

PENGEMBANGAN BUTIR SOAL MATA KULIAH PENGUKURAN DENGAN PENDEKATAN MODEL RASCH

Rengga Hardiansyah

S1 Pendidikan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya

E-mail : mochammadhildad.20004@mhs.unesa.ac.id

Rachmad Syarifudin Hidayatullah

Pendidikan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya

E-mail : rachmadhidayatullah@unesa.ac.id

Abstrak

Pendidikan tinggi saat ini menuntut mahasiswa untuk memiliki kemampuan berpikir kritis dan kreatif dalam menghadapi tantangan yang semakin kompleks di masyarakat. Oleh karena itu, diperlukan pengembangan soal ujian yang tidak hanya menguji pemahaman konsep, tetapi juga kemampuan mahasiswa dalam menerapkan pengetahuan mereka pada situasi nyata. Dalam mata kuliah Alat Ukur, misalnya, mahasiswa diharapkan tidak hanya memahami teori dasar, tetapi juga menguasai teknik-teknik praktis dalam pengukuran. Penelitian ini bertujuan untuk mengembangkan butir soal yang dapat mengukur kedua aspek tersebut dengan tepat, khususnya dalam konteks mata kuliah Alat Ukur di Program Studi Pendidikan Teknik Mesin UNESA. Pengembangan butir soal dilakukan dengan menggunakan model ADDIE yang terdiri dari lima tahapan, yaitu analisis, desain, pengembangan, implementasi, dan evaluasi. Soal yang dikembangkan berupa soal pilihan ganda dan isian singkat yang menguji pemahaman mahasiswa tentang pengukuran dasar menggunakan jangka sorong dan micrometer yang sebanyak 20 butir. Sampel dipilih menggunakan metode purposive sampling dengan sampel sebanyak 50 mahasiswa Program Studi S-1 Pendidikan Teknik Mesin yang me program mata kuliah alat ukur. Hasil penelitian menunjukkan bahwa butir soal yang dikembangkan telah memenuhi kriteria validitas melalui validasi oleh expert judgement dan hasil uji terbatas, reliabilitas, serta mampu mengukur kompetensi mahasiswa sesuai dengan tuntutan kurikulum.

Kata Kunci: alat ukur, evaluasi, pengembangan butir soal, rasch model.

Abstract

Higher education today demands students to possess critical and creative thinking skills to face the increasingly complex challenges in society. Therefore, it is essential to develop exam questions that assess conceptual understanding and evaluate students' ability to apply their knowledge to real-life situations. In the case of the Measuring Instruments course, for example, students are expected not only to understand basic theory but also to master practical measurement techniques. This research aims to develop question items that accurately assess both aspects, specifically in the context of the Measuring Instruments course in the Mechanical Engineering Education Program at UNESA. The development of the question items follows the ADDIE model, consisting of five stages: analysis, design, development, implementation, and evaluation. The developed questions include multiple-choice and short-answer items that assess students' understanding of essential measurement using calipers and micrometers, totaling 20 items. The sample was selected using purposive sampling with 50 students from the S-1 Mechanical Engineering Education Program enrolled in the Measuring Instruments course. The research results show that the developed question items meet validity criteria through expert judgment validation, limited testing results, and reliability and can measure students' competencies according to curriculum requirements.

Keywords: measuring tools, evaluation, test item development, rasch model.

PENDAHULUAN

Perkembangan pemikiran manusia dalam memberikan batasan tentang makna dan pengertian pendidikan, setiap saat selalu menunjukkan adanya perubahan (Mubarak et al., 2021). Perubahan itu didasarkan atas berbagai temuan dan perubahan di lapangan yang berkaitan dengan semakin bertambahnya komponen sistem pendidikan yang ada. Berkembangnya pola pikir para ahli pendidikan, pengelola pendidikan dan pengamat pendidikan yang

membuahkan teori-teori baru (Ajiban, 2024). Kemajuan alat teknologi turut andil dalam mewarnai perubahan makna dan pengertian pendidikan tersebut. Pada saat yang sama, proses pembelajaran dan pendidikan selalu menjadi tren dan terus berlangsung. Karena itu, bisa jadi pandangan seseorang tentang makna atau pengertian pendidikan yang dianut oleh suatu negara tertentu, pada saat yang berbeda dan di tempat yang berbeda makna dan pengertian pendidikan itu justru tidak relevan (Suyana et al., 2024). Namun demikian, selama belum ada teori dan

temuan baru tentang makna dan pengertian pendidikan, maka teori dan temuan yang telah ada masih relevan untuk dimanfaatkan sebagai acuan.

Pendidikan merupakan usaha secara sadar untuk mewujudkan sesuatu pewarisan budaya dari satu generasi ke generasi yang lain (Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 103 Tahun 2014 Tentang Pembelajaran Pada Pendidikan Dasar Dan Pendidikan Menengah, 2014). Pendidikan menjadikan generasi ini sebagai sosok panutan dari pengajaran generasi yang terdahulu. Sampai sekarang ini, pendidikan tidak mempunyai batasan untuk menjelaskan arti pendidikan secara lengkap karena sifatnya yang kompleks seperti sasarannya yaitu manusia. Sifatnya yang kompleks itu sering disebut ilmu pendidikan (Nur, 2022). Ilmu pendidikan merupakan kelanjutan dari pendidikan. Ilmu pendidikan lebih berhubungan dengan teori pendidikan yang mengutamakan pemikiran ilmiah. Pendidikan dan ilmu pendidikan memiliki keterkaitan dalam artian praktik serta teori. Sehingga, dalam proses kehidupan manusia keduanya saling berkolaborasi.

