

Pengembangan Instrumen Tes Keterampilan Pemecahan Masalah pada Materi Kelistrikan dan Kemagnetan

Rahmawati^{1,2a)}, Nuryani Y. Rustaman^{1,b)}, Ida Hamidah^{1,c)} dan Dadi Rusdiana^{1,d)}

¹Program Studi Pendidikan IPA,
Sekolah Pascasarjana,, Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudhi no. 229 Bandung, Indonesia, 40154

²Program Studi Pendidikan Fisika,
Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Makassar,
Jl. Sultan Alauddin no. 259 Makassar, Indonesia, 90221

a) rahmawatisyam@unismuh.ac.id (corresponding author)

b) nuryanirustaman@upi.edu

c) idahamidah@upi.edu

d) dadirusdiana@upi.edu

Abstrak

Keterampilan pemecahan masalah merupakan salah satu dari tujuh keterampilan berpikir yang penting di kehidupan abad 21. Proser pembekalan keterampilan pemecahan masalah dapat dilakukan melalui pembelajaran dan asesmen. Peran asesmen dalam melatih keterampilan pemecahan masalah dapat dilakukan dengan mengembangkan instrumen penilaian yang baik dan dapat mengukur keterampilan pemecahan masalah, khususnya dalam fisika. Penelitian ini bertujuan untuk mengembangkan instrumen tes keterampilan pemecahan masalah pada materi kelistrikan dan kemagnetan. Instrumen tes disusun berdasarkan kerangka kerja indikator pemecahan masalah model IDEAL oleh Bransford & Stein (1993) yaitu identifikasi masalah, penentuan tujuan, representasi masalah, eksplor strategi, jalankan strategi, dan melihat serta mempelajari kembali solusi pemecahan masalah. Penelitian ini menggunakan model pengembangan instrumen tes Oriondo & Antonio (1984) yang terdiri dari lima tahap, yaitu planning, developing preliminary of product, preliminary field testing, dan main field testing. Sebanyak tiga puluh enam butir soal dikembangkan berdasarkan enam indikator keterampilan pemecahan masalah yang dikelompokkan menjadi enam pokok soal terkait materi hukum Ohm, energi dan daya penghantar arus listrik, hukum Kirchhoff, rangkaian RC tentang pengisian dan pengosongan kapasitor, medan magnet, dan gaya magnet. Validasi isi ditentukan berdasarkan penilaian para ahli dan dianalisis untuk menentukan nilai koefisien CVR dan I-CVI. Validasi empirik bertujuan untuk menentukan validitas dan reliabilitas instrumen tes yang dianalisis menggunakan pemodelan Rash berbantuan aplikasi program Winsteps version 3.68.2. Hasil pengembangan menunjukkan secara keseluruhan butir soal bersifat valid dengan tingkat reliabilitas tinggi (nilai Alpha Cronbach = 0,92). Berdasarkan penilaian para ahli dan ujicoba pada beberapa sampel maka disimpulkan bahwa instrumen tes keterampilan pemecahan masalah topik kelistrikan dan kemagnetan dapat digunakan.

Kata-kata kunci: Peluruhan alfa, instrumen tes keterampilan pemecahan masalah. Koefisien CVR dan I-CVI, model Rasch, program Winsteps, materi kelistrikan dan kemagnetan

PENDAHULUAN

Struktur analisis kerangka kerja keterampilan abad 21 menyebutkan bahwa terdapat empat kelompok keterampilan yang perlu dikuasai oleh manusia dalam menjalani kehidupan di abad 21, yaitu *ways of thinking*, *ways of working*, *tools for working*, dan *living in the world* [1, 2]. Salah satu kelompok keterampilan tersebut adalah kelompok keterampilan *ways of thinking*, yaitu kelompok keterampilan berpikir. Beberapa jenis yang termasuk dalam kelompok keterampilan berpikir diantaranya 1) kreatif dan inovatif; 2) berpikir kritis, pemecahan masalah, pengambilan keputusan; dan 3) belajar bagaimana untuk belajar, dan metakognitif [1–4].

Uraian di atas menunjukkan bahwa salah satu keterampilan berpikir yang penting dikuasai di abad ke-21 adalah keterampilan pemecahan masalah. Pernyataan ini sejalan dengan gagasan paradigma pembelajaran sains menurut kerangka National Science Teachers Association (1985) yang menekankan bahwa keterampilan pemecahan masalah merupakan keterampilan yang harus dikembangkan dalam pembelajaran sains [5]. Sejumlah penelitian terkait upaya pengembangan keterampilan pemecahan masalah telah dilakukan beberapa tahun belakangan ini dalam dunia pendidikan sains [6–10]. Keterampilan pemecahan masalah memiliki peranan penting pada pembelajaran sains termasuk fisika. Komponen dasar dari suatu permasalahan didasarkan pada adanya informasi yang diterima atau suatu kondisi (keadaan) tertentu (*givens*), suatu tujuan akhir yang diharapkan (*goal*), dan alat untuk memperoleh informasi dari satu keadaan tertentu ke keadaan lainnya (*operation*). *Problem solving* bertujuan tidak hanya dalam rangka menemukan jawaban yang benar tetapi juga merupakan suatu tindakan yang mencakup periode mental dan kemampuan [11]. Oleh karena itu, keterampilan pemecahan masalah perlu dilatihkan kepada peserta didik di setiap jenjang pendidikan.

