



INSTRUMEN PENILAIAN TERINTEGRASI : PENGUKURAN KEMAMPUAN BERPIKIR KRITIS DAN KETERAMPILAN SAINS SISWA PADA MATERI LARUTAN ELEKTROLIT DAN NONELEKTROLIT

Sri Rejeki Dwi Astuti¹, Anggi Ristiyana Puspita Sari², Rizki Nor Amelia³, Karlinda⁴

^{1,4} Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

² Universitas Palangka Raya, Palangkaraya, Indonesia

³ Universitas Negeri Semarang, Semarang, Indonesia

Email : ¹ srirejeki.dwiastuti@yahoo.com

DOI : <http://dx.doi.org/10.30829/tar.v29i2.1858>

ARTICLE INFO

Article History

Received : September 30, 2022

Reviewed : December 2, 2022

Accepted : December 14, 2022

Keywords

ABSTRACT

This study aimed to explore the characteristics of the tests used in measuring students' abilities such as item fit, item difficulty level, and test reliability and to measure students' ability in electrolyte test takers' reliability, and person misfit. This research involved 260 students from two senior high schools in Bantul, Yogyakarta. The research instruments was integrated assessment instrument for electrolyte and nonelectrolyte solution matter consisted of 8 essay question. The data were analyzed with Rasch model using Winstep software version 3.73. The results revealed that all items on the integrated assessment instrument have been proven to be fit with Rasch model, have good test reliability, and have good difficulty level parameters. Moreover, students' abilities were obtained very diverse, i.e. very low, low, medium, high and very high abilities. Meanwhile, the person reliability were very low and 48 students were identified as person misfit. Thus, further examination can be conducted, so that students who are identified as misfit s can obtain guidance from the teacher.

Pendahuluan

Pembelajaran merupakan suatu proses komunikasi timbal balik antara guru dan siswa dimana mengajar dilakukan oleh guru sebagai pendidik dan belajar dilakukan oleh siswa sebagai peserta didik. Pembelajaran kimia merupakan proses interaksi antara siswa dengan

lingkungannya dalam rangka mencapai tujuan pembelajaran kimia. Suatu proses pembelajaran meliputi tahapan-tahapan dalam pembelajaran, salah satunya tahapan penilaian. Penilaian merupakan suatu proses evaluasi untuk menentukan kualitas atau keberhasilan dalam proses pembelajaran. Penilaian mencakup tiga aspek dalam pembelajaran, yaitu aspek pengetahuan (*knowledge*), sikap (*attitude*), dan keterampilan (*skill*). Dalam ilmu kimia, ketiga aspek tersebut sangat penting dalam pembelajaran kimia karena didasari pada sifat ilmu kimia sebagai produk dan kimia sebagai proses.

Produk dari suatu pembelajaran terdiri atas dua macam pengetahuan, yaitu pengetahuan deklaratif (*declarative knowledge*) dan pengetahuan prosedural (*procedural knowledge*) (Marzano, 2007). Pengetahuan deklaratif adalah pengetahuan kognitif dari suatu ilmu yaitu meliputi fakta, prinsip, konsep, teori dan hukum. Pengetahuan prosedural adalah pengetahuan yang berorientasi pada keterampilan proses atau dengan kata lain pengetahuan prosedural adalah pengalaman sains (Ismail & Jusoh, 2001). Proses pembelajaran tidak dapat terlepas dari pengembangan kognitif siswa dan pengembangan keterampilan siswa dikarenakan ada keterkaitan antara kognitif dengan ketrampilan proses. Keterampilan proses sangat esensial untuk memahami, mengaplikasikan dan menguasai suatu konsep materi pembelajaran (Marasigan & Espinosa, 2014; Darmaji, Kurniawan, & Irdianti, 2019). Oleh karena itu, dalam pembelajaran tidak boleh hanya mengedepankan aspek kognitifnya saja, tetapi harus seimbang antara kognitif dan keterampilan proses sehingga penilaian dalam pembelajaran kimia haruslah mencakup aspek kognitif dan keterampilan proses.

Siswa dituntut memiliki kompetensi yang diperlukan untuk memenuhi tuntutan global. Kompetensi ini disebut sebagai kompetensi global. Kompetensi global mencakup tiga aspek yaitu *empirical based knowledge and skills, high-order cognitive and metacognitive skills* serta *global disposition, prespectives, and attitudes* (Balistreri et al., 2012). Ketiga aspek tersebut merupakan cerminan dari aspek kognitif, afektif dan psikomotorik yang merupakan produk dari suatu pembelajaran. Oleh karena itu, jika pengembangan ketiga aspek tersebut dalam pembelajaran benar-benar diperhatikan, maka siswa akan memiliki kompetensi global yang baik pula.

Sekolah harus dapat menyiapkan siswa agar memiliki kompetensi global untuk menjawab tantangan global tersebut. Salah satu cara untuk mempersiapkan siswa agar memiliki kompetensi global adalah mempersiapkan aspek kognitif siswa. Aspek kognitif yang dimaksudkan adalah *high-order cognitive and metacognitive skills* atau kemampuan kognitif tingkat tinggi dan metakognitif. Kemampuan kognitif tingkat tinggi dan metakognitif dalam

kompetensi global meliputi kemampuan berpikir kritis, berpikir kreatif dan pemecahan masalah (Balistreri *et al.*, 2012; Rumahlatu, Sangur, & Liline, 2020). Kemampuan berpikir kritis merupakan berpikir secara menyeluruh dan beralasan untuk memutuskan apa yang dipercaya dan apa yang akan dilakukan (Brookhart, 2010).

