	-0.526
11	0.232	26	0.175	11	0.172	26	0.066
12	0.291	27	0.120	12	0.260	27	-0.073
13	0.088	28	-0.097	13	0.558	28	-0.207
14	-0.087	29	-0.117	14	-0.328	29	0.126
15	0.066	30	-0.447	15	0.025	30	-0.532
		Min	-0,546			Min	-0.543
		Maks	0,532			Maks	0.558

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). Hasil estimasi parameter butir (Indeks Kesukaran) disajikan dalam tabel 3.

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. Rangkuman indeks taraf sukar berdasarkan Tabel 4 dikelompokkan ke dalam lima kelompok dapat dilihat pada tabel 4.

Tabel 4. Rangkuman indeks taraf sukar

Taraf Sukar	CS3	PS3	Kategori
(-2,00) - (-1,30)	0	0	Sangat Mudah
(-1,20) - (-0,41)	4	3	Mudah
(-0,40) - (0,39)	24	25	Sedang
0,40 - 1,20	2	2	Sukar
1,19 - 2,00	0	0	Sangat Sukar
Jumlah	30	30	

Untuk butir kategori sedang (-0,40 - (0,39)) cukup banyak pada kelompok PS3 sebanyak 25 butir atau sekitar 83,33% dan pada kelompok CS3 sebanyak 24 butir atau sekitar 80%. Untuk kategori butir sukar sebanyak dua butir atau sekitar 6,67% terdapat pada kelompok CS3 dan PS3. Tabel 5 disajikan sebaran butir-butir dilihat dari taraf sukar. Sedangkan hasil estimasi parameter kemampuan (θ) menggunakan program Bilog disajikan pada tabel 6.

Tabel 5. Sebaran butir berdasarkan taraf sukar

Kategori Soal	CS3	PS3
sangat mudah	-	-
mudah	9,24,30,25	9,25,30
sedang	2,3,4,5,6,7,8, 10,11,12,13, 14,15,17,18, 19,20,21,22, 23,26,27,28, 29	1,2,3,4,5,7,8, 10,11,12,14, 15,16,17,18,1 920,21,22,23, 24, 26,27, 28,29
sukar	1,16	6,3
sangat sukar	-	-
jumlah	30	30

Hasil estimasi parameter kemampuan (θ) menggunakan program Bilog disajikan pada tabel 6.

Tabel 6. Sebaran estimasi kemampuan siswa

Kemampuan	CS3	PS3
(-2,4) - (-1,6)		
(-1,5) - (-0,0)	16	8
(-0,6) - (-0,2)	202	185
0,3 - 1,1	32	54
1,2 - 2,0		3
jumlah	250	250
minimum	-1,335	-1,255
maksimum	0,448	1,430
rerata	-0,324	-0,195

Berdasarkan tabel di atas bahwa kemampuan siswa dimulai dari kemampuan paling rendah -2,312 sampai kemampuan yang paling tinggi 2,062.

Tabel 7. Sebaran tingkat kemampuan berdasarkan kelompok kemampuan

Kemampuan	CS3	PS3
Rendah	16 (6,4%)	8 (3,2%)
Sedang	202 (80,8%)	185 (74%)
Tinggi	32 (12,8%)	57 (22,8%)

Tabel 8. Sebaran indeks kewajaran skor kelompok CS

Skor	Indeks Kewajaran	Frekuensi	Jumlah/Proporsi
		θ tinggi	θ tinggi
Tak wajar	< (-2,000)	1	0,438
	(-2,000) – (-1,001)	2	
	(-1,000) – 0,000	11	
	Jumlah	14	
Wajar	0,001 – 1,000	14	0,562
	1,001 – 2,000	3	
	> 2,000	1	
	Jumlah	18	
Total		32	1,000
Minimum		-2,775	
Maksimum		2,493	
Rerata		$-6,26 \times 10^{-15}$	

Tabel 9. Sebaran indeks kewajaran skor kelompok PS

Skor	Indeks Kewajaran	Frekuensi	Jumlah/Proporsi
		θ tinggi	θ tinggi
Tak wajar	< (-2,000)	1	0,737
	(-2,000) – (-1,001)	-	
	(-1,000) – 0,000	42	
	Jumlah	42	
Wajar	0,001 – 1,000	12	0,263
	1,001 – 2,000	-	
	> 2,000	3	
	Jumlah	15	
Total		57	1,000
Minimum		-0,996	
Maksimum		4,393	
Rerata		$-1,714 \times 10^{-15}$	

Perhitungan indeks kewajaran l_{gz} diperoleh 18 siswa mempunyai skor wajar dan 14 siswa mempunyai skor tidak wajar dari 32 siswa dengan kemampuan tinggi. Sebaran indeks kewajaran l_{gz} untuk kelompok CS dapat diikuti pada tabel 8. Sedangkan indeks kewajaran skor

siswa yang mengerjakan soal empat pilihan dengan model *punishment score* terlihat pada tabel 9.

