



Karakteristik Butir Tes Pengantar Statistika Sosial Berdasarkan Teori Respon Butir

Agus Santoso

¹ Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Terbuka, Jl. Cabe Raya, Pondok Cabe, Pamulang, Tangerang Selatan, 15418, Indonesia.

* Korespondensi Penulis. E-mail: aguss@ecampus.ut.ac.id

Abstrak

Analisis karakteristik butir tes dilakukan untuk menghasilkan butir-butir tes yang berkualitas (terstandar). Analisis karakteristik butir tes dapat dilakukan menggunakan pendekatan klasik maupun pendekatan modern (Teori Respon Butir). Penelitian ini bertujuan untuk mendeskripsikan karakteristik butir tes pengantar statistika sosial yang dikembangkan oleh Universitas Terbuka menggunakan pendekatan teori respon butir. Penelitian ini merupakan penelitian deskriptif eksploratif menggunakan pendekatan kuantitatif. Pengumpulan data dilakukan melalui dokumentasi 269 butir tes Pengantar Statistik Sosial dan respon 82966 mahasiswa UT yang mengerjakan tes tersebut. Analisis karakteristik butir dilakukan melalui tahapan organisasi data, uji asumsi, uji kecocokan model, estimasi parameter butir. Estimasi parameter dilakukan menggunakan bantuan software Bilog MG. Hasil analisis kecocokan model menunjukkan bahwa tes cocok diestimasi menggunakan model 3 Parameter Logistik (3 PL). Hasil estimasi parameter butir menunjukkan bahwa dari 269 butir hanya terdapat 81 butir yang memiliki kualitas baik dengan rerata indeks kesukaran (b) yaitu 2,149 dengan standar deviasi 2,451; rerata daya pembeda (a) yaitu 0,852 dengan standar deviasi 0,491; dan rerata tebakan semu (c) yaitu 0,288 dengan standar deviasi 0,093.

Kata Kunci: Analisis karakteristik butir, teori respon butir, model 3 parameter logistik.

The Characteristics of Basic Social Statistics Test Items Based on Item Response Theory

Abstract

Analysis of the characteristics of test items was carried out to produce test items that were of high quality (standardized). Analysis of the characteristics of test items could be done using the classical approach as well as the modern approach (Item Response Theory). This study aimed to describe the characteristics of basic social statistics test items developed by the Universitas Terbuka using the item response theory approach. This study was an explorative descriptive by using a quantitative approach. Data collection was carried out through documentation of 269 test items of Basic Social Statistics and responses of 82966 students of Universitas Terbuka who took the test. Analysis of the characteristics of test items was carried out through steps: organizational data, assumption test, model fit test, items parameter estimate. Parameter estimation was done using Bilog MG software. The results of model fit test show that test was suitable estimated using 3 Parameters Logistics model (3 PL). The items parameter estimation results show that from 269 items only 81 items which has good quality and difficulty index (b) average was 2.149 with standard deviation was 2.451; discriminant index (a) average was 0.852 with standard deviation was 0.491; and pseudo-guessing (c) average was 0.288 with standard deviation was 0.093.

Keywords: *Analysis of item characteristics, item response theory, 3-parameter logistic model*

How to Cite: Susanto, A. (2018). Karakteristik butir tes pengantar statistik sosial berdasarkan teori respon butir. *Jurnal Pendidikan Matematika dan Sains, VI(2)*, 1-11. doi:<http://dx.doi.org/10.21831/jpms.v4i1.10111>

Permalink/DOI: DOI: <http://dx.doi.org/10.21831/jpms.v4i1.10111>

PENDAHULUAN

Dalam bidang pendidikan, peran ilmu pengukuran (*measurement*) tidak dapat dipisahkan. Melalui pengukuran dapat diketahui berbagai pencapaian dibidang pendidikan, khususnya yang berkaitan prestasi peserta didik. Pengukuran umumnya diimplementasikan melalui suatu tes. Melalui tes tersebut dapat diketahui bagaimana capaian atau kemampuan subjek yang hendak diukur (Phopam, 2009). Agar hasil pengukuran tersebut benar-benar menggambarkan capaian atau kemampuan subjek yang sesungguhnya, maka dibutuhkan alat ukur berupa tes yang berkualitas (Kartowagiran, Munadi, Retnawati, & Apino, 2018). Untuk menghasilkan tes yang berkualitas tentunya bukanlah pekerjaan yang mudah. Selain diperlukan keahlian khusus dari pembuat tes, dibutuhkan juga teknik-teknik pengukuran lainnya yang dapat membantu menghasilkan tes yang berkualitas.

Salah satu cara yang dapat dilakukan untuk menghasilkan tes yang berkualitas yaitu melalui analisis karakteristik butir (Gronlund, 1998; Siri & Freddano, 2011; Retnawati, 2016; Quaigrain & Arhin, 2017). Secara umum analisis karakteristik butir soal tersebut adalah mengetahui bagaimana kualitas tes secara empiris (Retnawati, 2016; Sutrisno, 2016). Dalam hal ini setelah tes selesai dirakit, tes tersebut akan diujicobakan kepada sejumlah sampel dan respon dari sejumlah sampel ujicoba tersebut digunakan untuk melakukan analisis karakteristik butir (Adedoyin & Mokobi, 2013). Selanjutnya, analisis karakteristik butir dapat dilakukan baik menggunakan pendekatan klasik maupun menggunakan pendekatan modern (teori respon butir) (Sudaryono, 2011; Ratnaningsih & Isfarudi, 2013; Retnawati, 2016). Namun seiring dengan perkembangan ilmu pengetahuan dan teknologi, penggunaan pendekatan modern saat ini lebih populer.