Kemampuan berpikir kritis sangat penting dimiliki siswa. Kemampuan berpikir kritis lebih menekankan agar siswa dapat mengemukakan pendapat secara kritis dan menghasilkan alasan yang dapat dipertanggungjawabkan sesuai dengan pengetahuan yang dimiliki siswa. Berpikir kritis dan analitis merupakan kemampuan berpikir yang esensial yang sangat diperlukan dalam pembelajaran yaitu kemampuan berpikir secara kritis dan objektif tentang masalah dalam pembelajaran dan menghasilkan argumen atau pendapat yang berbobot (Judge, Jones, & McCreery, 2009; Schulz & FitzPatrick, 2016)). Kemampuan berpikir kritis siswa dapat dilihat dari indikator kemampuan berpikir kritis, seperti mengidentifikasi masalah, merekonstruksi argumen, mengevaluasi argumen., menentukan solusi dan menarik kesimpulan dari suatu masalah.

Siswa yang memiliki kemampuan berpikir kritis yang baik akan menggunakan alasan dan fakta-fakta yang jelas untuk mendukung argumennya. Alasan dan bukti yang jelas ini diperoleh dari pengetahuan yang baik. Kemampuan berpikir kritis merupakan tujuan pendidikan yang sangat penting, karena keterampilan berpikir kritis dapat membantu siswa dalam memecahkan masalah secara efisien dan sebagai cara untuk belajar mandiri (Mahapoonyanont *et al.*, 2010). Oleh karena itu, kemampuan berpikir kritis merupakan salah satu tujuan pembelajaran di era globalisasi ini, sehingga siswa harus dibiasakan menggunakan kemampuan berpikir kritis dalam pembelajaran

Aspek keterampilan juga merupakan salah satu aspek dalam kompetensi global, yaitu *empirical based knowledge and skills*. *Empirical based knowledge and skills* adalah suatu keterampilan dan seperangkat pengetahuan yang bersifat empiris dan diperoleh secara praktikal (Balistreri *et al.*, 2012). Salah satu keterampilan dalam *empirical based knowledge and skills* adalah *science process skills* atau keterampilan proses sains. Keterampilan proses sains merupakan keterampilan yang diperlukan dalam memperoleh dan mengembangkan ilmu pengetahuan.

Keterampilan proses sangat penting dalam suatu langkah pembelajaran untuk memperoleh pengetahuan (Rauf *et al.*, 2013). Salah satu tujuan utama dalam pendidikan sains adalah untuk mempersiapkan siswa untuk mengakumulasi pengetahuan akan alam dan dunia serta menemukan sendiri pengetahuan saintifik melalui penerapan keterampilan proses

sains (Li & Klahr, 2006; Mutisya, Too, & Rotich, 2014). Keterampilan proses sains siswa dapat dilihat dari indikator keterampilan proses sains, seperti observasi, komunikasi, klasifikasi, prediksi, inferensi, organisasi data dalam tabel analisis data percobaan, identifikasi variabel dan merancang eksperimen (Rusmini, Suyono, & Agustini, 2021). Keterampilan proses dipercaya dapat memastikan siswa memiliki pengetahuan yang bermakna karena siswa menemukan sendiri pengetahuan, sehingga akan lebih mudah dipahami dan diaplikasikan oleh siswa. Selain itu, keterampilan proses dapat membantu siswa untuk meningkatkan kemampuan berpikir tingkat tinggi. Oleh sebab itu pengembangan keterampilan proses bagi siswa sangat penting untuk diperhatikan. Sains merupakan proses untuk memahami dan mencari solusi dari suatu masalah gejala alam yang melibatkan pengujian terhadap solusi yang ditawarkan (Chebii, Wachanga, & Kiboss, 2012).

Kimia merupakan salah satu ilmu yang memfasilitasi kemampuan berpikir kritis di setiap materinya dan juga memegang peranan penting dalam pengembangan kemampuan berpikir kritis siswa (Zhou, Huang, & Tian, 2013). Pernyataan tersebut didasarkan pada ilmu kimia mudah dikaitkan dalam permasalahan di kehidupan sehari-hari serta penerapannya nyata di kehidupan masyarakat. Salah satu materi dalam pembelajaran kimia adalah larutan elektrolit dan nonelektrolit. Materi larutan elektrolit dan nonelektrolit merupakan materi kimia yang bersifat kontekstual, sehingga mudah dikaitkan dan diaplikasikan dalam kehidupan sehari-hari. Fenomena tentang larutan elektrolit banyak dijumpai pada kehidupan sehari-hari, seperti pemanfaatan sel aki dan kebutuhan tubuh akan larutan elektrolit. Fenomena-fenomena tersebut dapat mengembangkan kemampuan berpikir kritis siswa terhadap sesuatu. Selain itu, materi larutan elektrolit merupakan materi prasyarat untuk materi elektrokimia (sel volta dan elektrolisis).

Permendikbud Nomor 66 tahun 2013 tentang standar penilaian pendidikan menyatakan bahwa penilaian dalam pembelajaran hendaknya mencakup ketiga aspek dalam pembelajaran, yaitu pengetahuan, sikap dan keterampilan (Kemendikbud, 2013). Akan tetapi penilaian terhadap hasil pembelajaran, masih terfokus hanya pada aspek kognitif. Proses pembelajaran yang berlangsung selama ini hanya menitikberatkan pada aspek kognitif saja sedangkan aspek keterampilan prosesnya kurang diperhatikan (Zoller, 2001). Selain itu, guru masih belum melakukan penilaian kemampuan berpikir kritis karena mereka masih belum yakin dalam menyusun instrumen yang benar-benar dapat menilai kemampuan berpikir kritis siswa (Schulz & FitzPatrick, 2016). Berdasarkan hasil wawancara terhadap beberapa guru kimia SMA di Kabupaten Bantul mengindikasikan bahwa guru belum optimal dalam melakukan penilaian

secara menyeluruh dalam pembelajaran. Guru masih mengedepankan penilaian terhadap aspek pengetahuan, sedangkan penilaian keterampilan proses sains hanya didasarkan pada asumsi subjektif guru. Pelaksanaan pembelajaran telah memfasilitasi siswa untuk mengembangkan kemampuan berpikir kritis, tetapi penilaian terhadap kemampuan berpikir kritis siswa dalam pembelajaran kimia kurang mendapat perhatian dari guru. Guru masih menggunakan instrumen tes berupa soal pilihan ganda dalam penilaian hasil belajar.

