

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

A. Hasil Pengembangan Produk Awal

Instrumen penilaian literasi kimia peserta didik berbasis etnosains Batik Pekalongan pada materi Larutan Elektrolit dan Non-Elektrolit merupakan hasil pengembangan produk pada penelitian ini. Model pengembangan instrumen yang digunakan terdiri dari lima tahap, yaitu 1) perencanaan pembuatan instrumen penilaian (*planning test*), 2) uji coba instrumen (*trying out the test*), 3) penetapan validitas (*establishing test validity*), 4) penetapan reliabilitas (*establishing test reliability*), dan 5) interpretasi skor (*interpreting the test scores*) (Oriondo & Dallo-Antonio, 1984: 34).

Instrumen yang dikembangkan berupa instrumen penilaian literasi kimia pada materi Larutan Elektrolit dan Non-elektrolit yang terintegrasi pada konteks lokal Batik Pekalongan sebagai bentuk etnosains. Bentuk etnosains pada Batik Pekalongan yang digunakan adalah bahan yang digunakan dalam pembuatan Batik Pekalongan, proses pewarnaan Batik Pekalongan, dan limbah yang dihasilkan. Produk divalidasi secara teoritis dan empiris. Validitas teori dilakukan untuk mendapatkan penilaian dari *expert*, sedangkan validitas empiris dilakukan untuk mendapatkan penilaian mengenai kualitas, karakteristik butir soal, dan kesimpulan kelayakan instrumen penilaian yang dikembangkan. Produk akhir yang didapatkan digunakan untuk mengukur kemampuan literasi kimia peserta didik pada aspek konteks lokal Batik Pekalongan, konten atau pengetahuan kimia, dan kompetensi kimia yang terdiri dari 17 butir soal uraian.

Tahap perencanaan pembuatan instrumen penilaian terdiri dari tujuan penelitian, sintesis, adopsi serta adaptasi aspek dalam literasi kimia, sintesis ilmu sains dalam etnosains Batik Pekalongan sebagai pedoman dalam membuat kisi-kisi instrumen, penentuan bentuk butir soal, penulisan butir soal disertai penulisan kunci jawaban dengan pedoman penskoran, kemudian dilakukan validasi serta perbaikan produk. Validasi teoritis instrumen dilakukan oleh *expert* yaitu Dr. Eli Rohaeti, M. Si., Dr. Insih Wilujeng, M. Pd., dua orang pendidik kimia SMA sebagai praktisi di sekolah yang memiliki pengalaman mengajar lebih dari 25 tahun, dan tiga orang teman sejawat program studi Pendidikan Kimia Program Pascasarjana. Butir pertanyaan yang telah dibuat kemudian dinilai oleh para ahli dan dibuktikan validitas teorinya.

Analisis skor validitas teori menggunakan formula Aiken yang disesuaikan dengan indeks Aiken. Berdasarkan analisis menggunakan indeks Aiken, butir soal valid dan layak digunakan apabila nilai $V \geq 0,75$ untuk tujuh *raters* serta nilai $V \geq 0,79$ untuk enam *raters* (Aiken, 1985). Hasil analisis skor validasi dengan formula Aiken untuk instrumen penilaian literasi kimia berbasis etnosains Batik Pekalongan disajikan dalam Tabel 10, 11, dan Tabel 12. Analisis skor validitas empiris yang didapatkan dari uji coba I, II, dan III menggunakan analisis model Rasch dengan bantuan program *Winstep*.

Tabel 10. Hasil Analisis Skor Validasi Keterbacaan Instrumen

Nomor soal	Respon Rater							Indeks Aiken	Kriteria
	I	II	III	IV	V	VI	VII		
1.	4	5	5	5	5	4	4	0,89	Valid
2.	4	4	5	5	4	4	4	0,82	Valid
3.	4	4	5	5	4	4	4	0,82	Valid
4.	4	4	4	5	4	4	4	0,78	Valid
5.	4	4	5	5	4	4	4	0,82	Valid
6.	4	4	4	4	4	4	4	0,75	Valid
7.	3	4	4	5	4	4	4	0,75	Valid
8.	3	4	5	5	4	3	4	0,75	Valid
9.	3	4	5	5	3	4	4	0,75	Valid
10.	3	4	5	5	4	4	4	0,78	Valid
11.	3	4	5	5	5	4	4	0,82	Valid
12.	4	4	5	4	5	4	4	0,82	Valid
13.	4	4	4	4	4	4	4	0,75	Valid
14.	4	4	4	4	4	4	4	0,75	Valid
15.	4	4	4	4	4	4	4	0,75	Valid
16.	4	4	4	4	4	4	4	0,75	Valid
17.	4	4	4	4	4	4	4	0,75	Valid
18.	3	4	5	5	4	4	4	0,78	Valid
19.	3	4	5	5	4	4	4	0,78	Valid
20.	3	4	4	4	5	4	4	0,75	Valid
21.	3	4	5	5	4	4	4	0,78	Valid
22.	3	4	5	5	4	4	4	0,78	Valid
23.	3	4	5	5	4	4	3	0,75	Valid
24.	4	4	5	5	4	4	4	0,82	Valid
25.	4	4	4	4	4	4	4	0,75	Valid

Berdasarkan Tabel 10, semua butir soal (25 soal uraian) dalam instrumen penilaian kemampuan literasi kimia berbasis etnosains Batik Pekalongan memenuhi kriteria valid dengan nilai indeks Aiken $\geq 0,75$ sehingga layak untuk digunakan sebagai instrumen penilain (Aiken, 1985). Butir-butir soal kemudian diperbaiki sesuai saran yang diperoleh. Secara rinci saran yang diberikan disajikan pada Tabel 11.

Tabel 11. Saran Validator pada Validasi Keterbacaan Instrumen

No.	Saran	Nomor Butir Soal
1.	Perbaiki kalimat perintah menjawab dalam soal	4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 21, 22, 24
2.	Perbaiki gambar	Bacaan 2
3.	Perbaiki penulisan dalam tabel	16, 17

Saran perbaikan kalimat perintah menjawab soal adalah saran yang paling banyak diberikan oleh validator. Tujuannya agar jawaban atau respon peserta didik tidak terlalu luas atau terbuka karena soal berbentuk uraian biasanya akan menghasilkan jawaban yang bervariasi (Latifah & Fiati, 2014). Butir-butir soal hasil validasi dibedakan menjadi tiga jenis atau kategori yaitu tanpa revisi, revisi, dan tidak dapat digunakan. Secara rinci kategori untuk semua butir soal tersaji pada Tabel 12.

Tabel 12. Kategori Butir Soal Hasil Validasi Keterbacaan Instrumen

	Nomor Butir Soal		
	Dapat Digunakan tanpa Direvisi	Dapat Digunakan dengan Revisi	Tidak Dapat Digunakan
	1, 2, 3, 13, 18, 19, 20, 23, 25	4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 21, 22, 24	-
Jumlah	9 butir soal	16 butir soal	-

Jumlah soal yang dapat digunakan tanpa revisi sebanyak sembilan butir, dan sebanyak 16 butir soal dapat digunakan dengan revisi. Perbaikan atau revisi pada butir soal dilakukan sesuai saran yang diberikan oleh validator.

Hasil analisis skor validitas konstruk pada instrumen penilaian literasi kimia berbasis etnosains Batik Pekalongan tersaji pada Tabel 13.

Tabel 13. Hasil Analisis Skor Validasi Konstruk Instrumen

Nomor soal	Respon Rater							Indeks Aiken	Kriteria
	I	II	III	IV	V	VI	VII		
1.	4	5	5	5	4	4	4	0,86	Valid
2.	3	4	5	5	4	4	3	0,75	Valid
3.	4	5	5	5	3	4	4	0,82	Valid
4.	4	5	5	5	4	4	3	0,82	Valid
5.	4	4	5	4	4	4	4	0,78	Valid
6.	4	4	4	4	4	4	4	0,75	Valid
7.	4	4	4	4	4	4	4	0,75	Valid
8.	4	4	4	4	4	4	4	0,75	Valid
9.	4	4	5	5	3	4	4	0,78	Valid
10.	4	4	5	5	4	4	4	0,82	Valid
11.	4	4	5	5	4	4	4	0,82	Valid
12.	4	4	4	5	4	4	4	0,78	Valid
13.	4	4	5	5	4	4	4	0,82	Valid
14.	4	4	5	5	4	4	4	0,82	Valid
15.	4	4	4	5	4	4	4	0,78	Valid
16.	4	4	4	5	4	4	4	0,78	Valid
17.	4	4	5	4	4	4	4	0,78	Valid
18.	4	4	5	4	4	4	4	0,78	Valid
19.	4	4	4	4	4	4	4	0,75	Valid
20.	4	4	4	4	4	4	4	0,75	Valid
21.	4	4	4	5	4	4	4	0,78	Valid
22.	4	4	4	4	4	4	4	0,75	Valid
23.	4	4	5	5	4	4	4	0,82	Valid
24.	4	4	4	5	4	4	4	0,78	Valid
25.	4	4	5	5	4	4	4	0,82	Valid

Berdasarkan Tabel 13, semua butir soal (25 soal uraian) sesuai dengan kisi-kisi yang telah dibuat. Kesesuaian tersebut dibuktikan dengan analisis skor validitas yang memenuhi kriteria valid berdasarkan indeks Aiken, yaitu perhitungan formula Aiken yang didapatkan $\geq 0,75$ (Aiken, 1985). Secara umum, yang perlu diperbaiki atas saran validator adalah rumusan kompetensi dasar dan indikator perlu dicermati kembali.

Hasil analisis skor validitas materi pada instrumen penilaian literasi kimia berbasis etnosains Batik Pekalongan tersaji pada Tabel 14.

