

**ANALISI BUTIR SOAL PENILIAN AKHIR SEMESTER (PAS) PADA MATA
PELAJARAN PPKN MENGGUNAKAN APLIKASI WINSTEP DAN ITEM AND
TES ANALYSIS (ITEMAN) 4.3**

Elfina Saely¹, Shaleh²

Pendidikan Guru Madrasah Ibtidaiyah, Fakultas Ilmu Tarbiyah Dan Keguruan,
Universitas Islam Negeri Sunan Kalijaga Yogyakarta
21204082038@student.uin-suka.ac.id

ABSTRACT

This research was conducted with the aim of determining the quality of Final Semester Assessment (PAS) question items in Civics subjects. The method used in this research study is quantitative descriptive. The research data used secondary data in the form of 24 polytomus questions in 4 categories and PAS results of 30 grade V T.A 2022/2023 students at the Ma'arif Education Institute NU Sidoarjo. Data analysis is done by empowering Winstep and Iteman 4.3 applications. The results of the analysis using the Winstep application show a validity index if it meets the requirements of the MNSQ outfit, ZSTD Outfit, and Point Measure correlation (PT measure Corr) and a reliability index of 0.83. The characteristics of the question items are shown by the difficulty data of the question items by 54% being in the difficult question category and 46% of the difficulty level of the question items is very easy, the discriminating power is 100%, and 58% of distractors have functioned properly. While the results of data analysis using ITEMAN 4.3 software show that 80% of the difficulty level of the question items is in the medium category, 90% of the question items have good discriminating power, and 90% of distractors have functioned properly. The results of this analysis show that the two results of the analysis using Winstep and Iteman on the Final Semester Assessment (PAS) question items in the PPKn subject show "similarity" and both are included in the good category.

Keywords: Question Item, Winstep, ITEMAN 4.3, PPKn.

ABSTRAK

Penelitian ini dilakukan dengan tujuan untuk mengetahui kualitas butir soal Penilaian Akhir Semester (PAS) pada mata pelajaran PKN. Metode yang dipakai dalam kajian penelitian ini ialah deskriptif kuantitatif. Data penelitian menggunakan data sekunder berupa 24 butir soal politomus 4 kategori dan hasil PAS 30 orang siswa kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo. Analisis data dilakukan dengan memberdayakan aplikasi Winstep dan Iteman 4.3. Hasil analisis menggunakan aplikasi Winstep menunjukkan indeks validitas apabila sudah memenuhi syarat outfit MNSQ, Outfit ZSTD, serta Point Measure correlation (PT measure Corr) dan indeks reliabilitas sebesar 0,83. Karakteristik butir soal ditunjukkan dengan data kesukaran butir soal sebesar 54% berada pada kategori soal sulit dan 46% tingkat kesukaran butir soal sangat mudah, daya pembeda sebesar 100%, dan 58% pengecoh (distractor) telah berfungsi dengan baik. Sedangkan hasil analisis data menggunakan software ITEMAN 4.3 menunjukkan bahwa sebesar 80% tingkat kesukaran butir soal berada pada kategori sedang, 90% butir soal memiliki daya pembeda yang baik, dan 90%

pengecoh (distractor) telah berfungsi dengan baik. Hasil analisis ini menunjukkan bahwa kedua hasil analisis menggunakan Winstep dan Iteman pada butir soal Penilaian Akhir Semester (PAS) pada mata pelajaran PPKn menunjukkan “kemiripan” dan keduanya termasuk dalam kategori baik.

Kata Kunci: Butir Soal, Winstep, ITEMAN 4.3, PPKn.

A. Pendahuluan

Penilaian belajar siswa ialah cara yang dilaksanakan oleh tenaga pengajar agar diperoleh informasi terkait pembelajaran yang telah dipaparkan pada peserta didik selama satu semester masa belajar. Paparan dari Ralph Tyler (Arikunto, 2017; dan Chellamani, 2013), menyatakan bahwa penilaian belajar ialah hal yang penting dilaksanakan untuk memberikan informasi terkait capaian pengajaran yang direncanakan untuk peserta didik telah merata atau masih butuh keadaan perbaikan. Hal yang juga dinilai dalam evaluasi ini ialah kelayakan soal yang disusun oleh tenaga pengajar untuk diselesaikan oleh peserta didik saat menghadapi evaluasi akhir belajar. Penilaian dalam belajar ialah cara untuk mengetahui sejauh apa pencapaian yang diraih oleh peserta didik serta kualitas dari soal yang disusun oleh tenaga pendidik untuk diujikan pada peserta didik. Pengadaan pengujian evaluasi akhir belajar dilaksanakan dua kali dalam satu semester, ialah ketika tengah semester serta evaluasi akhir

untuk penilaian final pengumpulan nilai peserta didik untuk mengetahui capaian dalam satu semester pembelajaran.

Penilaian terhadap peserta didik selama melaksanakan pembelajaran ialah memantau serta memberikan nilai berupa angka agar dijadikan apresiasi atau motivasi untuk terus memajukan tingkat pendidikan masa berikutnya. Penilaian yang dilakukan terhadap peserta didik memakai soal yang terjamin dan baik kualitas untuk dipecahkan oleh peserta didik (Sopiah, 2019). Soal yang akan diujikan pada peserta didik harus memiliki kualitas serta valid untuk lanjut pada tahapan pengujian yang telah selesai melalui proses objektif, standar kompetensi, tingkat kevalidan, reliabilitas serta soal dalam konteks mendidik (Puskur, 2008; Kubiszyn & Borich, 2013, Arikunto, 2013) serta pemaparan dari kementerian pendidikan nasional (Fitri, 2019). Soal yang dilakukan analisa kelayakan juga harus lolos tahap kelayakan dan pantas untuk dilaksanakan pengujian pada peserta didik. Pemantauan serta

analisis untuk soal yang layak dipakai menguji kemantapan ilmu pengetahuan peserta didik butuh untuk dilaksanakan secara berkala agar hasil pantauan sesuai serta tidak akan meleset dari perhitungan yang dilaksanakan. Pelaksanaan kegiatan ini dinilai sesuai dengan C.E.S 2022 Conference of Elementary Studies 394 394 ialah dengan tujuan mengadakan perbaikan kualitas pembelajaran dalam tingkatan peserta didik serta tingkatan tenaga pendidik yang akan melaksanakan transfer ilmu secara langsung.