Penilaian terhadap hasil pembelajaran berupa aspek kognitif dan keterampilan haruslah berjalan beriringan. Siswa yang memiliki keterampilan proses yang baik dalam pembelajaran, memiliki kemampuan kognitif yang lebih baik pula, dari pada siswa yang tidak memiliki keterampilan proses dan hanya mengandalkan hafalan (Austin et al., 2015). Hal ini menunjukkan bahwa ada hubungan antara aspek kognitif dan keterampilan. Saat ini penilaian terhadap aspek kognitif dan aspek keterampilan dilakukan secara terpisah (Keil, Haney, & Zoffel, 2009; Saribaz & Bayram, 2009; Tosun & Taskesenligil, 2013).

Oleh karena itu dilakukan pengukuran kemampuan berpikir kritis dan keterampilan proses sains pada materi larutan elektrolit dan nonelektrolit. Penelitian ini bertujuan untuk mengeksplorasi karakteristik tes yang digunakan dalam pengukuran kemampuan siswa seperti *item fit*, tingkat kesukaran butir dan reliabilitas tes serta mengeksplorasi karakteristik kemampuan siswa, reliabilitas peserta tes, dan *person misfit*.

Metode Penelitian

Desain Penelitian

Penelitian ini merupakan penelitian lanjutan dari penelitian pengembangan instrumen penilaian terintegrasi "*critical-science process skills instrument*" untuk mengukur kemampuan berpikir kritis dan keterampilan proses sains pada materi larutan elektrolit. Penelitian sebelumnya menyatakan bahwa instrumen penilaian terintegrasi telah berhasil dikembangkan berdasarkan model pengembangan tes McIntire dan terbukti valid berdasarkan validitas konten dan validitas konstruk (Astuti et al., 2017). Penelitian ini berfokus pada pengukuran kemampuan berpikir kritis terintegrasi kemampuan proses sains siswa kelas X pada materi larutan elektroli dan nonelektrolit. Oleh karena itu penelitian ini menggunakan pendekatan kuantitatif dalam pengumpulan dan analisis data.

Sampel

Penelitian ini melibatkan 260 siswa kelas X yang berasal dari dua SMA Negeri di Bantul, Yogyakarta. Pemilihan sampel penelitian menggunakan teknik *purposive sampling*

berdasarkan peringkat sekolah yang berada di posisi empat besar data Penerimaan Peserta Didik Baru (PPDB) Kabupaten Bantul tahun pelajaran 2020/2021. Pemilihan sampel tersebut didasari asumsi bahwa sekolah yang memiliki peringkat tinggi telah memfasilitasi kemampuan berpikir tingkat tinggi siswa selama proses pembelajaran. Hal ini dikarenakan jika akan dilakukan pengukuran kemampuan berpikir tingkat tinggi siswa maka siswa harus sudah dibiasakan untuk mengembangkan kemampuan berpikir tingkat tinggi selama proses pembelajaran. Dengan asumsi tersebut, maka kemampuan berpikir kritis siswa dapat diukur dengan instrumen penilaian terintegrasi kemampuan berpikir kritis terintegrasi kemampuan proses sains.

Prosedur Penelitian

Penelitian ini bertujuan untuk mengukur kemampuan berpikir kritis dan keterampilan proses sains siswa secara terintegrasi. Pengukuran dilakukan dengan cara siswa diminta untuk mengerjakan soal-soal instrumen penilaian terintegrasi yang telah dikembangkan pada penelitian sebelumnya. Instrumen ini dijadikan sebagai ulangan harian untuk materi larutan elektrolit dan nonelektrolit, sehingga siswa dapat mengerjakan soal dalam keadaan seperti pembelajaran biasa. Penelitian ini dilaksanakan pada bulan Februari 2020, dan pengerjaan soal dilakukan selama 2 jam pelajaran.

Instrumen Penelitian

Instrumen penelitian yang digunakan pada penelitian ini adalah instrumen yang telah dikembangkan pada penelitian sebelumnya, yaitu instrumen penilaian terintegrasi untuk mengukur kemampuan berpikir kritis dan keterampilan proses sains pada materi larutan elektrolit dan nonelektrolit (Astuti et al., 2017). Instrumen ini berupa soal esai dan terdiri atas 8 konstruk soal dan 20 subsoal. Tes tersebut dikembangkan berdasarkan kisi-kisi indikator materi larutan elektrolit dan nonelektrolit serta indikator keterampilan berpikir kritis dan indikator keterampilan proses sains. Tes telah terbukti valid berdasarkan validitas isi dan validitas konstruk, sehingga siap digunakan untuk pengukuran.

Teknik Analisis Data

Data hasil pengukuran dianalisis menggunakan pemodelan Rasch dengan bantuan *software* Winstep versi 3.73. Informasi yang didapatkan dari analisis ini adalah identifikasi karakteristik instrumen seperti tingkat kesukaran butir soal (*b*), *item fit*, dan reliabilitas tes serta karakteristik kemampuan berpikir kritis terintegrasi keterampilan proses sains siswa, reliabilitas peserta tes, dan *person misfit*. Selain itu, peneliti juga yang melakukan analisis *person misfit*.

Hasil Penelitian

Pengukuran kemampuan berpikir kritis terintegrasi keterampilan proses sains dilakukan kepada 268 siswa kelas X SMA di Kabupaten Bantul. Pengukuran ini dilaksanakan untuk mengetahui sejauh mana kemampuan berpikir kritis dan keterampilan proses sains. Sebelum dilakukan analisis terhadap kemampuan siswa, terlebih dahulu dilakukan analisis karakteristik instrumen penilaian, dikarenakan kemampuan siswa dapat ditentukan setelah karakteristik instrumen diketahui. Hasil ringkasan analisis karakteristik instrumen penilaian terintegrasi (tingkat kesukaran (b) dan *item fit* disajikan pada Tabel 1.