Beragam penelitian terkait cara meningkatkan keterampilan pemecahan masalah peserta didik telah banyak dilakukan dengan memberikan berbagai inovasi dalam pembelajaran dan teknik asesmen [1,4]. Agar dapat memantau perkembangan tingkat keterampilan pemecahan masalah peserta didik, maka perlu dilakukan pengukuran terhadap capaian hasil keterampilan pemecahan masalah sebagai efek perlakuan dalam pembelajaran sains peserta didik, termasuk pembelajaran fisika [12–15]. Sebagian besar cakupan materi dalam pembelajaran fisika bersifat abstrak yang dapat dikontekstualkan dalam kehidupan sehari-hari peserta didik melalui pembelajaran dan asesmen sehingga hal ini berpotensi dapat digunakan dalam mengembangkan keterampilan pemecahan masalah peserta didik [16–18]. Beberapa diantaranya adalah materi kelistrikan dan kemagnetan.

Materi kelistrikan dan kemagnetan merupakan konsep penting yang harus diajarkan kepada peserta didik dalam melatih keterampilan pemecahan masalah mereka karena interaksi materi kelistrikan dan kemagnetan memainkan peran sentral dalam menentukan struktur kehidupan dan menjadi pondasi perkembangan teknologi penting dalam kehidupan [19, 20]. Oleh sebab itu, materi kelistrikan dan kemagnetan penting untuk dikembangkan dalam asesmen pembelajaran.

Instrumen tes untuk mengukur keterampilan pemecahan masalah pada konsep kelistrikan dan kemagnetan masih sangat kurang dikembangkan. Kebanyakan instrumen tes terkait materi kelistrikan lebih menitikberatkan pada pengetahuan konseptual kelistrikan dan kemagnetan. Beberapa contoh instrumen tes pengetahuan konseptual terkait materi kelistrikan dan kemagnetan yang telah dikembangkan di antaranya adalah Conceptual Survey of Electricity and Magnetism (CSEM), Brief Electricity and Magnetism Assessment (BEMA), Electricity and Magnetism Conceptual Assessment (EMCA), Colorado Upper-division Electrostatics (CUE), and Critical Thinking skills in Electricity and Magnetism (CTEM) [20]. Instrumen tes tersebut lebih fokus mengukur tingkat kesulitan peserta didik pada materi kelistrikan dan kemagnetan [11,20, 21, 22]. Sementara, instrumen tes untuk mengukur keterampilan pemecahan masalah terkait materi kelistrikan dan kemagnetan masih sangat sedikit dijumpai.

Hasil penelusuran terkait keterbatasan dan kelemahan instrumen tes kelistrikan dan kemagnetan yang sudah ada menjadi dasar tujuan penelitian ini, yaitu mengembangkan instrumen tes yang dapat mengukur keterampilan pemecahan masalah mahasiswa calon guru fisika. Instrumen tes disusun berdasarkan enam indikator yang dikembangkan dari kerangka pemecahan masalah model *IDEAL* [23]. Indikator keterampilan pemecahan masalah tersebut adalah identifikasi masalah, penentuan tujuan, representasi masalah, eksplor strategi, jalankan strategi, dan lihat dan pelajari kembali solusi yang ditemukan. Materi yang dipilih untuk dikembangkan dalam instrumen keterampilan pemecahan masalah pada penelitian ini adalah hukum Ohm, energi dan daya penghantar arus listrik, hukum Kirchhoff, rangkaian RC tentang pengisian dan pengosongan kapasitor, medan magnet, dan gaya magnet.

Suatu instrumen tes dikatakan baik jika memenuhi tiga karakteristik, yaitu valid, reliabel, dan usable [24–28]. Sejumlah pendekatan untuk menguji validasi instrumen tes, yaitu validasi konten, konstruk, dan validasi kriteria [25, 29]. Oleh karena itu, penelitian ini mencoba menjabarkan proses pengembangan dan validasi instrumen tes keterampilan pemecahan masalah materi kelistrikan dan kemagnetan.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian dan pengembangan model pengembangan R&D oleh Borg & Gall (2003) Borg & Gall (2003) [30] yang dimodifikasi menggunakan metode pengembangan instrumen yang dikembangkan oleh Oriondo & Antonio (1984) [31] dengan tahapan proses pengembangan terdiri dari tahap *planning*, *developing preliminary of product*, *preliminary field testing*, dan *main field testing*. Bentuk tes yang dikembangkan adalah uraian sebanyak tiga puluh enam butir soal berdasarkan enam indikator keterampilan pemecahan masalah yang dikelompokkan menjadi enam pokok soal terkait materi hukum Ohm, energi dan daya penghantar arus listrik, hukum Kirchhoff, rangkaian RC tentang pengisian dan pengosongan kapasitor, medan magnet, dan gaya magnet. Subjek penelitian ini adalah mahasiswa calon guru fisika sebanyak 63 orang. Durasi waktu yang diberikan dalam mengerjakan soal adalah 120 menit.

Teknik analisis data yang digunakan adalah analisis data kuantitatif berdasarkan data hasil uji keterbacaan instrumen tes pada lima orang mahasiswa calon guru fisika, validasi *expert judgement*, dan uji coba empirik. The instruments used in this study were readability assessment sheet, and validation sheet for expert judgment.