Ketika melakukan analisis karakteristik butir dengan pendekatan klasik, kebanyakan statistik yang digunakan untuk mengestimasi parameter butir sangat bergantung pada rerata tingkat kemampuan, rentang dan sebaran kemampuan sampel yang digunakan (Retnawati, 2016). Sebagai contoh jika sampel yang digunakan memiliki rerata kemampuan tinggi, maka indeks kesukaran butir juga akan tinggi yang menandakan bahwa butir tersebut mudah. Namun sebaliknya, jika sampel yang digunakan memiliki rerata kemampuan yang rendah, maka indeks kesukaran butir tersebut juga akan

rendah, sehingga butir yang sama jika diberikan kepada dua kelompok sampel dengan kemampuan yang berbeda, akan menghasilkan nilai parameter yang berbeda. Selain kelemahan tersebut, tentunya masih terdapat beberapa kelemahan lain dari pendekatan klasik, dan kelemahan-kelemahan tersebut dapat diatasi menggunakan pendekatan modern atau teori respon butir (Sudaryono, 2011; Ratnaningsih & Isfarudi, 2013; Retnawati, 2016)

Pada teori respon butir, estimasi parameter dilakukan dengan cara yang berbeda dengan teori tes klasik. Dalam teori respon butir, hubungan antara probabilitas menjawab benar pada suatu skala kemampuan dinyatakan dengan suatu hubungan dengan parameter butir yang digunakan. Banyaknya parameter butir yang digunakan tersebut menentukan model persamaannya. Jika model persamaan hanya memuat parameter indeks kesukaran (b) dikenal dengan model 1 Parameter Logistik (PL) atau model Rasch. Jika model memuat parameter b dan indeks daya pembeda (a), dikenal dengan model 2 PL. Adapun jika model persamaan memuat parameter b , a , dan tebakan semu (c), maka dikenal dengan model 3 PL. Model-model tersebut digunakan untuk penskoran dikotomi dimana skor 1 diberikan untuk jawaban benar, dan skor 0 untuk jawaban salah.

Menurut Hambleton, Swaminathan, dan Rogers (1991), model persamaan untuk 1 PL atau model Rasch dinyatakan sebagai

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}}$$

dengan $i = 1, 2, 3, \dots, n$; $P_i(\theta)$ merupakan probabilitas peserta tes yang memiliki kemampuan θ yang dipilih secara acak dapat menjawab butir i dengan benar; θ merupakan tingkat kemampuan subjek; b_i merupakan indeks kesukaran butir ke- i ; e merupakan bilangan natural yang nilainya mendekati 2,718; dan n merupakan banyaknya butir tes. Selanjutnya, jika menggunakan 2 PL, maka persamaannya dinyatakan sebagai

$$P_i(\theta) = \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}}$$

dengan a_i merupakan indeks daya pembeda butir ke- i . Pada model persamaan 2 PL, jika $a_i = 1$, maka secara otomatis persamaan berubah menjadi model 1 PL. Adapun jika model persamaan memuat 3 PL, maka persamaan matematisnya dapat dinyatakan sebagai

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}}$$

dengan c_i merupakan parameter tebakan semu (*pseudo guessing*) butir ke- i .

Untuk menentukan apakah data yang akan diestimasi cocok dengan model 1 PL, 2 PL, atau 3 PL, maka perlu dilakukan uji kecocokan model. Uji kecocokan model dapat dilakukan dengan 2 cara, yaitu menggunakan pendekatan statistik (chi-square (χ^2) atau nilai probabilitas) dan pendekatan kurva karakteristik butir (Retnawati, 2014). Jika menggunakan pendekatan statistik, suatu butir dikatakan cocok dengan model jika nilai χ^2 hitung $< \chi^2$ tabel, atau nilai probabilitas (p-value/signifikansi [sig.]) $< \alpha$. Jika menggunakan pendekatan kurva karakteristik butir, jika plot distribusi data berada disekitar kurva modelnya, maka butir tersebut dikatakan cocok dengan model persamaan logistiknya. Distribusi data pada kurva karakteristik butir dapat dibuat menggunakan bantuan software. Jika salah satu model persamaan logistik (1 PL, 2 PL, atau 3 PL) menghasilkan paling banyak butir yang cocok dengan salah satu model, maka model persamaan logistik tersebutlah yang digunakan untuk mengestimasi parameter butir.

Sebelum mengestimasi parameter butir, ada beberapa asumsi yang harus dipenuhi dalam teori respon butir. Terdapat 3 asumsi yang harus dipenuhi dalam teori respon butir yaitu, unidimensi, independensi lokal, dan invariansi parameter (Hambleton, Swaintham, & Rogers, 1991). Asumsi unidimensi dimaksudkan untuk membuktikan apakah setiap butir tes hanya mengukur satu kemampuan. Asumsi unidimensi dapat dibuktikan melalui analisis faktor. Asumsi independensi lokal dimaksudkan untuk menyelidiki apakah jawaban peserta tes terhadap suatu butir tidak mempengaruhi jawaban peserta tes terhadap butir lain. Asumsi independensi lokal dapat pula dibuktikan melalui asumsi unidimensi (De Mars, 2010), sehingga ketika asumsi unidimensi terpenuhi, maka asumsi independensi lokal juga terpenuhi. Asumsi invariansi parameter dimaksudkan untuk menyelidiki apakah karakteristik butir tidak tergantung pada distribusi parameter kemampuan peserta dan sebaliknya. Invariansi parameter terdiri dari invariansi parameter butir dan invariansi parameter kemampuan. Invariansi parameter butir dibuktikan dengan mengestimasi parameter butir (indeks kesukaran, daya beda, dan tebakan semu) pada kelompok tes yang berbeda, misalnya kelompok berdasarkan jenis kelamin. Dari hasil estimasi parameter butir pada kedua kelompok kemudian disajikan diagram pencar.

Jika korelasinya tinggi atau titik-titik pada diagram pencar menyebar mendekati garis lurus dengan gradien 1, maka invariansi parameter butir terpenuhi. Sedangkan untuk invariansi parameter kemampuan, prosedurnya kurang lebih sama, hanya saja hasil estimasi yang digunakan adalah estimasi kemampuan peserta (θ) dalam mengerjakan butir ganjil dan butir genap (bisa juga butir diklasifikasikan berdasarkan separuh butir bagian atas dan separuh butir bagian bawah).

