

**ANALISIS TINGKAT KESESUAIAN BUTIR SOAL (ITEM FIT)
MENGUNAKAN MODEL RASCH**

Yessica Mega Aprita, Tri Haryati
Universitas Bina Sarana Informatika
(Naskah diterima: 1 Januari 2021, disetujui: 30 Januari 2021)

Abstract

The results of the research using Winsteps Software indicates that the Person Reliability in the Accounting Final Examination is 0.81 while the Item Reliability is 0.93. The amount of Alpha Cronbach is 0.82. Person Measure 0.94 logit shows that the average ability of students is above the item, which means that the response patterns of the students are appropriate. Additional information can be obtained are the value of logit person, for the Student-1 with logit +4.07 indicates the person with the highest ability (able to do almost all of the questions); Student-61 with logit -0.98 showed the students with the lowest ability (fewest amount in answering the questions correctly). Misfit occurs for Students-1 because $MNSQ > 1.5$ and PT Measure corr < 0.4 this case happens because he could not answer the easy questions correctly, but can answer the difficult item correctly. Additionally, Student-31 also experienced a misfit, where $MNSQ < 0.5$ and PT Measure corr < 0.4 this case happens because he was less careful in work on the items numbers 15 and 26. Based on the research results, it can be concluded that most of the items in Accounting Final Examination instruments, have a fit item pattern, so it is assumed that in this Final Examination of Accounting is able to measure the measurement objectives.

Keywords: Item Fit, Rasch Model

Abstrak

Hasil penelitian dengan bantuan software Winsteps menunjukkan bahwa *Person Reliability* pada UAS Akuntansi sebesar 0,81 sedangkan *Item Reliability* sebesar 0,93. Besarnya Alpha Cronbach yaitu 0,82. *Person Measure* 0,94 logit menunjukkan rata-rata kemampuan siswa di atas item. Hasil analisis juga menunjukkan bahwa semua item fit, sehingga soal tidak perlu direvisi atau dieliminasi. Informasi lain yang dapat diperoleh yaitu nilai logit person, untuk person-1 dengan +4,07 logit menunjukkan person dengan kemampuan tertinggi (mampu mengerjakan hampir semua soal); person-61 dengan -0,98 menunjukkan siswa dengan kemampuan paling rendah (paling sedikit dalam menyelesaikan soal dengan benar). Untuk person 1 terjadi misfit karena $MNSQ > 1,5$ dan PT Measure corr $< 0,4$ hal tersebut dikarenakan dia tidak berhasil menjawab soal mudah dengan benar, namun dapat menjawab dengan benar pada soal sulit. Selain itu, Siswa-31 juga mengalami *misfit*, dimana $MNSQ < 0,5$ dan PT Measure corr $< 0,4$ hal tersebut dikarenakan dia kurang teliti dalam mengerjakan soal nomor 15 dan 26. Berdasarkan hasil penelitian, dapat disimpulkan bahwa sebagian besar soal UAS SMK Program Keahlian

Akuntansi Keuangan memiliki pola item yang fit, sehingga diasumsikan bahwa dalam UAS Akuntansi kali ini secara keseluruhan, instrumen mampu mengukur apa yang menjadi tujuan ukur.

Kata Kunci: Item Fit, Rasch Model

I. PENDAHULUAN

Pendidikan merupakan salah satu upaya yang dilakukan untuk meningkatkan kualitas sumber daya manusia melalui upaya pengajaran dan pelatihan. Peningkatan kualitas pendidikan menjadi bagian terpenting dalam membina dan membangun sumber daya manusia yang tangguh dan dapat diandalkan. Pendidikan yang berkualitas tersebut akan tercapai apabila kegiatan belajar mengajar dan evaluasi terhadap hasil belajar dilaksanakan secara efektif sehingga hasil belajar lebih optimal.

Evaluasi merupakan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja, atau produktivitas suatu lembaga dalam melaksanakan suatu program (Mardapi, 2012: 4). Evaluasi dapat digunakan untuk mengetahui tingkat pencapaian keberhasilan guru dalam proses pembelajaran. Selain untuk mengukur sejauh mana tujuan pembelajaran tercapai, evaluasi juga dapat digunakan untuk membuat keputusan dalam peningkatan pembelajaran.

Evaluasi belajar akhir semester merupakan kegiatan rutin setiap tahun yang diselenggarakan di berbagai jenjang sekolah, khususnya sekolah menengah kejuruan. Pada dasarnya ujian akhir semester diselenggarakan untuk mengukur pencapaian kompetensi peserta didik yang dilakukan oleh satuan pendidikan dalam memperoleh pengakuan atas prestasi belajar peserta didik. Idealnya, pelaksanaan ujian tidak sekedar memantau seberapa besar kemampuan peserta dalam menjawab butir soal sebanyak mungkin, akan tetapi upaya untuk memahami apakah instrumen lembar soal tersebut mampu mengukur kemampuan peserta dengan baik juga perlu diamati. Untuk mengatasi permasalahan yang telah dikemukakan tersebut, maka pada penelitian ini dilakukan penerapan metode *item fit*. Item fit menjelaskan apakah item soal kita berfungsi normal melakukan pengukuran atau tidak. Jika ada item yang tidak fit, hal ini mengindikasikan adanya miskonsepsi subjek dalam menjawab soal tersebut (Sumintono, B., & Widhiarso, W. 2015:70)