Tabel 1. Ringkasan Analisis Karakteristik Instrumen Penilaian Terintegrasi

No	Konstruk soal	Outfit -MNSQ	b (logit)
1	1	1.11	-0.27
2	2	0.95	0.58
3	3	1.00	1.17
4	4	0.96	-1.46
5	5	0.93	-0.67
6	6	1.28	1.98
7	7	0.84	0.65
8	8	1.26	-1.99
Reliabilitas Tes		0.98	
<i>Separation</i>		7.48	

Tabel 1 menunjukkan bahwa nilai *outfit* -MNSQ untuk semua konstruk soal berkisar antara 0.84 sampai 1.28 nilai tingkat kesukaran butir soal berkisar antara -1.99 logit sampai 1.98 logit dan nilai reliabilitas tes sebesar 0.98 dengan *separation* 7.48. Selanjutnya, hasil analisis kemampuan siswa (θ) disajikan pada tabel ringkasan statistik *person measure* pada Tabel 2.

Tabel 2. Ringkasan analisis karakteristik instrumen penilaian terintegrasi

No	Jumlah siswa	θ (logit)
1	3	2.47
2	9	1.46
3	29	0.70
4	55	0.01
5	70	-0.68
6	59	-1.45
7	35	-2.46
Reliabilitas Siswa		0.28
<i>Separation</i>		0.62

Tabel 2 menunjukkan bahwa nilai kemampuan siswa (θ) untuk semua siswa berkisar antara 3.88 logit sampai -3.97 logit dan reliabilitas siswa sebesar 0.28 dengan *separation* 0.62. Nilai θ yang beragam menunjukkan bahwa kemampuan siswa sangat beragam, yaitu ada yang

memiliki kemampuan tinggi, sedang dan rendah. Setelah itu, dilakukan analisis statistik untuk *person misfit* disajikan pada Tabel 3. Berdasarkan Tabel 3, terdapat 48 siswa yang *misfit* dengan statistik *outfit* MNSQ berkisar antara 1.55 sampai 9.90 sedangkan kemampuan siswa yang mengalami *misfit* berkisar antara -2.49 logit sampai 1.46 logit.

**Tabel 3. Ringkasan Statistik *Outfit* MNSQ dan Kemampuan Siswa (Θ)
 untuk Person *Misfit***

No	Kode siswa	Θ (logit)	MNSQ	No	Kode siswa	Θ (logit)	MNSQ
1	168	1.46	4.63	25	117	-0.68	1.61
2	180	1.46	5.42	26	140	-0.68	2.18
3	29	0.70	1.61	27	151	-0.68	2.50
4	50	0.70	2.42	28	192	-0.68	1.79
5	52	0.70	2.28	29	201	-0.68	1.89
6	182	0.70	1.51	30	228	-0.68	2.23
7	188	0.70	1.74	31	235	-0.68	2.37
8	197	0.70	2.10	32	245	-0.68	1.89
9	219	0.70	2.43	33	246	-0.68	1.89
10	221	0.70	2.10	34	250	-0.68	1.99
11	247	0.70	1.65	35	128	-1.45	1.79
12	24	0.01	2.12	36	130	-1.45	1.79
13	204	0.01	2.02	37	144	-1.45	4.22
14	211	0.01	2.36	38	163	-1.45	1.71
15	244	0.01	2.01	39	165	-1.45	3.14
16	249	0.01	2.03	40	176	-1.45	3.14
17	253	0.01	2.03	41	214	-1.45	1.64
18	16	-0.68	2.50	42	234	-1.45	2.42
19	26	-0.68	2.50	43	238	-1.45	1.71
20	41	-0.68	1.55	44	260	-1.45	1.71
21	57	-0.68	2.39	45	102	-2.49	9.90
22	66	-0.68	2.39	46	196	-2.49	5.04
23	68	-0.68	1.61	47	215	-2.49	3.06
24	109	-0.68	3.05	48	257	-2.49	5.04

Pembahasan

Penelitian ini bertujuan untuk mengukur kemampuan berpikir kritis terintegrasi keterampilan proses sains pada materi larutan elektrolit dan nonelektrolit. Pemodelan Rasch digunakan untuk menganalisis tingkat kesukaran butir soal, kemampuan siswa dan *person misfit*. Pemodelan Rasch setara dengan IRT 1-PL, yaitu hanya menggunakan 1 parameter logistik, yaitu tingkat kesukaran pada proses analisisnya.

Sebelum dilakukan analisis pengukuran kemampuan siswa, terlebih dahulu dilakukan analisis karakteristik butir soal seperti tingkat kesukaran butir dan *item fit*. *Item fit*

menunjukkan kesesuaian butir dengan model, artinya apakah butir berfungsi normal melakukan pengukuran atau tidak (Sumintono & Widhiarso, 2015). Melalui *item fit*, konsistensi pola respon jawaban siswa terhadap suatu butir soal dapat diketahui (Razak, Khairani, & Thien, 2012). Jika respon jawaban terhadap suatu butir soal tidak konsisten, artinya butir soal tersebut *misfit* (tidak cocok) dianalisis menggunakan Pemodelan Rasch.

Butir soal dikatakan *fit* dengan model apabila telah memenuhi kriteria nilai *outfit means-square*, *outfit z-standard* dan *point measure correlation* (Bond & Fox, 2015). Jika butir soal tidak memenuhi ketiga kriteria tersebut, maka butir soal dinyatakan tidak sesuai (*misfit*) dengan model analisis. Adapun ukuran statistik yang digunakan untuk kriteria *item fit* pada penelitian ini hanya *Outfit*-MNSQ, dimana suatu butir dikatakan *fit* jika berada pada rentang MNSQ yang ditentukan, yakni $0.5 \leq \text{MNSQ} \leq 1.5$ (Linacre, 2002).