Pendidikan merupakan unsur yang sangat penting dalam membangun dan mengembangkan potensi diri setiap manusia. Seperti yang disebutkan dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional pada pasal 1 mengemukakan bahwa pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara. Selain itu di dalam Undang-Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional pasal 15 mengemukakan bahwa pendidikan kejuruan merupakan pendidikan yang menyiapkan terbentuknya keterampilan, kecakapan, pengertian, perilaku, sikap, kebiasaan kerja dan apresiasi terhadap bidang-bidang yang dibutuhkan oleh masyarakat dunia usaha/industri (Undang - Undang Republik Indonesia Nomor 20 Tahun 2003, 2003). Dunia pendidikan khususnya pada pendidikan kejuruan, baik menengah kejuruan maupun perguruan tinggi di bidang teknik dimana siswa atau mahasiswa dituntut untuk dapat berfikir kritis dan kreatif dalam memecahkan sebuah permasalahan yang diberikan oleh dosen atau guru.

Dengan berkembangnya pola pikir manusia, maka harus diimbangi dengan sesuatu yang dapat menyesuaikan pola pikir yang dimiliki oleh manusia. Untuk di dunia pendidikan seperti contoh yaitu soal atau tes yang dapat menyesuaikan pola pikir manusia di era saat ini, soal atau tes perlu di kembangkan atau di

modifikasi agar siswa dapat berpikir secara kritis dan kreatif (Saely & Shaleh, 2023; Umaroh et al., 2022).

Secara harfiah kata evaluasi berasal dari bahasa inggris yaitu evaluation. Evaluasi adalah suatu kegiatan atau proses yang sistematis, berkelanjutan menyeluruh dalam rangka pengendalian penjaminan, dan penetapan kualitas (nilai dan arti) berbagai komponen pembelajaran berdasarkan pertimbangan dan kriteria tertentu.

Ada tiga istilah yang sering digunakan dalam evaluasi yaitu tes, pengukuran dan penilaian. Tes merupakan salah satu alat untuk melakukan pengukuran, yaitu alat untuk mengumpulkan informasi karakteristik suatu objek. Pengukuran merupakan kuantifikasi atau penetapan angka tentang karakteristik atau keadaan individu menurut aturan-aturan tertentu. Penilaian merupakan kegiatan menafsirkan atau mendeskripsikan hasil pengukuran sedangkan evaluasi merupakan penetapan nilai atau implikasi perilaku.

Soal atau tes yang digunakan untuk melatih siswa atau mahasiswa agar dapat berfikir secara kritis dan kreatif yaitu di berikan soal berupa soal analisis atau yang biasa disebut dengan soal HOTS (High Order Thinking Skill) (Miharja et al., 2019; Novriandami et al., 2023; Pradana & Mawardi, 2021; Syadiah & Hamdu, 2020). Kriteria untuk soal HOTS dapat mencakup aspek berpikir kritis, berpikir kreatif, dalam menguasai keterampilan abad 21. Penilaian merupakan akhir dari proses pembelajaran yang dapat dilakukan melalui ujian (Syadiah & Hamdu, 2020).

Instrumen tes lebih sering digunakan di sekolah baik pada Penilaian Tengah Semester maupun Penilaian Akhir Semester. Instrumen ini dinilai praktis karena mengandung ringkasan dari materi yang telah diajarkan oleh seorang guru atau dosen dalam bentuk soal-soal. Soal yang bermutu yaitu soal yang dapat memberikan informasi yang sesuai dengan tujuan pembelajaran diantaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru atau dosen, maka dari itu diperlukan instrumen yang ideal.

Instrumen tes dapat dikatakan ideal apabila memiliki validitas dan reliabilitas yang tinggi (Saely & Shaleh, 2023; Simatupang & Maysarah, 2024). Validitas yang tinggi dapat dilihat dari ketepatan sebuah instrumen tes dalam mengukur kemampuan pembelajar dan reliabilitas yang tinggi dapat dilihat pada ketetapan instrumen tersebut selama digunakan untuk menguji kemampuan belajar. Pada kenyataannya, kebanyakan instrumen tes masih belum diketahui kualitasnya karena jarang dilakukan uji kualitas soal. Akibatnya terjadilah penilaian semu yang berdampak tidak terukur nya kemampuan peserta didik sebenarnya.

Berdasarkan permasalahan di atas maka pada penelitian ini perlu mengembangkan soal yang menyesuaikan era saat ini serta pola pikir manusia pada

era saat ini. Yang mana manusia dituntut untuk berfikir kritis dan kreatif dalam memecahkan masalah.

Di pihak lain, perkembangan ilmu evaluasi dan pengukuran diiringi pula dengan perkembangan berbagai program yang dapat digunakan untuk menunjang sistem evaluasi dan pengukuran. Analisis kualitas instrumen dalam pendidikan dapat dilakukan melalui dua pendekatan yaitu klasik dan modern. Pada pendekatan klasik dipergunakan program iteman, program *Quest*, SPSS, dan lain-lain. Selanjutnya, pada pendekatan modern dipergunakan IRT (*Item Response Theory*), bilog, rascal, dan oscal. Analisis butir dapat dilakukan secara manual atau dengan bantuan aplikasi (menggunakan analisis butir soal).