Pada penelitian ini, peneliti ingin mengetahui tingkat kesesuaian instrumen tes yang digunakan untuk mengetahui kemampuan siswa pada soal UAS Akuntansi dengan pendekatan model Rasch. Kualitas ini diukur berdasarkan beberapa indikator yaitu butir soal yang fit dengan model rasch dan reliabilitas butir soal. Oleh karena itu, dirancang suatu instrumen tes kemudian ditentukan butir soal mana saja yang fit serta yang tidak fit dengan model rasch. Selain itu, dengan bantuan software Winstep akan ditentukan nilai alfa cronbach untuk mengetahui reliabilitas butir soal

II. KAJIAN TEORI

Model Teori Tes

Untuk dapat mengungkap karakteristik instrumen tes, diperlukan pemahaman berkaitan dengan model teori tes. Pada dasarnya perkembangan model teori tes tersebut berdasarkan dua pendekatan besar yakni pendekatan *classic* yang melahirkan teori tes klasik dan pendekatan *latent trait* yang melahirkan teori respon butir dan *rasch model*.

a. Teori Tes Klasik

Teori tes klasik merupakan teori yang dikembangkan dan diaplikasikan sejak lama, teori ini telah memberikan banyak kontribusi utamanya dalam dunia pendidikan dan psiko-

metri. Azwar (2015: 26) menyatakan bahwa performansi individu yang diungkap oleh suatu skala pengukuran atau tes dinyatakan dalam bentuk angka yang disebut skor. Skor ini merupakan skor perolehan (*observed score* atau skor tampak) dan disimbolkan dengan huruf (X), setiap peserta juga mendapatkan skor lain yang merupakan skor sesungguhnya. Skor ini tidak dapat diungkap secara langsung oleh tes, sehingga tidak dapat diketahui besarnya. Skor sesungguhnya (*true score*) merupakan performansi yang benar dan murni, dan disebut skor murni (T).

Selain diperoleh skor tampak (X) dan skor murni (T), pada setiap hasil pengukuran juga terdapat kesalahan pengukuran (*error*) yang besarnya bagi setiap peserta tes juga tidak dapat diketahui. Allen dan Yen (1979: 57) mengungkapkan hubungan antara skor tampak (X), skor murni (T), dan kesalahan pengukuran (E) sebagai berikut:

$$X = T + E$$

Dengan demikian skor tampak (X) akan tergantung pada besarnya kesalahan pengukuran (E), sedangkan besarnya skor murni pada setiap pengukuran yang sama diasumsikan tetap.

Allen dan Yen (1979: 57-60) menyatakan bahwa ada tujuh asumsi pada teori tes klasik sebagai berikut:

- 1) Skor yang diperoleh peserta tes terdiri dari skor murni (*true score*) dan kesalahan pengukuran. Dengan demikian, skor yang diperoleh dari suatu pengukuran pada umumnya tidak menunjukkan keadaan yang sebenarnya. Melesetnya skor perolehan dari keadaan yang sebenarnya merupakan kesalahan pengukuran (*error of measurement*). Skor murni dianggap tetap, sementara skor kesalahan dapat terjadi secara dinamis. Sebenarnya yang diperlukan adalah skor murni yaitu skor yang mencerminkan secara tepat besaran atribut yang diukur, namun pada umumnya tidak ada jalan untuk mengetahui skor murni itu secara langsung. Skor murni hanya dapat diketahui dengan cara tidak langsung, yaitu melalui kesalahan baku.
- 2) Nilai harapan skor yang diperoleh sama dengan skor murni. Skor murni merupakan harga rata-rata distribusi teoretik skor tampak apabila peserta yang sama dikenai tes yang sama berulang kali, dengan asumsi pengulangan tes itu dilakukan secara independen. Ketika dikenai tes, peserta diasumsikan tidak dipengaruhi oleh faktor kelelahan, dan hasil tes yang satu tidak saling mempengaruhi dengan hasil tes yang lain.
- 3) Skor murni dan kesalahan pengukuran tidak saling berkorelasi. Hal ini berarti tidak ada hubungan sistematis antara skor murni dan skor kesalahan pengukuran. Implikasinya adalah peserta tes dengan skor murni yang tinggi tidak akan mempunyai skor kesalahan pengukuran yang selalu positif ataupun selalu negatif daripada siswa dengan skor murni yang rendah.
- 4) Kesalahan pengukuran pada dua tes yang mengukur kemampuan yang sama tidak berkorelasi. Seorang peserta tes, dimana skor pada tes pertama mengandung kesalahan pengukuran besar, belum tentu akan memiliki kesalahan pengukuran yang besar pula pada tes yang kedua. Asumsi ini berlaku dengan pengertian bahwa pada tes pertama dan kedua terjadi pengaruh kelelahan, pengaruh latihan, dan proses belajar.
- 5) Pada dua tes yang mengukur kemampuan yang sama, kesalahan pengukuran pada tes pertama tidak berkorelasi dengan skor sebenarnya pada tes kedua. Asumsi ini tidak akan bertahan apabila tes yang kedua itu mengukur aspek yang mempengaruhi

kesalahan pengukuran pada pengukuran yang pertama.

- 6) Dua tes yang menghasilkan skor yang memenuhi kelima asumsi pertama disebut *parallel test* jika skor murni dan variansi kesalahan pengukuran yang diperoleh peserta tes sama.
- 7) Dua tes yang menghasilkan skor yang memenuhi kelima asumsi pertama disebut *essentially t-equivalent* jika selisih skor murni yang diperoleh peserta tes pada tes pertama dan pada tes kedua merupakan bilangan konstan.