Terkait dengan penerapan teori respon butir dalam pengembangan instrumen, beberapa peneliti telah mempublikasikan hasil penelitiannya. Adegoke (2013) juga melaporkan bahwa statistik butir yang dihasilkan melalui analisis teori respon butir 2 PL cenderung lebih stabil dibandingkan dengan analisis menggunakan teori tes klasik, dan butir yang dieleminasi juga lebih sedikit ketika menggunakan teori respon butir. Al-Khader dan Albursan (2017) melaporkan bahwa penggunaan teori respon butir dengan 2 PL lebih akurat dari pada teori tes klasik dalam mengases kemampuan partisipan. Eleje, Onah, dan Abanobi (2018) melaporkan bahwa terdapat perbedaan statistik parameter butir dan parameter tes antara pendekatan teori tes klasik dan teori respon butir 3 PL, dimana hasil analisis data *Diagnostik Quantitative Economics Skill Test* (DQUEST) menggunakan teori respon butir lebih valid dan reliabel dibandingkan pendekatan teori tes klasik. Hasil penelitian Ayanwale, Adeleke, dan Mamadelo (2018) juga melaporkan bahwa penerapan teori respon butir lebih superior dibandingkan penerapan teori tes klasik dalam menganalisis butir *Basic Education Certificate Examination*. Dalam hasil penelitian tersebut dilaporkan bahwa reliabilitas instrumen lebih tinggi ketika menggunakan teori respon butir dan jumlah butir yang dieliminasi (*poor criteria*) juga lebih sedikit dibandingkan dengan analisis menggunakan teori tes klasik.

Berdasarkan kajian teori dan penelitian relevan yang telah dikemukakan sebelumnya, maka penerapan teori respon butir dalam menganalisis karakteristik butir terlihat lebih menguntungkan dibandingkan dengan teori tes klasik, baik dalam melakukan estimasi parameter butir maupun estimasi parameter tes. Dengan demikian tujuan dari penelitian ini adalah untuk mendeskripsikan karakteristik butir tes Pengantar Statistika Sosial yang dikembangkan oleh Universitas Terbuka menggunakan pendekatan teori respon butir. Dengan adanya penelitian ini

diharapkan dapat diperoleh instrumen tes yang berkualitas untuk keperluan pengembangan perangkat tes selanjutnya.

METODE

Penelitian ini merupakan penelitian deskriptif eksploratif menggunakan pendekatan kuantitatif yang bertujuan untuk mendeskripsikan karakteristik butir tes pengantar statistika sosial yang dikembangkan oleh Universitas Terbuka. Pengumpulan data dilakukan melalui dokumentasi terhadap 269 butir tes Ujian Akhir Semester pada mata kuliah Pengantar Statistik Sosial yang disebar dalam 13 paket soal. Selanjutnya pengumpulan data juga dilakukan melalui dokumentasi terhadap respon atau jawaban dari 82966 mahasiswa UT yang telah mengerjakan perangkat tes tersebut. Butir soal dalam tes tersebut berbentuk tes objektif dengan 4 pilihan jawaban dimana salah satu pilihan jawaban merupakan kunci jawaban dan pilihan lainnya sebagai pengecoh (*distraktor*).

Analisis karakteristik butir dilakukan menggunakan pendekatan teori respon butir melalui tahapan organisasi data, uji asumsi, uji kecocokan model, estimasi parameter butir, dan interpretasi. Estimasi parameter butir dan tes dilakukan menggunakan bantuan software Bilog MG. Selain itu juga digunakan software SPSS untuk menguji asumsi unidimensional.

HASIL DAN PEMBAHASAN

Uji Asumsi

Butir-butir soal tes Pengantar Statistik Sosial dikembangkan berdasarkan kisi-kisi. Butir soal yang dikembangkan berbentuk pilihan ganda dengan satu jawaban tunggal dan tidak ada keterkaitan antar butir tes. Oleh karena itu, butir pertama tidak ada kaitan dengan butir kedua, butir kedua tidak berkaitan dengan butir ketiga dan seterusnya. Dengan kata lain, peluang peserta tes untuk menjawab benar antar butir soal tidak saling mempengaruhi, sehingga secara logis membuktikan asumsi independensi lokal. Berdasarkan kisi-kisi pula, butir-butir soal dikembangkan untuk mengukur kemampuan mahasiswa pada bidang Statistik Sosial. Butir juga dikembangkan dengan dasar sebuah modul, dengan kisi-kisi dan spesifikasi tertentu, sehingga dapat dikatakan perangkat tes ini hanya mengukur kemampuan Statistik Sosial saja. Hal ini menjadi bukti logis bahwa butir soal hanya mengukur kemampuan mahasiswa dalam belajar Statistik Sosial saja, bukan

kemampuan lainnya. Melalui analisis faktor tidak tampak adanya faktor yang dominan yang berarti secara empiris data kurang dapat dikatakan mengukur dimensi dominan. Namun demikian, karena kisi-kisi hanya memuat satu kemampuan dominan yang diukur, yaitu kemampuan statistik mahasiswa, maka secara teoritis tes yang dikembangkan memenuhi asumsi unidimensional.