Tabel 14. Hasil Analisis Skor Validasi Materi Instrumen

Nomor Soal	Respon Rater						Indeks Aiken	Kriteria
	I	II	III	IV	V	VI		
1.	4	5	5	4	4	4	0,83	Valid
2.	4	5	5	5	4	5	0,92	Valid
3.	4	5	5	5	4	4	0,87	Valid
4.	4	5	5	4	4	4	0,83	Valid
5.	4	5	5	5	4	4	0,87	Valid
6.	4	4	4	5	4	5	0,83	Valid
7.	4	4	4	5	4	4	0,79	Valid
8.	4	4	4	4	4	5	0,79	Valid
9.	4	4	4	5	4	5	0,83	Valid
10.	4	4	4	5	4	5	0,83	Valid
11.	4	4	4	5	4	5	0,83	Valid
12.	4	5	5	5	4	4	0,87	Valid
13.	4	4	5	5	4	4	0,83	Valid
14.	4	5	5	5	4	4	0,87	Valid
15.	4	4	4	5	4	4	0,79	Valid
16.	4	4	4	4	4	5	0,79	Valid
17.	4	5	4	4	4	4	0,79	Valid
18.	4	5	3	5	4	5	0,83	Valid
19.	4	4	4	5	4	5	0,83	Valid
20.	4	4	4	5	4	5	0,83	Valid
21.	4	4	4	5	4	4	0,79	Valid
22.	4	5	4	4	4	4	0,79	Valid
23.	4	5	4	4	4	4	0,79	Valid
24.	4	5	5	4	4	4	0,83	Valid
25.	4	4	5	5	4	4	0,83	Valid

Berdasarkan Tabel 14, 25 butir soal dapat disimpulkan valid yaitu memenuhi kesesuaian materi larutan elektrolit dan non elektrolit karena nilai dari perhitungan formula Aiken yang diperoleh memenuhi kriteria $\geq 0,79$ pada indeks Aiken (Aiken, 1985). Saran umum perbaikan yang diberikan tersaji pada Tabel 15.

Tabel 15. Saran Perbaikan dari Hasil Validasi Materi Instrumen

No.	Saran	Nomor Butir Soal
1.	Penambahan nama senyawa dan perbaikan rumus kimia	4, 17
2.	Penambahan rumus kimia dalam bacaan	Bacaan 1

Saran perbaikan yang diberikan oleh validator berupa penulisan senyawa kimia pada soal nomor 17 dari senyawa $(\text{NH}_2)_2\text{CO}$ menjadi $\text{CO}(\text{NH}_2)_2$, dan nomor 4 dengan penambahan nama senyawa asam klorida pada keterangan air keras yang kurang familiar bagi peserta didik. Rumus kimia pewarna indigosol pada Batik Pekalongan ditambahkan pada bacaan 1 sesuai dengan saran validator.

Produk awal instrumen penilaian kemampuan literasi kimia berbasis etnosains Batik Pekalongan secara umum dinyatakan valid dan layak digunakan setelah dilakukan perbaikan sesuai saran validator. Butir-butir soal kemudian disusun kembali untuk dijadikan produk awal dan dilakukan uji coba pertama pada peserta didik untuk memperoleh kualitas serta karakteristik butir soal.

B. Pembahasan dan Penemuan

Uji coba produk instrumen penilaian dilakukan pada bulan Februari 2019 di tiga sekolah, yaitu SMA Negeri 1 Pekalongan, SMA Negeri 2 Pekalongan, dan SMA Negeri 3 Pekalongan. Uji coba dilakukan sebanyak tiga kali, yaitu uji coba I dan II sebagai uji coba secara empiris dan uji coba III sebagai implementasi produk. Uji coba ke I melibatkan 156 peserta didik, uji coba ke II melibatkan 255 peserta didik, dan uji coba ke III atau implementasi produk melibatkan 197 peserta didik.

1. Uji Coba

Uji coba dilakukan untuk menguji butir soal secara empiris sehingga didapatkan bukti kualitas dan karakteristik butir soal pada instrumen penilaian literasi kimia berbasis etnosains Batik Pekalongan. Uji coba dalam penelitian pengembangan ini dilakukan sebagai berikut.

a. Uji Coba I

Produk awal instrumen penilaian kemampuan literasi kimia berbasis etnosains Batik Pekalongan terdiri dari 25 butir soal uraian pada materi larutan elektrolit dan non-elektrolit. Peserta didik yang telah selesai mempelajari materi kimia larutan elektrolit dan non-elektrolit diuji kemampuan literasi kimianya dengan mengerjakan instrumen penilaian literasi kimia yang dikaitkan dengan etnosains Batik Pekalongan. Waktu mengerjakan soal dibutuhkan 90 menit.

b. Uji Coba II

Uji coba II dilakukan setelah analisis uji coba I dilakukan. Butir soal yang dianggap tidak valid secara empiris pada tahap uji coba I akan dieliminasi dan diganti butir soal yang lebih baik. Penggantian butir soal yang tidak valid disajikan pada Lampiran 15. Analisis pada uji coba II menghasilkan produk yang siap diimplementasikan untuk mengukur kemampuan literasi kimia peserta didik kelas X pada materi larutan elektrolit dan non-elektrolit. Waktu mengerjakan soal pada uji coba II adalah 90 menit atau setara dengan 2 jam pelajaran untuk kelas X.

c. Uji Coba III (Implementasi Produk)

Uji coba III adalah proses uji coba yang terakhir atau implementasi produk akhir dari pengembangan instrumen penilaian. Hasil uji coba III pada peserta didik akan menunjukkan profil kemampuan literasi kimia peserta didik pada materi larutan elektrolit dan non-elektrolit.

2. Analisis Kualitas Instrumen Penilaian Literasi Kimia

a. Uji Asumsi Unidimensi Uji Coba I, II, dan III

Hasil uji coba kemudian dianalisis menggunakan respon butir teori atau *Modern Item Response Theory* (IRT), model Rasch dengan bantuan program *Winstep 3.73*. Bond & Fox (2015: 307) menjelaskan bahwa IRT menggunakan konsep unidimensi dari *item* yang digunakan untuk mengukur sesuatu yang bersifat laten atau dalam penelitian ini adalah kemampuan literasi kimia peserta didik. Konsep unidimensi dalam instrumen ditunjukkan dengan setiap butir soal atau *item* mengukur satu indikator. Unidimensionalitas instrumen dapat dilihat dari validasi oleh *expert* atau analisis kuantitatif dengan bantuan program *Winstep* pada *output tables 23*.

Uji unidimensionalitas instrumen pada uji coba I didapatkan sebesar 31,9%, uji coba II 32,6%, dan uji coba III atau implementasi sebesar 36,1%. Ketiga uji coba instrumen telah memenuhi persyaratan unidimensionalitas dengan batas minimal prasyarat unidimensionalitas sebesar 20% (Brentani & Golia, 2007; Sumintono & Widhiarso, 2014: 122). Tujuan dilakukan analisis instrumen untuk mendapatkan karakteristik setiap butir soal sehingga instrumen yang digunakan memiliki kualitas yang baik. Kualitas instrumen dapat dilihat dari

analisis 1) unidimensi uji coba I, II, dan III, 2) tingkat kesesuaian butir soal terhadap model Rasch sebagai fungsi normal butir soal yang dilihat dari *output tables 10 item column: fit order*, 3) Reliabilitas dan separasi atau kelompok *item* dan peserta didik, serta 4) analisis *bias* pada butir soal.

b. Uji Coba I

1) Tingkat Kesesuaian Butir Soal terhadap Model Rasch (*Item Fit*)

Item fit menggambarkan apakah butir soal yang digunakan dapat berfungsi normal dalam melakukan pengukuran atau tidak serta untuk mengetahui seberapa baik *item* dapat dipahami responden (peserta tes). Analisis ketepatan pengukuran dapat dilakukan dengan melihat nilai *outfit mean square*, *outfit z-standard*, dan *PT measure correlation*. Butir soal yang memenuhi ketepatan pengukuran dapat diterima atau digunakan sebagai instrumen untuk melakukan pengukuran yang dikehendaki yaitu mengukur kemampuan literasi kimia peserta didik. Jika butir soal tidak *fit (misfit)*, maka sangat dimungkinkan bahwa butir soal mengindikasikan terjadinya miskonsepsi pada responden (peserta tes).

Kriteria nilai *outfit mean square*, *outfit z-standard*, dan *PT measure correlation* dapat ditentukan dari: 1) nilai *outfit mean square (MNSQ)* dapat diterima jika berada di antara 0,5 sampai 1,5; 2) nilai *outfit z-standard (ZSTD)* diterima jika berada di antara -2,0 sampai +2,0; dan 3) nilai *point measure correlation (Pt Mean Corr)* berada di antara 0,4 sampai 0,85 (Bond & Fox, 2015; Sumintono & Widhiarso, 2014; Park & Liu, 2019). Jika butir soal tidak memenuhi minimal dua kriteria tersebut, maka butir soal dikatakan tidak *fit (misfit)* sehingga dieliminasi atau perlu diganti.

Hasil analisis *item fit* butir soal pada uji coba I disajikan pada Tabel 16 dan nilai yang tidak memenuhi kriteria dicetak tebal dan miring.

Tabel 16. Hasil Analisis Kesesuaian Butir Soal terhadap Model Rasch

No.	Nomor Butir Soal	Kriteria Ketepatan Pengukuran			Keputusan
		<i>Outfit MNSQ</i>	<i>Outfit ZSTD</i>	PT-MEAN CORR	
1.	S20	2,32	6,6	0,20	Ditolak
2.	S6	1,59	4,9	0,42	Ditolak
3.	S1	1,34	2,9	0,28	Ditolak
4.	S15	1,33	1,5	0,49	Diterima
5.	S24	1,20	1,5	0,56	Diterima
6.	S19	1,21	1,6	0,40	Diterima
7.	S10	1,11	0,6	0,55	Diterima
8.	S3	1,06	0,6	0,44	Diterima
9.	S25	0,97	-0,1	0,50	Diterima
10.	S9	1,00	0,1	0,44	Diterima
11.	S21	0,90	-0,5	0,55	Diterima
12.	S17	1,01	0,1	0,42	Diterima
13.	S2	1,00	0,1	0,43	Diterima
14.	S16	0,91	-0,8	0,49	Diterima
15.	S11	0,87	-1,2	0,45	Diterima
16.	S14	0,89	-0,6	0,54	Diterima
17.	S8	0,86	-0,9	0,52	Diterima
18.	S13	0,78	-1,9	0,63	Diterima
19.	S4	0,82	-1,5	0,41	Diterima
20.	S7	0,66	-2,5	0,50	Diterima
21.	S5	0,72	-3,1	0,37	Ditolak
22.	S12	0,71	-2,0	0,63	Diterima
23.	S18	0,64	-3,4	0,33	Ditolak
24.	S23	0,64	-2,9	0,665	Diterima
25.	S22	0,61	-2,0	0,13	Diterima

Berdasarkan Tabel 16, dari 25 butir soal sebanyak 5 butir soal yang tidak memenuhi kriteria nilai *MNSQ* berada di antara 0,5 sampai 1,5; nilai *ZSTD* berada di antara -2,0 sampai +2,0; dan/atau nilai *Pt Mean Corr* berada di antara 0,4 sampai 0,85 (Bond & Fox, 2015; Sumintono & Widhiarso, 2014; Park & Liu,

2019). Kelima butir soal tersebut adalah butir soal nomor 20, 6, 1, 5, dan 18. Penggantian dan/atau eliminasi pada butir soal yang tidak valid disajikan pada Lampiran 15. Sebanyak 20 butir soal yang dinyatakan valid dapat digunakan untuk mengukur. Soal kemudian disusun kembali menjadi instrumen penilaian dan diujicobakan pada uji coba II.