Terdapat tiga parameter butir yang diestimasi pada teori tes klasik yaitu tingkat kesukaran, daya pembeda, dan keberfungsian pengecoh. Tingkat kesukaran soal merupakan proporsi peserta tes yang menjawab benar dengan banyaknya peserta tes. Daya beda butir tes merupakan besarnya kemampuan butir tes untuk membedakan peserta tes yang pandai dengan yang kurang pandai. Daya beda menunjuk kepada selisih proporsi yang menjawab benar pada kelompok atas dan proporsi yang menjawab benar pada kelompok bawah. Daya beda butir dihitung dengan formulasi korelasi poin biserial. Daya beda yang tinggi terjadi pada butir dengan tingkat kesukaran optimal, tetapi tingkat kesukaran opti-

mal tidak menjamin daya beda butir soal akan tinggi (Sax, 1980: 278). Butir yang baik adalah butir yang mempunyai tingkat kesukaran sedang daya beda tinggi, dan pengecoh yang berfungsi efektif (Purwanto, 2011: 97).

Keberfungsian pengecoh menunjuk kepada seberapa banyak peserta yang memilih pengecoh tersebut. Pengecoh berfungsi sebagai pengidentifikasi peserta tes yang berkemampuan tinggi. Pengecoh dikatakan berfungsi efektif apabila banyak dipilih oleh peserta tes yang berasal dari kelompok bawah, sebaliknya apabila pengecoh itu banyak dipilih oleh peserta tes yang berasal dari kelompok atas, maka pengecoh itu tidak berfungsi sebagaimana mestinya.

Beberapa kelebihan dari teori tes klasik adalah mudah dalam pemahaman konsep. Selain itu ukuran sampel yang digunakan tidak banyak, minimal 30 orang. Pada akhirnya teori ini hanya dapat diaplikasikan pada skala kecil. Dalam teori tes klasik, tingkat kesukaran dan daya beda sangat menentukan kualitas butir soal. Namun, karakteristik butir soal yang dihasilkan teori tes klasik inkonsisten (berubah) bergantung pada kelompok peserta ujian. Hal inilah yang menjadikan tes teori klasik memiliki kelemahan dalam mengukur kompetensi peserta ujian. Selain itu, kelema-

han lainnya adalah tingkat kesukaran butir dan daya beda butir bergantung pada kelompok peserta ujian. Pada kenyataannya, kemampuan seseorang untuk menjawab benar dari suatu butir soal bergantung pada kemampuan individu peserta ujian itu sendiri, bukan berdasarkan kemampuan kelompok peserta ujian. Artinya, seseorang yang belajar dan memahami mata kuliah yang dipelajari akan dapat mengerjakan soal dengan baik. Hal ini berarti, peluang menjawab benar soal yang diujikan lebih tinggi dibanding peserta yang tidak belajar (Ratnaningsih, 2013: 99).

b. Teori Respon Butir

Berbeda dengan teori tes klasik, teori respon butir menggunakan asumsi yang relatif kuat namun menghasilkan skala pengukuran yang memiliki sejumlah keunggulan bila dibandingkan dengan teori tes klasik dan sekarang memegang tempat sentral dalam teori pengukuran pendidikan (MacCann & Stanley, 2008: 2). Teori respon butir (item response theory) dikembangkan oleh para ahli pengukuran bidang psikologi dan pendidikan sebagai upaya meminimalkan kekurangan-kekurangan yang ada dalam teori tes klasik (Suwanto, 2011:75). Model pengukuran pada teori respon butir didasarkan pada dua kaidah, yaitu: (1) kinerja peserta tes pada suatu butir soal

dapat diprediksi oleh sekumpulan faktor yang disebut *traits* atau kemampuan (*abilities*), dan (2) hubungan antara kinerja peserta tes pada suatu butir soal dan sekumpulan *traits* dapat digambarkan dalam sebuah fungsi monoton naik yang disebut fungsi karakteristik butir (*item characteristic function*) atau kurva karakteristik butir (*item characteristic curve*) (Hambleton, Swaminatan, & Rogers, 1991: 7). Fungsi karakteristik butir ini menyatakan bahwa semakin meningkat level kemampuan seseorang, semakin meningkat pula peluangnya menjawab benar suatu butir tertentu. Selain itu, teori respon butir dinilai dapat melakukan kajian parameter butir lebih lengkap, meliputi tidak hanya daya beda, taraf kesukaran, dan peluang tebakan semu, tetapi juga dapat menganalisis parameter fungsi informasi butir dan kemungkinan adanya bias butir (Azwar, 2015: 151-152).

Dikatakan bahwa dalam teori respon butir, parameter butir tes dan parameter kemampuan disebut invarian. Maknanya adalah ketidaktergantungan parameter kemampuan yang menjadi ciri peserta terhadap butir tes yang dipakai untuk mengestimasi kemampuan peserta tes. Dengan kata lain, untuk peserta tes yang sama, karakteristik peserta tes selalu tetap, meskipun butir tes berubah. Invariansi

parameter butir tes adalah tetapnya karakteristik butir tes meskipun kelompok peserta tes yang menjawab butir tes berubah. Parameter butir tes yang diperoleh dari kelompok peserta tes yang berbeda akan selalu sama (Hambleton, Swaminatan, & Rogers, 1991: 18). Dalam teori respon butir, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir (Retnawati, 2014: 1)

Berlakunya teori respon butir dilandasi oleh tiga asumsi yakni unidimensi dan independensi lokal (Hambleton, Swaminatan, & Rogers, 1991: 9). Asumsi unidimensi artinya tes dibuat hanya untuk mengukur satu kemampuan (Suwanto, 2013: 69). Tes yang butirnya mengukur lebih dari satu kemampuan tidak dapat memberi informasi mengenai jawaban peserta tes. Unidimensi dalam hal ini adalah adanya faktor-faktor dominan yang mempengaruhi hasil suatu tes. Faktor-faktor dominan itulah yang disebut kemampuan yang diukur oleh suatu tes.