Selain adanya konsep teori tes klasik dengan program ITEMAN, juga berkembang model mutakhir analisis butir soal dengan paradigma teori respon butir (*Item Response Theory*) dikenal adanya program quest yang digunakan khususnya untuk analisis butir soal dengan 1 parameter (1 PL) atau dikenal dengan Rasch Model serta ada juga program Bilog yang dapat digunakan untuk menganalisis butir soal sampai dua bahkan tiga parameter logistik (2 PL atau 3 PL). Program Bilog memang memiliki keunggulan dalam hal kalibrasi item karena seleksi yang digunakan lebih ketat.

Berdasarkan uraian di atas, peneliti tertarik untuk menginvestigasi pengembangan butir soal mata kuliah pengukuran dengan pendekatan model rasch.

METODE

Rancangan Penelitian

Penelitian ini menggunakan model pengembangan atau dalam bahasa inggris disebut dengan *Research and Development* (R&D). Penelitian pengembangan ialah metode penelitian yang digunakan untuk menghasilkan produk tertentu, dan untuk menguji keefektifan produk tersebut. Penelitian pengembangan ini menggunakan model ADDIE. Model ADDIE terdiri dari lima tahap diantaranya adalah Analisis (*Analyze*), Desain (*Design*), Pengembangan (*Development*), Pelaksanaan (*Implementation*) dan Evaluasi (*Evaluation*) (Creswell, 2009).

Populasi dan Sampel

Populasinya adalah mahasiswa Program Studi S-1 Pendidikan Teknik Mesin Universitas Negeri Surabaya dan sampelnya adalah 50 mahasiswa di Program Studi S-1 Pendidikan Teknik Mesin Universitas Negeri Surabaya tahun akademik 2023/2024 yang memprogram mata kuliah alat ukur. Teknik pengambilan sampel yang digunakan adalah *purposive sampling* (Fraenkel, Jack R., Wallen, 2022).

Pengumpulan Data

Pada penelitian ini pengumpulan data dilakukan menggunakan metode tes berupa soal pilihan ganda sebanyak 20 butir yang diberikan kepada 50 mahasiswa di Program Studi S-1 Pendidikan Teknik Mesin Universitas Negeri Surabaya tahun akademik 2023/2024 yang memprogram mata kuliah alat ukur. Penelitian ini juga menggunakan dokumentasi sebagai sarana untuk mengumpulkan data terkait implementasi model pembelajaran tersebut. Selain itu, validasi instrumen oleh tiga *expert judgement* juga dilakukan untuk mengukur validitas instrumen sebelum digunakan.

Teknik Analisis Data

Pengambilan data validitas teoritis menggunakan metode observasi berupa lembar telah butir soal sesuai dengan kaidah penulisan soal yang divalidasi oleh validator ahli. Selanjutnya hasil jawaban tes dianalisis secara kuantitatif deskriptif menggunakan bantuan *software* minstep versi 5.8.1.0 sehingga mendapatkan data reliabilitas dan validitas (Engelhard, 2013).

Tabel 1. Kategori nilai person reability dan item reability pengujian butir soal winstep (Saely & Shaleh, 2023)

Nilai	Kategori
< 0,50	Buruk
0,50 – 0,60	Jelek
0,60 – 0,70	Cukup
0,70 – 0,80	Bagus
> 0,80	Bagus Sekali

Tabel 2. Kategori nilai cronbach alpha reability pengujian butir soal winstep (Saely & Shaleh, 2023)

Nilai	Kategori
< 0,67	Lemah
0,67 – 0,80	Cukup
0,81 – 0,90	Bagus
0,91 – 0,94	Bagus Sekali
> 0,94	Istimewa

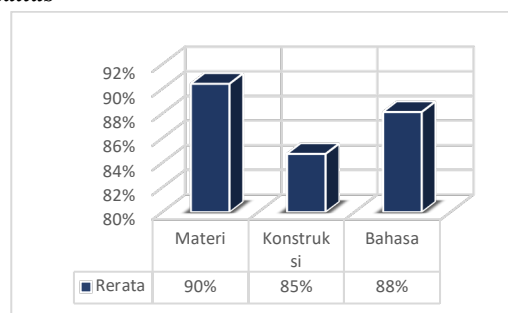
Tabel 3. Kriteria validitas kelayakan butir soal pemodelan rasch (Saely & Shaleh, 2023)

Aspek	Interval Nilai
<i>Outfit Mean Square</i> (MNSQ)	0,5 < MNSQ < 1,5
<i>Outfit Z-Standard</i> (ZSTD)	-2,0 < ZSTD < +2,0
<i>Point Measure Correlation</i> (Pt Measure Corr)	0,4 < Pt Measure Corr < 0,85