2) Uji Reliabilitas dalam *Summary Statistics*

Nilai reliabilitas *person* (peserta didik) dari skor yang didapat pada uji coba I sebesar 0,75. Nilai 0,75 menunjukkan bahwa reliabilitas *person* atau peserta didik pada kategori sedang. Nilai reliabilitas *item* diperoleh 0,95 dengan kategori tinggi. Secara keseluruhan rata-rata nilai reliabilitas atau *Cronbach Alpha* sebesar 0,86 berkategori sangat baik (Cronbach, 1951; Cortina, 1993). Nilai reliabilitas menunjukkan keajegan, keandalan atau konsistensi tes dan kesesuaian tes pada saat diujicobakan pada peserta didik. Sehingga dapat disimpulkan bahwa skor yang diperoleh dari instrumen penilaian tersebut reliabel karena telah memenuhi nilai koefisien minimal 0,6 (Sujarwanto & Rusilowati, 2015). Nilai reliabilitas dan kategori keajegan tersaji pada Tabel 17 dan Gambar 8.

Nilai *separation* menunjukkan tingkatan pembagian kelompok kemampuan peserta didik dalam menjawab instrumen penilaian. Berdasarkan Gambar 8 dihasilkan pembagian kelompok atau *separation* peserta didik, yaitu sebanyak dua kelompok kemampuan yang diambil dari pembulatan nilai separasi sebesar 1,73 dan kelompok *item* terbagi menjadi empat kelompok dari pembulatan 4,49.

Tabel 17. Nilai Reliabilitas pada Uji Coba I

Reliabilitas	Nilai	Kategori
Reliabilitas <i>Item</i>	0,95	Tinggi
Reliabilitas <i>Person</i>	0,75	Sedang
<i>Cronbach Alpha</i>	0,86	Sangat Baik

TABLE 3.1 C:\Users\admin\Desktop\2.prn ZOU733WS.TXT May 25 10:40 2019
 INPUT: 156 Person 25 Item REPORTED: 156 Person 25 Item 5 CATS WINSTEPS 3.73

SUMMARY OF 156 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	51.5	20.5	-.20	.18	1.00	.0	.99	.0
S.D.	16.8	4.6	.40	.03	.29	1.0	.36	1.0
MAX.	108.0	25.0	1.20	.33	2.02	3.6	2.89	3.6
MIN.	16.0	7.0	-1.26	.15	.51	-2.0	.50	-1.8
REAL RMSE	.20	TRUE SD	.34	SEPARATION	1.73	<u>Person RELIABILITY</u>		.75
MODEL RMSE	.19	TRUE SD	.35	SEPARATION	1.86	Person RELIABILITY		.78
S.E. OF Person MEAN = .03								

VALID RESPONSES: 82.1% (APPROXIMATE)

Person RAW SCORE-TO-MEASURE CORRELATION = .64 (approximate due to missing data)
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .86 (approximate due

SUMMARY OF 25 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	321.2	128.0	.00	.07	1.01	-.1	1.01	-.1
S.D.	96.7	25.6	.36	.01	.35	2.8	.36	2.3
MAX.	528.0	155.0	.60	.10	2.20	6.9	2.32	6.6
MIN.	171.0	81.0	-.53	.06	.56	-4.9	.61	-3.4
REAL RMSE	.08	TRUE SD	.36	SEPARATION	4.49	<u>Item RELIABILITY</u>		.95
MODEL RMSE	.07	TRUE SD	.36	SEPARATION	4.78	Item RELIABILITY		.96
S.E. OF Item MEAN = .07								

Gambar 8. Summary Statistics Uji Coba I

c. Uji Coba II

Butir soal yang digunakan pada uji coba II sebanyak 20 butir. 20 butir soal dipilih berdasarkan hasil analisis *item fit* pada uji coba I yang kemudian di susun ulang. Rincian nomor butir soal yang digunakan pada uji coba II tersaji pada Tabel 18.

Tabel 18. Soal yang Digunakan untuk Uji Coba II

Uji Coba I		Uji Coba II	
Nomor Butir Soal	Keterangan	Keterangan	Nomor Butir Soal
1	Digunakan	Diganti	1
2	Digunakan	Diganti	2
3	Digunakan	Dibuang	-
4	Digunakan	Digunakan	3
5	Digunakan	Dibuang	-
6	Digunakan	Diganti	4
7	Digunakan	Digunakan	5
8	Digunakan	Digunakan	6
9	Digunakan	Digunakan	7
10	Digunakan	Digunakan	8
11	Digunakan	Digunakan	9
12	Digunakan	Dibuang	-
13	Digunakan	Dibuang	-
14	Digunakan	Digunakan	10
15	Digunakan	Digunakan	11
16	Digunakan	Digunakan	12
17	Digunakan	Digunakan	13
18	Digunakan	Diganti	14
19	Digunakan	Digunakan	15
20	Digunakan	Diganti	16
21	Digunakan	Digunakan	17
22	Digunakan	Dibuang	-
23	Digunakan	Digunakan	18
24	Digunakan	Digunakan	19
25	Digunakan	Digunakan	20

1) Tingkat Kesesuaian Butir Soal terhadap Model Rasch (*Item Fit*)

Hasil analisis kesesuaian butir soal pada uji coba II disajikan pada Tabel

19. Nilai yang tidak memenuhi kriteria dicetak tebal dan miring.

Tabel 19. Hasil Analisis Kesesuaian Butir Soal terhadap Model Rasch

No.	Nomor Butir Soal	Kriteria Ketepatan Pengukuran			Keputusan
		<i>Outfit MNSQ</i>	<i>Outfit ZSTD</i>	PT-MEAN CORR	
1.	S19	1,40	3,4	0,34	Ditolak
2.	S7	1,37	3,4	0,34	Ditolak
3.	S20	1,20	1,9	0,42	Diterima
4.	S4	1,17	1,4	0,39	Diterima
5.	S3	1,16	1,5	0,47	Diterima
6.	S8	1,15	1,7	0,49	Diterima
7.	S5	1,06	0,8	0,34	Diterima
8.	S6	1,02	0,3	0,45	Diterima
9.	S16	1,02	0,2	0,47	Diterima
10.	S11	0,94	-0,7	0,56	Diterima
11.	S9	0,99	-0,1	0,52	Diterima
12.	S13	0,89	-0,8	0,50	Diterima
13.	S2	0,84	-1,8	0,72	Diterima
14.	S12	0,93	-0,6	0,51	Diterima
15.	S10	0,86	-1,6	0,55	Diterima
16.	S15	0,90	-1,0	0,53	Diterima
17.	S18	0,87	-1,4	0,53	Diterima
18.	S17	0,90	-0,9	0,46	Diterima
19.	S1	0,88	-1,1	0,43	Diterima
20.	S14	0,71	-3,6	0,61	Ditolak

Berdasarkan Tabel 19, dari 20 butir soal sebanyak 3 butir soal dieliminasi atau dibuang karena baik butir 19 (S19), 7 (S7) dan 14 (S14) tidak memenuhi kriteria nilai *ZSTD* antara -2,0 sampai +2,0; dan nilai *Pt Mean Corr* berada di antara 0,4 sampai 0,85 (Bond & Fox, 2015; Sumintono & Widhiarso, 2014; Park & Liu, 2019).

Sebanyak 17 butir soal yang memenuhi kriteria nilai *MNSQ* dan nilai *ZSTD* dapat digunakan untuk uji coba III. Butir soal kemudian disusun kembali menjadi seperangkat instrumen penilaian literasi kimia sebagai produk akhir untuk diimplementasikan.

2) Uji Reliabilitas dalam *Summary Statistics*

Nilai reliabilitas atau *cronbach alpha* pada uji coba II sebesar 0,92 yang memiliki reliabilitas tinggi dan berkategori baik sekali (Cronbach, 1951; Cortina, 1993). Berdasarkan Gambar 9, nilai reliabilitas peserta didik sebesar 7,4 yang tergolong dalam kategori sedang. Konsistensi peserta didik dalam menjawab soal adalah cukup serta kualitas *item* atau soal yang istimewa yaitu dibuktikan dengan nilai reliabilitas soal sebesar 0,97. Tingginya nilai reliabilitas menunjukkan tingkat konsistensi yang baik ketika melakukan sebuah pengukuran. Nilai reliabilitas juga disajikan pada Tabel 20 dan Gambar 9.

Tabel 20. Nilai Reliabilitas pada Uji Coba II

Reliabilitas	Nilai	Kategori
Reliabilitas <i>Item</i>	0,97	Tinggi
Reliabilitas <i>Person</i>	0,74	Sedang
<i>Cronbach Alpha</i>	0,92	Baik Sekali

TABLE 3.1 C:\Users\admin\Documents\K1_X1.prn ZOU797WS.TXT Jul 12 10:47 2019
INPUT: 255 Person 20 Item REPORTED: 255 Person 20 Item 5 CATS WINSTEPS 3.73

SUMMARY OF 255 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	52.6	17.0	.09	.20	1.04	.1	1.05	.1
S.D.	18.5	4.1	.43	.04	.32	1.1	.40	1.0
MAX.	95.0	20.0	1.42	.39	1.91	2.5	3.04	2.6
MIN.	10.0	5.0	-1.13	.16	.33	-3.2	.38	-2.9
REAL RMSE	.22	TRUE SD	.37	SEPARATION	1.70	Person	RELIABILITY	.74
MODEL RMSE	.20	TRUE SD	.38	SEPARATION	1.89	Person	RELIABILITY	.78
S.E. OF Person MEAN = .03								

VALID RESPONSES: 84.8% (APPROXIMATE)
Person RAW SCORE-TO-MEASURE CORRELATION = .78 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .92 (approximate due to missing data)

SUMMARY OF 20 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	670.5	216.2	.00	.05	1.01	.0	1.01	.0
S.D.	162.8	28.6	.31	.00	.14	1.8	.18	1.7
MAX.	994.0	254.0	.55	.06	1.35	3.9	1.40	3.4
MIN.	429.0	163.0	-.54	.05	.76	-3.5	.71	-3.6
REAL RMSE	.05	TRUE SD	.30	SEPARATION	5.56	Item	RELIABILITY	.97
MODEL RMSE	.05	TRUE SD	.30	SEPARATION	5.74	Item	RELIABILITY	.97
S.E. OF Item MEAN = .07								

UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -.86 (approximate due to missing data)
4324 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 11521.81 with 4047 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): 1.3005

Gambar 9. Summary Statistics Uji Coba II

Pengelompokan *person* dan *item* dapat dilihat dari nilai *separation*. Semakin besar nilai *separation* maka kualitas instrumen dan peserta didik sebagai responden adalah semakin bagus karena instrumen dapat mengidentifikasi kelompok kemampuan peserta didik dan butir soal (*item*) yang digunakan secara lebih teliti (Sumintono & Widhiarso, 2014). Berdasarkan Gambar 9 didapatkan besarnya pembagian kelompok kemampuan peserta didik yaitu dua kelompok. Dua kelompok ini diambil berdasarkan nilai *separation* 1,70. Sedangkan untuk *item* terbagi menjadi enam kelompok yang diambil berdasarkan nilai *separation* sebesar 5,56.

d. Uji Coba III (Implementasi Instrumen)

Uji coba III dilakukan kepada 197 peserta didik. Instrumen penilaian yang digunakan terdiri 17 butir soal yang dinyatakan valid pada uji coba II berdasarkan permodelan Rasch. Susunan ulang instrumen penilaian untuk uji coba III disajikan pada Tabel 21. Berdasarkan kisi-kisi instrumen penilaian, butir soal yang dieliminasi adalah butir soal nomor 19 dan 7 yang menguji pemecahan masalah dan nomor 14 merupakan interpretasi data ilmiah.

