

ANALISIS KUALITAS BUTIR SOAL UJIAN SATUAN PENDIDIKAN (USP) MATERI C2 TEKNIK PEMESINAN KELAS XII DI SMK PGRI 1 LAMONGAN MELALUI MODEL RASCH

Achmad Faizin Ramadhan

S1 Pendidikan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya
Email : achmadramadhan16050524016@mhs.unesa.ac.id

Rachmad Syarifudin Hidayatullah

Email : rachmadhidayatullah@unesa.ac.id

Abstrak

Tes merupakan alat evaluasi bagi guru sebagai informasi terkait kemampuan peserta didik. Dalam memberikan informasi secara tepat maka kualitas tes yang diberikan harus baik. Namun kebanyakan instrumen tes masih belum diketahui kualitasnya, sehingga perlu dilakukan analisis butir soal. Penelitian ini bertujuan untuk mengetahui kualitas soal USP Materi C2 Teknik Pemesinan Kelas XII dilihat dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda melalui model rasch. Jenis penelitian merupakan penelitian deskriptif kuantitatif. Soal yang diberikan berupa 25 butir soal pilihan ganda dan diujikan kepada 36 siswa. Teknik pengumpulan data menggunakan metode observasi dan tes untuk selanjutnya dianalisis melalui permodelan rasch menggunakan program Winstep versi 3.73. Berdasarkan hasil penelitian menunjukkan bahwa: 1) Validitas butir soal diperoleh 20 soal dinyatakan valid dan 5 soal yang tidak valid. 2) Realibilitas pada alpha cornbach menunjukkan bahwa instrumen memiliki kategori yang jelek, terdapat 22 soal (non extreme) dengan reliabilitas item yang bagus sedangkan untuk keseluruhan dari butir soal termasuk kategori cukup. 3) Tingkat kesukaran menunjukkan soal yang layak digunakan berjumlah 14 soal sedangkan untuk butir soal yang memiliki tingkat kesukaran paling tinggi yaitu sebanyak 4 soal dan tingkat kesukaran paling rendah sebanyak 2 soal. Terdapat juga kategori outlier sebanyak 4 soal. 4) Daya beda dilihat dari nilai standar error (S.E) bahwa soal dengan daya beda yang baik sebanyak 19 soal, 3 soal dengan daya beda yang cukup, dan 3 soal dengan daya beda yang buruk. Nilai separasi keseluruhan item butir soal bernilai baik karena membagi soal dalam 3 kelompok yaitu tinggi, sedang, dan rendah.

Kata Kunci: analisis butir soal, model rasch, kualitas soal

Abstract

A test is an evaluation tool for teachers as information related to the ability of students. In providing precise information, the quality of the test given must be good. However, most test instruments are still of unknown quality, so it is necessary to analyze the items. This study aims to determine the quality of USP questions C2 Material of Class XII Machining Techniques seen from the aspects of validity, reliability, difficulty level, and differentiability through the Rasch model. This type of research is quantitative descriptive research. The questions given were 25 multiple choice questions and tested on 36 students. Data collection techniques using observation and test methods and then analyzed through Rasch modeling using the Winstep version 3.73 program. Based on the results of the study showed that: 1) The validity of the items obtained was 20 questions declared valid and 5 invalid questions. 2) Reliability on alpha cornbach shows that the instrument has a bad category, there are 22 questions (non-extreme) with good item reliability while for the whole of the items including the sufficient category. 3) The level of difficulty shows that the questions that are feasible to use are 14 questions, while the items that have the highest level of difficulty are 4 questions and the lowest level of difficulty is 2 questions. There are also 4 outlier categories. 4) Distinguishing power seen from the standard error (S.E) value that questions with good differentiation are 19 questions, 3 questions with sufficient differentiation, and 3 questions with poor differentiation. The overall separation value of the items is good because it divides the questions into 3 groups, namely high, medium, and low.

Keywords: item analysis, rasch model, item quality

PENDAHULUAN

Kegiatan edukatif tidak dapat dipisahkan dari aktivitas belajar dan pembelajaran (Pane & Dasopang, 2017:333).

Menurut (Rachman & Fahdiansyah, 2018) variabel yang mempengaruhi pembelajaran yaitu metode, kondisi dan hasil belajar. Menurut (Muhammad Ali, 1987: 30) komponen pembelajaran mencakup tujuan

pembelajaran, materi atau bahan ajar, metode dalam pengajaran, media dan alat yang digunakan, serta pelaksanaan evaluasi. Seorang pendidik memiliki pengaruh untuk mengoptimalkan sistem pengajaran termasuk keseluruhan komponen di dalamnya (Rindawan & Syahrir, 2019). Menurut (Anggraini dkk., 2021). Peran seorang pendidik mempengaruhi keberhasilan pembelajaran terutama pada aktivitas belajar mengajar.

Hasil belajar merupakan indikasi keberhasilan tercapainya proses pembelajaran (Nuryanti dkk., 2018). Hasil belajar pada peserta didik merupakan gambaran kemampuan pemahaman dalam kegiatan proses belajar yang diperlihatkan dalam bentuk nilai sebagai indikator keberhasilan. Untuk mengetahui hasil belajar maka perlu dilakukan kegiatan penilaian biasa disebut dengan kegiatan evaluasi. Menurut (Widyaningsih & Yusuf, 2018) perlu dilakukan evaluasi untuk menilai ketercapaian pembelajaran yang ditentukan oleh kurikulum. Tujuan evaluasi adalah untuk melihat apakah tujuan pada sistem pembelajaran yang diberikan sudah efektif dan efisien. Menurut (Idrus, 2019) Evaluasi dibentuk sebagai pendorong pendidik untuk memaksimalkan kualitas proses pembelajaran yang direncanakan. (Pratiwi dkk., 2021) mengatakan bahwa kegiatan evaluasi yang baik dapat mengukur kualitas pendidikan dan kemampuan siswa yang sebenar - benarnya. Seorang guru yang berkompeten diharuskan memiliki kompetensi untuk memahami teknik dan prosedur evaluasi pembelajaran (Riyadi, 2017). Menurut Riyadi (2017) Evaluasi dimulai dari pelaksanaan perencanaan, pembuatan soal tes, menganalisis hasil tes, serta menindaklanjuti hasil evaluasi yang didapatkan.

Secara umum, terdapat dua macam kategori alat evaluasi yaitu tes dan nontes. Biasanya, untuk mengukur kemampuan peserta didik umumnya guru sering menggunakan jenis instrumen tes. (Hariyanto & Basuki, 2014) mengatakan bahwa tes sendiri merupakan alat yang digunakan untuk menilai atau metode penilaian yang sistematis, faktual, serta dapat dipercaya dalam menetapkan tingkat pengetahuan dan keterampilan peserta didik terhadap materi. Sehingga dari tes tersebut diharapkan mampu mendapatkan informasi mengenai kemampuan siswa dan keberhasilan teknik pembelajaran yang diterapkan oleh guru tersebut.