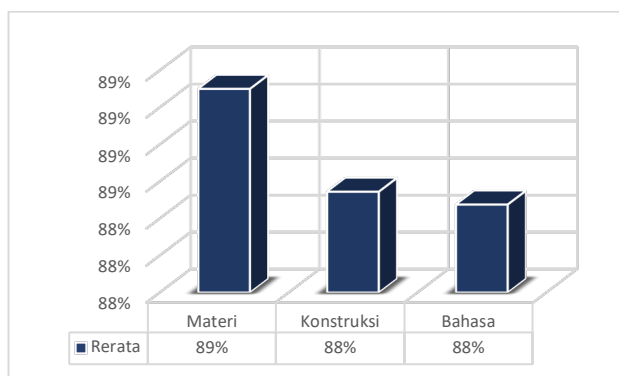
HASIL DAN PEMBAHASAN

Hasil

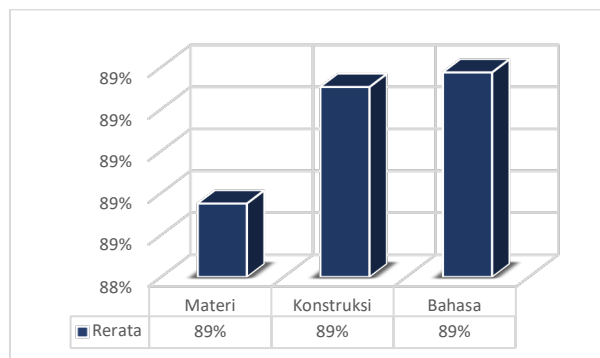
Validitas



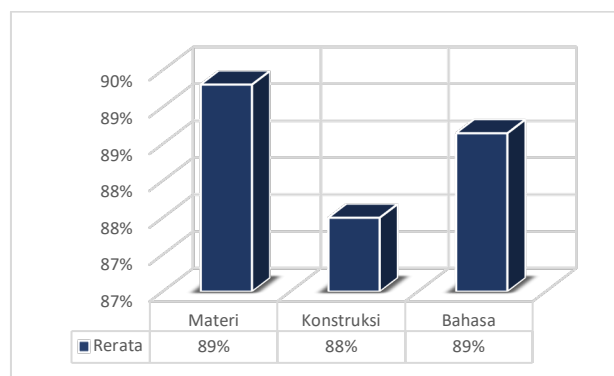
Gambar 1. Hasil validasi validator 1



Gambar 2. Hasil validasi validator 2



Gambar 3. Hasil validasi validator 3



Gambar 4. Akumulasi hasil validasi dari seluruh validator
Validasi ini berfungsi untuk memastikan bahwa instrumen butir soal yang dikembangkan dapat digunakan untuk melakukan penelitian. Hasil validasi oleh *expert judgement* menunjukkan bahwa total validasi aspek materi, konstruksi, dan bahasa dalam predikat sangat baik. Hal ini menunjukkan bahwa instrumen butir soal layak atau dapat digunakan untuk diterapkan pada proses pembelajaran untuk mengukur kemampuan peserta didik (Ajiban & Hidayatullah, 2024).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
14	2	50	5.54	.77	1.40	.82	5.05	2.00	A-.08	.30	96.0	95.9	Q14
4	47	50	-2.32	.63	1.20	.55	3.31	1.77	B-.06	.27	94.0	94.0	Q4
8	45	50	-1.68	.51	1.20	.69	2.08	1.25	C-.17	.34	88.0	90.3	Q8
19	18	50	2.09	.35	1.11	.73	1.57	1.54	D-.45	.53	76.0	74.8	Q19
2	44	50	-1.44	.48	1.12	.50	1.50	.80	E-.27	.36	90.0	88.5	Q2
1	32	50	.44	.35	1.31	1.71	1.22	.65	F-.37	.52	70.0	76.9	Q1
10	34	50	.19	.36	1.26	1.39	1.25	.66	G-.38	.51	64.0	77.5	Q10
20	37	50	-.22	.38	1.11	.63	1.20	.52	H-.40	.48	78.0	79.1	Q20
3	43	50	-1.22	.45	1.19	.78	.73	-.07	I-.33	.38	84.0	86.7	Q3
6	39	50	-.51	.39	.97	-.07	1.16	.45	J-.44	.45	86.0	81.2	Q6
18	36	50	-.08	.37	.95	-.21	.75	-.36	j-.53	.49	80.0	78.4	Q18
17	39	50	-.51	.39	.88	-.57	.68	-.36	i-.53	.45	82.0	81.2	Q17
13	36	50	-.08	.37	.87	-.66	.73	-.42	h-.56	.49	80.0	78.4	Q13
12	41	50	-.84	.42	.85	-.61	.53	-.51	g-.52	.42	86.0	83.6	Q12
7	42	50	-1.02	.43	.84	-.60	.49	-.49	f-.51	.40	82.0	84.9	Q7
5	48	50	-2.79	.75	.83	-.09	.35	-.30	e-.31	.22	96.0	96.0	Q5
11	31	50	.56	.35	.81	-1.12	.75	-.62	d-.62	.53	84.0	76.5	Q11
9	16	50	2.34	.36	.79	-1.37	.65	-.93	c-.64	.53	84.0	76.1	Q9
15	25	50	1.26	.34	.79	-1.42	.65	-1.26	b-.66	.54	80.0	74.7	Q15
16	33	50	.31	.36	.54	-3.12	.41	-1.74	a-.75	.51	90.0	77.2	Q16
MEAN	34.4	50.0	.00	.44	1.00	-.10	1.25	.13			83.5	82.6	
P.SD	11.3	.0	1.80	.13	.22	1.10	1.10	.99			7.8	6.9	

Gambar 5. Item misfit order

Item Fit Order adalah tingkat kesesuaian butir soal (validitas) yang digunakan untuk menjelaskan apakah butir soal berfungsi normal atau tidak normal melakukan pengukuran.

Gambar 5 menggambarkan hasil *item misfit order*, 20 butir soal dikategorikan dalam keadaan *fit* dan dipertahankan. Hal ini dikarenakan batas item dinyatakan *fit* jika memenuhi salah satu atau keduanya (OUTFIT MNSQ, ZSTD, dan PTMEASUR-AL CORR.).