Soal yang dianalisa terdiri dari beberapa pengujian ialah pengujian pada tingkat kesulitan yang dimiliki oleh soal, keunikan dan perbedaan dengan soal yang telah disusun sebelumnya serta mengadakan bentuk soal dalam bentuk kecoh (Mardapi, 2017) soal yang diuji kesulitannya dipaparkan penilaiannya dengan angka 0,00-1,00. Nilai angka yang semakin banyak jumlahnya akan menyatakan tingkat kemudahan soal untuk diselesaikan, tapi jika angka yang ditunjukkan kecil berarti soal tersebut sulit untuk diselesaikan (Kaplan & Sacuso, 2005; Sujana, 2011). Daya beda yang terdapat pada soal ialah soal dengan keadaan tingkat sulit yang sama dengan

diadakan penilaian akan ditemui hasil peserta didik masuk golongan mudah memecahkan soal atau peserta didik yang sulit memecahkan soal. Daya beda soal dinyatakan dengan angka 0,00-1,00. Soal dengan kualitas yang baik ialah capaian angkanya 0,3 atau angka atasnya (Nunnally & Berstein, 2009; Azwar, 2012), keadaan serta kualitas soal dengan indikasi sedang dinyatakan dengan angka 0,20-0,29 (Crocker & Algina, 2008). Kegunaan distraktor ialah untuk mengadakan timbulnya rasa suka peserta didik dalam bidang soal yang sebelumnya tidak dipahami. Soal dengan kualitas yang baik ialah dengan capaian pengukuran lebih banyak nilainya dari 5% (Mardapi, 2017; dan Ailen & Yen 1979; serta Mahjabeen (2017).

Pernyataan yang dilakukan setelah mengadakan pengujian ialah memakai analisa kuantitatif serta kualitatif. Soal yang dianalisa ialah soal dalam wujud pilihan ganda dan akan dilaksanakan pengujian pada peserta didik pada masa tengah semester atau akhir semester. Pengujian yang dilakukan ialah dengan bahan distraktor beserta kunci dari jawaban soal yang akan diujikan untuk evaluasi peserta didik (Burud, 2019). Evaluasi akhir terhadap peserta didik ialah cara yang

dilaksanakan untuk menemukan informasi seberapa baik siswa telah menguasai pelajaran yang diajarkan tenaga pendidik selama masa pembelajaran diadakan tatap muka dalam lingkungan kelas.

Sesuai dengan uraian yang telah dipaparkan dalam latar belakang diatas, peneliti tertarik untuk melakukan studi penelitian dengan judul **“ANALISI BUTIR SOAL PENILAIAN AKHIR SEMESTER (PAS) PADA MATA PELAJARAN PPKN MENGGUNAKAN APLIKASI WINSTEP DAN ITEM AND TES ANALYSIS (ITEMAN) 4.3 ”**. Software yang digunakan untuk analisis butir soal pada penelitian ini ialah aplikasi *Item And Tes Analysis (Iteman)* versi 4.3 dan rasch model dengan bantuan aplikasi Winstep. Aplikasi *Item And Tes Analysis (Iteman)* versi 4.3 dan rasch model dengan bantuan aplikasi Winstep ialah aplikasi yang dipakai peneliti untuk membantu melangsungkan kajian penelitian ini dan dipakai untuk melakukan uji terhadap soal yang akan dipergunakan dalam evaluasi akhir untuk peserta didik yang dinamakan Teori Tes Klasik (TTK). Parameter yang diuji dalam kajian penelitian ini ialah pengujian empirik dengan hal yang mencakup ialah kerumitan soal,

beda soal, soal yang butirannya efektif dalam kajian distraktor, angka reliabilitas serta tingkat eror yang mungkin ada dalam perjalanan pembuatan soal hingga tahapan akhir.

B. Metode Penelitian

Metode yang dipakai dalam kajian penelitian ini ialah deskriptif kuantitatif. Menurut Sugiyono (2017) metode penelitian deskriptif kuantitatif bertujuan untuk mendeskripsikan suatu fenomena, peristiwa, gejala dan kejadian yang terjadi secara factual, sistematis serta akurat. Data penelitian ini menggunakan data sekunder berupa 24 butir soal politomus 4 kategori dari hasil PAS 30 orang siswa kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo. Analisis butir soal dilakukan ialah soal yang di buat oleh tenaga pendidik Mata Pelajaran Pendidikan Pancasila dan Kewarganegaraan dan telah dilaksanakan pengujian pada peserta didik serta akan dikaji ulang untuk menemukan kelayakan yang terdapat dalam soal yang telah disusun. Analisis data pada penelitian ini dilakukan dengan memberdayakan aplikasi Winstep dan Iteman 4.3.

C.Hasil Penelitian dan Pembahasan

Hasil Analisis menggunakan Winstep

Model Rasch mudah digunakan dan memberikan hasil analisis yang akurat. Ini juga dapat digunakan untuk memeriksa peluang siswa untuk menjawab pertanyaan dengan benar dengan membandingkan keterampilan mereka dengan tingkat kesulitan pertanyaan (Khotimah & Sri, 2014). Model Rasch memiliki tingkat kesukaran soal tetap varian, bagaimanapun sampel yang terlibat dalam variasi awal (Wei dkk, 2012). Rasch mengembangkan model pengukuran data yang dapat menentukan hubungan antara tingkat kemampuan mahasiswa itu sendiri (person ability) dan pada tingkat kesukaran item (item difficulty) dengan menggunakan fungsi logaritma untuk dapat menghasilkan pengukuran dengan nilai interval yang sama (Bambang, 2014). Adapun kelebihan model Rasch menurut Sumintono & Widhiarso (2015: 57), Model Rasch menyelesaikan masalah butir soal, menangani data yang hilang, dan memenuhi standar pengukuran.