Soal merupakan cara yang sering digunakan oleh para guru untuk mengukur kemampuan siswa melalui tes, sehingga soal yang diberikan harus baik kualitasnya. Soal yang berkualitas ialah soal yang mampu memberikan sebuah informasi secara tepat sesuai dengan tujuan yang direncanakan. Terdapat kriteria untuk menjadikan instrumen tes soal dapat dikatakan baik atau berkualitas, menurut (Erfan dkk., 2020) kriteria tersebut diantaranya adalah validitas yang sesuai, reliabilitas yang bagus,

mempunyai tingkat kesukaran soal yang bervariasi, dan juga memiliki daya beda pada butir soal yang dapat membedakan kemampuan peserta didik dalam menjawab soal.

Sebagai karakteristik pertama soal yang berkualitas, menurut (Kurniawan, 2015) validitas ialah menentukan keakuratan dari keseluruhan tes dalam menjalankan fungsi ukurnya. Kedua adalah Realibilitas, menurut (Gunawan, 2018 : 112) reliabilitas merupakan ukuran konsistensi peserta didik dalam menjawab soal tes. Selanjutnya adalah tingkat kesukaran, menurut (Rahayu dan Djazari, 2016) tingkat kesukaran adalah perbandingan jumlah seluruh peserta tes terhadap peserta didik yang dapat menjawab dengan benar. Keempat ialah daya beda, menurut (Muluki, 2020) daya beda merupakan kemampuan butir soal untuk membedakan siswa yang mampu dan kurang mampu dalam memahami materi. Sehingga untuk mendapatkan instrumen tes soal yang berkualitas maka perlu dilakukan pengujian butir soal dari segi validitas, reliabilitas, tingkat kesukaran, serta daya beda butir soal (Perdana, 2018).

Namun pada kenyataannya, menurut (Fauziana dan Dessy Wulansari, 2021) pada kebanyakan instrumen tes yang digunakan oleh para guru masih belum diketahui kualitasnya karena jarang dilakukan analisis kualitas soal, sehingga penilaian yang didapatkan hasilnya masih semu.. Menurut (Pohan dan Dafit, 2021) kelemahan guru disebabkan oleh kurangnya pelatihan mengenai pembelajaran sesuai kurikulum yang didapatkan. Kegiatan analisis butir soal yang jarang dilakukan menyebabkan aspek yang mempengaruhi kualitas soal sering dianggap rendah.

Berdasarkan hasil observasi melalui wawancara di SMK PGRI 1 Lamongan, didapatkan informasi bahwa pelaksanaan evaluasi yang dilakukan selama ini masih belum memperhatikan kualitas butir soal yang diberikan, sehingga belum diketahui apakah kualitas butir soal sudah sesuai dengan tujuan. Upaya yang digunakan guru untuk mengetahui kualitas soal baik atau buruk masih dengan cara yang sederhana, yaitu dengan melihat perbandingan jawaban benar dan salah dari keseluruhan siswa yang menjawab soal tersebut. Sehingga mengakibatkan tidak terukurnya kemampuan siswa yang sebenarnya. Sedangkan instrumen penilaian yang baik, harus memiliki kualitas butir soal yang bisa mengukur tingkat kemampuan siswa yang sebenarnya. Menurut (Ariyana, 2011) Salah satu faktor yang mempengaruhi kualitas soal adalah guru yang menyusun soal. Disini analisis butir soal perlu dilakukan demi memperoleh informasi mengenai kualitas butir soal tersebut, sehingga kesalahan pada butir soal yang diketahui bisa diperbaiki. Berdasarkan dari permasalahan tersebut maka perlu dilakukan analisis butir soal.

Terdapat dua pendekatan mengenai tes yang biasa digunakan untuk melakukan analisis butir soal yaitu pendekatan secara klasik / CTT (*Classical Test Theory*) serta pendekatan secara modern / IRT (*Item Response Theory*). Pendekatan modern IRT adalah cara yang efektif digunakan pada saat ini. (Fisher, 1993) mengatakan bahwa permodelan rasch lebih efektif untuk diterapkan daripada model analisis klasik. Menurut (Fajrianti dkk., 2016) Hal tersebut terjadi karena pada IRT pengukuran kemampuan peserta lebih akurat, begitu juga dalam mengevaluasi butir soal. Salah satu model IRT ialah permodelan rasch atau Rasch Model, menurut (Rasyid dkk., 2007) model rasch adalah konsep penilaian modern yang dapat digunakan untuk mengklasifikasi perhitungan item dan person dalam bentuk peta distribusi. Menurut (Stenbeck dkk., 1992 : 7) Model rasch didasarkan pada dua prinsip yaitu kemampuan peserta didik atau subjek terhadap butir soal serta hubungan antara kemampuan peserta pada suatu pertanyaan terhadap kemampuan lain. Menurut (Aziz, 2015) Permodelan rasch lebih unggul dari teori tes klasik dilihat dari kemampuan prediksi terhadap data yang hilang berdasarkan pada pola respon yang sistematis.

Instrument tes yang baik ialah instrumen yang mampu memberi informasi yang tepat terkait kemampuan peserta didik atas keterampilan yang diujikan (Azizah & Wahyuningsih, 2020). Berdasarkan penelitian yang dilakukan oleh (Erfan dkk., 2020) terkait “Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri Dan Paralel Melalui Teori Tes Klasik Dan Model Rasch” melalui penelitian yang dilakukan secara deskriptif kualitatif pada Model Rasch dengan bantuan software Winstep didapatkan hasil analisis instrumen tes yang mencakup aspek validitas, reabilitas, tingkat kesukaran soal, serta daya beda soal pada butir soal yang diujikan sehingga dapat diketahui butir soal yang tidak sesuai dengan kriteria kualitas butir soal untuk selanjutnya bisa dilakukan perbaikan pada instrumen tes.

Berdasarkan pada penelitian tersebut, diketahui bahwa penerapan permodelan rasch dapat digunakan untuk mengetahui kualitas butir soal sehingga dapat dijadikan sebagai dasar pertimbangan untuk mengadakan perbaikan soal. Menurut (Ida & Musyarofah, 2021) seorang guru harus melakukan kegiatan menganalisis butir soal untuk meningkatkan kualitas butir soal yang telah disusun. Melalui kegiatan analisis butir soal dapat diperoleh informasi tentang kekurangan suatu soal dan petunjuk untuk melakukan perbaikan (Pratiwi dkk., 2021). Pada penelitian ini, peneliti bertujuan untuk melakukan analisis kualitas instrumen tes untuk mengetahui kemampuan hasil belajar siswa Teknik Pemesinan yaitu pada soal Ujian Satuan Pendidikan (USP) materi C2 Teknik Pemesinan kelas XII TPM di SMK PGRI 1 Lamongan tahun

pelajaran 2022/2023 menggunakan model Rasch dengan bantuan software Winstep. Menurut (Mair & Hatzinger, 2007) software komputer mempermudah dalam melakukan analisis secara langsung.

Soal Ujian Satuan Pendidikan (USP) yang digunakan adalah tes pilihan ganda yang termasuk dalam jenis tes sumatif. Menurut (Widoyoko, 2014) manfaat dari tes sumatif adalah untuk memperoleh informasi tentang parameter kemampuan siswa dalam pemahaman suatu pembelajaran. Dalam hal ini informasi yang didapatkan melalui evaluasi pengetahuan dan keterampilan peserta didik kelas XII TPM SMK PGRI 1 Lamongan terhadap materi C2 Teknik Pemesinan tahun 2022/ 2023. Sehingga sebagai bentuk evaluasi pembelajaran maka peneliti bertujuan untuk melakukan penelitian mengenai “Analisis Kualitas Butir Soal Ujian Satuan Pendidikan (USP) Materi C2 Teknik Pemesinan Kelas XII di SMK PGRI 1 Lamongan Melalui Model Rasch”.