Reliabilitas

SUMMARY OF 50 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	13.8	20.0	1.29	.71	.92	-.07	1.25	.26	
SEM	.5	.0	.22	.03	.05	.12	.24	.15	
P.SD	3.8	.0	1.56	.24	.33	.83	1.66	1.07	
S.SD	3.8	.0	1.57	.24	.33	.83	1.68	1.09	
MAX.	19.0	20.0	4.65	1.43	1.67	2.11	7.75	4.28	
MIN.	4.0	20.0	-2.02	.53	.20	-1.66	.04	-.95	
REAL RMSE	.78	TRUE SD	1.35	SEPARATION	1.74	Person RELIABILITY	.75		
MODEL RMSE	.75	TRUE SD	1.36	SEPARATION	1.82	Person RELIABILITY	.77		
S.E. OF Person MEAN	.22								

Gambar 6. Person Reability

Gambar 4.6 menunjukkan tingkat reliabilitas siswa berdasarkan hasil uji terbatas pada butir soal. Reliabilitas siswa, yang mencerminkan konsistensi mereka dalam menjawab butir soal, berada pada rentang 0,75–0,77 dan dikategorikan sebagai bagus.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	34.4	50.0	.00	.44	1.00	-.10	1.25	.13
SEM	2.6	.0	.41	.03	.05	.25	.25	.23
P.SD	11.3	.0	1.80	.13	.22	1.10	1.10	.99
S.SD	11.6	.0	1.85	.13	.22	1.13	1.13	1.02
MAX.	48.0	50.0	5.54	.77	1.40	1.71	5.05	2.00
MIN.	2.0	50.0	-2.79	.34	.54	-3.12	.35	-1.74
REAL RMSE	.49	TRUE SD	1.73	SEPARATION	3.55	Item	RELIABILITY	.93
MODEL RMSE	.46	TRUE SD	1.74	SEPARATION	3.79	Item	RELIABILITY	.93
S.E. OF Item MEAN	= .41							

Gambar 7. Item Relability

Gambar 4.7 menampilkan reliabilitas item (butir soal) berdasarkan hasil uji terbatas. Reliabilitas item, yang mencerminkan kualitas butir soal, mencapai nilai 0,93 dengan kategori bagus sekali.

Person RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .81 SEM = 1.64
STANDARDIZED (50 ITEM) RELIABILITY = .89

Gambar 8. Cronbach Alpha Relability

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT MATCH OBS%	EXP%	Item
14	2	50	5.54	.77	1.40	.82	5.05	2.00	-.08	.30	96.0	95.9	Q14
9	16	50	2.34	.36	.79	-1.37	.65	-.93	.64	.53	84.0	76.1	Q9
19	18	50	2.09	.35	1.11	.73	1.57	1.54	.45	.53	76.0	74.8	Q19
15	25	50	1.26	.34	.79	-1.42	.65	-1.26	.66	.54	80.0	74.7	Q15
11	31	50	.56	.35	.81	-1.12	.75	-.62	.62	.53	84.0	76.5	Q11
1	32	50	.44	.35	1.31	1.71	1.22	.65	.37	.52	70.0	76.9	Q1
16	33	50	.31	.36	.54	-3.12	.41	-1.74	.75	.51	90.0	77.2	Q16
10	34	50	.19	.36	1.26	1.39	1.25	.66	.38	.51	64.0	77.5	Q10
13	36	50	-.08	.37	.87	-.66	.73	-.42	.56	.49	80.0	78.4	Q13
18	36	50	-.08	.37	.95	-.21	.75	-.36	.53	.49	80.0	78.4	Q18
20	37	50	-.22	.38	1.11	.63	1.20	.52	.40	.48	78.0	79.1	Q20
6	39	50	-.51	.39	.97	-.07	1.16	.45	.44	.45	86.0	81.2	Q6
17	39	50	-.51	.39	.88	-.57	.68	-.36	.53	.45	82.0	81.2	Q17
12	41	50	-.84	.42	.85	-.61	.53	-.51	.52	.42	86.0	83.6	Q12
7	42	50	-1.02	.43	.84	-.60	.49	-.49	.51	.40	82.0	84.9	Q7
3	43	50	-1.22	.45	1.19	.78	.73	-.07	.33	.38	84.0	86.7	Q3
2	44	50	-1.44	.48	1.12	.50	1.50	.80	.27	.36	90.0	88.5	Q2
8	45	50	-1.68	.51	1.20	.69	2.08	1.25	.17	.34	88.0	90.3	Q8
4	47	50	-2.32	.63	1.20	.55	3.31	1.77	.06	.27	94.0	94.0	Q4
5	48	50	-2.79	.75	.83	-.09	.35	-.30	.31	.22	96.0	96.0	Q5
MEAN	34.4	50.0	.00	.44	1.00	-.10	1.25	.13			83.5	82.6	
P.SD	11.3	.0	1.80	.13	.22	1.10	1.10	.99			7.8	6.9	

Gambar 5. Item measure

Pembahasan

Pengembangan Butir Soal

Pengembangan butir soal untuk sistem pendidikan era 5.0 pada mata kuliah Alat Ukur di Jurusan Teknik Mesin UNESA dilakukan dengan memperhatikan karakteristik mahasiswa dan tuntutan kompetensi di era industri 4.0. Proses pengembangan soal ini mengikuti model ADDIE (Gagne, Briggs, & Wager, 1992) yang terdiri dari lima tahapan, yaitu Analysis, Design, Development, Implementation, dan Evaluation. Tahap analisis diawali dengan mengidentifikasi kebutuhan pembelajaran pengukuran dasar. Mahasiswa diharapkan menguasai teknik penggunaan alat ukur seperti jangka sorong dan

Gambar 4.8 menyajikan hasil Cronbach Alpha (KR-20) dari butir soal yang telah diuji pada uji terbatas. Nilai Cronbach Alpha, yang merepresentasikan koefisien reliabilitas instrumen butir soal, diperoleh sebesar 0,81 dan termasuk dalam kategori bagus.