Independensi lokal bermakna antara tes dan peserta tes tidak saling mempengaruhi. Peluang seorang individu menjawab benar suatu soal tidak bergantung pada jawaban soal lainnya. Urutan soal dalam teks tidak mempe-

ngaruhi posisi seseorang dalam dimensi yang diukur. Demikian juga jawaban seseorang terhadap suatu soal tidak dipengaruhi oleh jawaban orang lain terhadap soal tersebut. Kedudukan butir tes tidak akan mempengaruhi skor peserta tes atau urutan butir tes tidak mempengaruhi posisi peserta dalam suatu tingkat kemampuan.

Kurva karakteristik butir dinyatakan dengan tiga fungsi matematik yang cocok untuk tes dikotomus. Ketiga fungsi matematik tersebut menghasilkan tiga model logistik yaitu model satu parameter, dua parameter, dan tiga parameter (Hambleton & Swaminatan, 1991: 35-49).

c. Model Rasch

Salah satu metode yang paling dapat diandalkan dan cocok untuk menilai kemampuan siswa menggunakan Model Rasch. Model Rasch digunakan untuk mengukur kemampuan, sikap dan sifat-sifat pribadi dari data penilaian (Nopiah, 2012: 1172). Salah satu fitur mencolok dari model Rasch adalah kemampuan untuk mempelajari distribusi dari setiap uji statistik independen pada parameter person dan item. Hal ini dapat dilakukan dengan menyelenggarakan tes pada skor person dan margin suatu item. Rasch sendiri menganjurkan bentuk pendekatan inferensi

bersyarat untuk menguji modelnya, tetapi tidak pernah dilaksanakan karena perhitungan yang terlalu rumit (Christensen, 2010:101).

Model Rasch pada mula pertama diajukan oleh George Rasch sekitar tahun 1966 dan kemudian dikembangkan di Amerika oleh Benjamin Wright di Universitas Chicago. Parameter butir pada model ini adalah tingkat kesukaran butir, sedangkan parameter lainnya seperti daya pembeda dianggap sama, dan dugaan atau *guessing* sama dengan nol. Berikut ini disajikan formula model Rasch (Wright & Stone, 1979: 16).

$$\pi_{vi} = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)}$$

π adalah peluang peserta tes yang memiliki kemampuan v menjawab butir i dengan benar, β_v adalah tingkat kemampuan peserta tes, parameter δ adalah tingkat kesulitan butir i , dan \exp adalah bilangan transedental yang memiliki nilai 2,718.

Ketika peserta tes v lebih memiliki kemampuan daripada tingkat kesulitan item i , kemudian β_v lebih besar dari δ_i , maka akan terjadi selisih positif dan probabilitas untuk menjawab item i lebih dari 0,5. Jika kemampuan seseorang jauh melebihi tingkat kesulitan item, maka akan terjadi selisih positif yang lebih besar, yang pada akhirnya probabilitas

menjawab benar mendekati 1. Tetapi ketika item terlalu sukar dijawab oleh peserta tes, kemudian β_v lebih kecil dari δ_i , selisihnya menghasilkan angka negatif sehingga probabilitas peserta tes untuk menjawab benar kurang dari 0,5. Selanjutnya, ketika tingkat kesukaran jauh melampaui kemampuan seseorang, maka probabilitas peserta tes menjawab benar akan mendekati nol.

Tingkat kesukaran pada model Rasch merupakan fungsi dari kemampuan seseorang. Seseorang yang mempunyai kemampuan tinggi akan merasa mudah mengerjakan butir soal, sebaliknya seseorang yang memiliki kemampuan rendah akan sulit menjawab butir soal. Asumsi yang digunakan pada model Rasch adalah: (1) semua butir memiliki daya pembeda yang sama, dan (2) peluang menjawab butir benar bagi yang memiliki kemampuan rendah sama dengan nol. Dengan kata lain semua kurva karakteristik butir soal adalah sejajar atau mendekati sejajar, karena besarnya slope kurva sama dan memotong sumbu y di titik 0 atau mendekati 0.

a. Ditinjau dari perspektif teori respon butir, model Rasch menawarkan sifat matematika yang berlaku jika model fit dengan data: 1) independensi lokal berarti bahwa probabilitas setiap item untuk dipecahkan adalah