Instrumen lembar penilaian keterbacaan digunakan untuk memperoleh respon dari lima mahasiswa tentang tingkat keterbacaan instrumen tes keterampilan pemecahan masalah (TKPM). Secara umum, validasi tingkat keterbacaan instrumen TKPM dinilai pada tiga aspek utama, yaitu: (1) penggunaan kata dan kalimat butir soals, (2) penggunaan bahasa; dan (3) kejelasan petunjuk instrumen tes [12]. Teknik penskalaan digunakan dalam menilai setiap aspek penilaian. Skala penilaian yang digunakan adalah skala Likert. Dalam penskalaan, mahasiswa membaca setiap kalimat pertanyaan setiap butir soal dan memberikan penilaian berdasarkan hasil penilaian dengan membubuhi salah satu kolom di lembar angket penilaian keterbacaan instrumen tes [32–34]. Penilaian dilakukan dalam bentuk pemberian skor pada masing-masing aspek yang dinilai. Rentang skor yang disediakan dari skor 1 sampai 5 dengan deskripsi kategori skor yang disediakan. Skor 1 mengindikasikan sangat buruk, kategori buruk diberi skor 2, pemberian skor 3 jika dinilai cukup, untuk penilaian baik diberi skor 4, dan sangat baik dengan skor 5. Tabulasi data dilakukan dengan menghitung skor rata-rata tiap aspek yang dinilai. Selanjutnya, data kuantitatif dikonversi ke dalam bentuk kualitatif dalam bentuk empat kategori, yaitu sangat buruk, buruk, baik dan sangat baik. Tabel 1 menunjukkan konversi interval skor rata-rata penilaian keterbacaan instrumen tes.

Tabel 1 Konfensi Interval Skor Rata-rata

Interval skor rata-rata	Kategori
$X \geq \bar{X} + 1. SBx$	Sangat baik
$\bar{X} + 1. SBx > X \geq \bar{X}$	Baik
$\bar{X} > X \geq \bar{X} - 1. SBx$	Buruk
$X < \bar{X} - 1. SBx$	Sangat buruk

Berdasarkan Tabel 1, dilakukan interpretasi skor rata-rata uji keterbacaan untuk menarik suatu kesimpulan terhadap penilaian skala rating yang diperoleh. Interpretasi tingkat keterbacaan ditunjukkan Tabel 2.

Tabel 2 Interpretasi Skor Rata-rata Uji Keterbacaan Instrumen Tes

Interval skor rata-rata	Kategori tingkat keterbacaan	Keputusan
≥ 4	Sangat jelas	Tanpa revisi
$4 > X \geq 3$	Jelas	Sedikit revisi
$3 > X \geq 2$	Cukup jelas	Revisi banyak
$X < 2$	Tidak jelas	Ditolak

Hasil validasi *expert judgement* diolah secara kuantitatif dengan menentukan nilai koefisien kesepakatan para ahli menggunakan persamaan CVR dan I-CVI yang dirumuskan oleh Lawshe (1975) [35]. Analisis validitas isi CVR menggunakan formulasi:

$$CVR = \frac{N_e - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

Keterangan:

CVR = *Content Validity Ratio*/Rasio Validitas Isi

N_e = Jumlah ahli yang menyatakan relevan

N = Jumlah total ahli

Selanjutnya, analisis I-CVI dihitung dengan menggunakan formula:

$$I - CVI = \frac{N_e}{N} \quad (2)$$

Keterangan:

I-CVI = *Item Content Validity Index*

N_e = Jumlah ahli yang menyatakan relevan

N = Jumlah total ahli

Suatu butir dikatakan layak apabila koefien *CVR* bernilai pada rentang 0-1. Meskipun demikian, penentuan butir soal diterima atau ditolak dilakukan dengan membandingkan nilai *CVR* hasil hitung dengan nilai kritis *CVR* [36]. Nilai kritis *CVR* bergantung pada jumlah *reviewer*. Penelitian ini menggunakan sebanyak lima *reviewer* sehingga nilai koefisien *CVR* kritisnya adalah 0,99

Selanjutnya, nilai indeks validitas isi *item (I-CVI)* diinterpretasi dengan sejumlah kategori. Indeks validitas konten butir soal berada pada rentang 1-0 dengan tingkat kategori dibedakan atas tiga kategori [37] seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Kategorisasi Indeks Validitas Isi Item

Interval Indeks <i>I-CVI</i>	Kategori Validitas
$I-CVI_{hitung} \geq 0.79$	Sesuai
$0 \leq I-CVI_{hitung} < 0.79$	Revisi
$I-CVI_{hitung} < 0$	Dieliminasi

Analisis data kuantitatif juga dilakukan pada hasil uji coba lapangan untuk memperoleh tingkat validitas dan reliabilitas tes. Penentuan validitas dan reliabilitas tes menggunakan metode analisis teori butir soal (IRT) Rasch model dengan bantuan program *Winsteps version 3. 68. 2*. Analisis validasi konstruk instrumen tes dilakukan pada dua aspek, yaitu ringkasan statistik dan tingkat kesesuaian butir soal (*item fit*).

Ada beberapa kriteria yang perlu diamati dalam menentukan kualitas instrumen tes melalui analisis ringkasan statistik, meliputi sebagai berikut:

- 1) Membandingkan nilai *person measure* dengan *item measure* (dalam satuan *logit*) (nilai *item measure* selalu bernilai 0,0 *logit*). Jika nilai *logit person measure* lebih tinggi dari *item measure*, hal ini menunjukkan bahwa kecenderungan abilitas (kemampuan) responden lebih tinggi dibandingkan tingkat kesulitan butir soal. Demikian pula sebaliknya.
- 2) Data nilai koefisien *Alpha Cronbach*. Nilai ini mengukur reliabilitas instrumen, yaitu interaksi antara *person* dan *item* secara keseluruhan [38]. Selanjutnya, nilai koefisien *Alpha Cronbach* dikategorisasikan berdasarkan rentang nilai koefisiennya. Kategori tingkat reliabilitas *Alpha Cronbach* ditunjukkan pada Tabel 4.