Ketidaksesuaian butir soal dengan model dapat mengindikasikan bahwa terjadi miskonsepsi pada peserta tes terhadap butir soal tersebut (Sumintono & Widhiarso, 2015). Butir soal yang *misfit* dapat dipastikan bahwa butir kurang bagus dalam melakukan pengukuran. Oleh karena itu, butir soal yang *misfit* dapat dihilangkan (jika butir soal dalam tes tersebut banyak) atau dapat diperbaiki supaya tidak terjadi lagi miskonsepsi pada butir soal tersebut (jika butir soal dalam tes sedikit).

Tabel 1 menunjukkan bahwa nilai *Outfit*-MNSQ untuk semua butir soal pada instrument penilaian terintegrasi masuk dalam rentang $0.5 \leq \text{MNSQ} \leq 1.5$, sehingga dapat dikatakan bahwa kedelapan butir soal tersebut *fit* dengan analisis Pemodelan Rasch. Butir yang *fit* dengan model dapat menjalankan fungsinya sebagai instrumen pengumpulan data untuk tujuan tertentu. Oleh karena itu, ke delapan butir soal tersebut mampu menjalankan fungsinya untuk melakukan pengukuran terhadap kemampuan berpikir kritis dan keterampilan proses sains siswa pada materi larutan elektrolit dan nonelektrolit

Setelah semua butir soal dinyatakan *fit* dengan model, maka analisis dilanjutkan dengan menentukan tingkat kesukaran butir untuk setiap butir soal. Tingkat kesukaran butir soal tidak ditentukan pada saat proses penyusunan butir soal, tetapi dianalisis setelah instrument tersebut digunakan dalam proses pengukuran. Hal ini dikarenakan tingkat kesukaran butir soal dianalisis berdasarkan respon jawaban siswa dalam menjawab setiap butir soal. Butir soal yang hampir semua siswa memberikan respon yang benar dianggap sebagai butir yang mudah, sedangkan butir soal yang hanya sedikit siswa memberikan respon jawaban benar dianggap butir yang sukar. Untuk menentukan bagaimana tingkat kesukaran butir soal yang sesuai dengan standar yang ada, maka dilakukan analisis dengan pemodelan Rasch.

Parameter tingkat kesukaran butir (b) merupakan perkiraan tingkat kesukaran suatu butir yang dinyatakan dalam satuan logit. Suatu butir dikatakan baik apabila memiliki indeks kesukaran berkisar antara -2.0 logit sampai $+2.0$ logit (Baker, 2001; Hambleton & Swaminathan, 1985). Butir soal yang memiliki tingkat kesukaran dibawah -2.0 logit dikategorikan terlalu mudah, sedangkan butir soal yang memiliki tingkat kesukaran di atas $+2.0$ logit dikategorikan terlalu sukar. Tingkat kesukaran butir soal merupakan faktor yang mempengaruhi kemampuan siswa, karena mempengaruhi kemungkinan jawaban siswa dalam merespon butir soal tertentu. Butir soal yang memiliki tingkat kesukaran yang relatif tinggi memerlukan kemampuan yang tinggi untuk merespon butir soal dengan benar, sedangkan butir soal yang mudah hanya memerlukan tingkat kemampuan yang rendah untuk merespon butir soal dengan benar.

Dalam penelitian ini, butir soal dikategorikan mudah apabila nilai b kurang dari -1 logit. Hal ini mengindikasikan bahwa banyak siswa kemampuan rendah dan kemampuan tinggi memiliki probabilitas menjawab benar. Butir dikategori sedang apabila memiliki tingkat kesukaran pada interval $-0.5 < b < +0.5$. Butir dikategorikan sangat sukar apabila nilai b lebih dari $+1$. Hal ini mengindikasikan bahwa siswa kemampuan rendah tidak dapat merespon jawaban dengan benar butir soal tersebut (Adedoyin & Mokobi, 2013). Tabel 1 menunjukkan bahwa tingkat kesukaran butir soal berada pada rentang -1.99 sampai $+1.98$ logit. Oleh karena itu, seluruh butir soal (8 butir) dalam instrumen penilaian terintegrasi memiliki tingkat kesukaran yang tergolong baik.

Tabel 1 memberikan informasi tentang tingkat kesukaran masing-masing butir soal dari butir paling sukar ke butir paling mudah ditinjau dari hasil pengukuran. Berdasarkan kriteria tingkat kesukaran butir Adedoyin & Mokobi (2013), butir soal yang tergolong sukar memiliki nilai b , butir soal 3 dan 6 merupakan butir tersukar ($b > 1$ logit), butir soal 2 dan 7 merupakan butir yang sukar (0.5 logit $< b < 1$ logit), butir soal 1 merupakan butir sedang (-0.5 logit $< b < +0.5$ logit), butir soal 5 merupakan butir mudah (-1 logit $< b < -0.5$ logit), serta butir 4 dan 8 merupakan butir termudah ($b < -1$ logit).

Hasil analisis menunjukkan bahwa butir 6 memiliki parameter b sebesar $+1.98$ logit, artinya diperlukan kemampuan peserta minimal $+1.98$ logit untuk merespon jawaban butir 6 dengan benar. Butir 8 yang merupakan butir paling mudah membutuhkan kemampuan peserta tes minimal -1.99 logit untuk menjawab butir soal 8 dengan benar. Semakin besar nilai parameter b , maka semakin besar kemampuan yang diperlukan untuk dapat merespon jawaban dengan benar. Oleh karena itu, semakin besar nilai parameter b maka butir soal semakin sukar.