Item STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL		MODEL	INFIT	OUTFIT	PT-MEASURE	EXACT	MATCH				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item
2	19	30	-.40	.43	1.86	3.8	2.23	3.1	A-.24	.45	43.3	72.7	S2
18	15	30	.29	.41	1.15	.9	1.57	2.1	B .28	.46	76.7	69.9	S18
20	18	30	-.22	.42	1.35	1.8	1.25	.9	C .21	.45	60.0	71.3	S20
7	15	30	.29	.41	1.14	.9	1.31	1.3	D .33	.46	70.0	69.9	S7
13	12	30	.81	.42	1.14	.8	1.25	1.0	E .34	.46	60.0	70.7	S13
4	11	30	.99	.43	1.20	1.1	1.13	.5	F .32	.45	56.7	71.6	S4
19	16	30	.12	.41	1.17	1.0	1.07	.4	G .35	.46	56.7	70.2	S19
14	20	30	-.58	.43	1.13	.7	1.15	.5	H .33	.44	73.3	73.9	S14
12	17	29	-.14	.43	1.13	.8	1.12	.5	I .37	.46	65.5	71.5	S12
1	16	30	.12	.41	1.09	.6	1.05	.3	J .40	.46	70.0	70.2	S1
5	21	30	-.77	.44	1.08	.5	.93	-.3	K .39	.43	70.0	74.9	S5
24	21	30	-.77	.44	1.04	.3	1.00	.1	L .40	.43	70.0	74.9	S24
17	16	30	.12	.41	1.00	.1	.92	-.2	M .47	.46	70.0	70.2	S17
23	19	30	-.40	.43	.99	.0	.91	-.2	N .46	.45	70.0	72.7	S23
6	11	30	.99	.43	.80	-1.1	.91	-.2	O .58	.45	76.7	71.6	S6
10	21	30	-.77	.44	.83	-.8	.71	-.7	P .56	.43	83.3	74.9	S10
21	20	30	-.58	.43	.81	-1.0	.69	-.9	Q .59	.44	80.0	73.9	S21
3	14	30	.46	.41	.79	-1.3	.68	-1.4	R .64	.46	73.3	69.8	S3
8	13	30	.64	.42	.78	-1.3	.68	-1.4	S .64	.46	83.3	70.1	S8
22	21	30	-.77	.44	.78	-1.1	.76	-.6	T .58	.43	90.0	74.9	S22
11	13	30	.64	.42	.70	-1.9	.60	-1.8	U .70	.46	83.3	70.1	S11
16	12	30	.81	.42	.69	-1.9	.58	-1.7	V .70	.46	80.0	70.7	S16
9	15	30	.29	.41	.67	-2.2	.58	-2.0	W .72	.46	83.3	69.9	S9
15	23	30	-1.19	.47	.64	-1.7	.43	-1.4	X .68	.40	76.7	78.4	S15
MEAN	16.6	30.0	.00	.43	1.00	.0	.98	-.1			71.8	72.0	
S.D.	3.5	.2	.63	.01	.27	1.4	.38	1.2			10.6	2.3	

a. Validitas

Validitas, reliabilitas, dan kesulitan soal dipertimbangkan dalam analisis kualitas soal untuk membedakan antara seri dan seri paralel. Pada Model Rasch, untuk melihat kualitas butir soal dari aspek validitas adalah jika memenuhi beberapa kriteria berikut ini (Sumintono & Widhiarso, 2015).

- Nilai Outfit MNSQ (Mean Square) yang diterima adalah: $0,5 < \text{Outfit} - \text{MNSQ} < 1,5$
- Nilai Outfit ZSTD (Z –Standard) yang diterima adalah: $-2,0 < \text{ZSTD} < +2,0$
- Nilai Pt Measure Corr (Point Measure Correlation): $0,4 < \text{Point Measure Corr} < 0,85$

Tabel 5
Hasil analisis validitas menggunakan
Winstep

SUMMARY OF 30 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	13.3	24.0	.30	.49	1.00	.0	.98	-.1	
S.D.	5.2	.2	1.15	.09	.11	.8	.18	.8	
MAX.	22.0	24.0	2.55	.75	1.22	1.1	1.47	1.2	
MIN.	3.0	23.0	-2.09	.43	.76	-2.0	.68	-1.9	

REAL RMSE	.51	TRUE SD	1.03	SEPARATION	2.02	Person RELIABILITY	.80		
MODEL RMSE	.50	TRUE SD	1.04	SEPARATION	2.07	Person RELIABILITY	.81		
S.E. OF Person MEAN = .21									

Person RAW SCORE-TO-MEASURE CORRELATION = .99									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83									
SUMMARY OF 24 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	16.6	30.0	.00	.43	1.00	.0	.98	-.1	
S.D.	3.5	.2	.63	.01	.27	1.4	.38	1.2	
MAX.	23.0	30.0	.99	.47	1.86	3.8	2.23	3.1	
MIN.	11.0	29.0	-1.19	.41	.64	-2.2	.43	-2.0	

REAL RMSE	.45	TRUE SD	.44	SEPARATION	.98	Item RELIABILITY	.49		
MODEL RMSE	.43	TRUE SD	.46	SEPARATION	1.08	Item RELIABILITY	.54		
S.E. OF Item MEAN = .13									

MEAN=.0000 USCALE=1.0000									
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00									

Pada penggunaan model rasch ini dianggap valid dan dirasa lebih akurat ketika sudah memenuhi ketiga kriteria tersebut. Maka dari hasil analisis kualitas soal dilihat dari segi validitas butir soal dengan menggunakan aplikasi winstep diperoleh 1 item karena tidak memenuhi persyaratan untuk outfit MNSQ, outfit ZSTD, dan korelasi Point Measure (PT mengukur Corr), pertanyaan nomor 15 adalah pertanyaan yang diduga tidak valid.

b. Reliabilitas

Pada model rasch menggunakan bantuan program Winsteps nilai reliabilitas dapat diukur dengan menggunakan nilai reliabilitas orang dan nilai

reliabilitas item, dan kriteria reliabilitas yang dapat digunakan adalah sebagai berikut: 0,67 (lemah); 0,67-0,8 (cukup); 0,8-0,9 (baik); 0,91-0,94 (sangat baik); >0,94 (Sumintono, B., & Widhiarso, 2015). Yang tergambar pada tabel berikut:

Tabel 2
Hasil analisis reliabilitas menggunakan
Winstep

Tabel 3. Kategori nilai person reliability dan item reliability butir soal	
Nilai Alpha Cronba	Kategori
< 0,50	Buruk
0,50 – 0,60	Jelek
0,60-0,70	Cukup
0,70-0,80	Bagus
>0,80	Bagus sekali

Tabel 4. Kategori nilai alpha cronbach (Reliabilitas) butir soal	
Nilai Person Reliability dan Item Reliability	Kategori
< 0,67	Lemah
0,67 – 0,80	Cukup
0,81 – 0,90	Bagus
0,91 – 0,94	Bagus Sekali
>0,94	Istimewa

Dengan bantuan Program Winsteps, temuan analisis reliabilitas menggunakan model Rasch menghasilkan nilai Cronbach Alpha sebesar 0.83 yang menunjukkan reliabilitas nya masuk kedalam kategori bagus. Sedangkan nilai person reliability berada pada kisaran 0.80 – 0.81 sehingga termasuk dalam kategori bagus sekali.