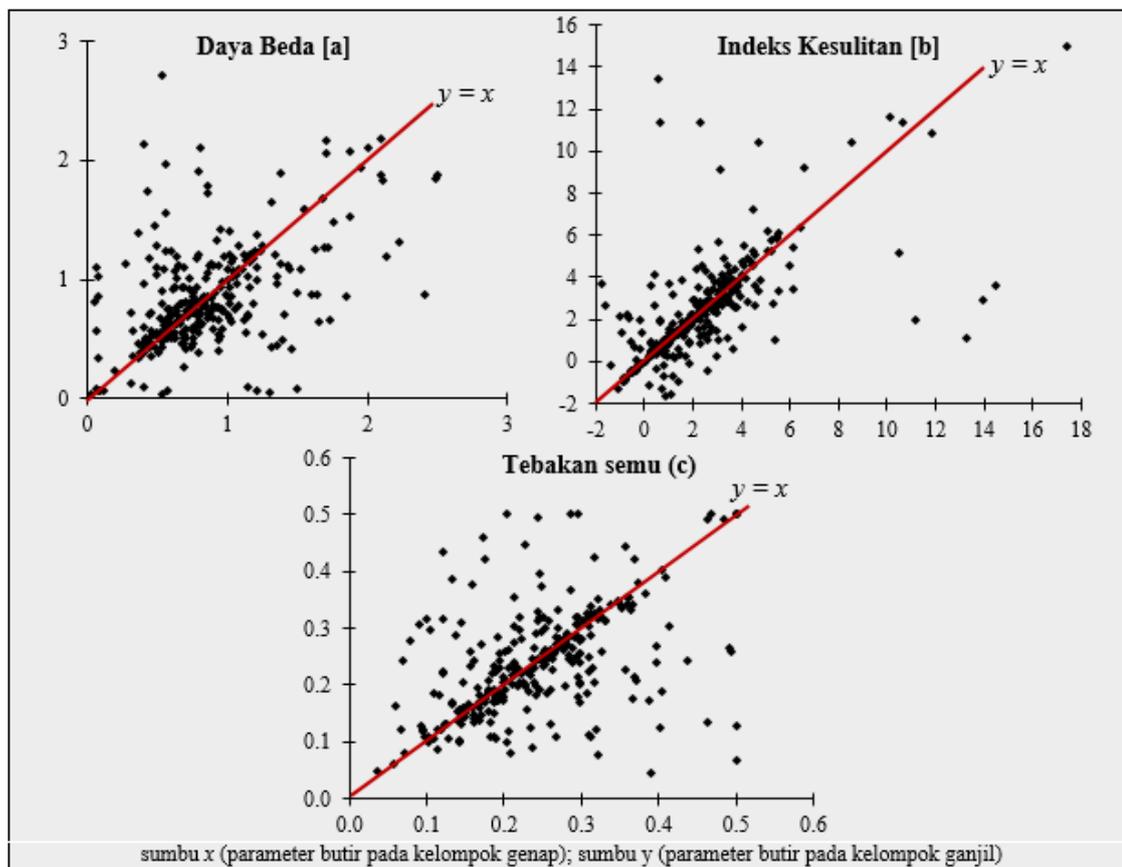
Selain unidimensional dan independensi lokal, asumsi lain yang perlu dibuktikan dalam teori respon butir yaitu invariansi parameter butir dan kemampuan. Invariansi parameter butir dibuktikan dengan mengelompokkan peserta tes menjadi dua kelompok yakni kelompok I dan II. Pengelompokkan peserta tes dilakukan secara acak (dengan undian). Sementara itu, invariansi parameter kemampuan dibuktikan dengan mengelompokkan butir soal menjadi dua bagian (ganjil dan genap). Setelah dikelompokkan, dilakukan estimasi parameter butir pada kelompok I dan II, dan estimasi parameter kemampuan pada kelompok soal ganjil dan genap. Hasil estimasi parameter butir dan kemampuan pada kedua kelompok tersebut kemudian disajikan dalam diagram pencar. Asumsi invariansi parameter akan terbukti jika sebaran titik-titik pada diagram pencar mendekati/mengikuti pola garis lurus $y = x$. Garis $y = x$ ditunjukkan oleh kurva berupa garis lurus yang melewati titik asal (0,0) dan memiliki kemiringan (gradien) = 1. Dengan kata lain, asumsi invarian terbukti jika terdapat korelasi positif yang tinggi (mendekati 1) pada parameter butir maupun parameter kemampuan pada dua kelompok yang dibandingkan.

Hasil analisis menunjukkan bahwa terdapat korelasi positif yang cukup tinggi (koefisien korelasi = 0,603) pada parameter tingkat kesulitan butir (b) pada kelompok I dan kelompok II. Jika kita cermati diagram pencar hubungan parameter tingkat kesulitan pada kedua kelompok tersebut (Gambar 1-b) terlihat bahwa sebaran titik-titik yang terbentuk cenderung mendekati pola garis $y = x$. Pada hasil estimasi parameter daya beda butir (a), koefisien korelasi yang diperoleh adalah sebesar 0,502. Artinya ada korelasi positif antara daya beda butir soal pada kedua kelompok peserta tes. Diagram pencar (Gambar 1-a) memperlihatkan bahwa sebaran titik yang terbentuk mendekati pola garis $y = x$. Hal yang sama juga ditunjukkan pada hasil estimasi peluang menebak jawaban (*guessing*) yang dilakukan oleh peserta tes. Koefisien korelasi yang diperoleh sebesar 0,48 yang artinya terdapat korelasi positif antara

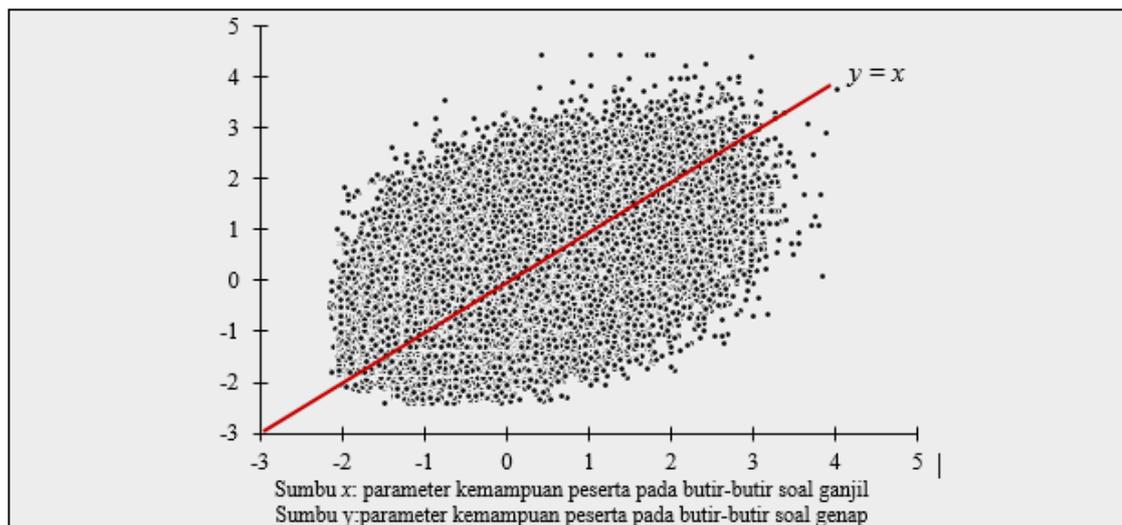
peluang menebak jawaban yang benar (*guessing*) pada kedua kelompok peserta tes. Diagram pencar yang menggambarkan sebaran titik yang terbentuk mendekati garis $y = x$ bisa dilihat pada Gambar 1-c.