Tabel 21. Susunan Instrumen Penilaian Uji Coba III

Uji Coba II		Uji Coba III	
Nomor Butir Soal	Keterangan	Keterangan	Nomor Butir Soal
1	Digunakan	Tetap digunakan	1
2	Digunakan	Tetap digunakan	2
3	Digunakan	Tetap digunakan	3
4	Digunakan	Tetap digunakan	4
5	Digunakan	Tetap digunakan	5
6	Digunakan	Tetap digunakan	6
7	Digunakan	Dibuang	-
8	Digunakan	Tetap digunakan	7
9	Digunakan	Tetap digunakan	8
10	Digunakan	Tetap digunakan	9
11	Digunakan	Tetap digunakan	10
12	Digunakan	Tetap digunakan	11
13	Digunakan	Tetap digunakan	12
14	Digunakan	Dibuang	-
15	Digunakan	Tetap digunakan	13
16	Digunakan	Tetap digunakan	14
17	Digunakan	Tetap digunakan	15
18	Digunakan	Tetap digunakan	16
19	Digunakan	Dibuang	-
20	Digunakan	Tetap digunakan	17

1) Tingkat Kesesuaian Butir Soal terhadap Model Rasch (*Item Fit*)

Analisis kesesuaian butir soal (*item fit*) pada uji coba III disajikan pada Tabel 22. *Item* dikatakan *fit* atau sesuai model Rasch jika memenuhi batasan nilai *outfit MNSQ*, *ZSTD*, dan/atau *CORR*. Dikatakan *misfit* apabila tidak memenuhi batasan kriteria nilai *outfit mean square (MNSQ)* berada di antara 0,5 sampai 1,5; nilai *ZSTD* berada di antara -2,0 sampai +2,0; dan/atau nilai *Pt Mean Corr* berada di antara 0,4 sampai 0,85 (Bond & Fox, 2015; Sumintono & Widhiarso, 2014; Park & Liu, 2019).

Tabel 22. Hasil Analisis Kesesuaian Butir Soal terhadap Model Rasch

No.	Nomor Butir Soal	Kriteria Ketepatan Pengukuran			Keputusan
		<i>Outfit MNSQ</i>	<i>Outfit ZSTD</i>	PT-MEAN CORR	
1.	S12	1,27	1,2	0,38	Diterima
2.	S17	1,10	0,6	0,45	Diterima
3.	S14	1,19	1,6	0,45	Diterima
4.	S5	1,21	1,8	0,34	Diterima
5.	S7	1,19	1,4	0,45	Diterima
6.	S4	1,17	1,5	0,58	Diterima
7.	S13	1,14	1,2	0,43	Diterima
8.	S6	1,05	0,5	0,37	Diterima
9.	S1	1,01	0,1	0,41	Diterima
10.	S16	0,98	0,0	0,34	Diterima
11.	S8	0,89	-1,0	0,55	Diterima
12.	S3	0,93	-0,6	0,50	Diterima
13.	S9	0,82	-1,2	0,54	Diterima
14.	S10	0,80	-1,9	0,65	Diterima
15.	S2	0,77	-2,0	0,69	Diterima
16.	S15	0,84	-1,5	0,50	Diterima
17.	S11	0,83	-1,7	0,60	Diterima

Secara umum instrumen yang diujikan pada uji coba III, terdapat 17 butir soal yang memenuhi kriteria nilai *MNSQ*, *ZSTD* dan/atau *Pt Mean Corr*. Syarat kesesuaian butir soal dengan model Rasch juga menunjukkan bahwa butir soal yang digunakan dapat mengukur kemampuan literasi kimia peserta didik dengan baik. Sebanyak 17 butir soal tersebut kemudian dijadikan produk akhir dari pengembangan instrumen penilaian.

2) Uji Reliabilitas dalam *Summary Statistics*

Nilai reliabilitas atau *cronbach alpha* uji coba III sebesar 0,89. Nilai reliabilitas tersebut dalam kategori sangat baik karena nilai reliabilitas yang dihasilkan lebih dari 0,6 (Sujarwanto & Rusilowati, 2015). Nilai reliabilitas responden yang didapatkan sebesar 0,78 dengan kategori sedang. Nilai reliabilitas soal didapatkan sebesar 0,97 dengan kategori tinggi (Boone, Stever, & Yale, 2014). Tingginya nilai reliabilitas menunjukkan tingkat konsistensi dalam sebuah pengukuran. Tabel 23 menunjukkan nilai reliabilitas uji coba III.

Tabel 23. Nilai Reliabilitas pada Uji Coba III

Reliabilitas	Nilai	Kategori
Reliabilitas <i>Item</i>	0,97	Tinggi
Reliabilitas <i>Person</i>	0,78	Sedang
<i>Cronbach Alpha</i>	0,89	Sangat Baik

TABLE 3.1 C:\Users\admin\Desktop\UJI 3 .prn ZOU487WS.TXT Jul 13 9:54 2019
 INPUT: 197 Person 17 Item REPORTED: 197 Person 17 Item 5 CATS WINSTEPS 3.73

SUMMARY OF 197 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	53.3	15.7	.22	.21	.99	.0	1.02	.1
S.D.	15.8	2.3	.48	.07	.31	1.0	.43	1.0
MAX.	84.0	17.0	2.27	.86	1.91	2.8	3.47	4.0
MIN.	10.0	5.0	-1.29	.17	.24	-3.2	.24	-2.5
REAL RMSE	.23	TRUE SD	.43	SEPARATION	1.86	Person	RELIABILITY	.78
MODEL RMSE	.22	TRUE SD	.43	SEPARATION	1.99	Person	RELIABILITY	.80
S.E. OF Person MEAN = .03								
VALID RESPONSES: 92.5% (APPROXIMATE)								
Person RAW SCORE-TO-MEASURE CORRELATION = .89 (approximate due to missing data)								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89 (approximate due to missing data)								
SUMMARY OF 17 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	617.9	182.2	.00	.06	1.01	.0	1.01	.0
S.D.	109.9	10.6	.35	.01	.16	1.7	.16	1.3
MAX.	841.0	196.0	.46	.08	1.26	2.9	1.27	1.8
MIN.	462.0	158.0	-.78	.05	.77	-3.1	.77	-2.0
REAL RMSE	.06	TRUE SD	.35	SEPARATION	5.67	Item	RELIABILITY	.97
MODEL RMSE	.06	TRUE SD	.35	SEPARATION	5.89	Item	RELIABILITY	.97
S.E. OF Item MEAN = .09								
UMEAN=.0000 USCALE=1.0000								
Item RAW SCORE-TO-MEASURE CORRELATION = -.95 (approximate due to missing data)								
3097 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 7589.47 with 2881 d.f. p=.0000								
Global Root-Mean-Square Residual (excluding extreme scores): 1.3046								

Gambar 10. Summary Statistics Uji Coba III

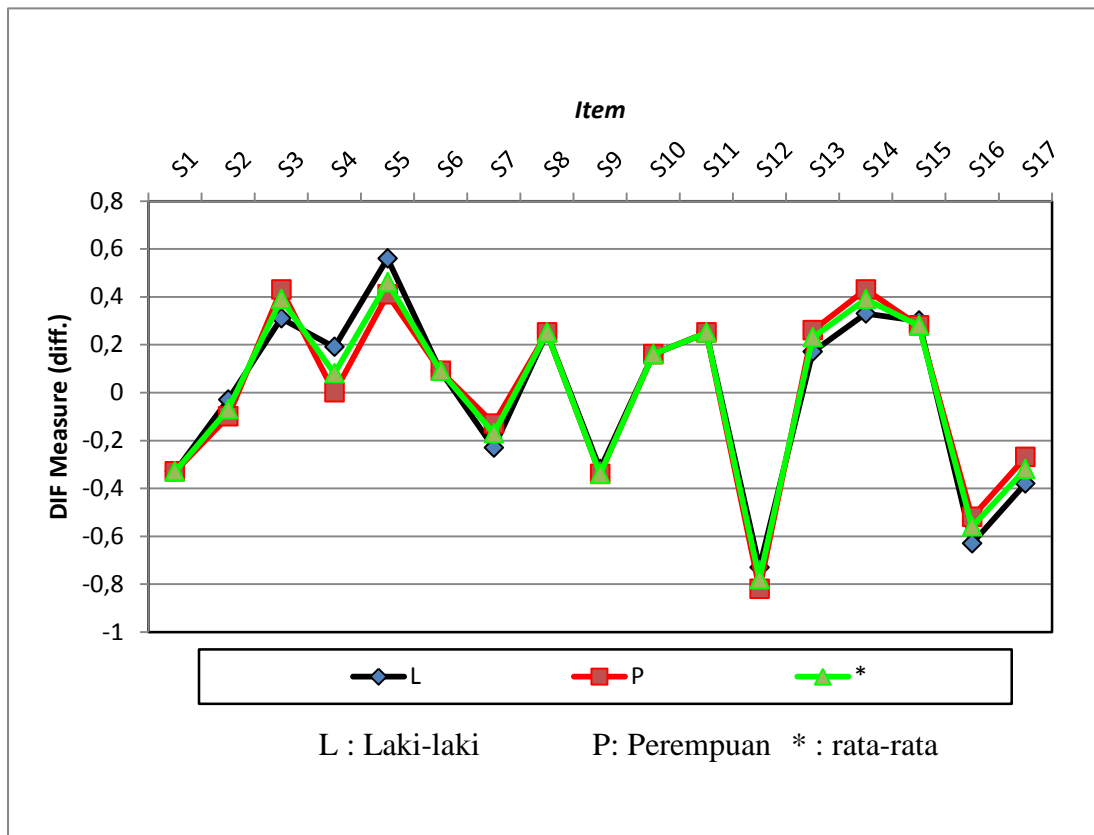
Berdasarkan Gambar 10, uji coba III menghasilkan dua kategori kemampuan siswa dalam menjawab instrumen. Dua kategori tersebut didapatkan dari nilai *separation* 1,86. Sedangkan untuk instrumen memiliki enam kategori atau tingkatan berdasarkan hasil nilai *separation item* 5,67. Semakin banyak kategori menunjukkan semakin baik pula instrumen karena mampu mengidentifikasi kelompok atau tingkatan kemampuan peserta didik dan kesulitan soal (*item*) sebagai alat ukur yang digunakan (Sumintono & Widhiarso, 2014: 112).