Tabel 4. Kategorisasi Tingkat Reliabilitas *Alpha Cronbach*

Interval Nilai Koefisien Alpha Cronbach (Reliabilitas)	Kategori
$0,8 \leq \alpha$	Bagus sekali
$0,7 \leq \alpha < 0,8$	Bagus
$0,6 \leq \alpha < 0,7$	Cukup
$0,5 \leq \alpha < 0,6$	Jelek
$\alpha < 0,5$	Buruk

- 3) Data nilai *person* dan *item reliability*. Tingkat kualitas reliabilitas *person (person reliability)* dan reliabilitas butir soal (*item reliability*) dapat dibedakan dalam beberapa kategori berdasarkan nilai koefisien reliabilitasnya. Kategorisasi tingkat reliabilitas *person* dan *item* disajikan pada Tabel 5. memperlihatkan kategorisasi tingkat reliabilitas *person* dan *item*.

Tabel 5. Kategorisasi Tingkat Reliabilitas Person dan Item

Interval Nilai Koefisien Reliabilitas Person dan Item (r)	Kategori
$0,94 \leq r$	Istimewa
$0,91 \leq r < 0,94$	Sangat bagus
$0,80 \leq r < 0,91$	Bagus
$0,67 \leq r < 0,80$	Cukup
$r < 0,67$	Lemah

- 4) Data *person* dan *item separation*. Separasi *person* dan *item* bertujuan untuk mengelompokkan *person* (responden) berdasarkan tingkat kemampuannya terhadap *item*. Sementara itu, separasi *item* digunakan untuk memverifikasikan hierarki *item*. Makin besar nilai *separation*, maka kualitas instrumen dalam hal keseluruhan responden dan butir soal makin bagus karena bisa mengidentifikasi kelompok responden dan kelompok butir [39, 40].
- 5) Data nilai *person fit* dan *item fit* pada kolom *infit mnsq* dan *outfit mnsq*, serta nilai *infit Zstd* dan *outfit Zstd* yang mengikuti nilai ideal model Rasch (yaitu 1,00). Sementara itu, Nilai Z standar (*Zstd*) pada *infit Zstd* dan *outfit Zstd*, baik pada *person fit* maupun *item fit* mengacu pada nilai ideal yaitu 0,0.

Analisis selanjutnya adalah analisis pada aspek tingkat kesesuaian butir soal (*item fit*). Untuk melihat tingkat kesesuaian butir soal, maka ada beberapa kriteria yang harus dipenuhi, yaitu 1) nilai *outfit mean square (mnsq)* yang diterima: $0,5 < mnsq < 1,5$; 2) nilai *outfit Z-standard (Zstd)* yang diterima: $-2,0 < zstd < +2,0$; dan 3) nilai *point measure correlation (Pt mean corr)*: $0,4 < Pt\ measure\ corr < 0,85$ [39–41].

HASIL DAN DISKUSI

Hasil dan diskusi penelitian ini dibagi menjadi lima bagian berdasarkan tahapan pengembangan instrumen tes model Oriondo & Antonio (1984) [31], yaitu tahap *planning*, *develop preliminary of product*, *revision of product preliminary*, *preliminary field testing*, dan tahap *operational product revision*.

Tahap Planning

Ada beberapa langkah yang perlu dilakukan di tahap perencanaan (*planning*) dari model tahapan pengembangan instrumen tes Oriondo & Antonio (1984) [31]. Adapun langkah-langkah tahap perencanaan dalam pengembangan instrumen tes yaitu sebagai berikut:

- 1) Penentuan tujuan asesmen, tujuan asesmen ini adalah untuk mengukur keterampilan pemecahan masalah mahasiswa calon guru fisika pada materi kelistrikan dan kemagnetan dengan indikator keterampilan pemecahan masalah model Bransford & Stein (1993) [23] yang meliputi identifikasi masalah, penentuan tujuan, representasi masalah, eksplor strategi, jalankan strategi, dan melihat serta mempelajari kembali solusi pemecahan masalah.
- 2) Penetapan bentuk instrumen tes, instrumen tes keterampilan pemecahan masalah dikembangkan dalam bentuk uraian tertulis. Penetapan jenis tes ini didasarkan pada kelebihan tes uraian dibandingkan jenis tes lainnya, yaitu dapat menggali lebih jauh proses berpikir peserta didik [12–14, 42].
- 3) Pembuatan kisi-kisi butir soal, instrumen tes keterampilan pemecahan masalah terdiri atas enam pokok soal yang disusun berdasarkan materi hukum Ohm, energi dan daya penghantar arus listrik, hukum Kirchhoff, rangkaian RC pengisian dan pengosongan kapasitor, medan magnet, dan gaya magnet. Setiap pokok soal dibagi menjadi enam item yang mewakili masing-masing indikator keterampilan pemecahan masalah sehingga terdapat tiga puluh enam butir soal dengan sebaran soal ditunjukkan pada Tabel 6.
- 4) Penulisan naskah dan pedoman penskoran, naskah soal dikembangkan mengacu pada kisi-kisi yang telah dikembangkan sebelumnya. Naskah soal dirumuskan dalam bentuk soal uraian sebanyak enam pokok soal yang memuat keenam indikator pemecahan masalah. Untuk memudahkan penilaian hasil tes tertulis keterampilan pemecahan masalah, maka dikembangkan rubrik penskoran instrumen tes. Jenis rubrik penskoran yang dikembangkan berupa rubrik holistik berdasarkan kerangka kerja rubrik Jonassen (2004, 2011) [14, 42].