Selanjutnya, untuk invariansi parameter kemampuan, diperoleh koefisien korelasi para-

meter kemampuan peserta pada kedua kelompok butir soal sebesar 0,543. Artinya parameter kemampuan peserta pada butir-butir soal ganjil memiliki korelasi positif dengan kemampuan peserta pada butir-butir soal genap. Diagram pencar pada Gambar 2 memperlihatkan sebaran titik berada disekitar garis lurus $y = x$.



Gambar 1. Diagram Pencar Invariansi Parameter Butir



Gambar 2. Diagram Pencar Invariansi Parameter Kemampuan

Berdasarkan hasil uji invariansi parameter, diperoleh kesimpulan bahwa parameter butir (tingkat kesulitan dan daya beda) pada perangkat tes Pengantar Statistik Sosial pada dua kelompok peserta tes tidak bervariasi. Demikian halnya dengan parameter kemampuan peserta dalam mengerjakan soal bernomor ganjil dan genap juga tidak bervariasi. Hal tersebut membuktikan bahwa parameter butir soal tes Statistik Sosial tidak akan mengalami perbedaan yang signifikan jika diujikan pada kelompok peserta tes yang berbeda. Dengan demikian, parameter butir soal tidak dipengaruhi oleh karakteristik peserta tes, dan kemampuan peserta tes tidak dipengaruhi oleh karakteristik butir.

Kecocokan Model

Model yang akan digunakan menunjukkan jumlah parameter yang dimuat dalam proses analisis. Model 1 PL jika hanya memuat satu parameter logistik yakni tingkat kesukaran (b), model 2 PL jika memuat dua parameter yakni tingkat kesukaran (b) dan daya beda (a), dan model 3PL jika memuat tiga parameter yakni tingkat kesukaran (b), daya beda (a) dan tebakan

semu (*pseudo-guessing* [c]). Kecocokan model ditentukan dengan uji khi-kuadrat yang dapat diperoleh dari *output phase 2* pada software BILOG. Untuk keperluan tersebut nilai khi-kuadrat dan probabilitas setiap butir soal pada masing-masing model ditabulasi. Butir dikatakan cocok (fit) model jika nilai probabilitas khi-kuadrat yang dihasilkan lebih besar dari taraf signifikansi α (prob. > 0,05). Model parameter logistik yang menghasilkan paling banyak butir soal yang cocok model (model fit) dipilih sebagai model yang akan digunakan untuk mengestimasi parameter butir. Adapun rekapitulasi jumlah butir yang cocok dengan model 1 PL, 2 PL, dan 3 PL disajikan pada Tabel 1. Berdasarkan Tabel 1 dapat dilihat bahwa model yang paling cocok adalah model 3 PL. Jumlah butir soal yang cocok model 3 PL yaitu sebanyak 73 butir, sedangkan model 1 PL hanya menghasilkan 24 butir yang cocok model, dan model 2 PL menghasilkan 42 butir soal yang cocok model. Dengan demikian model 3 PL akan digunakan untuk mengestimasi karakteristik tes Statistik Sosial.

Tabel 1. Distribusi Butir Soal Berdasarkan Kecocokan Model

Model dan Parameter yang Diestimasi	Butir yang Cocok Model
1 PL [tingkat kesulitan (b)]	29 butir
2 PL [tingkat kesulitan (b) & daya beda (a)]	89 butir
3 PL [tingkat kesulitan (b), daya beda (a), dan tebakan semu (c)]	105 butir

Karakteristik Butir

Sesuai dengan hasil yang telah diperoleh pada uji kecocokan model, maka karakteristik butir soal dideskripsikan berdasarkan hasil estimasi menggunakan model 3 PL. Dalam analisis karakteristik butir menggunakan model 3 PL, suatu butir dikatakan memiliki kualitas yang baik jika memiliki indeks kesukaran (b) -2 hingga +2, indeks daya beda (a) antara 0 hingga 2, dan nilai tebakan semu (c) kurang dari $1/k$, dimana k merupakan banyaknya pilihan jawaban. Dalam hal ini tes yang digunakan terdiri dari 4 pilihan jawaban, sehingga nilai tebakan semu yang baik adalah ketika $c < 0,25$. Ketika hal tersebut harus terpenuhi, jika ada salah satu parameter butir yang tidak memenuhi kriteria tersebut, maka butir memiliki kualitas yang kurang baik.

Hasil estimasi menghasilkan rerata indeks kesukaran (b) sebesar 2,149 dengan standar deviasi 2,451, rerata indeks daya beda (a)

sebesar 0,852 dengan standar deviasi 0,491, dan rerata indeks tebakan semu (c) sebesar 0,228 dengan standar deviasi 0,093. Karena rerata indeks kesukaran yang dihasilkan lebih dari 2, maka perangkat tes secara umum memiliki tingkat kesulitan yang tinggi. Sementara itu, jika ditinjau dari rerata daya beda, secara keseluruhan perangkat tes termasuk dalam kategori baik. Adapun jika ditinjau dari peluang menebak, kategori peluang menebak pada perangkat tes secara keseluruhan berada pada kategori baik, karena rerata parameter c tidak melebihi 0,25 (untuk 4 pilihan jawaban). Disamping itu, jika kita mencermati standar deviasi dari ketiga parameter, dapat disimpulkan bahwa nilai-nilai parameter daya indeks kesukaran lebih menyebar dari reratanya. Dengan kata lain sebaran nilai parameter indeks kesukaran lebih bervariasi dibandingkan sebaran nilai parameter daya beda dan tebakan semu. Dari 269 butir yang diestimasi diperoleh 81 butir atau hanya sekitar 30,11% butir yang memiliki kualitas yang baik.