Selain itu juga dilakukan analisis reliabilitas tes dan reliabilitas person (peserta tes) dengan pemodelan Rasch. Analisis reliabilitas angket digunakan untuk mengetahui keajegan atau konsistensi angket dalam memberikan hasil pengukuran. Angket yang bersifat reliabel akan memberikan hasil pengukuran yang relatif sama ketika diujikan berulang-ulang pada sampel yang berbeda. Indeks reliabilitas ditentukan sebagai bukti statistik yang digunakan untuk menentukan konsistensi internal suatu instrumen dalam melakukan pengukuran (Taber, 2018). Tabel 1 menunjukkan bahwa hasil analisis reliabilitas instrumen penilaian terintegrasi memiliki indeks reliabilitas sebesar 0.98 dengan indeks *separation* 7.48. Indeks reliabilitas sebesar 0.98 ($\alpha > 0.90$) memiliki artian bahwa instrumen penilaian terintegrasi memiliki konsistensi internal yang sangat bagus (Tavakol & Dennick, 2011). Dalam pemodelan Rasch, indeks reliabilitas yang tinggi tidak serta merta memiliki reliabilitas yang bagus, karena indeks reliabilitas yang tinggi disebabkan oleh ukuran sampel yang besar, sehingga diperlukan suatu indeks pemisah (*separation index*) untuk memperkuat penjelasan tentang reliabilitas tes (Bond & Fox, 2015). Indeks *separation* yang baik harus lebih dari 3.0 (Linacre, 2004; Yasin et al., 2015). Indeks *separation* menunjukkan seberapa jauh instrumen tes dapat menghasilkan rentang *pengukuran* pada garis logit atau membagi instrumen tes dalam menjadi beberapa level pengukuran (Bond & Fox, 2015; Yasin et al., 2015). Indeks *separation* sebesar 7.48 (> 2.0) mengindikasikan bahwa instrumen ini memiliki indeks *separation* yang sangat bagus. Hal tersebut menjelaskan bahwa seluruh butir soal dalam instrumen penilaian terintegrasi dapat dibagi mejadi 7 level pengukuran. Oleh karena itu, hampir seluruh butir soal dalam instrumen tersebut mengukur level kemampuan yang berbeda.

Selain analisis reliabilitas tes, reliabilitas siswa juga dianalisis untuk mengetahui apakah siswa yang berperan sebagai peserta tes memiliki kemampuan yang konsisten (sama) atau beragam. Tabel 2 menunjukkan bahwa hasil reliabilitas siswa sebesar 0.28 dengan nilai *separation* 0.62. Hasil ini menunjukkan bahwa reliabilitas siswa rendah karena berada pada rentang 0.200 sampai 0.399 (Taber, 2018). Reliabilitas yang relatif kecil menunjukkan bahwa variasi posisi kemampuan peserta tes pada garis logit hanya sedikit, sehingga banyak peserta tes yang terukur memiliki kemampuan (θ) yang sama. Selain itu, indeks *person separation*nya juga sangat kecil (< 2.0), sehingga kemungkinan dengan sejumlah peserta tes tersebut menyiratkan bahwa instrumen mungkin tidak cukup sensitif untuk membedakan antara yang berkemampuan tinggi, sedang, dan rendah.

Selanjutnya, dilakukan analisis kemampuan siswa untuk mengetahui seberapa baik kinerja siswa dalam merespon butir soal dalam instrumen penilaian terintegrasi. Kemampuan

(θ) ini diperoleh berdasarkan respon siswa dalam menjawab butir soal pada instrumen penilaian terintegrasi. Karena instrumen penilaian yang dikembangkan adalah instrumen penilaian terintegrasi, maka respon siswa dalam menjawab butir soal dalam instrumen menunjukkan kemampuan berpikir kritis dan keterampilan proses sains yang dimiliki siswa pada materi larutan elektrolit.

Berdasarkan hasil estimasi kemampuan siswa (θ) yang disajikan pada Tabel 2 terlihat bahwa sejumlah siswa memiliki kemampuan (θ) yang sama. Pada pengukuran kemampuan siswa menggunakan instrumen penilaian terintegrasi, didapatkan bahwa kemampuan siswa tertinggi adalah +2.47 logit dan kemampuan siswa terendah adalah -2.46 logit. Penggolongan kemampuan (θ) hampir sama dengan penggolongan tingkat kesukaran butir siswa karena siswa yang memiliki besar θ yang sama dengan b , maka siswa tersebut pasti dapat memberikan respon jawaban benar untuk butir soal tersebut ataupun butir soal di bawahnya (Sumintono & Widhiarso, 2015). Oleh karena itu, kemampuan siswa dapat digolongkan menjadi kemampuan sangat tinggi apabila $\theta > +1$, kemampuan tinggi apabila θ berada pada rentang $+0.5 < \theta < +1$, kemampuan sedang apabila θ berada pada rentang $-0.5 < \theta < +0.5$, kemampuan rendah apabila $-1 < \theta < -0.5$, dan kemampuan sangat rendah apabila $\theta < -1$.

Tabel 2 menunjukkan bahwa 11 siswa memiliki kemampuan sangat tinggi karena berada pada rentang $\theta > +1$. Sebanyak 29 siswa memiliki kemampuan tinggi karena termasuk dalam rentang $+0.5 < \theta < +1$. Sebanyak 55 siswa memiliki kemampuan sedang karena termasuk rentang $-0.5 < \theta < +0.5$. Sebanyak 70 siswa memiliki kemampuan rendah karena berada pada rentang $-1 < \theta < -0.5$. sebanyak 94 siswa memiliki kemampuan sangat rendah karena berada pada rentang $\theta < -1$.

Sebanyak 3 siswa yang memiliki kemampuan +2.47 logit memiliki kemampuan yang sangat tinggi. Siswa dengan kemampuan tersebut akan mampu menjawab semua butir soal dalam instrumen penilaian terintegrasi dengan benar dikarenakan semua butir soal dalam instrumen tersebut memiliki indeks tingkat kesukaran lebih rendah dari pada kemampuan siswa ($b < \theta$). Siswa yang memiliki kemampuan -2.46 logit memiliki kemampuan yang sangat rendah dan tidak mampu memberikan respon jawaban dengan benar. Dikarenakan instrumen penilaian terintegrasi berupa soal esai terbuka, maka dimungkinkan siswa yang memiliki kemampuan rendah tidak dapat memberikan respon dengan benar, sehingga mereka tidak mendapatkan skor sempurna untuk semua butir soal.