Tabel 6. *Blue Print* Instrumen Tes Keterampilan Pemecahan Masalah (TKPM)

Indikator KPM	Sebaran item pada tiap pokok bahasan						Jumlah
	Hukum Ohm	Energi & Daya Penghantar Arus Listrik	Hukum Kirchhoff	Rangkaian RC: Pengisian & Pengosongan Kapasitor	Medan Magnet	Gaya Magnet	
1. Identifikasi Masalah	1	1	1	1	1	1	6
2. Penentuan Tujuan	1	1	1	1	1	1	6
3. Representasi Masalah	1	1	1	1	1	1	6
4. Eksplor Strategi	1	1	1	1	1	1	6
5. Jalankan Strategi	1	1	1	1	1	1	6
6. Lihat dan Pelajari Kembali Solusi	1	1	1	1	1	1	6
Jumlah Item							36

Tahap *Developing Preliminary Form of Product*

Kisi-kisi instrumen tes keterampilan pemecahan masalah dan rubrik penskorannya yang telah dikembangkan selanjutnya dilakukan review oleh para ahli. Pada tahap ini, proses validasi isi instrumen tes dilakukan panel review oleh lima orang ahli dengan latar belakang pendidikan minimal doktor. Kelima validator ini memiliki keahlian dibidang asesmen pembelajaran fisika, konten fisika murni, dan ahli di bidang pembelajaran fisika. Kelima validator ini memberikan penilaian terhadap isi instrumen tes yang memuat tiga aspek utama, yaitu ketepatan konten, ketepatan konstruksi soal, dan unsur kebahasaan.

Skor yang diberikan untuk setiap aspek yang dinilai adalah 0 dan 1. Jika butir soal sesuai dengan aspek yang dinilai, maka diberi skor 1 dengan memberi tanda centang (\checkmark) pada kolom yang disediakan dan memberi skor 0 jika tidak sesuai dengan aspek penilaian dengan memberi tanda kali (X) pada kolom yang disediakan pada lembar *judgement*. Hasil *judgement* para ahli dianalisis menggunakan persamaan CVR dan I-CVI. Hasil analisis perhitungan menunjukkan bahwa nilai CVR sebesar 0,99 diterima untuk jumlah SME (Subject Matter Expert) atau validator sebanyak 5 berdasarkan ketentuan terkait nilai kritis CVR yang diperbolehkan. Sementara, nilai koefisien I-CVI diperoleh sebesar 0,99 dengan kategori sesuai. Dari hasil validasi isi CVR dan I-CVI, dapat disimpulkan bahwa instrumen tes TKPM memiliki validitas konten yang tepat untuk seluruh butir soal.

Tahap *Preliminary Field Testing*

Pada tahap ini, instrumen tes yang telah divalidasi ahli kemudian dilakukan uji keterbacaan kepada lima mahasiswa calon guru fisika tingkat kedua yang telah meluluskan mata kuliah Fisika Dasar 2. Terdapat lima aspek penilaian instrumen tes [12, 25]. Kelima aspek penilaian tersebut yaitu: 1) penggunaan kata tanya atau perintah soal singkat, jelas, dan tegas; 2) rangkaian kalimat soal merupakan pernyataan yang diperlukan saja; 3) penggunaan bahasa sesuai dengan kaidah Bahasa Indonesia; 4) Penggunaan kata mudah dipahami; dan 5) Kalimat pertanyaan/soal komunikatif. Hasil uji keterbacaan menunjukkan bahwa skor rata-rata penilaian untuk seluruh item berada pada rentang 3-5 dengan skor minimal 1 dan maksimal 5. Peroleh skor rata-rata hasil penilaian uji keterbacaan selanjutnya diinterpretasikan ke dalam kategori. Interpretasi hasil analisis uji keterbacaan berada pada kategori sangat jelas (65%) dengan keputusan tanpa revisi dan jelas (35%) yang berarti revisi kecil (revisi ringan).

Setelah melewati tahap validasi ahli dan uji keterbacaan butir soal, selanjutnya instrumen tes dibenahi sesuai dengan saran dan masukan para ahli sampai instrumen tes layak digunakan. Namun, uji coba instrumen tes perlu dilakukan terlebih dahulu untuk menganalisis validitas berdasarkan struktur internal (validitas empiris) dan tingkat reliabilitas butir soal.

Tahap Main Field Testing

Pada tahap *main field testing*, instrumen tes yang telah revisi berdasarkan hasil validasi ahli dan uji keterbacaan butir soal selanjutnya diujicobakan pada sejumlah mahasiswa calon guru fisika sebanyak 63 orang. Tujuan pelaksanaan ujicoba adalah untuk memperoleh informasi tingkat validitas empiris dan reliabilitas tes. Hasil ujicoba empiris selanjutnya dianalisis menggunakan bantuan software program *Winsteps version 3.68.2*. Ada dua aspek yang dianalisis untuk menentukan tingkat validitas internal dan reliabilitas tes, yaitu analisis ringkasan dan analisis tingkat kesesuaian butir soal (*item fit*). Berikut ini diuraikan pada masing-masing aspek yang dianalisis.

Analisis Ringkasan Statistik

Berikut ini disajikan ringkasan statistik hasil analisis data (*output*) yang diolah menggunakan aplikasi program *Winsteps version 3.68.2*.