Tingkat Kesulitan Butir Soal

Item measure adalah suatu pengukuran yang dihitung berdasarkan besar kecilnya nilai logit measure. Nilai ini digunakan untuk mengukur tingkat kesukaran butir soal. Didapatkan bahwa soal sangat sulit adalah nomor 14 dan soal sangat mudah adalah nomor 5. Namun, nilai logit dianggap ekstrem jika besarnya nilai logit measure diatas 2 atau di bawah -2. Berdasarkan gambar 4.6, butir soal nomor 2, 4, 5, 16, dan 18 terindikasi sebagai butir soal dengan nilai logit ekstrem yang artinya butir soal tersebut tidak memenuhi kriteria tingkat kesukaran yang baik.

micrometer, serta memahami prosedur kalibrasi dan teknik pengukuran yang tepat. Analisis ini memperhatikan kurikulum dan tujuan pembelajaran yang ditetapkan, sehingga soal yang dikembangkan mencakup topik-topik pengukuran dasar yang sesuai dengan kompetensi yang diharapkan, sebagaimana ditekankan dalam pendekatan Understanding by Design (Wiggins & McTighe, 2005). Selain itu, analisis juga menunjukkan bahwa pengukuran menggunakan alat-alat tersebut memerlukan pemahaman yang mendalam terhadap prinsip pengukuran serta keterampilan praktis dalam membaca hasil pengukuran dengan akurat. Masukan dari validator pada tahap ini juga diperhatikan, seperti penekanan pada contoh aplikasi

langsung, soal pengukuran praktis, dan referensi yang relevan.

Tahap desain melibatkan penyusunan kisi-kisi soal yang mencakup berbagai aspek pengukuran dasar, seperti pengukuran menggunakan jangka sorong, micrometer, serta prosedur kalibrasi alat ukur. Soal-soal dirancang dengan memperhatikan Taksonomi Bloom (Bloom et al., 1956) untuk memastikan soal mencakup berbagai tingkatan kognitif. Soal-soal dirancang dalam format pilihan ganda dan isian singkat untuk menguji pemahaman teori maupun keterampilan praktikum. Soal dirancang untuk mencakup semua topik yang relevan, seperti cara penggunaan alat ukur, pengukuran panjang dan ketebalan, serta pentingnya kalibrasi sebelum melakukan pengukuran. Setiap soal disusun untuk memastikan mahasiswa dapat mengaplikasikan teori pengukuran dalam praktik yang benar. Masukan validator pada tahap ini mengarahkan pada penambahan soal yang menguji pemahaman tentang pengaruh kesalahan kalibrasi, penggunaan gambar pada soal, dan soal pilihan ganda yang menguji aplikasi pengukuran dalam situasi nyata. Tahap pengembangan difokuskan pada perbaikan gambar-gambar yang digunakan dalam soal. Perbaikan ini bertujuan untuk meningkatkan kualitas visual dan mempermudah mahasiswa dalam memahami instruksi pengukuran yang diberikan.

Gambar diperbaiki dengan meningkatkan resolusi, menambah ketajaman dan detail, serta memberi label untuk menunjukkan bagian-bagian penting dari alat ukur. Penggunaan gambar dalam soal didasarkan pada teori belajar visual (Ware, 2008) yang menyatakan bahwa informasi visual dapat meningkatkan pemahaman dan retensi siswa. Masukan validator pada tahap ini meliputi pemeriksaan kontras warna gambar, penyajian alat ukur dalam kondisi yang lebih realistis, dan penambahan label atau keterangan pada gambar. Soal yang telah dikembangkan kemudian diimplementasikan pada 50 mahasiswa Program Studi Pendidikan Teknik Mesin dalam ujian praktik. Hasil ujian menunjukkan bahwa mahasiswa dapat menjawab soal-soal pengukuran dengan baik, namun terdapat sedikit kesulitan pada soal yang menguji kemampuan membaca skala micrometer. Evaluasi soal dilakukan dengan merujuk pada prinsip-prinsip pengukuran dan evaluasi pembelajaran (McMillan, 2013). Evaluasi soal menunjukkan bahwa tingkat keberhasilan pada soal tersebut rendah, yang mengindikasikan perlu adanya perbaikan lebih lanjut.

Pengembangan butir soal ini sejalan dengan prinsip-prinsip sistem pendidikan era 5.0 yang menekankan pada pengembangan keterampilan berpikir kritis, pemecahan masalah, kreativitas, dan inovasi untuk menyiapkan individu menghadapi tantangan *Society 5.0* (Kemdikbudristek, 2022). Soal-soal yang dikembangkan

tidak hanya menguji pemahaman konsep, tetapi juga menguji kemampuan mahasiswa dalam menerapkan pengetahuan tersebut dalam situasi nyata. Penggunaan gambar dan simulasi pada soal juga mendukung pembelajaran yang lebih interaktif dan menarik, sesuai dengan karakteristik mahasiswa di era digital. Pengembangan butir soal untuk sistem pendidikan era 5.0 pada mata kuliah Alat Ukur di Jurusan Teknik Mesin UNESA telah berhasil dilakukan dengan mengikuti model ADDIE. Soal yang dikembangkan telah memenuhi kriteria validitas dan reliabilitas, serta mampu mengukur kompetensi mahasiswa sesuai dengan tuntutan kurikulum.