Adapun rincian hasil analisis karakteristik butir disajikan pada Tabel 2.

Tabel 2. Rekapitulasi Hasil Analisis Karakteristik Butir dengan Model 3 PL

Parameter	Jumlah Butir		Rerata	Standar Deviasi
	Kualitas Baik	Kurang Baik		
Indeks Kesukaran (<i>b</i>)	146 (54,28%)	123 (45,72%) ¹	2,149	2,451
Daya Beda (<i>a</i>)	259 (96,28%)	10 (3,72%) ²	0,852	0,491
Tebakan Semu (<i>c</i>)	164 (60,97%)	105 (39,03%) ³	0,228	0,093
Keseluruhan (3 PL)	81 (30,11%)	188 (69,99%)		

¹ butir-butir yang memiliki indeks kesukaran > 2 (terlalu sulit)

² butir-butir yang memiliki indeks daya beda > 2

³ butir-butir yang memiliki indeks tebakan semu > 0,25

Berdasarkan Tabel 2, dapat dilihat bahwa kualitas butir yang kurang baik disebabkan karena banyaknya butir yang terlalu sukar, dimana terdapat 45,72% butir yang teridentifikasi memiliki

kesukaran tinggi. Dari 123 butir tersebut terdapat beberapa butir yang memiliki indeks kesukaran terlalu tinggi, sebagaimana disajikan pada Tabel 3.

Tabel 3. Butir-Butir dengan Parameter Tingkat Kesukaran Sangat Tinggi

Kode Soal		Nilai Parameter		
Paket	Butir	Indeks Kesukaran (<i>b</i>)	Daya Beda (<i>a</i>)	Tebakan Semu (<i>c</i>)
KN26	514	11,350	0,054	0,094
KN32	609	11,043	0,080	0,115
KN37	625	17,051	0,043	0,043
KN40	696	13,200	0,080	0,810
KN44	674	14,896	0,058	0,060
KN25	512	11,924	0,065	0,096
KN37	715	7,303	0,095	0,259
KN17	521	8,801	0,152	0,185

Ditinjau dari parameter indeks kesukaran, butir yang memiliki indeks kesukaran -2 hingga +2 merupakan butir yang tingkat kesulitannya baik. Jika nilai *b* < -2, maka butir tersebut tergolong mudah, sedangkan jika *b* > +2 maka butir tersebut tergolong sulit. Tingkat kemudahan dan kesukaran butir dapat dilihat dari nilai parameternya, semakin tinggi nilai *b* maka butir tersebut semakin sukar bagi peserta tes. Begitu-pun sebaliknya, semakin rendah nilai *b* maka butir tersebut semakin mudah. Berdasarkan Tabel 3, dapat dilihat bahwa terdapat 8 butir yang terindikasi sangat sukar, karena nilai parameter *b*-nya jauh di atas +2. Butir tersukar ada-

lah butir dengan kode soal KN37-625 yang memiliki nilai parameter *b* yaitu 17,051. Selain memiliki tingkat kesukaran yang sangat tinggi, butir dengan kode soal KN40-696 dan KN37-715 juga memiliki indeks tebakan semu yang tidak baik.

Jika ditinjau dari parameter daya beda butir, terdapat 10 butir yang memiliki daya beda yang kurang baik. Butir-butir tersebut teridentifikasi kurang baik karena memiliki indeks daya beda > 2. Tabel 4 menyajikan nilai parameter dari beberapa butir yang memiliki indeks daya beda kurang baik.

Tabel 4. Butir-Butir dengan Parameter Daya Beda Terkategori Kurang Baik

Kode Soal		Nilai Parameter		
Paket	Butir	Daya Beda (<i>a</i>)	Indeks Kesukaran (<i>b</i>)	Tebakan Semu (<i>c</i>)
KN26	497	2,128	0,913	0,152
KN37	589	2,192	0,566	0,318
KN45	475	2,179	2,580	0,320
KN23	793	3,614	2,292	0,288
KN23	824	2,320	1,986	0,221

Tabel 4 memperlihatkan bahwa butir dengan kode soal KN23-793 memiliki indeks daya beda tertinggi dibandingkan butir lainnya. Selain itu, butir tersebut juga memiliki indeks kesukaran dan tebakan semu terkategori kurang baik. Disamping itu, ada juga butir yang indeks kesukaran dan tebakan semunya berada pada kategori baik, meskipun parameter daya bedanya lebih dari 2, yaitu butir KN26-497. Namun secara keseluruhan, sebaran indeks daya beda butir-butir pada Tabel 4 tidak terlalu jauh dari nilai ideal (0 hingga 2), sehingga kualitas

butir-butir tersebut jika ditinjau dari kemampuan butir untuk membedakan antara peserta tes dengan kemampuan tinggi dan peserta tes dengan kemampuan rendah tidak terlalu buruk.

Selain itu, parameter lain yang juga berkontribusi menyebabkan banyaknya butir yang memiliki kualitas kurang baik adalah tebakan semu, dimana terdapat 39,03% butir yang indeks tebakan semunya melebihi 0,25 (untuk 4 pilihan jawaban). Tabel 5 menyajikan beberapa butir yang memiliki parameter tebakan semu terkategori kurang baik.