Selanjutnya dilakukan analisis pada person *fit* (kesesuaian peserta tes dengan model pengukuran). Tabel 3 memaparkan hasil analisis *outfit* MNSQ dan kemampuan (θ) khusus

untuk siswa yang termasuk dalam kategori person *misfit*. Person *fit* menggunakan kategori yang sama dengan item *fit*, yaitu dengan menggunakan statistik *Outfit*-MNSQ, dimana peserta tes dikatakan *fit* jika berada pada rentang MNSQ yang ditentukan, yakni $0.5 \leq \text{MNSQ} \leq 1.5$ (Linacre, 2002). Berdasarkan Tabel 2, terdapat 48 siswa yang termasuk dalam person *misfit* dengan statistik *outfit* MNSQ berkisar antara 1.51 sampai 9.90 sedangkan kemampuan siswa berkisar antara -2.94 logit. sampai 1.46 logit. Ke 48 siswa tersebut memiliki kemampuan yang beragam mulai dari kemampuan sangat rendah, sedang, dan sangat tinggi. Hal ini menyatakan bahwa 48 siswa tersebut memiliki pola respons yang tidak dapat diprediksi oleh pemodelan Rasch (Smith, 2001).

Person *misfit* dapat disebabkan oleh lima faktor, yaitu *cheating*, *lucky guessing*, *careless responding*, *creative responding*, dan *random responding* (Karabatsos, 2003; Meijer, 1996). *Cheating* atau menyalin jawaban dari siswa lain merupakan pada perilaku yang tidak adil atas pemberian respon jawaban benar pada butir yang sebenarnya tidak dapat dia jawab dengan benar. *Careless responding* dapat terjadi ketika siswa memberikan respon jawaban benar pada butir sukar namun dengan cara yang tidak jelas, tetapi memberikan respon jawaban salah pada butir mudah. *Lucky guessing* terjadi jika siswa menebak dengan benar soal yang sebenarnya dia tidak tau jawaban yang benar. Biasanya *lucky guessing* terjadi untuk tipe soal pilihan ganda, *Creative responding* terjadi pada siswa yang memiliki kemampuan tinggi saat memberikan respon salah pada butir yang sebenarnya mudah karena mereka mengartikan butir soal tersebut dengan cara yang tidak biasa. *Random responding* terjadi ketika siswa memilih opsi pilihan ganda secara acak dalam merespon butir soal. Pengukuran *person fit* tidak hanya mengidentifikasi pola respon yang tidak mungkin, tetapi juga pola respon yang terlalu mungkin (Meijer & Sitsma, 2001).

Pada penelitian ini, person misfit disebabkan oleh adanya faktor *careless responding* dan *creative responding*. Siswa yang memiliki kemampuan tinggi dapat memberikan respon yang salah terhadap butir soal yang relatif mudah dikarenakan cara berpikir siswa tersebut yang tidak biasa. Sebaliknya, siswa dengan kemampuan rendah dapat memberikan respon yang benar terhadap soal yang sukar dikarenakan cara berpikir mereka yang lebih simpel. Oleh karena itu, pemodelan Rasch dapat mengalami kendala karena adanya terlalu banyak ketidakpastian dan kepastian dalam menganalisis suatu data (Amelia, Sari, & Astuti, 2021). Dengan demikian, peneliti harus cermat dalam melakukan penelitian sampel, supaya sampel yang digunakan dapat memberikan hasil yang representative jika diuji atau dites menggunakan suatu instrumen penelitian.

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa semua butir soal pada instrument penilaian terintegrasi pada materi larutan elektrolit dan nonelektrolit terbukti *fit* dengan pemodelan Rasch, memiliki reliabilitas tes yang baik, serta memiliki parameter butir tingkat kesukaran yang baik (sangat mudah, mudah, sedang, sukar dan sangat sukar). Selain itu, berdasarkan hasil pengukuran kemampuan siswa, diperoleh kemampuan siswa sangat beragam, yaitu kemampuan sangat rendah, rendah, sedang, tinggi dan sangat tinggi. Sementara itu reliabilitas person masih sangat rendah dan 48 siswa teridentifikasi sebagai person *misfit*, sehingga dapat dilakukan pemeriksaan lebih lanjut untuk mendapatkan bimbingan dari guru.

References

- Adedoyin, O. O., & Mokobi, T. (2013). Using IRT Psychometric Analysis in Examining The Quality of Junior Certificate Mathematic Multiple Choice Examination Test Items. *International Journal of Asian Social Science*, 3(4), 992-1011.
- Amelia, R. N., Sari, A. R., & Astuti, S. R. (2021). Chemistry learning outcomes assessment: How is the quality of the tests made by the teacher? *Journal of Educational Chemistry*, 3(1), 11-22. doi:http://10.21580/jec.2021.3.1.6582
- Astuti, S. R., Suyanta, Widjajanti, E., & Rohaeti, E. (2017). An integrated assessment instrument: Developing and validating instrument for facilitating critical thinking abilities and science process skills on electrolyte and nonelectrolyte solution matter. *AIP Conference Proceedings* (pp. 050007-1–050007-6). College Park: AIP Publishing.
- Austin, A. C., Ben-Daat, H., Zhu, M., Atkinson, R., Barrows, N., & Gould, I. R. (2015). Measuring student performance in general. *Chemistry Education Research and Practice*, 16(1), 168-178. doi:https://doi.org/10.1039/C4RP00208C
- Baker, F. B. (2001). *The basic of item response theory*. New York: ERIC Clearinghouse on Assessment and Evaluation.
- Balistreri, S., Di Giacomo, F., Noisette, I., & Ptak, T. (2012). *Global education: Connections, concepts, and careers*. New York: College Board.
- Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model fundamental measurement in the human sciences third edition*. New York: Routledge.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria: ASCD.
- Chebie, R., Wachanga, S., & Kiboss, J. (2012). Effects of Science Process Skills Mastery Learning Approach on Students' Acquisition of Selected Chemistry Practical Skills in School. *Creative Education*, 3(8), 1291-1296.
- Darmaji, D., Kurniawan, D. A., & Irdianti, I. (2019). Physics education students' science process skills. *International Journal of Evaluation and Research in Education*, 8(2), 293-298. doi:http://10.11591/ijere.v8i2.28646
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory*. New York: Kluwer Inc.
- Ismail, Z. H., & Jusoh, I. (2001). Relationship between science process skills and logical thinking abilities of malaysian students. *Journal of Science and Mathematics Education in S.E Asia*, xxiv(2), 67-77.