3) *Bias pada Item (Differential Item Function/DIF)*

DIF merupakan salah satu hasil analisis untuk mengetahui apakah butir soal atau *item* memiliki *bias* atau kecenderungan pada kategori responden tertentu atau tidak, misalnya kecenderungan terhadap laki-laki atau perempuan (Sumintono & Widhiarso, 2014). Cara mengetahuinya dapat dilihat dari nilai probabilitas butir soal. Butir soal *bias* jika nilai probabilitas kurang dari 5% atau 0,05 (Rome & Zhang, 2018). Hasil analisis *bias* instrumen disajikan pada Tabel 24 atau dapat dilihat pada Gambar 11.

Tabel 24. Hasil Analisis *Bias* pada Instrumen terhadap Respon Peserta Didik

Nomor Soal	Nilai Probabilitas	Keputusan	Nomor Soal	Nilai Probabilitas	Keputusan
1	1,00	Tidak <i>Bias</i>	10	1,00	Tidak <i>Bias</i>
2	0,56	Tidak <i>Bias</i>	11	1,00	Tidak <i>Bias</i>
3	0,29	Tidak <i>Bias</i>	12	0,57	Tidak <i>Bias</i>
4	0,08	Tidak <i>Bias</i>	13	0,40	Tidak <i>Bias</i>
5	0,18	Tidak <i>Bias</i>	14	0,34	Tidak <i>Bias</i>
6	1,00	Tidak <i>Bias</i>	15	0,81	Tidak <i>Bias</i>
7	0,37	Tidak <i>Bias</i>	16	0,48	Tidak <i>Bias</i>
8	1,00	Tidak <i>Bias</i>	17	0,42	Tidak <i>Bias</i>
9	0,81	Tidak <i>Bias</i>			



Gambar 11. Grafik Analisis *Bias* Instrumen

Berdasarkan Gambar 11 disimpulkan bahwa butir soal termudah adalah butir soal nomor 12. Jika dilihat berdasarkan respon dalam bentuk grafik tersebut, peserta didik berjenis kelamin perempuan lebih mudah dan memiliki skor maksimal lebih banyak dibandingkan peserta didik berjenis kelamin laki-laki. Sedangkan untuk soal yang dianggap paling sukar yaitu butir soal nomor 5. Peserta didik jenis kelamin laki-laki lebih menguasai soal tersebut sehingga lebih mudah untuk menjawab dan mendapatkan skor maksimal. Butir soal nomor 5 lebih membutuhkan pemikiran atau logika yang lebih mendalam, karena kompetensi yang diukur adalah kompetensi mengevaluasi penyelidikan ilmiah dari larutan elektrolit kuat, lemah, atau non-elektrolit. Peserta didik diminta untuk

mengevaluasi, mengumpulkan data dan dapat menarik kesimpulan yang tepat dari gambar larutan senyawa kimia secara mikroskopis. Sedangkan, pada butir soal nomor 12 kompetensi yang diujikan adalah kompetensi dalam membaca grafik tentang daya hantar larutan.

3. Analisis Karakteristik Instrumen

a. Uji Coba I

Setiap butir soal memiliki tingkat kesukaran berdasarkan respon kemampuan peserta didik yang ditunjukkan dengan skala *logit* dalam model Rasch. Tingkat kesukaran butir soal dihasilkan pada *output tables 13 item: measure*. Tingkat kesalahan dapat juga dilihat sekaligus. Skala *logit* membagi tingkat atau kelompok kesukaran butir soal dan responden (peserta didik). Informasi tentang tingkat kesukaran *item* dan abilitas atau kemampuan peserta didik dalam menjawab semakin baik jika nilai *logit* semakin besar. *Item* yang baik adalah *item* yang tingkat kesalahannya semakin kecil. Serta butir soal yang baik adalah butir soal yang dapat digunakan untuk mengukur sekaligus membedakan kemampuan setiap peserta didik. Bagus tidaknya butir soal dapat dilihat dari nilai *standard error (SE)*. Butir soal (*item*) dikatakan baik atau ideal jika $SE < 0,5-1,00$. Tingkat kesukaran item dan tingkat kesalahan *item* tersaji dalam Tabel 25.

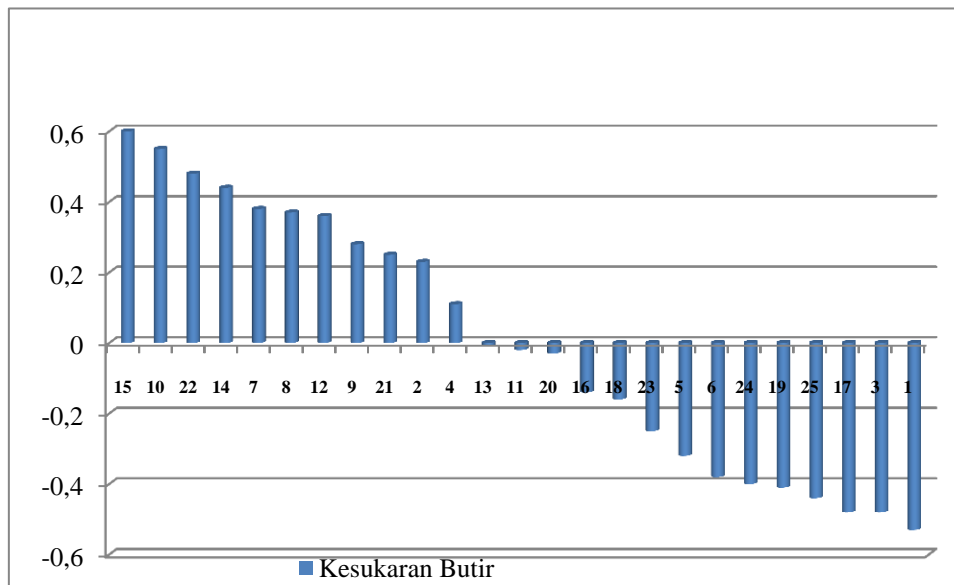
Tabel 25. Tingkat Kesukaran dan Kesalahan Butir Soal Uji Coba I

Butir soal (item)	Tingkat Kesukaran	Tingkat Kesalahan	Butir soal (item)	Tingkat Kesukaran	Tingkat Kesalahan
S15	+0,60	0,10	S20	-0,03	0,08
S10	+0,55	0,08	S16	-0,14	0,06
S22	+0,48	0,10	S18	-0,16	0,07
S14	+0,44	0,08	S23	-0,25	0,08
S7	+0,38	0,08	S5	-0,32	0,06
S8	+0,37	0,08	S6	-0,38	0,06
S12	+0,36	0,08	S24	-0,40	0,08
S9	+0,28	0,07	S19	-0,41	0,07
S21	+0,25	0,09	S25	-0,44	0,08
S2	+0,23	0,07	S17	-0,48	0,07
S4	+0,11	0,06	S3	-0,48	0,06
S13	-0,01	0,06	S1	-0,53	0,06
S11	-0,02	0,06			

Berdasarkan Tabel 25 butir soal di kategorikan dalam tiga kategori tingkat kesukaran yaitu mudah, sedang, dan sukar. Kriteria penerimaan indeks tingkat kesukaran butir soal berdasarkan teori respon butir model Rasch adalah -2 sampai +2 (Hambleton, & Swaminathan, 1985: 107; Mardapi, 2012: 191). Sehingga dapat disimpulkan bahwa secara keseluruhan butir soal memenuhi indeks tingkat kesukaran menurut respon butir teori model Rasch. Semua butir soal memiliki diskriminasi yang baik karena memiliki nilai $SE < 0,5-1,00$ sehingga dapat disimpulkan soal memiliki ketelitian yang baik dalam mengukur kemampuan literasi kimia peserta didik. Kategori setiap butir soal tersaji juga pada Tabel 26 atau Gambar 12.

Tabel 26. Kategori Butir Soal Uji Coba I

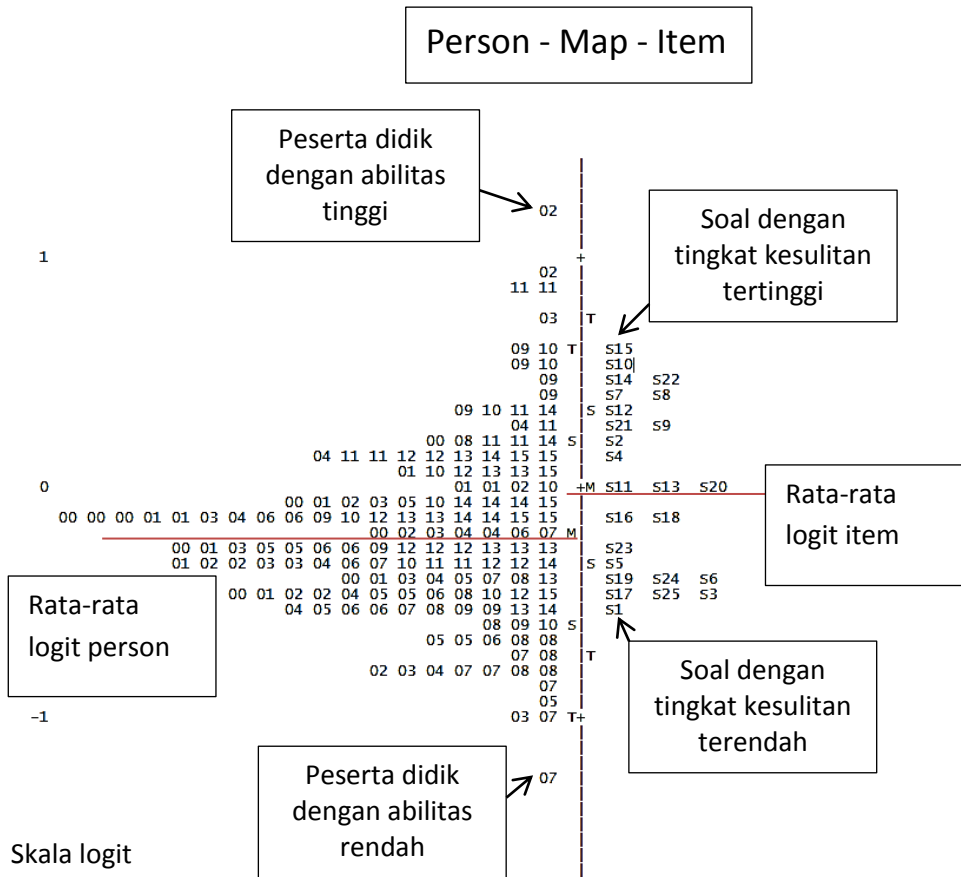
Nomor butir soal	Kategori kesulitan butir soal
S7, S8, S14, S22, S10, S15	Sukar
S12, S21, S9, S2, S4, S11, S13, S20, S16, S18, S23, S5	Sedang
S19, S24, S6, S17, S25, S3, S1	Mudah