SISWA_5	011111011111101011001011000100101010
SISWA_6	1111100111100110000100101001010111110
SISWA_7	1111111111111111111111111111111111110
SISWA_8	111010100110010000001110010111101100000
SISWA_9	1011100011000101000010001001100001110111
SISWA_10	1110000011011111011111000110110111100
SISWA_11	01110001111001010111110011110111010011
SISWA_12	1110101111010101001010011010001111110
SISWA_13	11110101110111111111110111011111111111
SISWA_14	11101011011010101111111011110101111010
SISWA_15	00011111110011110111101001011101100100
SISWA_16	11110001110101111111111101111101101010
SISWA_17	111100100110100111001000111111001001
SISWA_18	0110101111101001111000110111100000100
SISWA_19	111110111111101110010001011010111010100
SISWA_20	11110111110110111101110111011110101111
SISWA_21	1111110101011101111110110011101001111
SISWA_22	111010011110010001101110101011011010100
SISWA_23	1011100010001101010100001011110010000101
SISWA_24	10111101111001111111111101111101111110
SISWA_25	01111001111011111101111101111111111110
SISWA_26	11111001111010010110100010111101110111
SISWA_27	11110011111111101001100011111111101110
SISWA_28	1111010101101010011110011110011001001
SISWA_29	11111111111101111111111111111111111111
SISWA_30	000111111011101011101101100111001111000
SISWA_31	11111111111101111111111011111111111111
SISWA_32	11101000010010110110001001011101010101
SISWA_33	01101001111011111111111100101111011100
SISWA_34	001110011110000101111101001111010001010
SISWA_35	0100010101111101011111111110101111010
SISWA_36	1101100111101101010010100011010000011100
SISWA_37	111011111111111111111101111111111111110
SISWA_38	001100111110010011110110000111100111110
SISWA_39	10100111111110001011100101111011111100
SISWA_40	111111011110110111110111011111111111010
SISWA_41	1001001000100101100000001000010010100010
SISWA_42	0010000011101100010000110000010001100010
SISWA_43	111011111111010010111100111111111111101
SISWA_44	111111111111011111011111011111111111001010
SISWA_45	111010010110111111111100100011111101100
SISWA_46	110010010010010111111110000101111101101
SISWA_47	0111111111101111111000101111111111111100
SISWA_48	000010110010000111100011001010100100011
SISWA_49	000100011110011010010101011010101111100
SISWA_50	1111100111110111111101110001111001011100
SISWA_51	101111010110111111110110011111001111000
SISWA_52	1101100000100101101110001001011101010101
SISWA_53	0110100111101111111111100101111011100
SISWA_54	001110011110000101111101001111010001010
SISWA_55	01001010111110101111111111110101111010
SISWA_56	1101100111101101010010100011010000011100
SISWA_57	111011111111111111111011111111111111110
SISWA_58	001100111110010011110110000111100111110
SISWA_59	101001111111100010111100101111101111100
SISWA_60	1111110111101101111101110111111111111010
SISWA_61	1001001000100101100000001000010010100010
SISWA_62	0010000111101100010000110000010001100010
SISWA_63	111011111111010111111100111111111111101
SISWA_64	111111111111011111011111011111111111001010
SISWA_65	1110100101101101111100100011111101100
SISWA_66	11001001001001011111110000101111101101
SISWA_67	0111111111101111111000101111111111111100
SISWA_68	0000101100100001111100011001010100100011
SISWA_69	000100011110011101001010101010101111100
SISWA_70	1111100111110111111101110001111001011100
SISWA_71	101111010101111111111011001111100111000
SISWA_72	00110010101011011101100110010101010110
SISWA_73	0111101111101111101110101101110100111100

SISWA_74	011010111110111111111110010111111001111
SISWA_75	0100110111110110000101000010110101011111
SISWA_76	111110111111101110110110011111111111101
SISWA_77	111010111111101110111011111111111101101
SISWA_78	011111111111011011111111111111111111110
SISWA_79	000111001011101101110100011111100101110
SISWA_80	0111110011111111111111011010111111110100
SISWA_81	011110011111010011111000111111100001100
SISWA_82	0111101010111011111011101110111011100110
SISWA_83	01101011111010111110000101011111100110
SISWA_84	01111111011110111011101111111111111110
SISWA_85	011010101111101110111111011111111111111
SISWA_86	01100000111111110111011100111110011100
SISWA_87	0110001011000101110011100110100001000001
SISWA_88	011111011111101111101111101011111110001100
SISWA_89	0110101100101101011000101001111000000110
SISWA_90	0100111101101101100000101000010100100101
SISWA_91	110000011110010010100101110111000101
SISWA_92	111111001010010111111111111101110111111
SISWA_93	000000111111010001001000001010101011100
SISWA_94	110100011110010111001001010101010100000
SISWA_95	1101111111101101111101011101111111010001
SISWA_96	11111011111011111111011010111111101110
SISWA_97	01101001011111111110101100111110000011
SISWA_98	11010101111110111011001000110110001000
SISWA_99	111101011110110011011001010110110001000
SISWA_100	011100011110110010000000100100111100101
SISWA_101	111101011110110011011010010110110001100
SISWA_102	11110101111011011111000010111111101000
SISWA_103	0011111110010010000000100000011100000100100

Berdasarkan tabel tersebut analisis kemampuan siswa dengan model Rasch dilakukan dengan menggunakan *software* *Winsteps*.

Berikut ini hasil output analisis *Ministep*:

TABLE 3.1 Analisis Person Fit Soal UAS Akuntansi ZOU167WS.TXT May 15 13:34 2016
INPUT: 103 Person 40 Item REPORTED: 103 Person 40 Item 2 CATS WINSTEPS 3.73

SUMMARY OF 103 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	26.4	40.0	.94	.41	1.00	.1	.98	.1	
S.D.	6.3	.0	1.00	.12	.12	.8	.35	.9	
MAX.	39.0	40.0	4.07	1.03	1.34	2.4	3.56	3.1	
MIN.	12.0	40.0	-.98	.35	.78	-1.6	.28	-1.2	

REAL RMSE	.43	TRUE SD	.90	SEPARATION	2.09	Person RELIABILITY	.81		
MODEL RMSE	.42	TRUE SD	.91	SEPARATION	2.15	Person RELIABILITY	.82		
S.E. OF Person MEAN	= .10								

Person RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82

SUMMARY OF 40 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	67.9	103.0	.00	.25	1.00	.1	.98	.0	
S.D.	16.5	.0	1.01	.07	.11	1.1	.28	1.2	
MAX.	99.0	103.0	1.75	.52	1.34	3.0	2.16	3.6	
MIN.	34.0	103.0	-2.65	.21	.87	-1.5	.56	-1.4	