METODE

Desain Penelitian

Penelitian ini menggunakan metode penelitian deskriptif kuantitatif dengan menggunakan data sekunder instrumen soal Ujian Satuan Pendidikan (USP) Materi C2 Teknik Pemesinan Kelas XII. Item soal yang diberikan berupa 25 butir soal pilihan ganda dengan 5 opsi jawaban (a,b,c,d, dan e) yang diujikan melalui Google Form. Subjek dalam penelitian ini adalah 36 siswa kelas XII Teknik Pemesinan SMK PGRI 1 Lamongan Tahun Ajaran 2022-2023.

Tempat dan Waktu Penelitian

Penelitian dilaksanakan di sekolah SMK PGRI 1 Lamongan pada bulan maret 2023 dimulai dengan observasi dan pengujian instrumen tes di kelas XII Teknik Pemesinan SMK PGRI 1 Lamongan

Variabel Penelitian

Pada penelitian ini digunakan pendekatan analisis kualitas soal dengan variabel berupa kualitas butir soal Ujian Satuan Pendidikan (USP) Materi C2 Teknik Pemesinan Kelas XII dengan indikator untuk menentukan kualitas instrumen soal yang diberikan yaitu validitas, reliabilitas, tingkat kesukaran, serta daya beda soal.

Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini yaitu dengan menggunakan dua teknik berupa instrumen tes dan observasi. Pengambilan data melalui tes untuk mencari kualitas soal dari aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda serta teknik pengambilan data secara observasi melalui lembar telaah untuk mencari validitas teoritis soal meliputi aspek materi, konstruksi, dan bahasa.

Teknik Analisis Data

Pengambilan data validitas teoritis menggunakan metode observasi berupa lembar telaah butir soal sesuai dengan kaidah penulisan soal yang divalidasi oleh validator ahli. Selanjutnya hasil jawaban tes dianalisis secara kuantitatif deskriptif menggunakan permodelan rasch dengan bantuan software Winstep versi 3.73.

HASIL DAN PEMBAHASAN

Validitas

Hasil analisis validitas butir soal yang dengan menggunakan model Rasch melalui program winstep didapatkan hasil berupa output *Misfit Order* yang dapat dilihat pada gambar berikut :

OUTFIT	PT-MEASURE	EXACT MATCH			Item	
MNSQ	ZSTD	CORR.	EXP.	OBS%		
1.62	2.5	A-.22	.35	50.0	69.6	Soal 25
1.56	3.0	B-.12	.37	47.2	66.7	Soal 6
1.40	1.6	C-.01	.34	61.1	71.3	Soal 22
1.39	.8	D .07	.23	88.9	88.8	Soal 20
1.34	.9	E .07	.29	80.6	80.5	Soal 8
1.22	.7	F .13	.29	80.6	80.5	Soal 2
1.18	.7	G .12	.33	63.9	73.1	Soal 15
1.09	.5	H .20	.35	63.9	70.1	Soal 19
1.06	.3	I .21	.28	83.3	83.2	Soal 3
1.02	.1	J .39	.34	77.8	71.3	Soal 17
.88	-.1	K .31	.28	83.3	83.2	Soal 14
.93	-.2	k .44	.33	75.0	73.1	Soal 23
.59	-.6	j .39	.23	88.9	88.8	Soal 5
-.43	-.5	i .35	.17	94.4	94.4	Soal 12
.84	-.8	h .50	.36	75.0	69.0	Soal 4
.78	-1.3	g .57	.36	77.8	68.0	Soal 16
.74	-1.4	f .58	.36	75.0	69.0	Soal 11
.72	-1.2	e .58	.34	83.3	71.5	Soal 18
.55	-1.1	d .55	.28	83.3	83.2	Soal 7
.70	-1.7	c .62	.36	75.0	69.0	Soal 13
.71	-2.1	b .66	.37	80.6	66.7	Soal 9
.63	-1.5	a .64	.33	86.1	73.1	Soal 21
.97	-.1			76.1	75.6	
.33	1.3			11.8	8.0	

Gambar 1. Output Misfit Order

Pada uji validitas yang dilihat adalah berdasarkan tingkat kesesuaian (*Item fit*). Tingkat kesesuaian memperlihatkan butir soal yang digunakan dapat berfungsi sesuai dengan fungsi ukurnya. Analisis ketepatan item fit dapat dilakukan dengan melihat nilai MNSQ atau Outfit Mean Square, ZSTD atau Outfit Z-Standard, dan Pt Mean Corr atau Point Measure Correlation. Butir soal yang sesuai dengan pengukuran dapat diterima sebagai instrumen dan dapat digunakan sedangkan untuk butir soal yang tidak sesuai (misfit), maka butir soal berpotensi tidak dapat digunakan. Menurut (Sumintono dan Widhiarso, 2015) kriteria nilai outfit dapat ditentukan dari

- a) Nilai Outfit MNSQ : $0,5 < \text{Outfit} - \text{MNSQ} < 1,5$; 2).
- b) Nilai Outfit ZSTD) : $-2,0 < \text{ZSTD} < +2,0$;
- c) Nilai Pt Measure Corr : $0,4 < \text{Point Measure Corr} < 0,85$

Tahap pertama adalah menganalisis nilai MNSQ untuk mengevaluasi kesesuaian, nilai ZSTD diabaikan selama

nilai MNSQ berada dalam rentang kesesuaian yang diterima (Boone dkk., 2014). Jika terdapat butir soal yang memiliki nilai outfit MNSQ diluar nilai yang diterima maka nilai ZSTD perlu dianalisis. Tahap kedua yaitu menganalisis nilai outfit ZSTD dengan melihat nilai yang sesuai dengan rentang yang diterima (Sumintono dan Widhiarso, 2015) . Jika nilai ZSTD berada diluar rentang yang diterima maka soal tersebut harus diganti dengan yang baru atau dibuang. Namun jika nilai Namun jika nilai Outfit MNSQ dan PT Measure Corr tidak memenuhi rentang yang diterima sedangkan nilai Outfit ZSTD masih memenuhi maka soal masih bisa diterima namun perlu diperiksa. Hasil analisis item fit butir soal dapat dilihat pada tabel 1 dengan nilai yang tidak sesuai dicetak tebal miring.