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.

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.

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.

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. Nilai korelasi parameter butir disajikan pada tabel 10.

Tabel 10. Nilai korelasi parameter butir

Parameter	CS3	PS3	r_{tabel}
Taraf Sukar	0,416	0,487	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.

Pengujian 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. Nilai korelasi parameter kemampuan siswa disajikan pada tabel 11.

Tabel 11. Nilai korelasi parameter kemampuan siswa

Parameter	CS3	PS3	r_{tabel}
Taraf Sukar	0,297	0,522	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.

Uji Hipotesis untuk kelompok siswa yang dikoreksi dengan model PS dan CS, pada siswa dengan θ tinggi.

$$H_0: \pi_{PS} \leq \pi_{CS}$$

$$H_a: \pi_{PS} > \pi_{CS}$$

Berdasarkan hasil perhitungan diperoleh harga-harga sebagai berikut:

$$x_1 = 15 \quad x_2 = 18$$

$$n_1 = 57 \quad n_2 = 32$$

$$p_1 = \frac{x_1}{n_1} = \frac{15}{57} = 0,263 \quad p_2 = \frac{x_2}{n_2} = \frac{18}{32} =$$

$$0,563$$

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{15 + 18}{57 + 32} = \frac{33}{89} = 0,371$$

$$q = 1 - 0,371 = 0,629$$

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

$$z = \frac{0,263 - 0,563}{\sqrt{0,371 \cdot 0,629 \left[0,018 + 0,031 \right]}} = -2,806$$

Ternyata Z_{hitung} sebesar -2,806, sedangkan Z_{tabel} adalah -1,65 sehingga Z_{hitung} berada dalam daerah penolakan H_0 . Hal ini menunjukkan bahwa pada siswa berkemampuan tinggi untuk model penskoran *punishment score* menjangkir skor lebih banyak dibandingkan dengan model penskoran *correct score* dengan perbedaan proporsi yang signifikan. Dapat dikatakan bahwa siswa yang dikoreksi dengan model penskoran *punishment score* mempunyai indeks kewajaran yang lebih baik dibandingkan dengan indeks kewajaran siswa yang dikoreksi dengan model penskoran *correct score*.

Hasil pengujian siswa yang berkemampuan tinggi menunjukkan bahwa model penskoran menghasilkan proporsi skor wajar berbeda. Dilihat dari proporsi untuk siswa berkemampuan rendah

pada model *correct score* sebesar 43,8% dan pada model *punishment score* sebesar 62,5%. Sedangkan 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

Siswa berkemampuan tinggi untuk model penskoran *punishment score* menjangkir skor lebih banyak dibandingkan dengan model penskoran *correct score* dengan perbedaan proporsi yang signifikan. Dapat dikatakan bahwa siswa yang dikoreksi dengan model penskoran PS memiliki skor wajar lebih tinggi dibandingkan dengan yang dikoreksi dengan model CS.

DAFTAR PUSTAKA

- Hambleton, R.K., Swaminathan, H. and Rogers, H.J. *Fundamental of Item Response Theory*. (London: Sage Publication, In. 1991). h. 12.
- Hulin, C. I., Drasgow, P. and Parsons, C.K. *Item Response Theory Application to Psychological Measurement*. (USA: Dow Jones-Irwin. 1983). h. 110.
- Nitko, A. J. *Educational Assessment of Students*. (2nd ed). (Columbus Ohio: Prentice Hall. 1996). hh. 91-94.
- Naga, D. S. Indeks Kehati-hatian Sekor Responden pada Model SHL: Suatu Bentuk Indeks Ketidakwajaran pada Sekor Ujian. (*Jurnal Ilmiah Psikologi "Arkhe"* . Tahun 6, Nomor 1. 2001). h. 43.
- _____. Teori Sekor pada Pengukuran Mental. (Jakarta: PT. Nagrani Citrayasa. 2013). hh. 391 - 397

Walpole, R. E. Pengantar Statistik. (Jakarta: PT Gramedia Pustaka Utama. 1992). h. 321

Wiersma, W. & Jurs, S.G. *Educational Measurement and Testing*. (2nd ed.). (Boston: Allyn and Bacon. 1990). h. 340