Gambar 12. Tingkat Kesukaran Butir Soal pada Uji Coba I

Persebaran kemampuan peserta didik dalam menjawab butir soal atau data sebaran *logit* juga dapat dilihat dari *person-item map*. *Person-item map* disajikan pada Gambar 13. Berdasarkan Gambar 13 soal nomor 13, 11, dan 20 memiliki tingkat kesulitan rata-rata dengan nilai *logit* masing-masing sebesar -0,01; -0,02; dan -0,03. Nilai *logit* yang semakin besar menunjukkan bahwa butir soal memiliki tingkat kesulitan yang lebih tinggi, begitu pula sebaliknya. Tingkat rata-rata abilitas peserta didik dalam menjawab soal uji coba I ditunjukkan peserta didik dengan nomor urut 037C, 048C, 067D, 076D, 025F, dan 042C dengan nilai *logit*

sebesar $-0,17$. Semakin besar nilai *logit* maka kemampuan peserta didik dalam menjawab butir soal didik lebih baik.



Gambar 13. *Person-Item Map* Uji Coba I

Berdasarkan respon dari peserta didik, butir soal yang dianggap sulit adalah butir soal yang menjelaskan pemahaman larutan elektrolit dan non elektrolit berdasarkan bacaan yang tersedia, pemecahan masalah, menginterpretasikan data ilmiah, dan mengevaluasi penyelidikan ilmiah. Butir yang dianggap mudah berkaitan dengan pendapat peserta didik untuk menjelaskan fenomena tentang larutan elektrolit dan non elektrolit dengan larutan yang sering dijumpai.

Butir soal nomor 15 dianggap paling sukar oleh peserta didik. Butir soal nomor 15 merupakan butir soal yang menguji atau mengukur kompetensi tingkat 1, yaitu kemampuan menjelaskan fenomena ilmiah tentang proses pewarnaan pada Batik. Meskipun demikian, peserta didik harus teliti dan fokus pada bacaan, karena peserta didik wajib menjelaskan dan menyimpulkan tentang larutan elektrolit kuat, lemah, atau non-elektrolit berkaitan dengan bacaan yang tersedia. Peserta didik tidak memperoleh skor maksimal karena kurangnya ketelitian dalam membaca dan menemukan kata kunci untuk dapat menyimpulkan jawaban yang seharusnya dijawab.

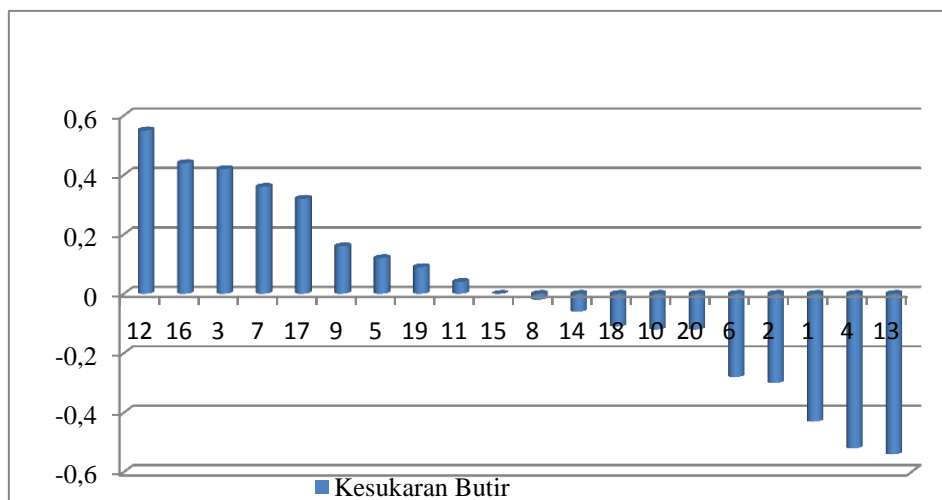
Butir soal nomor 1 merupakan butir soal yang dianggap peserta didik paling mudah dijawab dan mendapatkan skor maksimal. Hal ini karena butir soal nomor 1, peserta didik diminta untuk menjelaskan keterkaitan materi yang mereka pelajari yaitu larutan elektrolit dan non-elektrolit dengan proses pewarnaan Indigosol Batik Pekalongan. Jawaban peserta didik sebagian besar mendapatkan skor maksimal karena peserta didik hanya perlu mencari kata kunci yang menunjukkan adanya keterkaitan larutan elektrolit dan non-elektrolit dalam proses pewarnaan Batik Pekalongan, yaitu penggunaan asam klorida untuk memunculkan warna yang diinginkan pada batik. Melalui kata kunci tersebut, peserta didik dianggap telah memahami konsep materi yang dipelajari.

b. Uji Coba II

Informasi tingkat kesalahan dan tingkat kesukaran butir soal (*item*) disajikan pada Tabel 27 dan Gambar 14.

Tabel 27. Tingkat Kesukaran dan Kesalahan Butir Soal Uji Coba II

Butir soal (item)	Tingkat Kesukaran	Tingkat Kesalahan	Butir soal (item)	Tingkat Kesukaran	Tingkat Kesalahan
12	+0,55	0,06	8	-0,02	0,05
16	+0,44	0,06	14	-0,06	0,05
3	+0,42	0,05	18	-0,11	0,05
7	+0,36	0,05	10	-0,12	0,05
17	+0,32	0,06	20	-0,12	0,06
9	+0,16	0,05	6	-0,28	0,05
5	+0,12	0,05	2	-0,30	0,05
19	+0,09	0,06	1	-0,43	0,05
11	+0,04	0,05	4	-0,52	0,05
15	+0,00	0,05	13	-0,54	0,06



Gambar 14. Tingkat Kesukaran Butir Soal pada Uji Coba II

Berdasarkan Tabel 27 soal memiliki ketelitian yang baik yang dibuktikan dari nilai standar *error* yang memenuhi kriteria nilai $SE < 0,5-1,00$. Gambar 15 merupakan *variable maps* untuk uji coba II. *Variable maps* tersebut menunjukkan kemampuan peserta didik dalam menjawab soal dan tingkat kesulitan setiap butir soal berdasarkan nilai *logit*. Semakin tinggi nilai *logit* maka butir soal dianggap semakin sulit bagi peserta didik. Sebaliknya, semakin rendah nilai *logit* maka

butir soal dianggap semakin mudah. Tingkat kesulitan soal uji coba II disajikan pada Tabel 28.

Tabel 28. Kategori Kesulitan Butir Soal Uji Coba II

Nomor Butir Soal	Kategori Kesulitan Butir Soal
S12, S16, S3, S17, S7	Sukar
S5, S9, S11, S19, S15, S8, S14, S10, S20, S2, S6	Sedang
S1, S13, S4	Mudah

Berdasarkan Tabel 28, butir soal yang dianggap paling sukar adalah butir soal nomor 12. Jika dilihat dari IPK, butir soal nomor 12 merupakan butir soal yang menguji pemahaman peserta didik dalam menginterpretasikan data berupa grafik. Jika peserta didik tidak cermat dalam melihat grafik dan wujud zat dari senyawa yang tertulis, maka peserta didik akan terjebak dalam jawaban yang kurang tepat. Analisis jawaban dari peserta didik membuktikan bahwa peserta didik kurang cermat dalam melihat wujud zat dari suatu senyawa yang tertulis dan grafik daya hantar yang ada. Hal ini menjadi faktor banyaknya peserta didik yang tidak mendapatkan skor maksimal.

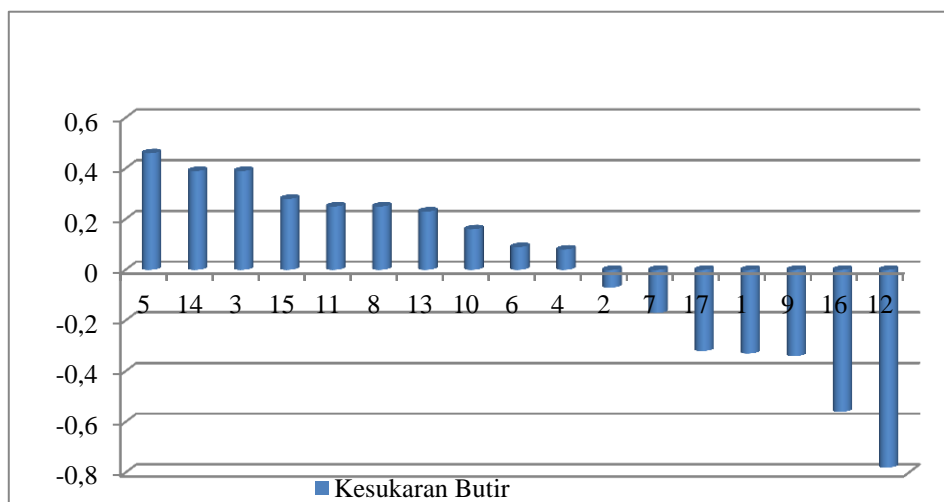
Butir soal nomor 13 dianggap paling mudah oleh peserta didik. Butir soal tersebut mengukur kompetensi tingkat tiga dalam literasi kimia, yaitu menginterpretasikan data untuk membedakan larutan elektrolit kuat, lemah, dan non-elektrolit. Dalam hal ini peserta didik dituntut lebih jeli. Meskipun demikian, peserta didik lebih tertolong dari pilihan senyawa yang menunjukkan larutan elektrolit kuat, lemah, atau non-elektrolit. Jika peserta didik sudah memahami konsep jenis ikatan senyawa, maka dengan mudah peserta didik menentukan

c. Uji Coba III (Implementasi Produk)

Analisis di setiap butir soal (*item*) dilakukan untuk melihat tingkat kesalahan atau standar eror dan tingkat kesulitan berdasarkan hasil empiris atau uji coba III. Hasil analisis standar eror dan tingkat kesulitan *item* disajikan pada Tabel 29.