INPUT: 63 PERSONS 36 ITEMS MEASURED: 63 PERSONS 36 ITEMS 128 CATS 3.68.2

SUMMARY OF 63 MEASURED PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	79.6	36.0	1.06	.19	1.05	.0	1.02	.0
S.D.	10.5	.0	.37	.02	.37	1.4	.76	1.1
MAX.	95.0	36.0	1.68	.24	2.43	3.9	5.26	4.3
MIN.	56.0	36.0	.30	-.17	.43	-3.0	.35	-1.8
REAL RMSE	.21	ADJ.SD	.30	SEPARATION	1.44	PERSON RELIABILITY	.67	
MODEL RMSE	.19	ADJ.SD	.31	SEPARATION	1.62	PERSON RELIABILITY	.72	
S.E. OF PERSON MEAN = .05								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .73

SUMMARY OF 36 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	139.3	63.0	.00	.25	1.01	.1	1.02	.2
S.D.	41.9	.0	1.27	.24	.14	.8	.27	.9
MAX.	188.0	63.0	1.70	1.01	1.46	2.3	1.86	2.6
MIN.	40.0	63.0	-3.13	-.10	.82	-1.1	.57	-1.1
REAL RMSE	.35	ADJ.SD	1.22	SEPARATION	3.45	ITEM RELIABILITY	.92	
MODEL RMSE	.35	ADJ.SD	1.22	SEPARATION	3.46	ITEM RELIABILITY	.92	
S.E. OF ITEM MEAN = .21								

Gambar 1. Ringkasan statistik analisis hasil ujicoba instrumen tes keterampilan pemecahan masalah materi kelistrikan dan kemagnetan

Informasi yang diperoleh dari Gambar 1 adalah nilai *logit person measure* (1,06) lebih tinggi dibanding nilai *logit item measure* (0,00) yang selalu bernilai tetap. Hal ini mengindikasikan bahwa kecenderungan kemampuan responden berada di atas tingkat kesulitan *item*. Informasi tambahan berikutnya adalah terkait nilai separasi dan reliabilitas pada *person* dan *item measure*. Nilai separasi dan reliabilitas *person* yang digunakan adalah nilai estimasi “REAL” dan bukan nilai estimasi “MODEL” dikarenakan nilainya lebih konserfatif [39]. Tingkat kualitas instrumen tes dapat diukur dari nilai reliabilitas *Alpha Cronbach*. Berdasarkan kategori tingkat reliabilitas tes pada Tabel 4, maka tingkat reliabilitas *Alpha Cronbach* instrumen tes Keterampilan Pemecahan Masalah berada pada kategori **bagus** (sebesar 0,73).

Informasi lain yang dapat digunakan dalam mengukur kualitas instrumen tes yaitu nilai *person* dan *item reliability*. Tingkat kualitas reliabilitas *person* (*person reliability*) dan reliabilitas butir soal (*item reliability*) dapat dibedakan dalam beberapa kategori berdasarkan nilai koefisien reliabilitasnya. Hasil analisis menunjukkan nilai *person reliability* dan *item reliability* berturut-turut adalah 0,67 dan 0,92. Berdasarkan kategori tingkat kualitas *person* dan *item reliability* pada Tabel 5, terlihat bahwa *person reliability* instrumen berada pada kategori **cukup bagus**. Sementara, *item reliability* berada pada kategori **sangat bagus**. Dengan demikian dapat disimpulkan bahwa konsistensi jawaban responden **cukup bagus** dengan kualitas butir soal instrumen tes aspek reliabilitasnya **sangat bagus**.

Tinjauan kualitas instrumen tes TKPM pada aspek separasi *person* dan *item* seperti yang diperlihatkan Gambar 1 menunjukkan bahwa nilai separasi *person* sebesar 1,44 (<3,0) termasuk dalam kategori dapat

diterima. Nilai separasi person dan reliabilitas person mengindikasikan bahwa instrumen tes sensitifnya cukup bagus (sedang) untuk membedakan antara person yang berkemampuan tinggi dan person yang berkemampuan rendah. Sama halnya dengan separasi *item* bernilai 3,45 ($>3,0$) dengan kategori sedang. Hal ini menyiratkan bahwa sampel person sudah cukup untuk mengkonfirmasi hierarki kesulitan butir soal [39].

Informasi penting lainnya adalah nilai *person fit* dan *item fit* pada kolom *infit mnsq* dan *outfit mnsq*, serta nilai *infit Zstd* dan *outfit Zstd*. Gambar 1 memperlihatkan bahwa rata-rata nilai *infit mnsq* dan *outfit mnsq* pada *person fit* secara berurutan adalah 1,05 dan 1,02. Kedua nilai *mnsq* ini sangat dekat dengan nilai ideal model Rasch (yaitu 1,00). Demikian pula pada *item fit*, rata-rata nilai *infit mnsq* dan *outfit mnsq* secara berurutan yaitu 1,01 dan 1,02 yang berarti keduanya sangat dekat dengan ekspektasi model Rasch. Nilai *infit Zstd* dan *outfit Zstd* pada *person fit* sama, yaitu 0,0. Sementara, pada *item fit* secara berurutan bernilai 0,1 dan 0,2. Hasil ini mengindikasikan bahwa baik pada *person fit* maupun *item fit* mendekati nilai Z standar 0,0 yang menunjukkan kualitas baik. Nilai standar deviasi *item measure* (*infit Zstd* = 0,8; *outfit Zstd* = 0,9) menunjukkan bahwa estimasi kemampuan person secara umum berada dalam rentang -2,0 sampai +2,0 yang merupakan kriteria diterimanya *item fit statistik*. Selain itu, hal ini mengindikasikan kualitas butir soal secara keseluruhan adalah ideal (baik).

Analisis Tingkat Kesesuaian Butir Soal (Item Fit)

Tabel 7 menampilkan ringkasan statistik output item fit instrumen tes keterampilan pemecahan masalah. Ada tiga kriteria yang digunakan dalam memeriksa item fit yang tidak sesuai (misfit) yaitu nilai outfit means-square (mnsq), outfit z-standard, dan point measure correlation yaitu: a) nilai outfit mean square (mnsq) yang diterima: $0,5 < mnsq < 1,5$; b) nilai outfit Z-standard (Zstd) yang diterima: $-2,0 < zstd < +2,0$; c) Nilai point measure correlation (Pt mean corr): $0,4 < Pt \text{ measure corr} < 0,85$. Jika *item* tidak memenuhi salah satu ketiga kriteria tersebut (*outfit mnsq*, *outfit z-std*, dan *Pt measure corr.*) maka item tersebut direvisi atau dibuang.