Tabel 1. Hasil analisis validitas butir soal terhadap model rasch

No	Nomor Butir Soal	Kriteria Ketepatan Pengukuran			Keputusan
		Outfit MNSQ	Outfit ZSTD	PT-MEAN CORR	
1	S-25	1,62	2,5	-0,22	Tidak Valid
2	S-6	1,56	3,0	-0,12	Tidak Valid
3	S-22	1,40	1,6	-0,01	Valid
4	S-20	1,39	0,8	0,07	Valid
5	S-8	1,34	0,9	0,07	Valid
6	S-2	1,22	0,7	0,13	Valid
7	S-15	1,18	0,7	0,12	Valid
8	S-19	1,09	0,5	0,20	Valid
9	S-3	1,06	0,3	0,21	Valid
10	S-17	1,02	0,1	0,39	Valid
11	S-14	0,88	-0,1	0,31	Valid
12	S-23	0,93	-0,2	0,44	Valid
13	S-5	0,59	-0,6	0,39	Valid
14	S-12	0,43	-0,5	0,35	Valid
15	S-4	0,84	-0,8	0,50	Valid
16	S-16	0,78	-1,3	0,57	Valid
17	S-11	0,74	-1,4	0,58	Valid
18	S-18	0,72	-1,2	0,58	Valid
19	S-7	0,55	-1,1	0,55	Valid
20	S-13	0,70	-1,7	0,62	Valid
21	S-9	0,71	-2,1	0,66	Valid
22	S-21	0,63	-1,5	0,64	Valid
23	S-1		0,00		Tidak Valid / Tidak Fit
24	S-10		0,00		
25	S-24		0,00		

Tabel diatas memperlihatkan hasil analisis validitas kesesuaian dari 25 butir soal diperoleh sebanyak 20 soal yang dinyatakan valid (fit) dan 5 soal dinyatakan tidak valid (misfit) , diantara 5 soal tersebut soal nomor 6 dan 25 tidak memenuhi kriteria *Outfit* MNSQ dan *Outfit* ZSTD sedangkan soal nomor 1, 10, dan 24 tidak memiliki nilai pengukuran dikarenakan validitas butir soal yang tidak sesuai karena diluar batas / outlier yang berarti soal tersebut sangat tidak valid dan perlu diganti dengan soal yang baru. Sehingga dapat disimpulkan apabila soal sudah dinyatakan valid (fit) maka pengujian melalui butir – butir soal tersebut sudah memenuhi kriteria serta terjamin kualitasnya untuk digunakan untuk mengukur tingkat pemahaman peserta didik.

Reliabilitas

Reliabilitas memberikan informasi apakah suatu tes dapat dipercaya sesuai dengan ketentuan. Data tersebut dilihat dari hasil analisis output summary statistics yang memberikan informasi keseluruhan mengenai kualitas pola respons, kualitas instrumen tes, serta interaksi antara responden dan butir soal . Berikut merupakan keluaran summary statistics :

TABLE 3.1 Analisis Soal C2 Teknik Pemesinan ZOU512WS.TXT May 28 17:08 2021
 INPUT: 36 Person 25 Item REPORTED: 36 Person 25 Item 2 CATS WINSTEPS 3.7

```

SUMMARY OF 36 MEASURED Person
-----
TOTAL SCORE   COUNT   MEASURE   MODEL   INFIT   OUTFIT
SCORE          COUNT   ERROR     ERROR   MNSQ    ZSTD   MNSQ    ZSTD
-----
MEAN          17.8    25.0      .95     .54     .98     .0   .97   -.1
S.D.          3.1     .0        .82     .07     .24     1.0  .38   .9
MAX.          23.0    25.0      2.73    .78     1.56    2.5  1.69  2.1
MIN.          12.0    25.0      -.46    .47     .52     -2.7 .30  -2.1
-----
REAL RMSE     .56 TRUE SD .60 SEPARATION 1.07 Item RELIABILITY .53
MODEL RMSE    .54 TRUE SD .62 SEPARATION 1.15 Person RELIABILITY .57
S.E. OF Person MEAN = .14
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE TEST RELIABILITY = .91

SUMMARY OF 22 MEASURED Item
-----
TOTAL SCORE   COUNT   MEASURE   MODEL   INFIT   OUTFIT
SCORE          COUNT   ERROR     ERROR   MNSQ    ZSTD   MNSQ    ZSTD
-----
MEAN          24.2    36.0      1.00    .43     1.00     .0   .97   -.1
S.D.          6.7     .0        1.06    .09     .21     1.3  .33   1.3
MAX.          34.0    36.0      1.89    .74     1.49    2.8  1.62  3.0
MIN.          11.0    36.0      -2.15   .36     .74     -2.2 .43  -2.1
-----
REAL RMSE     .45 TRUE SD .96 SEPARATION 2.12 Item RELIABILITY .83
MODEL RMSE    .44 TRUE SD .97 SEPARATION 2.21 Item RELIABILITY .83
S.E. OF Item MEAN = .23
MINIMUM EXTREME SCORE: 3 Item
UMEAN=.0000 USCALE=1.0000

SUMMARY OF 25 MEASURED (EXTREME AND NON-EXTREME) Item
-----
TOTAL SCORE   COUNT   MEASURE   MODEL   INFIT   OUTFIT
SCORE          COUNT   ERROR     ERROR   MNSQ    ZSTD   MNSQ    ZSTD
-----
MEAN          25.6    36.0      1.49    .59     .74     -2.2 .43  -2.1
S.D.          7.4     .0        1.67    .46     .74     -2.2 .43  -2.1
MAX.          36.0    36.0      1.89    .74     1.49    2.8  1.62  3.0
MIN.          11.0    36.0      -4.12   .36     .74     -2.2 .43  -2.1
-----
REAL RMSE     .76 TRUE SD 1.48 SEPARATION 1.95 Item RELIABILITY .80
MODEL RMSE    .75 TRUE SD 1.49 SEPARATION 1.98 Item RELIABILITY .80
S.E. OF Item MEAN = .34
    
```

■ Variabel ■ Separasi ■ Reliabilitas
■ Measure ■ Reliabilitas

Gambar 2. Output summary statistic

Dari hasil analisis output *summary stastistic* pada Winstep didapatkan keluaran variabel data berupa tabel yang terbagi menjadi tiga jenis. Tabel pertama berisi tentang ringkasan pengukuran dari *person* atau peserta didik, kedua adalah ringkasan pengukuran item butir soal dilihat dari kesesuaian item dimana hanya terdapat 22 item yang sesuai dan mengabaikan 3 item yang tidak sesuai karena masuk kategori *extreme score* sehingga tidak masuk pada ringkasan, ketiga merupakan ringkasan pengukuran seluruh item butir soal (*extreme and non extreme*) yang berjumlah 25 butir soal. Di dalam output summary stastistic tersebut terdapat nilai *Separation*, *Reability*, dan *Alpha cornbach* yang diringkaskan pada tabel berikut :

Variabel	Separation	Reability	Alpha Cornbach
Peserta didik	1,07	0,53	0,59
Butir soal (<i>non extreme</i>)	2,12	0,82	
Butir soal (<i>extreme dan non extreme</i>)	1,95	0,79	

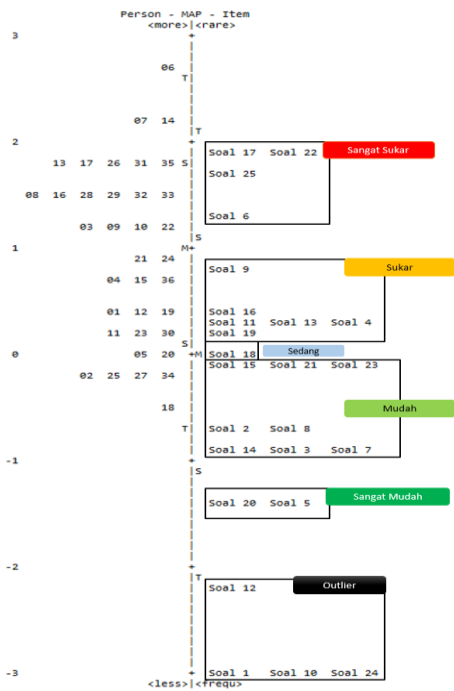
Gambar 3. Nilai Reliabilitas peserta didik dan item soal