Kelayakan Butir Soal

Kelayakan butir soal untuk sistem pendidikan era 5.0 pada mata kuliah Alat Ukur di Jurusan Teknik Mesin UNESA dievaluasi berdasarkan beberapa kriteria, yaitu validitas, reliabilitas, tingkat kesulitan, dan daya beda. Validitas soal diukur melalui *expert judgement* dan analisis *item fit order*, reliabilitas diukur melalui *person reliability*, *item reliability*, dan *Cronbach Alpha*, tingkat kesulitan diukur melalui *item measure*, dan daya beda dianalisis melalui *wright map*.

Hasil validasi *expert judgement* menunjukkan bahwa instrumen butir soal yang dikembangkan telah dinilai "sangat baik" oleh para ahli pada aspek materi, konstruksi, dan bahasa. Hal ini menunjukkan bahwa instrumen butir soal layak digunakan untuk diterapkan pada proses pembelajaran untuk mengukur kemampuan peserta didik. Selain itu, analisis *item fit order* menunjukkan bahwa 20 butir soal yang dikembangkan berada dalam keadaan fit dan dipertahankan. Hal ini menunjukkan bahwa butir-butir soal tersebut telah berfungsi dengan baik dalam mengukur kemampuan mahasiswa.

Analisis reliabilitas menunjukkan bahwa instrumen butir soal memiliki tingkat reliabilitas yang baik. *Person reliability* berada pada rentang 0,75-0,77 yang dikategorikan sebagai "bagus", *item reliability* mencapai nilai 0,93 dengan kategori "bagus sekali", dan *Cronbach Alpha* sebesar 0,81 yang termasuk dalam kategori "bagus". Hasil ini menunjukkan bahwa instrumen butir soal memiliki konsistensi internal yang baik dan dapat diandalkan dalam mengukur kemampuan mahasiswa.

Berdasarkan hasil tersebut dapat disimpulkan bahwa butir soal ini layak untuk digunakan. Hal ini sejalan dengan penelitian oleh (Zibar Parisu et al., 2024), dimana berdasarkan analisis kelayakan butir soal dengan pendekatan Rasch, instrumen yang dikembangkan memiliki reliabilitas tinggi (0,94). Dari 25 soal, hanya soal nomor 7 yang harus diganti karena tidak valid. Penelitian selaras juga dilakukan oleh (Novinda et al., 2019), bahwa instrumen tes pilihan ganda yang dikembangkan dalam penelitian ini tergolong layak digunakan untuk mengukur kemampuan peserta didik pada materi Gerak Lurus.

Instrumen ini memiliki validitas isi yang tinggi, reliabilitas istimewa, dan unidimensionalitas yang cukup. Sebanyak 26 butir soal memenuhi kriteria Model Rasch dan dapat digunakan untuk mengukur kemampuan siswa secara akurat. Hasil serupa lainnya oleh (Wibowo et al., 2023), dimana Instrumen soal IPA berbasis STEM yang dikembangkan dalam penelitian ini dinyatakan valid dan reliabel. Meskipun ada beberapa soal yang belum ideal, instrumen ini layak digunakan untuk mengukur kemampuan berpikir kritis siswa kelas IV materi kemagnetan dan gaya pada fase B.

Ucapan Terima Kasih

Peneliti ucapkan terimakasih kepada Koordinator, staff dosen, dan jajaran Program Studi S-1 Pendidikan Teknik Mesin Universitas Negeri Surabaya yang telah memfasilitasi penelitian ini.

PENUTUP

Simpulan

Berdasarkan hasil dan pembahasan diatas, dapat disimpulkan bahwa kelayakan butir soal berdasarkan validasi oleh *expert judgement* dan validitas serta reabilitas dari uji terbatas yang dilakukan, butir soal ini layak untuk digunakan. Materi dan konstruksinya valid dan reliabel, disajikan dengan bahasa yang mudah dipahami, serta dilengkapi dengan visualisasi seperti realita untuk pengukuran, menjadikannya butir soal yang efektif.

Saran

Berdasarkan hasil penelitian ini, diajukan beberapa saran sebagai berikut:

1. Pengembangan butir soal dikembangkan pada aspek materi yang lebih meluas lagi (tidak terbatas di alat ukur jangka sorong dan micrometer).
2. Penelitian dilanjutkan dengan analisis lebih mendalam tentang apakah terdapat faktor lain yang mempengaruhi hasil dari pengujian (seperti faktor asal sekolah dan faktor beda akelas pengujian).

DAFTAR PUSTAKA

Ajiban, M. H. (2024). Efektivitas Penerapan Media Pembelajaran Powerpoint Interaktif Terhadap Hasil Belajar Di Kelas Merdeka Belajar Teknik Kendaraan Ringan SMK Negeri 1 Jabon. *Innovative: Journal Of Social Science Research*, 4(1), 1141–1155. <http://j-innovative.org/index.php/Innovative/article/view/7539>
<https://j-innovative.org/index.php/Innovative/article/download/7539/5393>

Ajiban, M. H., & Hidayatullah, R. S. (2024). Pengaruh Model Discovery Learning pada Kelas Diferensiasi Merdeka Belajar di SMKN 1 Jabon. *Ejournal.Unesa.Ac.Id*, 03(01), 1–7.