REAL RMSE	.27	TRUE SD	.97	SEPARATION	3.67	Item RELIABILITY	.93		
MODEL RMSE	.26	TRUE SD	.97	SEPARATION	3.73	Item RELIABILITY	.93		
S.E. OF Item MEAN	= .16								

Gambar 1. Output Summary Statistics

Informasi penting yang dapat diperoleh dari tabel *Summary Statistics* adalah *Person Reliability* dan *Item Reliability*, nilai alpha Cronbach dan *Person Measure*. Dapat diketahui besarnya *Person Reliability* pada UAS Matematika sebesar 0,821 sedangkan *Item Reliability* sebesar 0,93. Besarnya Alpha Cronbach yaitu 0,82. *Person Measure* 0,94 logit menunjukkan rata-rata kemampuan siswa di atas item.

Selanjutnya, berikut ini adalah gambar output dari statistik butir soal:

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFLT /MNSQ	INFIT ZSTD	OUTFIT /MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXACT EXP.	HATCH OBS%	MATCH EXP%	Item
40	34	103	1.75	.23	1.34	3.0	1.63	3.6	.05	.41	64.1	72.3	10040
12	37	103	1.60	.23	.99	-1.1	.81	-1.6	-.41	.68.9	70.8	10041	
15	39	103	1.50	.22	.95	-1.5	.89	-1.9	-.46	.61.9	69.9	10015	
7	46	103	1.16	.22	1.11	1.6	1.22	1.8	.28	65.0	67.6	10007	
27	46	103	1.16	.22	.98	-1.2	.95	-1.4	-.42	.60	67.0	10027	
1	48	103	1.11	.22	1.09	1.3	.89	1.5	-.59	.64.0	70.8	10006	
26	49	103	1.02	.22	.96	-1.5	.98	-2.3	-.43	.60.9	69.7	10026	
24	50	103	.97	.22	.97	-1.3	.97	-2.2	-.42	.40	74.8	10024	
39	52	103	.88	.21	1.11	1.4	1.14	1.2	.29	39.9	61.2	10039	
22	54	103	.79	.21	.95	-1.7	.91	-1.7	-.44	39.9	67.0	10022	
36	58	103	.60	.22	.97	-1.4	.91	-1.7	-.42	38.8	68.0	10036	
1	60	103	.51	.22	1.12	1.6	1.13	1.9	.26	38.0	60.2	10001	
25	61	103	.46	.22	1.21	2.6	1.18	1.3	.19	37.9	61.5	10025	
13	61	103	.46	.22	.97	-1.3	.92	-1.5	-.41	37.9	65.0	10013	
13	62	102	.37	.21	1.06	1.8	1.28	2.1	.28	37.9	69.9	10013	
4	65	103	.27	.22	1.24	2.7	1.39	2.2	.12	36.6	59.2	10004	
35	65	103	.27	.22	1.16	1.9	1.13	1.9	.22	36.6	61.2	10035	
19	66	103	.22	.22	.93	-1.8	.91	-1.5	-.42	36.6	75.7	10019	
23	67	103	.17	.22	1.04	1.5	1.16	1.9	.31	36.6	67.0	10023	
31	69	103	.07	.23	.87	-1.5	.77	-1.4	-.49	35.5	72.8	10031	
34	71	103	.03	.23	.96	-1.4	.89	-1.5	-.39	34.4	70.9	10034	
37	72	103	-.08	.23	.88	-1.2	.85	-1.7	-.45	34.4	78.6	10037	
38	75	103	-.86	.24	.98	-1.7	.98	-2.3	-.34	34.4	86.8	10038	
5	75	103	.25	2.4	1.00	0.9	.98	0.9	.33	33.3	75.3	10005	
17	75	103	.25	2.4	.96	-1.3	.84	-1.7	.38	33.3	71.8	10017	
29	75	103	.25	2.4	.91	-1.0	.84	-1.7	.42	33.3	75.7	10029	
16	76	103	.30	2.4	1.13	1.1	1.13	1.7	.21	32.2	72.8	10016	
20	76	103	.30	2.4	.88	-1.1	.88	-1.5	-.42	32.2	78.6	10020	
2	78	103	-.42	2.4	.88	-1.0	.77	-1.9	-.44	32.2	80.6	10002	
9	78	103	-.42	2.4	1.02	1.2	.98	0.9	.30	32.2	76.7	10009	
3	79	103	-.48	2.5	.92	-1.6	.76	-1.9	-.41	31.1	75.7	10003	
2	80	103	-.54	2.5	.90	-1.7	.78	-1.8	-.41	31.1	82.5	10002	
21	81	103	-.61	2.5	.90	-1.7	.85	-1.5	-.39	30.9	83.5	10021	
10	85	103	-.89	2.7	.96	-2.2	1.16	2.9	.26	28.8	84.5	10010	
18	85	103	-.89	2.7	.94	-2.3	.82	-1.5	-.34	28.8	84.5	10018	
28	85	103	-.89	2.7	.87	-1.8	.71	-1.9	-.41	28.8	84.5	10028	
8	89	103	-1.21	3.0	.95	-2.2	.84	-2.3	.21	25.8	86.4	10008	
14	98	103	-2.41	4.7	1.01	1.2	.75	-1.2	.18	16.9	95.1	10014	
30	98	103	-2.41	4.7	.98	-1.1	.56	-1.6	.24	16.9	95.1	10030	
11	99	103	-2.65	5.2	1.00	1.1	.62	-1.4	.19	14.1	96.1	10011	
<hr/>													
MEAN	67.9	103.0	.00	.25	1.00	1.1	.98	0.0		73.8	74.1		
S.D.	16.5	.0	1.01	.07	.11	1.1	.28	1.2		9.9	9.8		