Reliabilitas adalah interaksi antara person atau peserta tes dengan item atau butir soal yang keseluruhannya dapat dilihat dari nilai *alpha cornbach*. Gambar 3 menunjukkan hasil dari nilai reliabilitas peserta didik yaitu 0,53 yang menunjukkan bahwa reliabilitas peserta didik masuk pada kategori lemah karena nilai tersebut kurang dari 0,67. Nilai reliabilitas butir soal (*non extreme*) berjumlah 22 soal dengan mengabaikan 3 soal kategori *extreme score* didapatkan nilai 0,82 dimana nilai tersebut berada pada rentang 0,80 - 0,90 yang berarti reliabilitas dari 22 soal tersebut bagus. Sedangkan untuk nilai reliabilitas keseluruhan dari 25 butir soal (*extreme dan non extreme*) adalah 0,79 dimana reliabilitas dari keseluruhan butir soal masuk pada kategori cukup karena berada pada rentang 0,67 – 0,80. Untuk *alpha crombach* didapatkan nilai sebesar 0,59 yang diartikan bahwa nilai reliabilitas antara peserta didik dan butir soal berkategori jelek karena berada pada rentang 0,50 - 0,60. Penilaian tersebut didasarkan pada nilai *alpha cronbach* yang dapat mengukur interaksi antara person dan butir-butir soal secara keseluruhan. Sehingga dapat disimpulkan reliabilitas peserta didik sangat rendah dan termasuk kategori lemah, hal itu menunjukkan bahwa tingkat konsistensi peserta didik dalam mengerjakan tes itu lemah. Sedangkan reliabilitas semua item butir soal termasuk kategori cukup, hal ini memberitahu bahwa item butir soal yang dikembangkan mempunyai tingkat konsistensi yang cukup untuk digunakan untuk mengukur kemampuan peserta didik.

Tingkat Kesukaran

Tingkat kesukaran butir soal menggambarkan seberapa sukar soal dikerjakan dengan benar oleh peserta didik. Pada pemodelan Rasch pada program Winstep didapatkan suatu peta yang menggambarkan sebaran kemampuan responden/ peserta didik serta tingkat kesulitan butir soal dengan skala yang sama, atau disebut dengan *Wright Map* (*person - item*). Gambar dibawah merupakan hasil output winstep berupa peta *Wright Map*. Sebelah kiri peta *Wright Map* menggambarkan sebaran kemampuan siswa dalam mengerjakan soal sedangkan sebelah kanan menggambarkan sebaran tingkat kesulitan item atau butir soal dengan skala yang sama. Peta pada *Wright Map Person - Item* terbagi menjadi 3 bagian yaitu pada menu *person*, *map*, dan *item*. *Person* menunjukkan abilitas siswa, *map* menunjukkan pemetaan abilitas siswa dan item menunjukkan tingkat kesukaran soal. Semakin tinggi posisi siswa atau butir soal dan semakin tinggi nilai logit, maka semakin tinggi pula abilitas siswa dan kesukaran butir soal. Menu *map* pada gambar 3 menunjukkan statistik tingkat kesukaran soal yang dikelompokkan

bedasarkan deviasi standar (S.D) dengan nilai mean (M) sebagai titik tengah.



Gambar 4. Peta Wright Map Person-Item

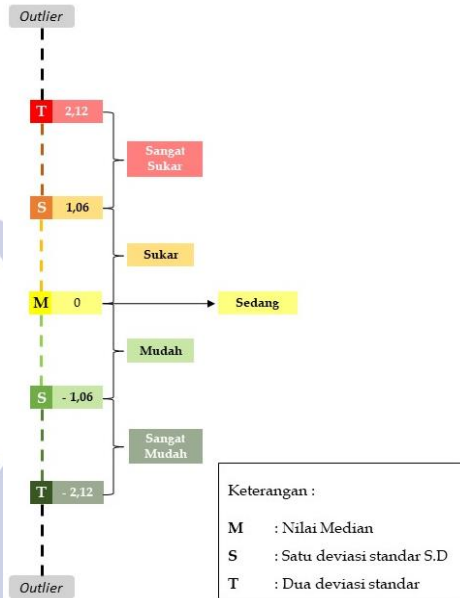
Tingkat kesukaran butir soal bisa dikelompokkan dengan meng-kombinasikan nilai rata-rata logit dan nilai Standar Deviasi (S.D). Hasil kominasi ini berguna untuk mengidentifikasi sebaran kelompok item (separation). Berdasarkan hasil analisis data *output summary statistic* butir soal (*non extreme*) dan (*extreme and non extreme*) pada gambar 4.2, Standar Deviasi (S.D) pada *item measure* / tingkat kesulitan butir soal diperoleh nilai logit yang diringkas dalam tabel berikut :

Tabel 2. Hasil item measure analisis kesukaran butir soal

Item Statistics	Measure (Logit)	
	(non extreme)	(extreme and non extreme)
Mean	0,00	-0,49
S.D.	1,06	1,67

Untuk mendapatkan klasifikasi soal berdasarkan tingkat kesukaran butir soal pada peta *Wright Map Person - Item* maka dapat ditentukan dengan menjumlahkan nilai rata – rata logit dengan nilai standar deviasi (S.D) (Sumintono & Widhiarso, 2015). Nilai yang diambil pada peta *Wright Map Person-Item* untuk mendapatkan nilai S.D adalah dari *item measure* (*non extreme*). Nilai logit tersebut digunakan karena dapat menunjukkan secara pasti tingkat kesulitan dalam interval yang sama dengan mengabaikan tiga item yang memiliki data karakteristik diluar dari batas nilai measure dan muncul dalam bentuk nilai yang ekstrim (*outlier*) sehingga tidak dapat digunakan untuk mengukur tingkat kesukaran butir soal

(Sumintono & Widhiarso, 2015). Dari hasil analisis nilai measure (*non extreme*) didapatkan nilai separasi hasil dari penjumlahan nilai rata-rata logit (0,00) + standar deviasi S.D (1,06) = 1,06 , sehingga didapatkan nilai jarak dari median (M) ke satu deviasi standar S.D (S). Nilai tersebut dapat digunakan sebagai informasi pegelompokan butir soal berdasarkan tingkat kesukaran sebagaimana pada gambar dibawah :



Gambar 5. Klasifikasi Tingkat Kesukaran Butir Soal

Berdasarkan gambar diatas, pengelompokan nilai measure logit dengan Interpretasi kesukaran butir soal diklasifikasikan menjadi lima kelompok yang dapat dilihat pada tabel berikut :

Tabel 3. Kriteria tingkat kesukaran butir soal

Nilai Measure (Logit)	Interpretasi Kesukaran Butir Soal
Lebih besar dari 1,06	Sangat Sukar
0,0 - 1,06	Sukar
0,0	Sedang
-1,06 - 0,0	Mudah
Lebih kecil dari -1,06	Sangat Mudah
Lebih dari -2,12 dan 2,12	Outliers