<https://ejournal.unesa.ac.id/index.php/jurnal-pendidikan-teknik-mesin/article/view/58090>

Creswell, J. W. (2009). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches*. In *Muqarnas* (Third Edit, Vol. 8). SAGE Publications. <https://doi.org/10.1163/22118993-90000268>

Engelhard, G. (2013). *Invariant Measurement Using Rasch Models in the Social, Behavioral, and Health Sciences*. Routledge.

Fraenkel, Jack R., Wallen, N. E. (2022). *How to Design and Evaluate Research in Education Eleventh Edition*. In *McGraw-Hill Higher Education* (Issue 0).

Miharja, F. J., Hindun, I., & Fauzi, A. (2019). Critical thinking, metacognitive skills, and cognitive learning outcomes: a correlation study in genetic studies. *Biosfer*, 12(2), 135–143. <https://doi.org/10.21009/biosferjpb.v12n2.135-143>

Mubarok, A. A., Aminah, S., Sukanto, S., Suherman, D., & Berlian, U. C. (2021). Landasan Pengembangan Kurikulum Pendidikan di Indonesia. *Jurnal Dirosah Islamiyah*, 3(1), 103–125. <https://doi.org/10.47467/jdi.v3i2.324>

Novinda, M. R. R., Silitonga, H. T. M., & Hamdani. (2019). Pengembangan tes pilihan ganda menggunakan model Rasch materi gerak lurus kelas X Pontianak. *Jurnal Pendidikan Dan Pembelajaran*, 8(6), 1–11. <https://doi.org/https://doi.org/10.26418/jppk.v8i6.33452>

Novriandami, A., Jannah, M., Arrahma, A., Randa, G., & Nasir*, M. (2023). Development of Microsoft Power Point Interactive Media Based on Visual Basic for Application as Middle School Science Learning Media. *Jurnal Pendidikan Sains Indonesia*, 11(3), 528–544. <https://doi.org/10.24815/jpsi.v11i3.29927>

Nur, A. (2022). Kapitalisme Pendidikan dan Reinventing Paradigma Pendidikan Indonesia. *Al-Ubudiyah: Jurnal Pendidikan Dan Studi Islam*, 3(1), 69–84. <https://doi.org/10.55623/au.v3i1.94>

Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 103 Tahun 2014 Tentang Pembelajaran Pada Pendidikan Dasar Dan Pendidikan Menengah. Menteri Kesehatan Republik Indonesia Peraturan Menteri Kesehatan Republik Indonesia 2004 (2014).

Pradana, F. A. P., & Mawardi, M. (2021). Pengembangan Instrumen Penilaian Sikap Disiplin Menggunakan Skala Likert dalam Pembelajaran Tematik Kelas IV SD. *Fondatia*, 5(1), 13–29. <https://doi.org/10.36088/fondatia.v5i1.1090>

Saely, E., & Shaleh. (2023). Analisis Butir Soal Penilaian Akhir Semester (Pas) Pada Mata Pelajaran PPKN Menggunakan Aplikasi Winstep Dan Item And Tes Analysis (Iteman) 4.3. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 08(03), 415–430. <https://doi.org/https://doi.org/10.23969/jp.v8i3.10613>

Simatupang, A. R., & Maysarah, S. (2024). Pengembangan Perangkat Pembelajaran Berbasis

- Model Discovery Learning Untuk Meningkatkan Kemampuan Berfikir Kreatif *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 5(1), 9–21. <https://doi.org/https://doi.org/10.46306/lb.v5i1.539>
- Suyana, N., Pendidikan, P. K., Global, T., & Pendidikan, S. (2024). Kebijakan Pendidikan Nasional Menghadapi Tantangan Global Sebuah Analisis Strategis dan Prioritas. *Journal of Education Research*, 5(1), 620–634. <https://www.jer.or.id/index.php/jer/article/view/888>
- Syadiah, A. N., & Hamdu, G. (2020). Analisis rasch untuk soal tes berpikir kritis pada pembelajaran STEM di sekolah dasar. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(2), 138. <https://doi.org/10.25273/pe.v10i2.6524>
- Umaroh, S. T., Soeryanto, S., Warju, W., & Marniati, M. (2022). Analysis of middle semester exam subjects for Automotive Vehicle Chassis (AVC) to improve the implementation management of learning evaluation. *Jurnal Pendidikan Vokasi*, 12(1), 64–75. <https://doi.org/10.21831/jpv.v12i1.45472>
- Undang - Undang Republik Indonesia Nomor 20 Tahun 2003, 1 (2003).
- Wibowo, F. S., Patonah, S., & Sukanto, S. (2023). Pengembangan Instrumen Soal IPA Berbasis STEM untuk Siswa Kelas IV Materi Kemagnetan dan Gaya Pada Fase B di Sekolah Dasar. *Jurnal Basicedu*, 7(3), 1609–1619. <https://doi.org/10.31004/basicedu.v7i3.4975>
- Zibar Parisu, C. L., Ekadayanti, W., Sisi, L., Juwairiyah, A., & Kasmawati. (2024). Analisis Butir Soal Pengetahuan Dasar Matematika Menggunakan Pendekatan Rasch. *SCIENCE TECH: Jurnal Ilmu Pengetahuan Dan Teknologi*, 10(1), 36–45. <https://doi.org/10.30738/st.vol10.no1.a17112>