Dengan bantuan Program Winsteps, temuan analisis keandalan menggunakan model Rasch menghasilkan nilai Cronbach Alpha. Nilai Cronbach Alpha

diperoleh dengan analisis keandalan menggunakan model Rasch dengan bantuan dari Program Winsteps.

c. Tingkat kesukaran winstep

Analisi pada model Rasch menggunakan bantuan program Winsteps tingkat kesukaran dapat dilihat dari nilai *measure*. Suminto dan Widhiarso (2015) memberikan panduan dalam menilai item tersebut menjadi empat kategori, yakni:

- Nilai *measure* < -1 = item sangat mudah
- Nilai *measure* -1 s.d. 0 = item mudah
- Nilai *measure* 0 s.d. 1 = item sulit
- Nilai *measure* > 1 = item sangat sulit

Tabel 3

Hasil analisis tingkat kesukaran
pada Winstep

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	OBSX	EXACT MATCH EXPX	Item
4	11	30	.99	.43	1.20	1.11	1.13	.5	.32	.45	56.7	71.6	S4
6	11	30	.99	.43	.80	-1.1	.91	-.2	.58	.45	76.7	71.6	S6
13	12	30	.81	.42	1.14	.8	1.25	1.0	.34	.46	60.0	70.7	S13
16	12	30	.81	.42	.69	-1.9	.58	-1.7	.70	.46	80.0	70.7	S16
8	13	30	.64	.42	.78	-1.3	.68	-1.4	.64	.46	83.3	70.1	S8
11	13	30	.64	.42	.70	-1.9	.60	-1.8	.70	.46	83.3	70.1	S11
3	14	30	.46	.41	.79	-1.3	.68	-1.4	.64	.46	73.3	69.9	S3
7	15	30	.29	.41	1.14	.9	1.31	1.3	.33	.46	70.0	69.9	S7
9	15	30	.29	.41	.67	-2.2	.58	-2.0	.72	.46	83.3	69.9	S9
18	16	30	.29	.41	1.15	.9	1.57	2.1	.28	.46	76.7	69.9	S18
1	16	30	.12	.41	1.09	.6	1.05	.3	.40	.46	70.0	70.2	S1
17	16	30	.12	.41	1.00	.1	.92	-.2	.47	.46	70.0	70.2	S17
19	16	30	.12	.41	1.17	1.0	1.07	.4	.35	.46	56.7	70.2	S19
12	17	29	-.14	.43	1.13	.8	1.12	.5	.37	.46	65.5	71.5	S12
20	18	30	-.22	.42	1.35	1.8	1.25	.9	.21	.45	60.0	71.3	S20
2	19	30	-.40	.43	1.86	3.8	2.23	3.1	-.24	.45	43.3	72.7	S2
23	19	30	-.40	.43	.99	.0	.91	-.2	.46	.45	70.0	72.7	S23
14	20	30	-.58	.43	1.13	.7	1.15	.5	.33	.44	73.3	73.9	S14
21	20	30	-.58	.43	.81	-1.0	.69	-.9	.59	.44	80.0	73.9	S21
5	21	30	-.77	.44	1.08	.5	.93	-.1	.39	.43	70.0	74.9	S5
10	21	30	-.77	.44	.83	-.8	.71	-.7	.56	.43	83.3	74.9	S10
22	21	30	-.77	.44	.78	-1.1	.76	-.6	.58	.43	90.0	74.9	S22
24	21	30	-.77	.44	1.04	.3	1.00	.1	.40	.43	70.0	74.9	S24
15	23	30	-1.19	.47	.64	-1.7	.43	-1.4	.68	.40	76.7	78.4	S15
MEAN	16.6	30.0	.00	.43	1.00	.0	.98	-.1			71.8	72.0	
S.D.	3.5	.2	.63	.01	.27	1.4	.38	1.2			10.6	2.3	

Sesuai dengan temuan diatas dapat di tarik kesimpulan bahwa soal yang termasuk dalam kategori item sulit terdapat pada nomor 4, 6, 13, 16, 8, 11, 3, 7, 9, 18, 1, 17, 19 dengan nilai *measure* 0 s.d. 1 . Sedangkan Item soal yang termasuk dalam kategori sangat mudah terdapat pada nomor 12, 20, 2, 23, 14, 21, 5, 10, 22, 24, 12 dengan nilai *measure* < -1 .

d. Daya beda

Pada model rasch menggunakan bantuan program Winsteps Boone, Staver, & Yale (2014) mengakatakan bahwa nilai outfit means-square, outfit z-standard, dan point measure correlation adalah kriteria yang digunakan untuk melihat tingkat kesesuaian butir. Jika item tersebut tidak memenuhi kriteria ada baiknya item tersebut diperbaiki atau diganti. Panduan untuk menilai kriteria kesesuaian butir menurut Boone, et al (2014) adalah sebagai berikut:

- Nilai Outfit Mean Square (MNSQ) yang diterima : $0,5 < \text{MNSQ} < 1,5$
- Nilai outfit Z-standard (ZSTD) yang diterima: $-2,0 < \text{ZSTD} < +2,0$

- Nilai Point Measure Correlation yang diterima: $0,4 < pt \text{ measure corr} < 0,85$