Pada model Rasch nilai logit (*logarithm odd unit*) yang tinggi menunjukkan soal tersebut memiliki kesukaran yang tinggi. Tingkat kesukaran butir soal direpresentasikan oleh nilai measure yang merupakan hasil dari pengolahan data ordinal berupa skor jawaban peserta didik menjadi logit yang bersifat *equal interval*. Setelah mengetahui kriteria tingkat kesukaran butir soal. Selanjutnya dilakukan analisis terhadap *output item measure* dari Winstep dengan Rasch model pada gambar

dibawah, berikut ini adalah *output item measure* dari program winstep :

Item STATISTICS: MEASURE ORDER			
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE
17	11	36	1.89
22	11	36	1.89
25	12	36	1.75
6	15	36	1.34
9	19	36	.83
16	22	36	.43
4	23	36	.30
11	23	36	.30
13	23	36	.30
19	24	36	.16
18	25	36	.01
15	26	36	-.14
21	26	36	-.14
23	26	36	-.14
2	29	36	-.66
8	29	36	-.66
3	30	36	-.86
7	30	36	-.86
14	30	36	-.86
5	32	36	-1.37
20	32	36	-1.37
12	34	36	-2.15
1	36	36	-4.12
10	36	36	-4.12
24	36	36	-4.12
MEAN	25.6	36.0	-.49
S. D.	7.4	.0	1.67

Gambar 6. *Output item measure*

Dari *output item measure* didapatkan nilai measure pada setiap butir soal. Nilai measure tersebut lalu dikelompokkan menjadi lima kelompok menurut nilai kriteria tingkat kesukaran pada tabel 4.8, sehingga didapatkan hasil analisis tingkat kesukaran butir soal sebagai berikut.

Tabel 4. Deskripsi hasil analisis tingkat kesukaran butir soal

Kategori	No. Butir Soal	Jumlah	Presentase
Sangat Sukar	6, 17, 22, 25	4	16%
Sukar	4, 9, 11, 13, 16, 19	6	24%
Sedang	18	1	4%
Mudah	2, 3, 7, 8, 14, 15, 21, 23	8	32%
Sangat Mudah	20, 5	2	8%
Outlier	1, 10, 12, 24	4	16%
Total		25	100%

Hasil analisis tingkat kesukaran butir soal didominasi oleh kelompok soal dengan kategori mudah sebanyak 8 soal menyusul kategori sukar sebanyak 6 soal. Menurut model Rasch pada butir soal tersebut memiliki tingkat kesukaran yang baik sebab soal tersebut termasuk soal yang tidak terlalu sukar dan tidak terlalu mudah. Menurut (Nuryanti dkk., 2018) Butir item soal yang diterima ialah jika butir item tersebut tidak terlalu sukar dan mudah, dengan kata lain tingkat kesukaran item masuk pada kategori sedang, mudah, dan sukar. Sehingga soal mempunyai tingkat kesukaran yang baik untuk dijangkau oleh kemampuan peserta didik. Selanjutnya untuk butir

soal yang memiliki tingkat kesukaran paling tinggi yaitu sebanyak 4 soal dan tingkat kesukaran paling rendah sebanyak 2 soal. Soal tersebut merupakan soal dengan tingkat kesukaran diluar batas satu Standar Deviasi (S.D) sehingga soal tersebut dinyatakan sebagai soal yang sangat sukar atau yang sangat mudah.

Kelompok butir soal yang sangat sukar yaitu soal nomor (6, 17, 22, 25) termasuk soal yang paling sulit dikerjakan dan tidak dapat dijangkau oleh kemampuan peserta didik. Butir soal tersebut tidak memiliki fungsi kelayakan untuk mengukur kemampuan peserta didik karena tingkat kesukaran yang sangat tinggi. Sama halnya dengan kelompok butir soal dengan kategori sangat mudah yaitu soal nomor (20 dan 5) termasuk soal yang memiliki measure kurang dari kemampuan peserta didik sehingga memiliki tingkat kesukaran paling mudah. Menurut (Sumintono & Widhiarso, 2015) Soal yang seperti itu tidak dapat menjalankan fungsinya sebagai alat ukur yang baik. Sedangkan untuk kategori outlier terdapat pada 4 butir soal yaitu nomor (1, 10, 12, 24), merupakan soal yang memiliki nilai melebihi batas T atau dua deviasi standar sehingga dapat dikatakan sebagai nilai extreme dan tidak dapat diketahui tingkat kesukarannya. Hal tersebut membuat butir soal tidak dapat digunakan dan dibuang saja. Dengan demikian berdasarkan hasil analisis tingkat kesukaran butir soal, soal yang layak digunakan berjumlah 14 butir soal yaitu yang berada pada kategori tingkat kesukaran mudah, sedang, dan sukar sehingga layak digunakan untuk mengukur kemampuan peserta didik karena memiliki tingkat kesukaran yang beragam dari yang paling mudah, sedang, dan sulit.

Daya Beda

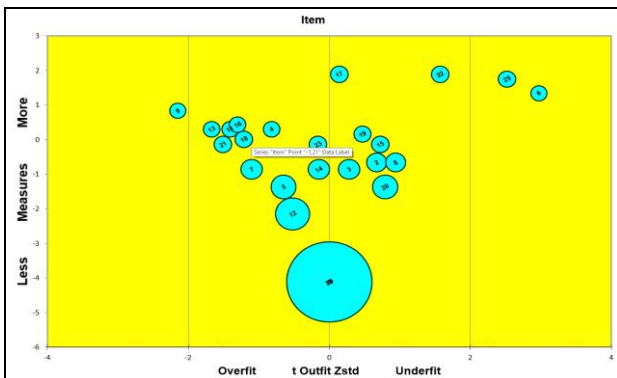
Indeks daya beda dapat diartikan sebagai kemampuan suatu butir soal membedakan kelompok peserta didik antara yang mampu mengerjakan tes dan yang tidak mampu mengerjakan suatu tes. Berikut ini adalah keluaran daya beda berupa nilai logit. Nilai daya beda soal dapat di-ketahui dari model S.E-nya (Standar Error). Nilai S.E memperlihatkan ketepatan suatu soal tersebut untuk membedakan tingkat pemahaman dan kemampuan peserta didik terhadap tes yang diujikan. Untuk mengetahui apakah variasi data baik atau buruk bisa dilihat dari nilai rata – rata (*Mean*) dan Standar Deviasi (S.D) pada SE. Jika nilai SD < dari rata – rata (*Mean*) maka variasi data bisa dikatakan baik dan jika nilai SD > rata – rata (*Mean*) maka variasi yang diperoleh buruk. Pada hasil analisis didapatkan nilai logit rata - rata sebesar 0,59 dan nilai S.D sebesar 0,46. Sehingga nilai SD 0,46 kurang dari nilai (*Mean*) 0,59 yang berarti variasi yang didapatkan bagus. Nilai (SE) yang diperoleh pada gambar 4.5 mengindikasikan kualitas daya beda setiap butir soal .