Tabel 2. Data referensi hasil eksperimen dan tingkat kesalahan relatif dari perhitungan

No.	Item	Infit		Outfit		Pt-Measure	
		Mnsq	Zstd	Mnsq	Zstd	Corr.	Exp.
1.	S1a	0,90	-0,6	0,85	-0,8	0,44	0,15
2.	S1b	0,86	-0,5	0,75	-1,0	0,52	0,17
3.	S1c	0,97	-0,1	1,24	0,9	0,35	0,34
4.	S1d	0,89	-0,4	1,04	0,2	0,43	0,32
5.	S1e	0,82	-1,0	0,80	-1,0	0,54	0,33
6.	S1f	0,86	-1,1	0,87	-0,8	0,54	0,42
7.	S2a	0,98	0,3	0,57	-0,1	0,22	0,05
19.	S4a	1,00	0,3	0,76	0,1	0,12	0,05
20.	S4b	0,99	0,2	0,76	-0,1	0,17	0,07
21.	S4c	1,46	2,3	1,46	2,6	-0,08	0,38
22.	S4d	0,98	0,0	0,97	0,0	0,32	0,30
23.	S4e	1,05	0,4	1,07	0,5	0,35	0,39
24.	S4f	1,07	0,6	1,09	0,6	0,31	0,39
25.	S5a	0,99	0,2	0,96	0,2	0,10	0,07
26.	S5b	1,01	0,2	1,02	0,2	0,08	0,10
27.	S5c	1,26	1,4	1,46	1,5	0,13	0,38
28.	S5d	1,04	0,2	1,14	0,6	0,28	0,33
29.	S5e	1,23	1,8	1,26	1,9	0,16	0,39
30.	S5f	1,14	0,9	1,26	1,0	0,23	0,37
31.	S6a	0,91	-0,4	0,86	-0,6	0,43	0,19
32.	S6b	0,91	-0,3	0,82	-0,7	0,45	0,25
33.	S6c	0,93	-0,5	0,87	-0,7	1,43	0,40
34.	S6d	0,92	-0,5	0,90	-0,5	0,49	0,37
35.	S6e	0,89	-0,9	0,92	-0,5	0,48	0,38
36.	S6f	0,86	-1,0	0,80	-1,0	0,54	0,33

Tabel 7 memperlihatkan bahwa butir soal S4c memiliki kecenderungan tidak fit karena tidak memenuhi syarat pada *Outfit Zstd* (nilainya 2,6) dan pada *Pt. measure corr.* (nilainya -0,08), namun untuk kriteria *Outfit Mnsq* masih dalam batas yang diperbolehkan sehingga butir soal S4c tetap dipertahankan. Untuk sejumlah butir soal lain yang hanya tidak memenuhi pada salah satu dari ketiga kriteria yaitu pada *Pt. measure corr.* Sementara, empat belas butir soal lainnya (S1a, S1b, S1d, S1e, S1f, S2e, S2f, S3d, S3e, S6a, S6b, S6d, S6e, dan S6f) telah memenuhi ketiga kriteria tersebut, sehingga tidak ada masalah pada keenam butir soal tersebut. Dengan demikian, kesimpulan akhirnya tidak ada soal yang perlu diubah atau diganti.

KESIMPULAN

Berdasarkan hasil pengembangan dan proses validasi dapat disimpulkan bahwa instrumen tes keterampilan pemecahan masalah yang dikembangkan telah memenuhi syarat validitas isi (*expert judgement*), validitas konstruk (validitas empiris), uji reliabilitas, dan uji tingkat kesesuaian butir soal. Dengan demikian, instrumen tes keterampilan pemecahan masalah layak digunakan untuk mengukur tingkat keterampilan pemecahan masalah pada materi kelistrikan dan kemagnetan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada berbagai pihak yang telah membantu dalam penulisan makalah ini. Ucapan terimakasih khususnya kepada para ahli (validator) di dibidang pembelajaran, asesmen, dan konten fisika serta kepada mahasiswa calon guru fisika dan dosen pengampu mata kuliah Fisika Dasar 2 program studi pendidikan Fisika Universitas Negeri Makassar yang telah membantu dalam proses pengambilan data penelitian sehingga makalah ini dapat terselesaikan.