Tabel 4

Hasil analisis daya beda pada
Winstep

Item STATISTICS: MEASURE ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MODEL MEASURE	S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
4	11	30	.99	.43	1.20	1.13	.51	.32	.45	56.7	71.6	54	
6	11	30	.99	.43	.80	-1.1	.91	-.2	.58	45	76.7	71.6	56
13	12	30	.81	.42	1.14	.8	1.25	1.0	.34	46	60.0	70.7	S13
16	12	30	.81	.42	.69	-1.9	.58	-1.7	.70	46	80.0	70.7	S16
8	13	30	.64	.42	.78	-1.3	.68	-1.4	.64	46	83.3	70.1	S8
11	13	30	.64	.42	.70	-1.9	.60	-1.8	.70	46	83.3	70.1	S11
3	14	30	.46	.41	.79	-1.3	.68	-1.4	.64	46	73.3	69.8	S3
7	15	30	.29	.41	1.14	.9	1.31	1.3	.33	46	70.0	69.9	S7
9	15	30	.29	.41	.67	-2.2	.58	-2.0	.72	46	83.3	69.9	S9
18	15	30	.29	.41	1.15	.9	1.57	2.1	.28	46	76.7	69.9	S18
1	16	30	.12	.41	1.09	.6	1.05	.3	.40	46	70.0	70.2	S1
17	16	30	.12	.41	1.00	.1	.92	-.2	.47	46	70.0	70.2	S17
19	16	30	.12	.41	1.17	1.0	1.07	.4	.35	46	56.7	70.2	S19
12	17	29	-.14	.43	1.13	.8	1.12	.5	.37	46	65.5	71.5	S12
20	18	30	-.22	.42	1.35	1.8	1.25	.9	.24	45	60.0	71.3	S20
2	19	30	-.40	.43	1.86	3.8	2.23	3.1	-.24	45	43.3	72.7	S2
23	19	30	-.40	.43	.99	.0	.91	-.2	.46	45	70.0	72.7	S23
14	20	30	-.58	.43	1.13	.7	1.15	.5	.33	44	73.3	73.9	S14
21	20	30	-.58	.43	.81	-1.0	.69	-.9	.59	44	80.0	73.9	S21
5	21	30	-.77	.44	1.08	.5	.93	-.1	.39	43	70.0	74.9	S5
10	21	30	-.77	.44	.83	-.8	.71	-.7	.56	43	83.3	74.9	S10
22	21	30	-.77	.44	.78	-1.1	.76	-.6	.58	43	90.0	74.9	S22
24	21	30	-.77	.44	1.04	.3	1.00	.1	.40	43	70.0	74.9	S24
15	23	30	-1.19	.47	.64	-1.7	.43	-1.4	.68	40	76.7	78.4	S15
MEAN	16.6	30.0	.00	.43	1.00	.0	.98	-.1			71.8	72.0	
S.D.	3.5	.2	.63	.01	.27	1.4	.38	1.2			10.6	2.3	

TABLE 13.3 C:\Users\HP\Desktop\nekap_jawaban sis ZOU968WS.TXT May 23 13:51 2023
 INPUT: 30 Person 24 Item REPORTED: 30 Person 24 Item 2 CATS WINSTEPS 3.73

Dari hasil temuan diatas, dapat kita lihat bahwa hasil analisis daya beda pada winstep menunjukkan bahwa diantara ke tiga kriteria pada tabel tidak ada yang tidak memenuhi kriteria kesesuaian butir. Artinya daya beda pada butir soal PAS mata pelajaran PPKn kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo sudah menunjukkan daya beda yang bagus dan dapat diterima.

e. Pengecoh

Pada analisis butir soal dengan model Rasch dengan bantuan

program Winsteps efektivitas pengecoh dilihat dari nilai Data Count %, apabila nilai tersebut lebih dari sama dengan 5 % maka pengecoh berfungsi. Selain itu, efektivitas pengecoh soal dikatakan berfungsi apabila nilai Average Ability bernilai negatif, artinya pengecoh dipilih oleh peserta dengan kemampuan rendah

Tabel 5

Hasil analisis pengecoh pada

Item CATEGORY/OPTION/DISTRATOR FREQUENCIES: MEASURE ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTFIT MNSQ	PTMEA CORR.	Item
4	D	0	1	3	-2.09	.1	-.39	.54	
	C	0	11	37	-.62	.16	.4	-.61	
	B	0	7	23	1.35	.50	4.2	.50	
	A	1	11	37	.78*	.11	.8	.32	
6	B	0	11	37	-.43	.25	.6	-.48	S6
	D	0	6	20	-.15	.33	.7	-.20	
	C	0	2	7	.86	.10	1.5	.13	
	A	1	11	37	1.18	.35	1.0	.58	
13	C	0	8	27	-.42	.27	.6	-.38	S13
	B	0	3	10	-.06	.12	.6	-.11	
	A	0	7	23	.47	.45	1.5	.06	
	D	1	12	40	.78	.39	1.5	.34	
16	C	0	3	10	-.82	.63	.4	-.32	S16
	A	0	10	33	-.74	.16	.4	-.64	
	B	0	5	17	.68	.13	1.4	.15	
	D	1	12	40	1.30	.24	.5	.70	
8	C	0	3	10	-1.34	.37	.2	-.48	S8
	B	0	10	33	-.20	.27	.8	-.31	
	A	0	4	13	.06	.36	.9	-.08	
	D	1	13	43	1.15	.26	.7	.64	
11	C	0	10	33	-.54	.15	.4	-.52	S11
	B	0	4	13	-.32	.52	.7	-.21	
	D	0	3	10	-.05	1.02	1.2	-.10	
	A	1	13	43	1.23	.23	.6	.70	
3	B	0	7	23	-.51	.36	.7	-.39	S3
	A	0	9	30	-.29	.33	.8	-.34	
	C	1	14	47	1.09	.23	.6	.64	

Winstep

7	C	0	9	30	-.31	.28	.7	-.35	S7
	D	0	3	10	-.09	.09	.7	-.11	
	A	0	3	10	.64	.27	1.5	.10	
	B	1	15	50	.68	.36	1.7	.33	
9	C	0	2	7	-.95	1.13	.5	-.29	S9
	B	0	6	20	-.53	.20	.5	-.36	
	D	0	7	23	-.41	.33	.7	-.34	
	A	1	15	50	1.13	.22	.6	.72	
18	B	0	7	23	-.59	.18	.5	-.43	S18
	A	0	8	27	.48	.71	4.0	.09	
	D	1	15	50	.63	.11	.8	.28	
1	C	0	11	37	-.25	.29	.9	-.37	S1
	A	0	3	10	.05	.51	1.0	-.07	
	B	1	16	53	.74	.30	1.2	.40	
17	D	0	3	10	-1.12	.16	.3	-.41	S17
	B	0	4	13	-.48	.64	.8	-.27	
	C	0	7	23	.20	.20	1.1	-.05	
	A	1	16	53	.82	.29	1.0	.47	
19	A	0	2	7	-1.20	.24	.2	-.35	S19
	D	0	3	10	-.05	1.02	1.4	-.10	
	B	0	9	30	.08	.24	1.0	-.13	
	C	1	16	53	.68	.30	1.2	.35	
12	A	0	6	21	-.78	.20	.4	-.48	S12
	C	0	6	21	.37	.28	1.4	.03	
	D	1	17	59	.67	.31	1.4	.37	
	MISSING ***		1	3#	.10			-.03	
20	D	0	3	10	-.61	.36	.5	-.26	S20
	C	0	9	30	.20	.39	1.5	-.06	
	B	1	18	60	.51	.29	1.3	.21	