Nilai tersebut lalu di kelompokkan menjadi tiga kategori kualitas daya beda yaitu bagus dengan nilai (<0,5), cukup (0,5-1), dan Buruk (>1)

Item STATISTICS: MEASURE ORDER				
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.
17	11	36	1.89	.39
22	11	36	1.89	.39
25	12	36	1.75	.38
6	15	36	1.34	.36
9	19	36	.83	.36
16	22	36	.43	.37
4	23	36	.30	.37
11	23	36	.30	.37
13	23	36	.30	.37
19	24	36	.16	.38
18	25	36	.01	.39
15	26	36	-.14	.39
21	26	36	-.14	.39
23	26	36	-.14	.39
2	29	36	-.66	.44
8	29	36	-.66	.44
3	30	36	-.86	.47
7	30	36	-.86	.47
14	30	36	-.86	.47
5	32	36	-1.37	.55
20	32	36	-1.37	.55
12	34	36	-2.15	.74
1	36	36	-4.12	1.82
10	36	36	-4.12	1.82
24	36	36	-4.12	1.82
MEAN	25.6	36.0	-.49	.59
S.D.	7.4	.0	1.67	.46

Gambar 7. Output daya beda (SE)

Dilihat dari nilai model SE pada hasil analisis didominasi oleh soal dengan daya beda yang baik/ bagus sebanyak 19 soal. Soal tersebut memiliki nilai S.E yang kecil atau dibawah <0,5 sehingga butir soal memiliki ketelitian yang mampu membedakan kemampuan peserta didik yang tinggi dan yang rendah. Terdapat juga soal yang memiliki daya beda yang cukup dengan nilai 0,5 – 1 yaitu nomor (5, 12, 20), sehingga masih bisa digunakan untuk mengukur abilitas siswa namun lebih baik jika dilakukan perbaikan. Untuk soal dengan daya beda yang buruk dengan nilai >1 yaitu nomor (1, 10, 24), butir soal tersebut tidak dapat digunakan untuk mengukur kemampuan peserta didik sehingga harus diganti dengan yang baru. Untuk memperjelas daya beda setiap butir soal maka digambarkan dalam sebuah grafik berupa bubble chart dari program winstep. Tingkat kesukaran soal dan daya beda soal dapat dilihat berdasarkan posisi lingkaran pada grafik bubble chart berikut :



Gambar 8. Bubble chart winstep

Gambar tersebut menunjukkan posisi lingkaran yang menunjukkan tingkat kesukaran butir soal. Jika posisi lingkaran semakin keatas maka butir soal semakin sukar, namun jika posisi lingkaran ke bawah maka butir soal semakin mudah. Sedangkan ukuran lingkaran menunjukkan daya beda butir soal. Lingkaran dengan ukuran yang semakin besar menunjukkan daya beda yang jelek suatu soal untuk membedakan abilitas siswa, dan sebaliknya untuk lingkaran yang semakin kecil ukurannya maka akan semakin bagus suatu soal membedakan abilitas peserta didik.

Berdasarkan gambar diatas, soal nomor (6, 9, 17, 22, 25) berada pada posisi paling atas dan memiliki ukuran lingkaran yang relatif kecil, artinya soal tersebut memiliki tingkat kesukaran paling tinggi diantara 25 soal dan memiliki daya beda yang baik. Sementara soal nomor (24, 1, 10) berada di posisi paling bawah dan bentuk lingkaranya paling besar dengan posisi yang sama dalam satu lingkaran yang artinya soal tersebut memiliki tingkat kesukaran yang rendah dan daya beda soal tersebut buruk. Selain itu daya beda bisa juga dilihat melalui identifikasi pada kelompok responden berdasarkan indeks separasi responden. Menurut (Sumintono & Widhiarso, 2015) Jika nilai separasi semakin besar maka kualitas instrumen termasuk kategori bagus, karena dapat mengidentifikasi keseluruhan dari kelompok responden dan butir soal. Disini penggunaan persamaan pemisahan strata (H) digunakan untuk melihat pengelompokan secara lebih teliti, persamaan tersebut dapat dilihat pada persamaan 1 dibawah ini :

$$H = \frac{[(4 \times \text{SEPARATION}) + 1]}{3} \tag{1}$$

Keterangan :
 H : Nilai strata (Pengelompokan)
 Separation : Nilai separation untuk responden yang dihasilkan

Berdasarkan hasil analisis diketahui bahwa nilai separasi butir soal sebesar 2,21 maka pengelompokan butir soal dapat ditentukan perhitungan separasi sebagaimana pada persamaan 2

$$H_{\text{Butir soal}} = \frac{(4 \times 2,12) + 1}{3} = \frac{9,48}{3} = 3,16 \tag{2}$$

Dari nilai separasi yang didapatkan nilai H Butir soal sebanyak 3,16 jika dibulatkan ke satuan terdekat menjadi 3, artinya separation soal bernilai baik karena membagi soal dalam 3 kelompok yaitu tinggi, sedang, dan rendah. Sehingga dapat disimpulkan daya beda tersebut sudah cukup mampu dalam membedakan pemahaman dari siswa yang memiliki kemampuan tinggi dan rendah. Semakin banyak separation menunjukkan bahwa instrumen yang dibuat bagus.

PENUTUP

Simpulan

Dari analisis kualitas soal USP menggunakan model Rasch dengan bantuan Winstep, didapatkan kesimpulan bahwa :

- Validitas soal menunjukkan bahwa tingkat kesesuaian dari 25 butir soal diperoleh sebanyak 20 soal yang dinyatakan fit/ valid dan 5 soal yang tidak fit/ tidak valid
- Realibilitas instrumen antara peserta didik dan butir soal menunjukkan bahwa instrumen memiliki kategori yang jelek. Reliabilitas person menunjukkan reabilitas peserta didik sangat rendah dengan kategori lemah. Soal (non extreme) menunjukkan reliabilitas item tersebut bagus. Sedangkan reliabilitas keseluruhan dari butir soal (extreme dan non extreme) masuk pada kategori cukup.
- Tingkat kesukaran butir soal didominasi oleh soal berkategori mudah sebanyak 8 soal menyusul kategori sukar sebanyak 6 soal. Sehingga soal yang layak digunakan berjumlah 14 butir soal yaitu yang berada pada kategori tingkat kesukaran mudah, sedang, dan sukar. Untuk butir soal yang memiliki tingkat kesukaran paling tinggi yaitu sebanyak 4 soal dan tingkat kesukaran paling rendah sebanyak 2 soal. Terdapat kategori outlier sebanyak 4 soal Untuk kategori soal yang sangat sulit, sangat mudah, dan outlier harus diganti atau dibuang.
- Daya beda butir soal dilihat dari nilai Standar Error (S.E) didominasi soal dengan daya beda yang baik sebanyak 19 soal, 3 soal yang memiliki daya beda yang cukup sehingga soal tersebut masih bisa digunakan dengan perbaikan, dan 3 soal dengan daya beda yang buruk sehingga soal tersebut tidak dapat digunakan dan harus diganti dengan yang baru. Selanjutnya nilai separasi keseluruhan item butir soal bernilai baik karena membagi soal dalam 3 kelompok yaitu tinggi, sedang, dan rendah.

Saran

- Perlu adanya evaluasi pada kualitas soal untuk mengetahui aspek yang tidak sesuai dengan kemampuan peserta didik sehingga bisa dilakukan perbaikan sebelum diujikan.
- Soal – soal yang dibuat oleh guru Teknik Pemesinan sebaiknya dilakukan evaluasi terlebih dahulu agar dapat mengetahui kriteria soal yang valid dan dapat dijangkau oleh peserta didik.
- Penelitian ini dapat digunakan bagi peneliti selanjutnya untuk digunakan sebagai acuan analisis butir soal yang lain pada mata pelajaran Teknik Pemesinan serta dengan populasi yang lebih besar.