Tabel 29. Tingkat Kesukaran dan Kesalahan Butir Soal Uji Coba III

Butir Soal (<i>item</i>)	Tingkat Kesukaran	Tingkat Kesalahan	Butir Soal (<i>item</i>)	Tingkat Kesukaran	Tingkat Kesalahan
S5	+0,46	0,05	S4	+0,08	0,05
S14	+0,39	0,06	S2	-0,07	0,05
S3	+0,39	0,05	S7	-0,17	0,06
S15	+0,28	0,05	S17	-0,32	0,07
S11	+0,25	0,05	S1	-0,33	0,06
S8	+0,25	0,05	S9	-0,34	0,06
S13	+0,23	0,05	S16	-0,56	0,07
S10	+0,16	0,05	S12	-0,78	0,08
S6	+0,09	0,05			



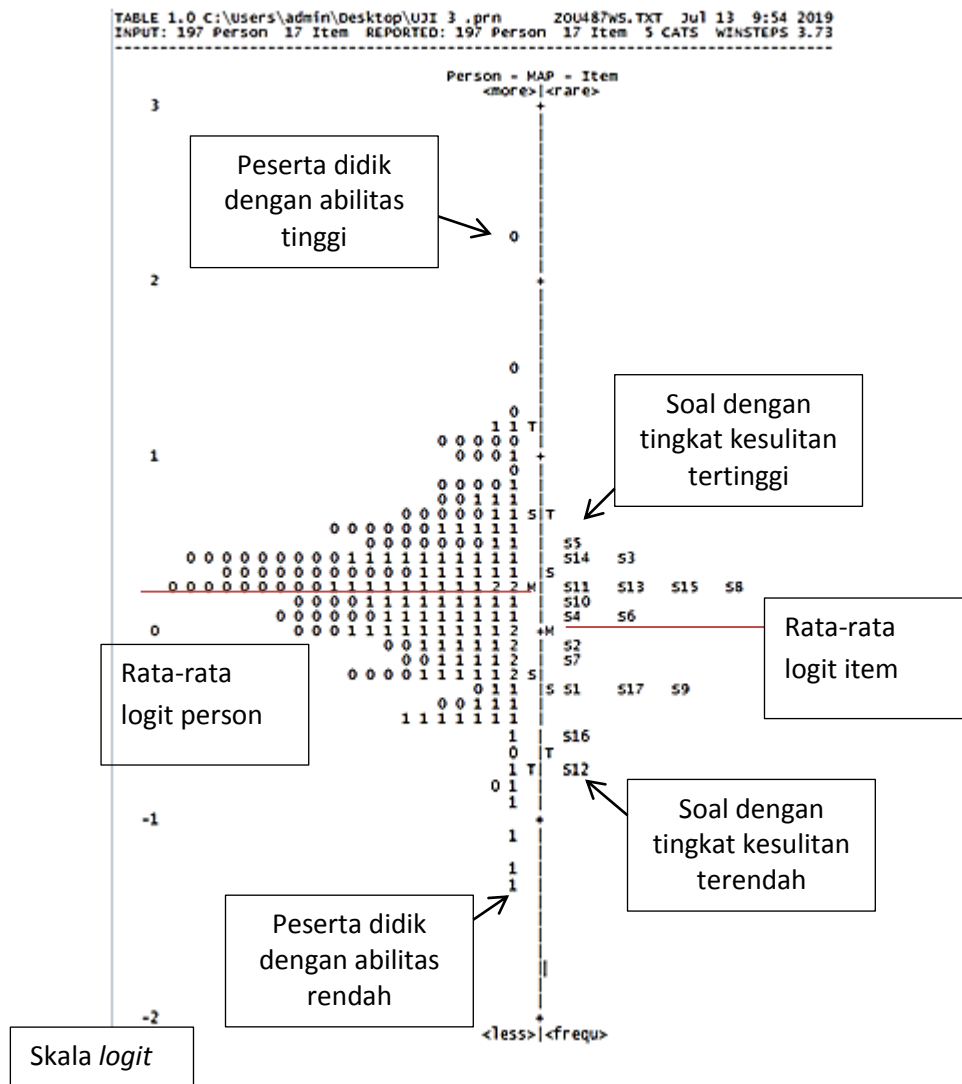
Gambar 16. Tingkat Kesukaran Butir Soal pada Uji Coba III

Nilai tingkat kesalahan setiap butir seperti yang tersaji pada Tabel 29 telah memenuhi kriteria nilai $SE < 0,5-1,00$ sehingga dapat disimpulkan bahwa semua butir soal dikatakan baik. Tingkat kesukaran butir soal dapat juga dilihat pada Gambar 16. *Person-item map* pada Gambar 17 menunjukkan bahwa butir soal yang dianggap paling sukar adalah butir soal nomor 5, yaitu simpulkanlah hasil percobaan pada $\text{Ca}(\text{ClO})_2(\text{aq})$ ketika diberikan arus listrik. Dalam bacaan tersedia ionisasi $\text{Ca}(\text{ClO})_2(\text{aq})$ secara mikroskopis yang dapat membantu peserta didik dalam menjawab. Namun, peserta didik masih merasa kesulitan karena ilustrasi mikroskopis jarang diberikan di kelas. Butir soal nomor 5 ini sukar karena soal tersebut mengukur kompetensi mengevaluasi penyelidikan ilmiah peserta didik.

Sedangkan butir soal yang dianggap paling mudah adalah butir soal nomor 12. Butir soal nomor 12 ini mengenai interpretasi data ilmiah berbentuk diagram persentase daya hantar listrik suatu larutan. Melalui diagram persentase daya hantar, dengan mudah peserta didik dapat menentukan jenis larutan. Secara umum berdasarkan hasil uji coba III, soal dikategorikan menjadi tiga yaitu sukar, sedang dan mudah. Secara lebih rinci, kategori butir soal disajikan pada Tabel 30.

Tabel 30. Kategori Butir Soal Uji Coba III

Nomor Butir Soal	Kategori Kesulitan Butir Soal
S5, S14, S3	Sukar
S11, S13, S15, S8, S10, S4, S6, S2, S7, S1, S17, S9	Sedang
S16, S12	Mudah



Gambar 17. *Person-Item Map* untuk Uji Coba III

C. Analisis Data Implementasi Produk

Pengukuran dilakukan pada peserta didik kelas X MIPA Tahun Ajaran 2018/2019 dengan materi larutan elektrolit dan non-elektrolit. Tujuannya adalah untuk mengukur kemampuan literasi kimia peserta didik. Tahap pengukuran melibatkan 197 peserta didik dari populasi SMA Negeri 1 Pekalongan dan SMA 3 Pekalongan. Tabel 31 menyajikan jumlah peserta didik secara rinci.

Tabel 31. Jumlah Peserta Didik sebagai Subjek Pengukuran

Nama Sekolah	Kelas	Jumlah Peserta
SMA Negeri 1 Pekalongan	X MIPA 1	36
	X MIPA 2	36
	X MIPA 3	35
	X MIPA 4	24
SMA Negeri 3 Pekalongan	X MIPA 1	32
	X MIPA 2	34
Jumlah		197

Hasil pengukuran kemampuan literasi kimia peserta didik terdiri dari tiga aspek, yaitu aspek konteks lokal Batik Pekalongan, pengetahuan kimia, dan kompetensi. Pengukuran dilakukan dengan menggunakan instrumen penilaian yang telah diujicobakan sebanyak dua kali sebelum akhirnya pada tahapan implementasi dinyatakan layak dan memiliki kulaitas yang baik untuk mengukur kemampuan literasi kimia peserta didik. Produk akhir instrumen penilaian literasi kimia terdiri dari 17 butir soal uraian yang didalamnya memuat aspek konteks lokal Batik Pekalongan, konten/pengetahuan kimia, dan kompetensi.

Penskoran pada respon peserta didik disesuaikan dengan rubrik atau pedoman penskoran yang telah dibuat. Skor kemudian dianalisis untuk mengetahui kemampuan literasi kimia peserta didik. Hasil analisis persentase penguasaan di setiap butir soal disajikan pada Tabel 32.

Tabel 32. Hasil Analisis Persentase Penguasaan di Setiap Butir Soal

No.	Rata-rata	Persentase	Keterangan
1.	3,954	79%	Baik
2.	2,459	49%	Kurang
3.	2,680	54%	Cukup
4.	3,016	60%	Cukup
5.	2,743	55%	Cukup
6.	3,258	65%	Cukup
7.	3,733	75%	Baik
8.	2,984	60%	Cukup
9.	3,995	80%	Baik
10.	3,138	63%	Cukup
11.	2,974	59%	Cukup
12.	4,497	90%	Sangat Baik
13.	3,035	61%	Cukup
14.	2,718	54%	Cukup
15.	2,976	60%	Cukup
16.	4,349	87%	Sangat Baik
17.	4,044	81%	Baik

Hasil pengukuran pada setiap aspeknya dijabarkan sebagai berikut:

1. Hasil Pengukuran Literasi Kimia pada Aspek Konteks

Pengukuran pada aspek konteks yang digunakan dalam penelitian ini adalah konteks Batik Pekalongan. Konteks Batik Pekalongan termasuk dalam ruang lingkup konteks lokal dalam literasi kimia. Hasil pengukurannya dilihat pada perhitungan semua butir soal karena mengandung aspek konteks. Hasil pengukuran aspek konteks lokal Batik Pekalongan disajikan pada Tabel 33.

Tabel 33. Kemampuan Literasi Kimia Peserta Didik pada Aspek Konteks

Aspek Konteks Kimia	Skor	Persentase Penguasaan	Kriteria
Lokal (Batik Pekalongan)	52,193	61%	Cukup

Pengukuran pada konteks lokal meliputi bidang lingkungan yaitu penggunaan bahan pewarna pada Batik Pekalongan dan pengolahan limbah pewarna Batik Pekalongan. Berdasarkan hasil pengukuran, aspek konteks yang dikuasai peserta didik terkategori cukup baik dengan skor 52,193. Hasil tersebut didasari dengan adanya muatan lokal batik pada ekstrakurikuler dan mata pelajaran prakarya serta pembudayaan literasi di sekolah SMA Negeri 1 Pekalongan dan SMA 3 Pekalongan.

