



**ANALISIS TINGKAT KESESUAIAN BUTIR SOAL (ITEM FIT)
MENGGUNAKAN 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 berulangkali, 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 menuhi 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 menuhi 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 (Suwarto, 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, Swaminathan, & 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 invariant. 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 (Suwarto, 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_i adalah tingkat kesulitan butir i , dan \exp adalah bilangan transendental 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

bebas dari probabilitas setiap item lain yang diberikan sifat laten; 2) undimensionalitas atau homogenitas item berarti bahwa secara umum semua item dari tes mengukur konstruk laten yang sama; 3) paralel dan ketat meningkatkan kurva karakteristik yang berarti bahwa dengan meningkatkan kemampuan seseorang, probabilitas untuk memecahkan item juga meningkat. 4) objektivitas tertentu menyiratkan bahwa suatu item tidak akan relevan apabila item tersebut digunakan untuk membandingkan dua orang dengan kemampuan yang berbeda. Item juga tidak akan relevan jika orang tersebut diuji dengan tingkat kesulitan tes yang berbeda (Ingrid Koller, 2013:2)

III. METODE PENELITIAN

Secara umum, penelitian ini menggunakan pendekatan kuantitatif. Penelitian ini dilaksanakan di SMK Muhammadiyah Wonosari pada Semester Ganjil Tahun Pelajaran 2015 / 2016. Penelitian ini difokuskan pada Soal Ujian Akhir Sekolah Mata Pelajaran Akuntansi yang berjumlah 40 butir soal. Soal Akuntansi Ujian Akhir Sekolah tersebut berbentuk tes objektif pilihan ganda dengan opsi sebanyak lima pilihan jawaban yang dikembangkan oleh Tim KKG yang ditunjuk oleh pihak yang

berwenang. Teknik pengumpulan data pada penelitian ini adalah teknik dokumentasi. Untuk menerapkan metode *item fit*, maka dibutuhkan pola respon peserta ujian yang berasal dari lembar jawaban peserta ujian. Pada penelitian ini lembar jawaban peserta ujian diperoleh sebanyak 103 lembar jawaban peserta ujian. Data penelitian yang telah dijaring kemudian dianalisis dengan metode *item fit* menggunakan model Rasch untuk menganalisis kemampuan siswa. Untuk mendukung pelaksanaan analisis, digunakan sebuah *software* yaitu Winsteps.

IV. HASIL PENELITIAN

Penelitian ini menganalisis dengan model Rasch menggunakan data Ujian Akhir Semester (UAS) Akuntansi di SMK Muhammadiyah Wonosari Program Keahlian Akuntansi untuk menguji instrumen lembar jawaban UAS Akuntansi Kelas XI. Soal UAS yang dibuat adalah soal pilihan ganda (dikotomi) dengan jumlah 40 soal dan jumlah siswa Program Keahlian Akuntansi kelas XI adalah 103 siswa. Berikut ini data hasil koreksi guru Akuntansi terhadap jawaban siswa.

Tabel 1. Respon siswa pada soal UAS
Akuntansi

YAYASAN AKRAB PEKANBARU
Jurnal AKRAB JUARA

SISWA_74 01101011111010111111111100101111111001111
SISWA_75 010011011110101000101000010110101101111111
SISWA_76 111110111111101110110110011111111111101
SISWA_77 1110101111111011101110111111111110111101
SISWA_78 0111111111101011111111111111111111111111110
SISWA_79 00011100101111011011101000111111001011110
SISWA_80 01111100111111111111111011010111111111101100
SISWA_81 0111110011111110100111110001111111100001100
SISWA_82 011110101101101111110110110111011011001110
SISWA_83 01101011111101011111100001010111111100110
SISWA_84 011111110111110111011110111111111111111110
SISWA_85 011010110111111011101111101011111111111111
SISWA_86 01100000011111111011101110011111110011100
SISWA_87 0110001011000101110011001100001000001
SISWA_88 0111110111111101111110101111111110001100
SISWA_89 0110101100010110101010000101001111000000110
SISWA_90 01001111011011011000001010000010100100101
SISWA_91 1100000011111100101100101101110111000101
SISWA_92 11111100101010111111111010110111011111111
SISWA_93 000000011111110000100100000010101011111100
SISWA_94 11010001111001011100100101010101010100000
SISWA_95 1101111111101011111010111011111111010001
SISWA_96 111110111111011111111110101101111111101110
SISWA_97 01101001011111111111010101001111100000011
SISWA_98 1101010111111101101100100011111110010001000
SISWA_99 11110101111101100110110010101110110001000
SISWA_100 01110001111101100100000000100100111100101
SISWA_101 111110101111011001101100101101110110001100
SISWA_102 11111010111101101111110000101111111101000
SISWA_103 001111110010010000000100000011100000100100

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	MNSQ	ZSTD	MNSQ	ZSTD	OUTFIT
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		

Person RAW SCORE-TO-MEASURE CORRELATION = .97
CROMBACH 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	
MAX.	99.0	103.0		1.75	.52	1.34	3.0	2.16	
MIN.	34.0	103.0		-2.65	.21	.87	-1.5	.56	
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	n = 16								

Gambar 1. Output Summary Statistics

YAYASAN AKRAB PEKANBARU
Jurnal AKRAB JUARA

```
87 +0100110101111001100101000000001001011001 SISWA_87
93 +1101110110010000111100000000010000011010 SISWA_93
42 +1111110001010000000101001100000110000000 SISWA_42
62 +1111110001010000000101001100000110000000 SISWA_62
103 +111000000000000101000000010100000001000110001 SISWA_103
41 +1110000000000001001000000010111001000010000 SISWA_41
61 +11100000000000010010000000010111001000010000 SISWA_61
|-----|
[113 1123 12 12333321 3123 32322 2114
[1408088123296057978413945353162946677520
```

Gambar 3. Hasil *output Guttman Scalogram of responses*

Analisis dari tabel *person measure* mendapati 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

YAYASAN AKRAB PEKANBARU
Jurnal AKRAB JUARA
Volume 6 Nomor 1 Edisi Februari 2021 (180-193)

- 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 raschuntuk 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.