Hasil Analisis Menggunakan Iteman

2	D	0	1	3	-2.09	.1	-.39	S2
	A	0	1	3	.10	1.0	-.03	
	C	0	9	30	1.03	.21	3.0	.42
	B	1	19	63	.09*	.27	1.7	-.24
23	C	0	4	13	-.70	.27	.5	-.34
	D	0	7	23	-.22	.45	1.1	-.25
	A	1	19	63	.71	.25	1.0	-.46
14	B	0	3	10	-.96	.00	.3	-.37
	D	0	7	23	.08	.50	1.6	-.11
	A	1	20	67	.57	.24	1.0	.33
21	C	0	7	23	-.69	.36	.7	-.48
	B	0	3	10	-.60	.42	.6	-.26
	A	1	20	67	.79	.22	.8	-.59
5	B	0	4	13	-.69	.31	.5	-.34
	D	0	3	10	-.58	.38	.6	-.26
	A	0	2	7	.56	.00	1.6	.06
	C	1	21	70	.60	.26	1.2	.39
10	D	0	2	7	-.99	1.09	.6	-.30
	A	0	7	23	-.60	.31	.7	-.43
	B	1	21	70	.73	.22	.8	-.56
22	A	0	1	3	-1.44	.2	-.28	S22
	C	0	3	10	-1.34	.37	.3	-.48
	B	0	5	17	-.19	.47	1.2	-.19
	D	1	21	70	.74	.21	.7	-.58
24	A	0	9	30	-.40	.36	1.0	-.40
	C	1	21	70	.60	.24	1.0	.40
15	B	0	7	23	-1.12	.16	.3	-.68
	A	1	23	77	.74	.20	.7	-.68

* Average ability does not ascend with category score
Missing % includes all categories. Scored % only of scored categories

Dari hasil analisis data dengan model rasch menggunakan bantuan program Winsteps, maka soal yang bersifat mengecoh pada 24 soal dengan pembuktian bahwa ada 14 soal memiliki kegunaan untuk mengecoh dengan kemampuan yang baik dengan angka 2,4,5,6,7,8,9,11,13,26.17.18.19, serta 22. Sesuai paparan diatas, 24 soal yang telah dianalisa, ditemukan bahwa terdapat soal untuk mengecoh serta ada yang tidak berguna sesuai dengan mestinya. Simpulannya ialah kegunaan soal dengan sifat mengecoh jika seluruh soal atau 3 soal yang berguna untuk mengecoh berguna serta ada yang tidak memiliki manfaat dengan 1, 2 serta 3 soal tipe mengecoh yang tidak memiliki guna.

Teori Tes Klasik (TTK) ini merupakan cara untuk melakukan pengujian pada soal Mata Pelajaran Pendidikan Pancasila dan Kewarganegaraan memakai bantuan dari aplikasi ITEMAN versi 4.3, informasi yang diterima setelah melaksanakan analisis ini ialah kualitas dari soal yang telah disusun oleh tenaga pendidik serta tingkatan kelayakan soal tersebut untuk bisa digunakan keberlanjutannya dalam pengaplikasian dalam masa evaluasi belajar untuk peserta didik yang diadakan dua kali dalam satu semester:

a. Reliabilitas

Reliabilitas berasal dari kata reliabilitas, yang berarti sesuatu yang dapat dipercaya. Sejalan dengan pendapat Drost bahwa *"reliability is a major concern when a psychological test is used to measure some attribute or behavior"*(Drost, 2012:106). Menurut definisi ini, reliabilitas didefinisikan sebagai kepercayaan, ketergantungan, konsistensi, atau stabilitas.

Pada penelitian kali ini penulis menguji reabilitas soal dengan

menggunakan aplikasi ITEMAN versi 4.3 dan temuan dari uji reabilitas menggunakan aplikasi ITEMAN versi 4.3 pada 24 soal ialah 0,839 yang digolongkan dalam hasil yang tinggi,

Score	Frequency
0,0 to 0,1	0
0,1 to 0,2	0
0,2 to 0,3	0
0,3 to 0,4	2
0,4 to 0,5	5
0,5 to 0,6	7
0,6 to 0,7	4
0,7 to 0,8	1
0,8 to 0,9	0
0,9 to 1,0	0

paparannya ialah:

Tabel 6
 Hasil analisis reliabilitas melalui
Iteman

Score	Alpha	SE M	Split - Half (Random)	Split-Half (First - Last)	Split-Half (Odd - Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0,839	2,144	0,636	0,648	0,614	0,777	0,786	0,761

Paparan diatas menyatakan bahwa temuan dari nilai Alpha sejumlah 0.839 dengan artian realibilitas soal yang diuji dengan kategori pilihan ganda dengan 4 opsi jawaban

tergolong pada soal dalam golongan reliabilitas angka capaian tinggi. Sesuai paparan tabel diatas, maka nilai dari pengujian soal pilihan ganda memiliki nilai 0.839. Koefisien realibitas soal ini ialah sejumlah 0,71-0,90 yang tergolong pada klasifikasi tinggi.

b. Tingkat kesukaran

Temuan dari pengujian Tingkat Kesukaran yang berhubungan dengan soal yang diuji menggunakan aplikasi ITEMAN versi 4.3 diperoleh dari hasil ialah:

Tabel 7
 Hasil analisis tingkat kesukaran pada
Iteman

Temuan dari tabel yang telah dipaparkan ialah ada 5 soal memiliki indikasi soal dengan golongan yang mudah untuk diselesaikan, 12 soal masuk pada golongan yang menengah dalam tingkat kesulitan pemecahannya serta 2 soal masuk dalam klasifikasi tingkat kerumitan tinggi untuk dipecahkan oleh peserta didik.

c. Daya beda

Daya pembeda soal merupakan kemampuan sebuah soal untuk membedakan antara peserta didik yang memiliki kemampuan tinggi dan memiliki

kemampuan rendah. Indeks daya pembeda soal yang tinggi menunjukkan semakin mampu soal untuk membedakan kemampuan peserta didik.. Kualitas soal akan mengungkapkan apakah soal yang dihasilkan memiliki berbagai keunggulan yang dapat digunakan untuk tes.