Tabel 5. Butir-Butir dengan Indeks Tebakan Semu Terkategori Kurang Baik

Kode Soal		Nilai Parameter		
Paket	Butir	Tebakan Semu (c)	Indeks Kesukaran (b)	Daya Beda (a)
KN26	429	0,500	1,312	0,517
KN26	435	0,459	2,262	0,331
KN26	482	0,403	0,946	1,169
KN32	543	0,500	0,453	0,643
KN37	548	0,437	-1,215	0,426
KN37	537	0,499	1,991	0,991
KN11	438	0,448	-0,627	0,655
KN11	494	0,439	0,804	0,884
KN31	547	0,500	0,848	0,731

Tabel 5 memperlihatkan beberapa butir yang memiliki indeks tebakan semu jauh di atas indeks tebakan semu ideal (0,25 untuk 4 pilihan jawaban). Berdasarkan Tabel 5, indeks tebakan semu tertinggi yaitu 0,5 dan terdapat 3 butir dengan indeks tebakan semu tersebut, yaitu butir KN26-429, KN32-543, dan KN31-547. Namun jika dikaitkan dengan parameter indeks kesukarannya, butir-butir pada Tabel 5 memiliki indeks kesukaran yang baik, kecuali butir KN26-435. Hal ini menunjukkan bahwa meskipun butir memiliki indeks kesukaran relatif baik (butir tidak terlalu mudah dan tidak terlalu sulit), peluang siswa untuk menjawab benar butir tersebut dengan cara menebak juga cukup tinggi. Dengan demikian, tidak ada jaminan bahwa butir yang memiliki indeks kesukaran baik, akan terbebas dari jawaban yang diperoleh dengan cara menebak. Akan tetapi, jika dikaitkan dengan parameter daya beda, meskipun indeks tebakan semu butir-butir pada Tabel 5 tinggi, indeks daya butir-butir tersebut masih berada pada kategori baik. Dengan demikian dapat dipahami bahwa peluang peserta tes menjawab dengan cara menebak tidak berhubungan dengan kemampuan butir untuk membedakan antara peserta tes dengan kemampuan tinggi dan rendah.

Pembahasan

Hasil analisis karakteristik butir menunjukkan bahwa hanya terdapat 81 butir (30,11%) dari 269 butir yang memiliki kualitas baik. Jika dipersentase, banyak butir yang memiliki kualitas baik tidak mencapai 50%, tetapi hanya 30,11%. Hasil analisis tersebut menunjukkan bahwa sebagian besar butir dari perangkat tes yang digunakan perlu ditelaah lagi, agar diperoleh butir-butir tes yang benar-benar berkualitas. Hal tersebut sesuai dengan salah satu tujuan dari analisis karakteristik butir yaitu untuk memperoleh informasi terkait kualitas butir (Sutrisno, 2016), sehingga dapat ditindaklanjuti dengan melakukan perbaikan jika butir tersebut dianggap kualitasnya kurang baik. Beberapa hal yang dapat dilakukan untuk meningkatkan kualitas butir tersebut diantaranya yaitu mengecek kesesuaian butir dengan kisi-kisi (Sutrisno, 2016), sehingga butir yang dihasilkan diharapkan sudah benar-benar sesuai dengan kemampuan yang hendak diuji.

Salah satu aspek yang perlu mendapat perhatian khusus bagi pengembang soal yaitu terkait dengan tingkat kesukaran butir. Hasil analisis dengan teori respon butir menggunakan model 3 PL menunjukkan bahwa 45,72% butir terkategori sangat sulit. Temuan tersebut mem-

berikan indikasi bahwa banyak butir dalam perangkat tes yang tidak sesuai dengan kemampuan peserta tes. Dengan kata lain, tingkat kesukaran yang telah dirumuskan oleh pembuat tes belum sesuai dengan tingkat kesukaran berdasarkan hasil analisis empiris. Hal tersebut dapat disebabkan karena pengklasifikasian tingkat kesukaran oleh pembuat tes hanya berdasarkan intuisi (Stanley & Wang, 1968), padahal suatu butir yang dianggap sulit atau mudah oleh pembuat soal belum tentu dirasakan sulit atau mudah oleh peserta tes (Baker, 2001). Selain itu, temuan tersebut juga dapat disebabkan oleh faktor lain seperti kesiapan peserta tes dalam mengerjakan ujian dan penguasaan terhadap materi yang diujikan. Ketika peserta tes tidak belajar dengan baik, maka kemungkinan peserta tes akan kesulitan mengerjakan butir-butir tes yang diujikan. Hal tersebut tentunya akan berdampak pada tingginya indeks kesukaran butir.

Ditinjau dari parameter daya beda, secara umum kemampuan perangkat tes untuk membedakan antara peserta yang memiliki kemampuan tinggi dengan peserta yang memiliki kemampuan rendah sudah baik. Hal tersebut terkonfirmasi dari hasil analisis yang menunjukkan bahwa 96,28% butir memiliki indeks daya beda pada kategori baik. Namun demikian masih terdapat butir-butir yang memiliki indeks daya beda yang kurang baik. Beberapa hal yang menyebabkan daya beda butir kurang baik yaitu tingkat kesukaran butir yang terlalu rendah atau terlalu tinggi dan terdapat pengecoh yang tidak masuk akal, meskipun tingkat kesukaran butir tersebut berada pada kategori baik (Thorndike, 2005). Dengan adanya pengecoh yang tidak masuk akal tersebut, akan memudahkan bagi peserta tes untuk menentukan bahwa ada pilihan jawaban yang salah, sehingga butir tersebut akan lebih mudah dijawab (Amelia & Kriswanto, 2017).

Berikutnya, untuk memperbaiki kualitas perangkat tes, parameter tebakan semu juga perlu mendapat perhatian serius dari pengembang tes. Persentase butir yang memiliki indeks tebakan semu lebih dari 0,25 (untuk 4 pilihan jawaban) yang mencapai 39,03%, menunjukkan bahwa masih banyak butir yang dapat dijawab benar oleh peserta tes dengan cara menebak. Dengan demikian pengembang soal perlu memperhatikan keberfungsian pengecoh atau distraktor pilihan jawaban pada setiap butir, mengingat butir yang pengecohnya kurang baik peluang terjawab benar dengan cara menebak akan lebih

tinggi. Selain itu, keberadaan pengecoh yang tidak masuk akal akan memudahkan peserta tes untuk memutuskan bahwa pengecoh tersebut salah, sehingga kemungkinan siswa menjawab benar dengan menebak sangat tinggi dan menyebabkan butir terlalu mudah (Amelia & Kriswanto, 2017). Sebaliknya, pengecoh yang terlalu dekat nilai kebenarannya dengan kunci jawaban akan menyebabkan butir menjadi terlalu sulit (Amelia & Kriswanto, 2017). Dengan demikian antara keberfungsian distraktor, tingkat kesukaran butir, dan tebakan semu secara logis memiliki keterkaitan satu sama lain. Hal tersebut tentunya perlu menjadi perhatian bagi pengembang soal ketika menyusun butir.