Gambar 2. Hasil *output item statistics*

Kolom *measure* menunjukkan nilai logit untuk setiap item yang diurutkan dari yang

tertinggi ke yang terendah. Untuk item ke 40 yaitu 1,75 logit menunjukkan butir soal yang tersulit, sedangkan item ke-11 merupakan item yang paling mudah sebesar -2,65 logit. Untuk memeriksa item yang tidak sesuai (*outliers* atau *misfits*) dengan meninjau hal-hal berikut ini.

- Nilai Outfit Mean Square (MNSQ) yang diterima berkisar 0,5 – 1,5
- Nilai Outfit Z-Standard (ZSTD) yang diterima berkisar -2,0 – 2,0
- Nilai Point Measure Correlation (Pt Mean Corr) berkisar 0,4 sampai 0,85

Dari hasil analisis tersebut dapat diketahui bahwa semua item fit, sehingga soal tidak perlu direvisi atau dieliminasi. Di bawah ini adalah gambar output dari skala Guttman dari respon siswa.

GUTTMAN SCALOGRAM OF RESPONSES:
 Person |Item
 |113 11223 12 12333321 3123 32322 2114
 |1408088123296057978413945353162946677520

[illegible]

```

47 +11111101111111111101110101001111100 SISWA_47
67 +1111110111111111111101110101001111100 SISWA_67
85 +1111111101111111111100110101111010011 SISWA_85
3 +1111111111111111110100010111111100010 SISWA_3
16 +111111111111011101111101111111000000 SISWA_16
74 +1111111111111111111100101001101010101 SISWA_74
80 +110111111111111111111101000011100110 SISWA_80
92 +1100111111111111110111011101111100001 SISWA_92
4 +111111011100111011111101110101110100 SISWA_4
14 +1110111111111111010110100011111011010 SISWA_14
21 +1110101111011111111111011010110110101 SISWA_21
27 +11111111111100011110011101101101011110 SISWA_27
88 +11111111111111110111101100001101010 SISWA_88
95 +11111101011111110011011011111101110001 SISWA_95
33 +11111111111101111100111011010000100 SISWA_33
35 +1111111101001111010100111001111101110 SISWA_35
53 +11111111111111110111110011101101000100 SISWA_53
55 +1111111101001111010100111001111101110 SISWA_55
50 +1111110011111111111111000011010000110 SISWA_50
51 +1111111010011111011111100110010101100 SISWA_51
70 +1111110011111111111111000011010000110 SISWA_70
71 +111111110100111110111111100110010101100 SISWA_71
73 +11110111111111111100111110010001010100 SISWA_73
82 +11110111110111110011111110000111011100 SISWA_82
26 +101111101111010101101111111110100001001 SISWA_26
39 +1111111101010011111001100111001110010 SISWA_39
59 +11111111101010011111001100111001110010 SISWA_59
102 +111111111111110110110111110100001100000 SISWA_102
10 +11101101111110111100100111111010000100 SISWA_10
11 +11111111111110010011110001011101001001 SISWA_11
19 +111111111111101100110001011111000011010 SISWA_19
83 +11111111111111001110101101000101010000 SISWA_83
12 +111111001111010011100101100111110101000 SISWA_12
15 +1111111100111100111111101011110100010001 SISWA_15
28 +111101101101100110101110101101001101001 SISWA_28
30 +0111110000111101011111110011001110010 SISWA_30
45 +11110011110111111111111010111101000000000 SISWA_45
65 +111100111101111111110101111101000000000 SISWA_65
81 +1111111111110111111010110010000001001010 SISWA_81
86 +110110111101101111010000111011011000110 SISWA_86
97 +111111011101111100010100110001110000111 SISWA_97
5 +110111111111010010000111110001100101010 SISWA_5
17 +111011111100110110110110001100010000111 SISWA_17
38 +111111011101010111101111000010110010000 SISWA_38
46 +111011110101111011111101001101000000001 SISWA_46
58 +111110111010101111011111000010110010000 SISWA_58
66 +111011110101111011111101001101000000001 SISWA_66
79 +11000111001111111101111100000100101010 SISWA_79
101 +1111111111010111100101010100001100000 SISWA_101
6 +111110101111011001100101101111010000100 SISWA_6
18 +111111111111011110101010001000001100010 SISWA_18
22 +111110111110010101101100011111000001000 SISWA_22
75 +111111010110110011000001111010100101001 SISWA_75
98 +111111111011110111000001010100010100010 SISWA_98
99 +11111111111010111000001010100011100000 SISWA_99
34 +1011111101011110110011110011001100000000 SISWA_34
54 +1011111101011110110011110011001100000000 SISWA_54
49 +1111111100110000111010110100100000001100 SISWA_49
69 +1111111100110000111010110100100000001100 SISWA_69
72 +110011111011101001100010110010110010000 SISWA_72
91 +1111101010110101101101100011100001000001 SISWA_91
32 +1100011101011110011101100010110000000001 SISWA_32
36 +1111111100111010011001010100110000001000 SISWA_36
52 +110001110101111001110110010110000000001 SISWA_52
56 +1111111100111010011001010100110000001000 SISWA_56
8 +111010111100010100111001000101001010000 SISWA_8
89 +111011001101010101011100110000100010000 SISWA_89
94 +1111111100111001000010011001100011000000 SISWA_94
9 +0100101101011010101100011010110100000001 SISWA_9
23 +0110011001011110101000010111100000001001 SISWA_23
90 +111100010101011001001001110000000110001 SISWA_90
100 +1101100011110001101100011101000001000001 SISWA_100
48 +1011011010001111000000101010000110010001 SISWA_48
68 +1011011010001111000000101010000110010001 SISWA_68