Temuan dari pengujian yang dilaksanakan terhadap Indeks Daya Beda (IDB) soal yang telah disusun oleh tenaga pendidik ialah soal dengan kualitas daya beda dalam kondisi baik memperoleh nilai lebih dari 0,41 - 0,70.

Tabel 3
 Hasil analisis daya beda menggunakan
Iteman

Score	Frequenc y
-0,4 to -0,3	1
-0,3 to -0,2	0
-0,2 to -0,1	0
-0,1 to 0,0	0
0,0 to 0,1	1
0,1 to 0,2	0
0,2 to 0,3	6
0,3 to 0,4	6
0,4 to 0,5	0
0,5 to 0,6	6
0,6 to 0,7	3
0,7 to 0,8	1
0,8 to 0,9	0
0,9 to 1,0	0

Dari hasil analisis pengujian soal memakai aplikasi ITEMAN menyatakan bahwa soal sejumlah 6 soal diklasifikasikan pada golongan baik. Soal dengan golongan daya beda menengah memiliki nilai 0,21 -0,40, ada 12 soal tergolong dalam kalsifikasi menengah, serta 6 soal tergolong dalam kalsifikasi soal yang tidak layak dipakai karena memiliki kualitas yang jelek dengan capaian angka 0,00 - 0,20.

d. Pengecoh

Pengujian yang dilaksanakan pada soal dengan karakteristik pengecoh ditemukan hasil yang dipaparkan dalam tabel, ialah

Tabel 4
Hasil analisis butir pengecoh pada
Iteman

Soal	Pengecoh	Hasil	Kategori
1.	A 0,100	3	Nilai sesuai
	B 0,533	16	Jawaban benar
	C 0,367	11	Nilai sesuai
	D 0,000	0	Nilai belum sesuai
2.	A 0,033	1	Nilai sesuai
	B 0,633	19	Jawaban benar
	C 0,300	9	Nilai sesuai
	D 0,033	1	Nilai sesuai
3.	A 0,300	9	Nilai sesuai
	B 0,233	7	Nilai sesuai
	C 0,467	14	Jawaban benar
	D 0,000	0	Nilai belum sesuai
4.	A 0,367	11	Jawaban benar
	B 0,233	7	Nilai sesuai
	C 0,367	11	Nilai sesuai
	D 0,033	1	Nilai sesuai
5.	A 0,067	2	Nilai sesuai
	B 0,133	4	Nilai sesuai
	C 0,700	21	Jawaban benar
	D 0,100	3	Nilai sesuai
6.	A 0,067	2	Nilai sesuai
	B 0,367	11	Nilai sesuai
	C 0,367	11	Jawaban benar
	D 0,200	6	Nilai sesuai
7.	A 0,100	3	Nilai sesuai
	B 0,500	15	Jawaban benar
	C 0,300	9	Nilai sesuai
	D 0,100	3	Nilai sesuai

8.	A 0,133	4	Nilai sesuai
	B 0,333	10	Nilai sesuai
	C 0,100	3	Nilai sesuai
	D 0,433	13	Jawaban benar
9.	A 0,500	15	Jawaban benar
	B 0,200	6	Nilai sesuai
	C 0,087	2	Nilai sesuai
	D 0,233	7	Nilai sesuai
10.	A 0,233	7	Nilai sesuai
	B 0,700	21	Jawaban benar
	C 0,000	0	Nilai belum sesuai
	D 0,087	2	Nilai sesuai
11.	A 0,433	13	Jawaban benar
	B 0,133	4	Nilai sesuai
	C 0,333	10	Nilai sesuai
	D 0,100	3	Nilai sesuai
12.	A 0,200	6	Nilai sesuai
	B 0,000	0	Nilai belum sesuai
	C 0,200	6	Nilai sesuai
	D 0,587	17	Jawaban benar
13.	A 0,233	7	Nilai sesuai
	B 0,100	3	Nilai sesuai
	C 0,287	8	Nilai sesuai
	D 0,400	12	Jawaban benar
14.	A 0,667	20	Jawaban benar
	B 0,100	3	Nilai sesuai
	C 0,000	0	Nilai belum sesuai
	D 0,233	7	Nilai sesuai
15.	A 0,767	23	Jawaban benar
	B 0,233	7	Nilai sesuai
	C 0,000	0	Nilai belum sesuai
	D 0,000	0	Nilai belum sesuai

16.	A 0,333	10	Nilai sesuai
	B 0,167	5	Nilai sesuai
	C 0,100	3	Nilai sesuai
	D 0,400	12	Jawaban benar
17.	A 0,533	16	Jawaban benar
	B 0,133	4	Nilai sesuai
	C 0,233	7	Nilai sesuai
	D 0,100	3	Nilai sesuai
18.	A 0,287	8	Nilai sesuai
	B 0,233	7	Nilai sesuai
	C 0,000	0	Nilai sesuai
	D 0,500	15	Jawaban benar
19.	A 0,087	2	Nilai sesuai
	B 0,300	9	Nilai sesuai
	C 0,533	16	Jawaban benar
	D 0,100	3	Nilai sesuai
20.	A 0,000	0	Nilai belum sesuai
	B 0,600	18	Jawaban benar
	C 0,300	9	Nilai sesuai
	D 0,100	3	Nilai sesuai
21.	A 0,667	20	Jawaban benar
	B 0,100	3	Nilai sesuai
	C 0,233	7	Nilai sesuai
	D 0,233	0	Nilai belum sesuai
22.	A 0,033	1	Nilai sesuai
	B 0,167	5	Nilai sesuai
	C 0,100	3	Nilai sesuai
	D 0,700	21	Jawaban benar
23.	A 0,633	19	Jawaban benar
	B 0,000	0	Nilai belum sesuai
	C 0,133	4	Nilai sesuai
	D 0,233	7	Nilai sesuai

Dari hasil analisis data menggunakan aplikasi ITEMAN versi 4.3, maka soal yang bersifat mengecoh ada 24 soal dengan pembuktian bahwa ada 21 soal memiliki kegunaan untuk mengecoh dengan kemampuan yang baik ditunjukkan pada angka berikut: 6,13,16,8,11,3,7,9,18,1,17,19,12,12, 20, 23, 14,21,5,10,24. Sesuai paparan diatas, 24 soal yang telah dianalisa, ditemukan bahwa terdapat soal untuk mengecoh serta ada yang tidak berguna sesuai dengan mestinya. Simpulannya ialah kegunaan soal dengan sifat mengecoh apabila seluruh soal atau 3 soal yang berguna untuk mengecoh berguna serta ada yang tidak memiliki manfaat dengan 1, 2 serta 3 soal tipe mengecoh yang tidak memiliki guna.