SIMPULAN

Hasil analisis menunjukkan bahwa perangkat tes Pengantar Statistik Sosial yang dikembangkan oleh Universitas Terbuka cocok diestimasi menggunakan model 3 PL. Hasil estimasi parameter butir dengan model 3 PL menunjukkan bahwa dari 269 butir hanya terdapat 81 butir yang memiliki kualitas baik. Rerata indeks kesukaran (b) perangkat tes yaitu 2,149 dan berada pada kategori kurang baik. Rerata daya beda (a) perangkat tes yaitu 0,852 dan berada pada kategori baik. Sedangkan rerata indeks tebakan semu (c) perangkat tes yaitu 0,288 dan berada pada kategori kurang baik untuk butir dengan 4 pilihan jawaban. Berdasarkan hasil estimasi parameter butir tersebut, secara keseluruhan kualitas perangkat tes perlu diperbaiki untuk menghasilkan perangkat tes yang memiliki karakteristik baik.

Terdapat beberapa saran untuk menindaklanjuti temuan dalam penelitian ini. *Pertama*, hasil penelitian ini hendaknya menjadi dasar bagi pengembang soal untuk memperbaiki kualitas tes, khususnya untuk butir-butir yang memiliki karakteristik kurang baik. *Kedua*, pengembang soal perlu memahami karakteristik peserta tes, sehingga butir yang dibuat tidak terlalu sulit dan sesuai dengan kemampuan peserta tes. *Ketiga*, pengembang soal harus memperhatikan keberfungsian distraktor butir, untuk meminimalisir peluang peserta tes menjawab butir dengan cara menebak. *Keempat*, pengembang soal perlu menelaah ulang kisi-kisi tes, untuk menjamin bahwa dimensi pengetahuan yang diukur adalah tunggal. *Kelima*, perlu adanya penelitian lainnya untuk menguji kesetaraan antar paket soal yang digunakan dalam perangkat tes ini.

DAFTAR PUSTAKA

- Adedoyin, O. O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asian Social Science*, 3(4), 992-1011.
- Adegoke, B. A. (2013). Comparison of item statistics of physics achievement test using classical test and item response theory frameworks. *Journal of Education and Practice*, 4(22), 87-96.
- Al-Khader, M. M. A., & Albursan, I. S. (2017). Accuracy of measurement in the classical and the modern test theory: An empirical study on a children intelligence test. *International Journal of Psychological Studies*, 9(1), 71-80.
- Amelia, R. N., & Kriswantoro, K. (2017). Implementasi item response theory sebagai basis analisis kualitas butir soal dan kemampuan kimia siswa Kota Yogyakarta. *JKPK (Jurnal Kimia dan Pendidikan Kimia)*, 2(1), 1-12.
- Ayanwale, M. A., Adeleke, J. O., & Mamadelo, T. I. (2018). An assessment of item statistics estimates of basic education certificate examination through classical test theory and item response theory approach. *International Journal of Educational Research Review*, 3(4), 55-67.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- De Mars, C. E. (2010). *Item response theory: Understanding statistics measurement*. Oxford, UK: Oxford University Press.
- Eleje, L. I., Onah, F. E., & Abanobi, C. C. (2018). Comparative study of classical test theory and item response theory using diagnostic quantitative economics skill test item analysis results. *European Journal of Educational & Social Sciences*, 3(1), 71-89.
- Gronlund, N. E. (1998). *Assessment of students achievement* (6th ed.). Boston, MA: Allyn and Bacon.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication.
- Kartowagiran, B., Munadi, S., Retnawati, H., & Apino, E. (2018). The equating battery test packages of mathematics national examination 2013-2016. *SHS Web of Conference*, 42, 1-6. doi: 10.1051/shsconf/20184200022
- Popham, W. J. (2009). *Instruction that measures up: Successful teaching in the age of accountability*. Alexandria, VA: ASCD.
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4, 1-11. doi:10.1080/2331186X.2017.1301013
- Ratnaningsih, D. J., & Isfarudi, I. (2013). Analisis butir tes objektif ujian akhir semester mahasiswa Universitas Terbuka berdasarkan teori tes modern. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, 14(2), 98-109.
- Retnawati, H. (2014). *Teori respon butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Parama Publishing.
- Retnawati, H. (2016). *Validitas, reliabilitas, dan karakteristik butir: Panduan untuk peneliti, mahasiswa, dan psikometrian*. Yogyakarta: Parama Publishing.
- Siri, A., & Freddano, M. (2011). The use of item analysis for the improvement of objective examinations. *Procedia-Social and Behavioral Sciences*, 29, 188-197.
- Stanley, J. C., & Wang, M. D. (1968). *Differential weighting: A survey of methods and empirical studies*. Baltimore, MD: Johns Hopkins.
- Sudaryono, S. (2011). Implementasi teori respon butir (*item response theory*) pada penilaian hasil belajar akhir di sekolah. *Jurnal Pendidikan dan Kebudayaan*, 17(6), 719-732.
- Sutrisno, H. (2016). An analysis of the mathematics school examination test quality. *Jurnal Riset Pendidikan*

Matematika, 3(2), 162-177.
doi:10.21831/jrpm.v3i2.11984

(7th ed.). Upper Saddle River, NJ: Pearson
Education

Thorndike, R. M. (2005). *Measurement and
evaluation in psychology and education*