```

```

87 +0100110101111001100101000000001001011001 SISWA_87
93 +1101110110010000111110000000010000011010 SISWA_93
42 +1111110001010000000101001100000110000000 SISWA_42
62 +1111110001010000000101001100000110000000 SISWA_62
103 +1111000001000010110010010000001000110001 SISWA_103
41 +1110000000001001000000011011100100010000 SISWA_41
61 +1110000000001001000000011011100100010000 SISWA_61
-----
|113 11223 12 12333321 3123 32322 2114
|1408088123296057978413945353162946677520

```

Gambar 3. Hasil *output Guttman Scalogram of responses*

Analisis dari tabel *person measure* mendapatkan bahwa person-1 tidak sesuai (misfit) dikarenakan dia tidak mampu mengerjakan soal nomor 5, padahal soal nomor 5 masih tergolong soal yang mudah, sedangkan soal dengan tingkat kesukaran tinggi, ia bisa mengerjakan (soal nomor 40). Selain siswa-1, siswa-31 juga mengalami misfit, hal tersebut dikarenakan dia kurang teliti dalam mengerjakan soal nomor 15 dan 26.

V. KESIMPULAN

Berdasarkan pembahasan di atas dapat disimpulkan bahwa Soal Ujian Akhir Semester Mata Pelajaran Akuntansi dengan bantuan *software Winsteps* menunjukkan bahwa *Person Reliability* pada UAS Akuntansi sebesar 0,81 sedangkan *Item Reliability* sebesar 0,93. Besarnya Alpha Cronbach yaitu 0,82. *Person Measure* 0,94 logit menunjukkan rata-rata kemampuan siswa di atas item. Hasil analisis juga menunjukkan bahwa semua item fit, sehingga soal tidak perlu direvisi atau dieliminasi. Informasi lain yang dapat diperoleh yaitu

nilai logit person, untuk person-1 dengan +4,07 logit menunjukkan person dengan kemampuan tertinggi (mampu mengerjakan hampir semua soal); person-61 dengan -0,98 menunjukkan siswa dengan kemampuan paling rendah (paling sedikit dalam menyelesaikan soal dengan benar). Untuk person 1 terjadi misfit karena $MNSQ > 1,5$ dan PT Measure corr $< 0,4$ hal tersebut dikarenakan dia tidak berhasil menjawab soal mudah dengan benar, namun dapat menjawab dengan benar pada soal sulit. Selain itu, Siswa-31 juga mengalami misfit, dimana $MNSQ < 0,5$ dan PT Measure corr $< 0,4$ hal tersebut dikarenakan dia kurang teliti dalam mengerjakan soal nomor 15 dan 26. Berdasarkan hasil penelitian, dapat disimpulkan bahwa sebagian besar soal UAS SMK Program Keahlian Akuntansi Keuangan memiliki pola item yang fit, sehingga diasumsikan bahwa dalam UAS Akuntansi kali ini secara keseluruhan, instrumen mampu mengukur apa yang menjadi tujuan ukur.

DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. 1979. *Introduction to measurement theory*, Monterey, California: Brooks/Cole Publishing Company.
- Azwar, S. 2015. *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Christensen, K.B & Kreiner, S. 2010. Monte Carlo tests of the Rasch model based on scalability coefficients. *British Journal of Mathematical and Statistical Psychology*, 63(1), 101-111.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamental of item response theory*. Newbury Park: Sage Publications Inc.
- Koller, Ingrid, & Hatzinger, R. 2013. Nonparametric tests for the Rasch model: explanation, development, and application of quasi-exact tests for small samples. *Interstat*, 11, 1-16.
- MacCann, Robert, G., & Stanley, G. 2008. The use of Rasch modeling to improve standard setting. *Journal of Practical Assessment Research & Evaluation*, 11(2).
- Mardapi, D. 2012. *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Litera
- Nopiah, Z.M., Jamalluddin, M.H., Othman, N.A., Ismail, N.A., Asshaari, I., & Osman, M.H. Reliability analisis of examination questions in a mathematics course using rasch measurement model. *Journal of Sains Malaysiana*. 2012 Sep 1;41(9):1171-1176.
- Purwanto. 2011. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar
- Ratnaningsih. 2013. Analisis Butir Tes Objektif UAS Mahasiswa UT

- Berdasarkan Teori Tes Modern. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, 14(2), 98-109.
- Retnawati, H. 2014. *Teori Respons Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Parama Publishing.
- Sax, G. 1980. *Principles of Educational and Psychological Measurement and Evaluation*. Belmont, CA: Wadsworth.
- Sumintono, B & Widhiarso, W. 2012. Aplikasi model rasch untuk ilmu-ilmu sosial. Cimahi: Trimkom Publishing House.
- Suwarto. 2013. *Pengembangan Tes Diagnostik dalam Pembelajaran: Panduan Praktis bagi Pendidik dan Calon Pendidik*. Yogyakarta: Pustaka Pelajar.
- Suwarto. 2011. Teori Tes Klasik dan Teori Tes Modern. *Jurnal Widyatama* 1(20), 69-78.