DAFTAR PUSTAKA

- Anggraini, M., Kasiyun, S., Mariati, P., & Sunanto, S. (2021). Analisis Keberhasilan Peserta Didik dalam Pembelajaran Tematik melalui Daring pada Masa Pandemi Covid-19 di Sekolah Dasar. *Jurnal Basicedu*, 5(5) : 3010–3019.
<https://doi.org/10.31004/basicedu.v5i5.1223>
- Ariyana, L. T.. (2011). *Analisis Butir Soal Ulangan Akhir Semester Gasal IPA Kelas IX SMP di Kabupaten Grobogan*. Skripsi. Semarang : UIN Walisongo.
<http://lib.unnes.ac.id/372/>
- Aziz, R. (2015). Aplikasi Model RASCH dalam Pengujian Alat Ukur Kesehatan Mental di Tempat Kerja. *Psikoislamika : Jurnal Psikologi Dan Psikologi Islam*, 12(2) : 29-39
<https://doi.org/10.18860/psi.v12i2.6402>
- Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen the Use of Rasch Model for Analyzing Test. *Jurnal Pendidikan Matematika*, 3(1) : 45–50.
<https://doi.org/10.30598/jupitekvols3iss1ppx45-50>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht : Springer Netherlands
<https://doi.org/10.1007/978-94-007-6857-4>
- Erfan, M., Mauliyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis Kualitas Soal Kemampuan Membedakan Rangkaian Seri dan Paralel Melalui Teori Tes Klasik dan Model Rasch. *Indonesian Journal Of Educational Research and Review*, 3(1) : 11-19
<https://doi.org/10.23887/ijerr.v3i1.24080>
- Fajrianti, F., Hendriani, W., & Septarini, B. G. (2016). Pengembangan Tes Berpikir Kritis Dengan Pendekatan Item Response Theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1) : 46-55
<https://doi.org/10.21831/pep.v20i1.6304>
- Fauziana, A., & Dessy Wulansari, A. (2021). Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch. *Ibriez : Jurnal Kependidikan Dasar Islam Berbasis Sains*. 6(1) : 11-19
<https://doi.org/10.21154/ibriez.v6i1.112>
- Fisher, A. G. (1993). The Assessment of IADL Motor Skills: An Application of Many-Faceted Rasch Analysis. *American Journal of Occupational Therapy* 47(4) : 319–29.
<https://doi.org/10.5014/ajot.47.4.319>
- Gunawan, C. (2018). *Uji Reabilitas*. Yogyakarta : Deepublish (Vol. 2, Issue 2).

- Hariyanto, & Basuki, I. (2014). *Asesmen Pembelajaran Bandung* : PT. Remaja Rosdakarya.
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *AL-MU'ARRIB: Journal Of Arabic Education*, 1(1) : 34-44
<https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar. *Journal of Elementary Education*, 4(1) : 1-6
<https://journal.unnes.ac.id/sju/index.php/jee/article/view/7488>
- L, Idrus. (2019). Evaluasi Dalam Proses Pembelajaran. *Jurnal Manajemen Pendidikan Islam*, 9(2), 920–935.
<https://doi.org/10.35673/ajmpi.v9i2.427>
- Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9) : 1-20
<https://doi.org/10.18637/jss.v020.i09>
- Muhammad Ali, H. (1987). *Guru dalam Proses Belajar Mengajar*. (Cet. 3). Bandung : Sinar Baru.
- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1) : 86 -96
<https://doi.org/10.23887/jisd.v4i1.23335>
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan Model Rasch pada Pengembangan Instrumen Kemampuan Berpikir Kritis Peserta Didik Sekolah Menengah Kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2) : 224-233
<https://journal.uny.ac.id/index.php/jipi/article/view/21442>
- Pane, A., & Darwis Dasopang, M. (2017). Belajar dan Pembelajaran. *FITRAH Jurnal Kajian Ilmu-Ilmu Keislaman*, 3(2) : 333 -352
<https://doi.org/10.24952/fitrah.v3i2.945>
- Perdana, S. A. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik dan Rasch Model. *Jurnal Kiprah*, 6(1). 41-48
<https://doi.org/10.31629/kiprah.v6i1.574>
- Pratiwi, R., Reflianti, S., Antini, S., & Walid, A. (2021). Analysis of Item Difficulty Index for Midterm Examinations in Junior High Schools 5 Bengkulu City. *Asian Journal of Science Education*, 3(1), 12–18. <https://doi.org/10.24815/ajse.v3i1.18895>
- Pohan, S. A., & Dafit, F. (2021). Pelaksanaan Pembelajaran Kurikulum 2013 di Sekolah Dasar. *Jurnal Basicedu*, 5(3), 1191–1197.
<https://doi.org/10.31004/basicedu.v5i3.898>
- Rachman, D. F., & Fahdiansyah, R. (2018). Pengembangan Model Pembelajaran My Trip My Adventure. *JPEK (Jurnal Pendidikan Ekonomi Dan Kewirausahaan)*, 2(2), 95-106.
<https://doi.org/10.29408/jpek.v2i2.935>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1) : 84-94
<https://doi.org/10.21831/jpai.v14i1.11370>
- Rasyid, R. A., Zaharim, A., & Masodi, M. S. (2007). *Appication of Rasch Measurement in Evaluation of Learning Outcome: A Case Study in Electrical Engineering. Regional Conference on Engineering Mathematics, Mechanics, Manufacturing & Architecture (EM3ARC)* 151-165
- Rindawan, R., & Syahrir, S. (2019). Kompetensi Guru Pendidikan Jasmani Olahraga dan Kesehatan Dalam Pelaksanaan Evaluasi Pembelajaran SMAN 1 Praya Barat Daya Tahun 2018. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 3(1) : 76-82
<https://doi.org/10.58258/jisip.v3i1.621>
- Riyadi, A. (2017). Kompetensi Guru Dalam Pelaksanaan Evaluasi pembelajaran. *Itihad Jurnal Kopertais Wilayah XI Kalimantan*, 15(28) : 52–67.
<https://jurnal.uinantasari.ac.id/index.php/ittihad/article/view/1933>
- Stenbeck, M., Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1992). *Fundamentals of Item Response Theory*. *Contemporary Sociology*, 21(2), 289.
<https://doi.org/10.2307/2075521>
- Sumintono, B., & Widhiarso, W. (2015). “Aplikasi Pemodelan Rasch Pada Assessment Pendidikan. In Aplikasi Rasch pemodelan Pada Assessment Pendidikan”. Makalah dipresentasikan dalam *Kuliah Umum pada Jurusan Statistika, Institut Teknologi Sepuluh November*, Surabaya, 17 Maret 2016.
- Widoyoko, E. P. (2014). *Penilaian Hasil Pembelajaran di Sekolah*, Pustaka Pelajar, Yogyakarta (1st ed.). Yogyakarta : Pustaka Pelajar.
- Widyaningsih, S. W., & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Rasch Model. *Gravity : Jurnal Ilmiah Penelitian Dan Pembelajaran Fisika*, 4(1) : 33-46
<https://doi.org/10.30870/gravity.v4i1.3116>