Perbandingan Hasil Analisis menggunakan Winstep dan Iteman

a. Validitas

Hasil perbandingan analisis validitas menggunakan aplikasi winstep dan iteman ialah diperoleh 1 soal yang dikatakan tidak valid dan validitas pada Iteman tidak tersedia.

b. Reliabilitas

Hasil perbandingan analisis reliabilitas soal PAS mata pelajaran PPKn kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo baik menggunakan winstep dan menggunakan iteman menghasilkan indeks reliabilitas yang sama yaitu sebesar 0,83

c. Tingkat kesukaran

Perbandingan tingkat kesukaran butir soal pada soal PAS mata pelajaran PPKn kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo pada aplikasi winstep menunjukkan 13 item soal dalam kategori item sulit dan 11 soal dalam kategori sangat mudah sedangkan menggunakan iteman ditemukan 5 soal dengan golongan yang mudah, 12 soal masuk pada golongan sedang serta 2 soal masuk dalam kategori tingkat kesukaran yang tinggi.

d. Daya Beda

Perbandingan daya beda pada soal PAS mata pelajaran PPKn kelas V T.A 2022/2023 pada Lembaga Pendidikan

Ma'arif NU Sidoarjo menggunakan winstep dan iteman terletak pada kriteria daya beda pada masing-masing program sehingga mendapatkan hasil daya pembeda butir soal sebesar 100% atau daya beda pada butir soal sudah bagus pada winstep dan 90% butir soal memiliki daya pembeda yang efektif pada iteman.

e. Pengecoh

Perbandingan pengecoh soal pada soal PAS mata pelajaran PPKn kelas V T.A 2022/2023 pada Lembaga Pendidikan Ma'arif NU Sidoarjo pada aplikasi winstep menunjukkan 21 item soal bersifat pengecoh dengan baik dan 3 soal yang pengecohnya tidak berfungsi sedangkan pada iteman menunjukkan 14 item dengan pengecoh yang efektif.

D. Kesimpulan

Kajian penelitian ini diadakan dengan tujuan agar dapat diketahui kualitas dari soal yang telah selesai dikerjakan oleh tenaga pendidik. Metode yang dipakai dalam kajian

penelitian ini ialah deskriptif kuantitatif. Siswa kelas V pada Lembaga Pendidikan Ma'arif NU Sidoarjo ialah objek dari kajian penelitian ini. Penilaian serta analisis terhadap soal yang dibuat oleh tenaga pendidik ini ialah Mata Pelajaran Pendidikan Pancasila dan Kewarganegaraan Kelas V untuk T.A 2022/2023 dengan menggunakan aplikasi Winstep dan Iteman 4.3.

Hasil analisis menggunakan aplikasi Winstep menunjukkan indeks validitas apabila sudah memenuhi syarat outfit MNSQ, Outfit ZSTD, serta Point Measure correlation (PT measure Corr) dan indeks reliabilitas sebesar 0,83. Karakteristik butir soal ditunjukkan dengan data kesukaran butir soal sebesar 54% berada pada kategori soal sulit dan 46% tingkat kesukaran butir soal sangat mudah, daya pembeda sebesar 100%, dan 58% pengecoh (*distractor*) telah berfungsi dengan baik. Sedangkan hasil analisis data menggunakan software ITEMAN 4.3 menunjukkan bahwa sebesar 80% tingkat kesukaran butir soal berada pada kategori sedang, 90% butir soal memiliki daya pembeda yang baik, dan 90% pengecoh (*distractor*) telah berfungsi dengan baik. Hasil analisis ini menunjukkan bahwa kedua hasil

analisis menggunakan Winstep dan Itean pada item soal Penilaian Akhir Semester (PAS) pelajaran PPKn menunjukkan “kemiripan” dan keduanya termasuk dalam kategori baik.

DAFTAR PUSTAKA

- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan (2nd ed.)*. Jakarta: Bumi Aksara.
- Azwar, S. (2012). *Tes prestasi fungsi pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar
- Bambang, S. 2014. *Model Rasch untuk Penelitian Sosial Kuantitatif*. Surabaya.
- Boone, W. J., Staver, R. J., & Yale, S. M. (2014). *Rasch Analysis in the Human Sciences*. London: Springer.
- Chellamani, C. &. (2013). *Analysisi of Test Items on Difficulty Level and Discrimination Index The Test For Reseach in Education*. International Journal of Social Science & Interdisciplinary Research, 189-193.
- Crocker, L. & Algina, J. 2008. *Introduction to Classical and Modern Test Theory*. Ohio: Cengage Learning
- Drost, E.A. (2012). *Validity and Reliability in Social Science Research*. Education Research and Perspectives, Vol.38, No.1,
- Fitri Alfarisa, Ima Ni'mah Chudari, & Firman Robiansyah. (2019). *Analisis Butir Soal IPS Kelas V Sekolah Dasar Menggunakan Software ITEMAN*. EduBasic Journal: Jurnal Pendidikan Dasar, 1(2): 100-106
- Kaplan, R.M. & Saccuzzo, D.P. (2005). *Psychological Testing: Principles, Applications, and Issues (6th edition)*. Belmont, CA: Thomson Wadsworth
- Khotimah, M & Sri W. (2014). *Analisis Kemampuan Peserta Didik dengan Model Rasch* : Universitas Negeri Yogyakarta
- Kubiszyn, T. & Borich, G.D. (2013). *Educational Testing and Measurement: Classroom Application and Practice (10th edition)*. Hoboken, NJ: John Wiley & Sons, Inc
- Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Nunnally, J.C. & Bernstein, I.H. 1994. *Psychometric Theory (Third Edition)*. New York: McGraw-Hill, Inc
- Puskur (2008). *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*. Jakarta: Balitbang Depdiknas
- Sudjana, N. (2011). *Penilaian Hasil Proses Belajar Meng-ajar*. Bandung: Remaja Rosdakarya
- Sugiyono. (2012). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabet
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment*

Pendidikan. Cimahi: Trim
Komunikata.

Wei, S., Liu, X., Wang, Z, & Wang, X.
2012. *Using rasch measurement
to develop a computer modeling-
based instrument to assess
students' conceptual
understanding of matter.* Journal
of Chemical Education, 89(3),
335-345