Faktor lain yang mendukung penguasaan aspek konteks adalah proses pembelajaran yang kontekstual dan *problem solving*. Kemampuan literasi sains yang di dalamnya termasuk kimia terbukti sangat dipengaruhi oleh aktivitas dan proses pembelajaran peserta didik. Salah satunya adalah keterlibatan alam sekitar dan pembelajaran kontekstual mampu menstimulus pengetahuan peserta didik dan melek terhadap ilmu pengetahuan (Srikandi, Sujana, & Aeni, 2017; Maimoona, Qasim, Qurrat-ul-ain, & Samina, 2018).

2. Hasil Pengukuran Literasi Kimia pada Sub Aspek Pengetahuan

Sub aspek konten/pengetahuan yang diukur dalam penelitian ini adalah konten, pengetahuan prosedural, dan pengetahuan epistemik. Pengetahuan konten yang dimaksud adalah pengetahuan dasar tentang konsep larutan elektrolit dan non-elektrolit seperti sifat larutan elektrolit dan non-elektrolit, membedakan larutan lektrolit kuat dan lemah, dan sifat senyawanya. Batasan pengetahuan prosedural yang digunakan adalah peserta didik mampu menggunakan pengetahuannya dalam mengevaluasi penyelidikan ilmiah dan menginterpretasi-kan hasil data ilmiah. Sedangkan pengetahuan epistemik adalah peserta didik mampu

menggunakan pengetahuannya dalam memecahkan masalah melalui pendekatan ilmiah. Hasil pengukuran pada sub aspek konten/pengetahuan dalam literasi kimia disajikan pada Tabel 34 dan secara umum persentase penguasaan sub aspek pengetahuan oleh peserta didik dapat digambarkan pada Gambar 18.

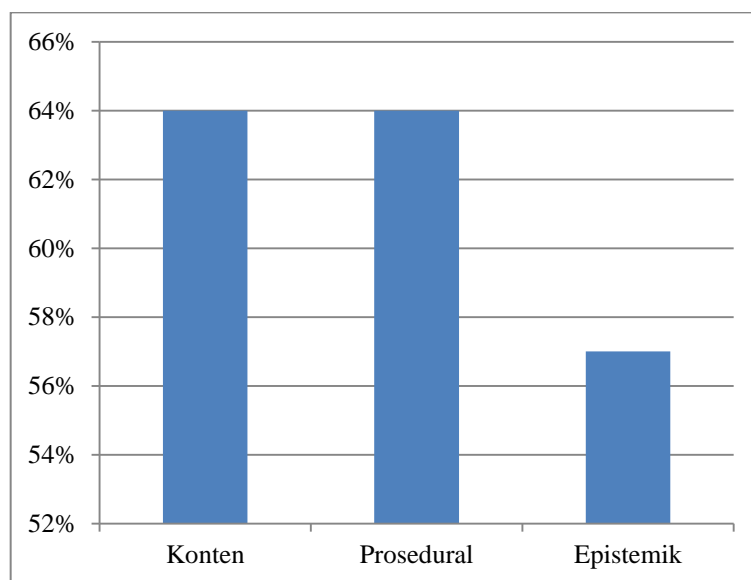
Tabel 34. Hasil Pengukuran Sub Aspek Pengetahuan

Sub Aspek Konten/Pengetahuan Kimia	Skor	Kriteria
Konten	12,84	Cukup
Pengetahuan Prosedural	22,24	Cukup
Pengetahuan Epistemik	17,11	Cukup

Berdasarkan Tabel 34 disimpulkan bahwa baik konten, prosedural, dan epistemik memiliki kriteria penguasaan kemampuan yang cukup. Hasil tersebut menunjukkan bahwa peserta didik cukup memahami pengetahuan kimia tentang konsep dasar larutan leketrolit dan non-elektrolit. Pengetahuan prosedural dan epistemik membutuhkan lebih dari pengetahuan dasar yang diketahui peserta didik. Pengetahuan prosedural membutuhkan pemahaman cara membangun dan tingkat kepercayaan terhadap ilmu pengetahuan (pengetahuan konten). Begitu pula dengan pengetahuan epistemik yang dibangun berdasarkan pemahaman tentang peran atau aplikasi khusus untuk membangun ilmu pengetahuan (Duschl, 2008).

Secara umum peserta didik telah memahami cukup baik pada tingkat pemahaman bagaimana ilmu pengetahuan didapatkan secara ilmiah dan aplikasi khususnya. Hal tersebut dibuktikan dari persentase respon peserta didik dalam menguasai sub aspek pengetahuan. Persentase penguasaan konten sebesar 64%, pengetahuan prosedural sebesar 64%, dan pengetahuan epistemic sebesar 57%.

Dengan begitu dapat disimpulkan bahwa penguasaan aspek pengetahuan yang dimiliki peserta didik cukup baik, meskipun masih tergolong relatif rendah karena kurang dari 70%. Hasil persentase juga tersaji pada Gambar 18.



Gambar 18. Persentase Penguasaan Sub Aspek Pengetahuan

3. Hasil Pengukuran Literasi Kimia pada Sub Aspek Kompetensi

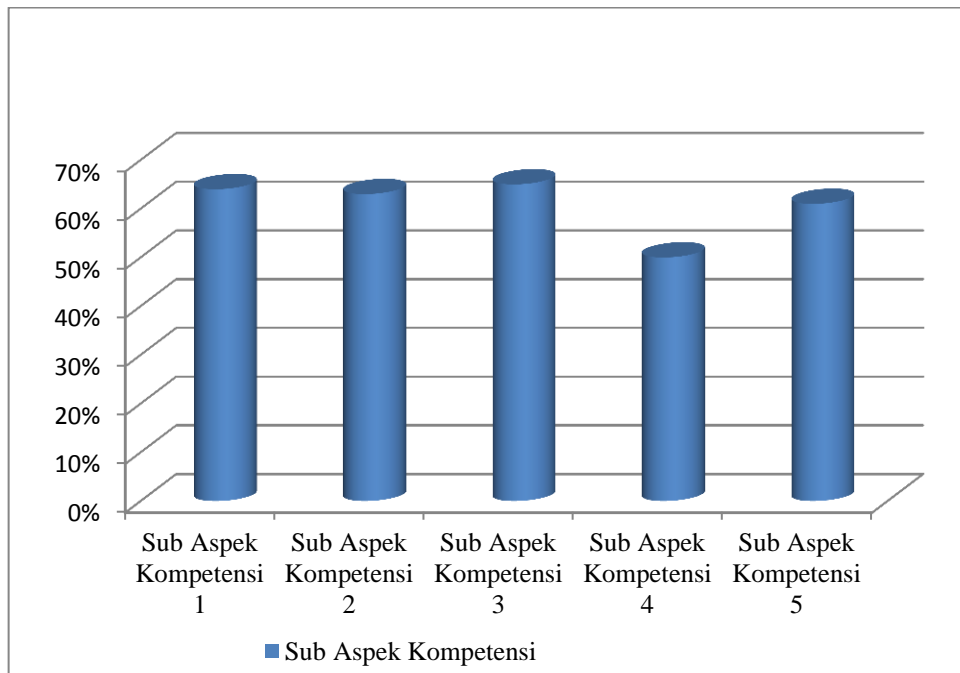
Aspek kompetensi terdiri dari lima sub aspek kontensi, yaitu 1) menjelaskan fenomena ilmiah, 2) mengevaluasi dalam penyelidikan ilmiah, 3) menginterpretasikan data ilmiah berbentuk grafik, 4) pemecahan masalah aritmatika, dan 5) pemecahan masalah dengan pendekatan ilmiah. Hasil pengukuran setiap sub aspek kompetensi disajikan pada Tabel 35.

Tabel 35. Hasil Pengukuran Sub Aspek Kompetensi

No.	Sub Aspek Kompetensi	Skor	Kriteria
1.	menjelaskan fenomena ilmiah	12,83	Cukup
2.	mengevaluasi dalam penyelidikan ilmiah	12,55	Cukup
3.	menginterpretasikasikan data ilmiah berbentuk grafik	9,69	Cukup
4.	pemecahan masalah aritmatika	4,96	Kurang
5.	pemecahan masalah dengan pendekatan ilmiah.	12,15	Cukup

Diskripsi hasil pengukuran sub aspek kompetensi berdasarkan Tabel 35 dalam kategori cukup baik dalam menjelaskan fenomena ilmiah pada materi larutan elektrolit dan non-elektrolit, mengevaluasi dalam penyelidikan ilmiah, menginterpretasikan data ilmiah berbentuk grafik, dan memecahkan masalah dengan pendekatan ilmiah.

Hasil tersebut juga mengindikasikan bahwa peserta didik cukup baik memahami ilmu sains terutama pada materi larutan elektrolit dan non-elektrolit. Persentase penguasaan terhadap kelima sub aspek kompetensi oleh peserta didik adalah sub aspek kompetensi 1 sebanyak 64%, sub aspek kompetensi 2 sebanyak 63%, sub aspek kompetensi 3 sebanyak 65%, sub aspek kompetensi 4 sebanyak 50%, dan sub aspek kopetensi 5 sebanyak 61%. Pesentase ini dapat juga dilihat pada Gambar 19.



Gambar 19. Persentase Penguasaan Sub Aspek Kompetensi

Sub aspek kompetensi pemecahan masalah aritmatika dalam kategori kurang dengan skor rata-rata 4,96. Kurangnya penguasaan pemecahan masalah aritmatika ini jika ditinjau dari respon peserta didik yang kurang dapat menyimpulkan nilai derajat ionisasi larutan elektrolit. Berdasarkan respon, sebagian besar peserta didik memahami konsep bahwa elektrolit kuat memiliki nilai derajat ionisasi sama dengan 1 dan elektrolit lemah memiliki nilai derajat ionisasi kurang dari 1. Hal tersebut dapat menjadi miskonsepsi. Kasus miskonsepsi merupakan kasus yang sering terjadi akibat dari pengamatan dan pemahaman yang kurang lengkap atau disebut sebagai *preconceive notions* (Gooding & Metz, 2011). Selain itu, butir soal yang dianggap paling sukar oleh peserta didik yaitu butir soal nomor 5 dengan kompetensi pemecahan masalah.