REFERENSI

- [1] Binkley, M *et al.*, Defining Twenty-First Century Skills, in *Assessment and Teaching of 21st Century Skills*, (California: Springer Science+Business Media B.V), p. 17–66 (2012)
- [2] Costa A L, *Developing Minds: A Resource Book for Teaching Thinking* Alexandria, Virginia: Association for Supervision and Curriculum Development (1985)
- [3] Koenig J A Elliott S Hilton M and Iverson K, *Assessing 21st Century Skills Summary of a Workshop* Washington D.C: The National Academies Press (2011)
- [4] McGregor D, *Developing Thinking Developing Learning: A Guided to Thinking Skills in Education USA* (2007)
- [5] National Science Teachers Association, *Science-Technology-Society: Science Education for the 1980's," in NSTA Handbook* Wasington, DC.: National Science Teachers Association (1985)
- [6] Dossey J A and Funke J, Canadian and United States students' performances on the OECD's PISA 2012 problem solving assessment *Can. J. Sci. Math. Technol. Educ.* **16**, 1 p. 92–108 (2016)
- [7] Freitas I M Jimenez R and Mellado V, Solving physics problems: The conceptions and practice of an experienced teacher and an inexperienced teacher *Res. Sci. Educ.* **34**, 1 p. 113–133 (2004)
- [8] Gok T, *The General Assessment of Problem Solving Processes and Metacognition in Physics Education* Eurasian Journla Phys. Chem. Educ. **2**, 2 p. 110–122 (2010)
- [9] Krulik S and Rudnick J, *The New Sourcebook for Teaching Reasoning and Problem Solving in Elementary School* United State of America: Allyn and Bacon (1995)
- [10] Pólya G, *How To Solve It. Qucosa.De*: New Jersey (1957)
- [11] Gok T, The Impact of Peer Instruction on College Students ' *Int. J. Sci. Math. Educ.* **10**, June 2011 p. 417–436 (2012)
- [12] Brookhart, S M, *How to Assess Higher-Order Thinking Skills in Your Calsroom*. ASCD Member Book, Alexandria (2010)
- [13] Nitco, J, *Educational Test and Measurement An Introduction* New York: Harcourt Brace Jovanovich, Inc. (1983)
- [14] Jonassen D H, *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Envirntments* New York and London: Routledge Taylor & Francis Group (2011)
- [15] Jonassen D H and Kwon II H, Communication patterns in computer mediated versus face-to-face group problem solving *Educ. Technol. Res. Dev.* **49**, 1 p. 35–51 (2001)

- [16] Reif, F., & Heller J I, Knowledge structures and problem solving in physics. *Educational Psychologist, Educ. Psychol.* **17**, 2 p. 102–127 (1982)
- [17] Baram-Tsabari A and Lewenstein B V., *An Instrument for Assessing Scientists' Written Skills in Public Communication of Science* **35**, 1 (2013)
- [18] Meyer C, Group Work Tests for Context-Rich Problems *Phys. Teach.* **54**, 5 p. 302–304 (2016)
- [19] Dega B G Kriek J and Mogese T F, Students' conceptual change in electricity and magnetism using simulations: A comparison of cognitive perturbation and cognitive conflict *J. Res. Sci. Teach.* **50**, 6 p. 677–698 (2013)
- [20] Tiruneh D T De Cock M Weldeslassie A G Elen J and Janssen R, Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism *Int. J. Sci. Math. Educ.* **15**, 4 p. 663–682 (2017)
- [21] Pollock S J, Longitudinal study of student conceptual understanding in electricity and magnetism *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **5**, 2 p. 020110 (2009)
- [22] Sadaghiani H R, Using multimedia learning modules in a hybrid-online course in electricity and magnetism *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **7**, 1 p. 1–7 (2011)
- [23] Bransford J D and Stein B S, *The Ideal Problem Solver: A Guide for Improving Thinking, Learning, and Creativity* 2nd ed. New York: W.H. Freeman (1993)
- [24] APA, *Standards for Educational and Psychological Testing* Washington, DC: American Psychological Association (1999)
- [25] Groundlund, N.E & Linn R L, *Measurement and Evaluation in Teaching* United Kingdom: Prentice Hall College Div (1990)
- [26] Haladyna T M and Rodriguez M C, *Developing and Validating Test* New York: Routledge Taylor & Francis Group (2013)
- [27] Russell M K and Airasian P W, *Classroom Assessment: Concepts and Applications* 7th ed. New York: McGraw-Hill Companies (1967)
- [28] Thorndike R L, *Educational Measurement* Second Edi United State of America: American Council on Education (1971)
- [29] Mehrens W . and Lehmann I ., *Measurement and Evaluation in Education and Psychology* Fourth Edi United State of America: Wadsworth Publishing Company (1991)
- [30] Borg W R Gall M D and Gall J P, *Educational Reasearch: An Introduction* Seventh Ed United State of America: Pearson Education, Inc. (2003)
- [31] Oriondo L L and Antonio E M D, *Evaluating Educational Outcome (Test, Measurement, and Evaluating)* Philippine: Rex Book Store, Inc. (1984)
- [32] Haladyna T M Downing S M and Rodriguez C, Applied Measurement in Education A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment May 2013 p. 37–41. (2010)
- [33] Svetina D and Levy R, A Framework for Dimensionality Assessment for Multidimensional Item Response Models *Educ. Assess.* **19**, 1 p. 35–57 (2014)
- [34] Tarrant M Ware J and Mohammed A M, An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis *BMC Med. Educ.* **9**, 1 p. 1–8. (2009)
- [35] Lawshe C, A Quantitative Approach to Content *Pers. Psychol.* **28** p. 563–575 (1975)
- [36] Wilson F R Pan W and Schumsky D A, Recalculation of the critical values for Lawshe's content validity ratio *Meas. Eval. Couns. Dev.* **45**, 3 p. 197–210 (2012)
- [37] Zamanzadeh V Ghahramanian A Rassouli M Abbaszadeh A Alavi-Majd H and Nikanfar A-R, Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication *J. Caring Sci.* **4**, 2 p. 165–178 (2015)
- [38] Sumintono B and Widhiarso W, *Aplikasi Pemodelan Rasch pada Assessment Pendidikan Cimahi*: Penerbit Trim Komunikasi (2015)
- [39] Boone W J Staver J R and Yale M S, *Rasch Analysis in the Human Sciences* New York, London: Springer (2014)
- [40] Boone W J and Scantlebury K, The role of rasch analysis when conducting science education research utilizing multiple-choice tests *Sci. Educ.* **90**, 2 p. 253–269 (2006)
- [41] Davidowitz B and Potgieter M, Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry *Int. J. Sci. Educ.* **38**, 9 p. 1483–1503 (2016)
- [42] Jonassen D H, *Learning to Solve Problems: An Instructional Design Guide* United States of America: Pfeiffer (2004)