- Judge, B., Jones, P., & McCreery, E. (2009). *Critical thinking skills for education students*. London: Learning Matters Ltd.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277-298. doi:https://doi.org/10.1207/S15324818AME1604_2
- Keil, C., Haney, J., & Zoffel, J. (2009). Improvements in student achievement and science process skills using environmental health science problem-based learning curricula. *Electronic Journal of Science Education*, 13(1), 1-18.
- Kemendikbud. (2013). *Permendikbud Nomor 66 tahun 2013 tentang Standar Penilaian Pendidikan*. Jakarta: Kemendikbud.
- Li, J., & Klahr, D. (2006). The psychology of scientific thinking: Implications for science teaching and learning. In J. Rhoton, & P. Shane, *Teaching Science in the 21st Century* (p. 2). Arlington: NSTA Press.
- Linacre, J. M. (2002). What do *infit* and *outfit* mean-square and standardized. *Rasch Measurement Transaction*, 16(2), 878.
- Linacre, J. M. (2004). Test Validity and Rasch Measurement: Construct, Content, etc. *asch Measurement Transactions*, 18(1), 970-971.
- Mahapoonyanont, N., Krahamwong, R., Kochakornjarupong, D., & Rachasong, W. (2010). Critical thinking abilities assessment tools: reliability generalization. *Procedia Social and Behavioral Science* (pp. 434–438). Amsterdam: Elsevier.
- Marasigan, A. C., & Espinosa, A. A. (2014). Modified Useful-Learning Approach: Effects on Students' Critical Thinking Skills and Attitude towards Chemistry. *International Journal of Learning, Teaching and Educational Research*, 1(1), 35-72.
- Marzano, R. J. (2007). *The art of science teaching*. Alexandria: ASDC.
- Meijer, R. R. (1996). Person-fit research: an introduction. *Applied Measurement in Education*, 9(1), 3-8. doi:https://doi.org/10.1207/s15324818ame0901_2
- Meijer, R. R., & Sitsma, K. (2001). Person fit statistic: What is their purpose? *Rasch Measurement Transactions*, 15(2), 823.
- Mutisya, S. M., Too, J. K., & Rotich. (2014). Performance in Science Process Skills: The Influence of Subject Specialization. *Asian Journal of Social Science & Humanities*, 3(1), 179-188.
- Rauf, R. A., Rasul, M. S., Mansor, A. N., Othman, Z., & Lyndon, N. (2013). Inculcation of science process skills in a science classroom. *Asian Social Science*, 9(8), 47-57.
- Razak, N. b., Khairani, A. Z., & Thien, L. M. (2012). Examining quality of mathematics test items using rasch model: Preliminary analysis. *Procedia Social Behavioral Science* (pp. 2205 – 2214). Amsterdam: Elsevier.
- Rumahlatu, D., Sangur, K., & Liline, S. (2020). The effect of complex instruction team product (CITP) learning model on increase student's skills. *International Journal of Instruction*, 13(1), 587-606.
- Rusmini, Suyono, & Agustini, R. (2021). Analysis of science process skills of chemical education students through self-project based learning (SJBL) in the covid-19 pandemic era. *Journal of Technology and Science Education*, 11(2), 371-387. doi:<https://doi.org/10.3926/jotse.1288>
- Saribaz, D., & Bayram, H. (2009). Is it possible to improve science process skills and attitudes towards chemistry through the development of metacognitive skills embedded within a motivated chemistry lab?: A self-regulate learning approach. *Procedia Social and Behavioral Sciences* (pp. 61-72). Amsterdam: Elsevier.

- Schulz, H., & Fitzpatrick, B. (2016). Teachers' understandings of critical and higher order thinking and what this means for their teaching and assessments. *Alberta Journal of Educational Research*, 62(1), 61-86.
- Smith, E. V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada asesment pendidikan*. Cimahi : Trim Komunikata Publishing House.
- Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296.
- Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. doi:DOI: 10.5116/ijme.4dfb.8dfd
- Tosun, C., & Taskesenligil, Y. (2013). The effect of problem-based learning on undergraduate students' learning about solution and their physical properties and scientific processing skills. *Chemistry Education Research and Practice*, 14(1), 36-50. doi:https://doi.org/10.1039/C2RP20060K
- Yasin, R. M., Yunus, F. A., Che Rus, R., Ahmad, A., & Rahim, M. B. (2015). Validity and reliability learning transfer item using rasch measurement model. *Procedia Social and Behavioral Science* (pp. 212 – 217). Amsterdam: Elsevier.
- Zhou, Q., Huang, Q., & Tian, H. (2013). Developing Students' Critical Thinking Skills by Task-Based Learning in Chemistry Experiment Teaching. *Crative Education*, 4(12A), 40-45.
- Zoller, U. (2001). Alternative assessment as (critical) means of facilitating HOCS-promoting teaching and learning in chemistry education. *Chemistry Education Research and Practice in Europe*, 2(1), 9-17.