

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian yang digunakan pada penelitian ini adalah *Development and Validation* (D&V) yaitu metode yang dikembangkan oleh Adams dan Wieman. Metode ini merupakan salah satu metode penelitian yang tepat untuk mengembangkan instrumen atau alat penelitian seperti instrumen Asesmen Kompetensi Minimum (AKM) literasi membaca dan numerasi yang baik dan layak ditinjau berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda.

Dalam metode penelitian ini, terdapat beberapa tahapan yang dilaksanakan, yang terdiri dari: (1) penetapan tujuan tes dan ruang lingkup materi atau kompetensi yang hendak diukur, (2) pengembangan dan evaluasi spesifikasi tes, (3) pengembangan, uji coba lapangan, evaluasi, dan pemilihan item beserta pedoman dan prosedur penilaian, dan (4) perakitan serta evaluasi tes untuk penggunaan operasional (Adams & Wieman, 2011).

Secara garis besar penelitian pengembangan dan validasi ini dilakukan menggunakan empat tahap yang terdiri dari:

1. Tahap perencanaan
2. Tahap pengembangan
3. Tahap uji coba dan analisis data
4. Tahap evaluasi

3.2 Partisipan dan Tempat Penelitian

Partisipan yang terlibat dalam penelitian ini terdiri atas para ahli di bidang kimia, yaitu tiga orang dosen dari Program Studi Pendidikan Kimia dan dua orang guru kimia dari sekolah yang berbeda, salah satunya berasal dari SMA tempat penelitian dilakukan. Para ahli ini berperan sebagai validator instrumen AKM yang dikembangkan, sedangkan guru kimia dari sekolah mitra juga berperan sebagai wali peserta didik. Populasi dalam penelitian ini mencakup seluruh peserta didik kelas XI di salah satu SMA di Kabupaten Garut. Sementara itu, responden atau sampel

penelitian ditetapkan secara khusus pada siswa kelas XI-3 yang telah mempelajari materi Keseimbangan Kimia. Sebanyak 30 peserta didik ditugaskan untuk menjawab instrumen AKM yang dikembangkan. Jawaban yang diperoleh kemudian dimanfaatkan sebagai data dalam pengujian kelayakan instrumen, meliputi validitas empiris, reliabilitas, tingkat kesukaran, serta daya pembeda. Selain itu, data tersebut juga digunakan untuk mengidentifikasi kemampuan literasi membaca dan numerasi peserta didik pada materi Keseimbangan Kimia.

3.3 Instrumen Penelitian

Instrumen yang digunakan untuk mengumpulkan data dalam penelitian ini, yaitu meliputi:

3.3.1 Lembar Uji Validitas Isi

Lembar validasi ini merupakan instrumen penilaian yang digunakan untuk memeriksa kesesuaian teks dengan butir soal dan kesesuaian indikator soal dengan butir soal. Lembar validasi ini divalidasi oleh validator yang merupakan para ahli di bidang pendidikan kimia, yang terdiri dari tiga dosen prodi pendidikan kimia dan dua guru kimia dari sekolah yang berbeda. Validator melakukan validasi isi instrumen dengan cara menuliskan tanda centang (✓) pada kolom “Ya” atau “Tidak” untuk menilai kesesuaian dari kedua aspek tersebut. Selain dari itu, para validator juga dapat memberikan komentar atau saran perbaikan dalam kolom yang telah disediakan. Berikut merupakan format validasi isi AKM pada materi keseimbangan kimia yang disajikan dalam Tabel 3.1.

Tabel 3.1 Format Lembar Validasi Isi Soal Literasi Membaca dan Numerasi

Teks ...									
No.	Indikator Butir Soal (1)	Butir Soal (2)	Pedoman Penskoran		Kesesuaian Teks dengan Butir Soal		Kesesuaian Indikator dengan Butir Soal		Saran (9)
			Skor (3)	Ket. (4)	Ya (5)	Tidak (6)	Ya (7)	Tidak (8)	

3.3.2 Kisi-Kisi Instrumen AKM

Kisi-kisi soal AKM dirancang dengan memperhatikan aspek-aspek literasi membaca dan numerasi pada setiap butir soal yang dilengkapi dengan judul teks informasi, capaian pembelajaran (CP), submateri, aspek-aspek literasi membaca dan numerasi yang disesuaikan dengan *framework* AKM, indikator butir soal, bentuk soal, dan nomor soal. Berikut merupakan format tabel kisi-kisi soal AKM yang disajikan pada Tabel 3.2.

Tabel 3.2 Format Kisi-Kisi Instrumen AKM

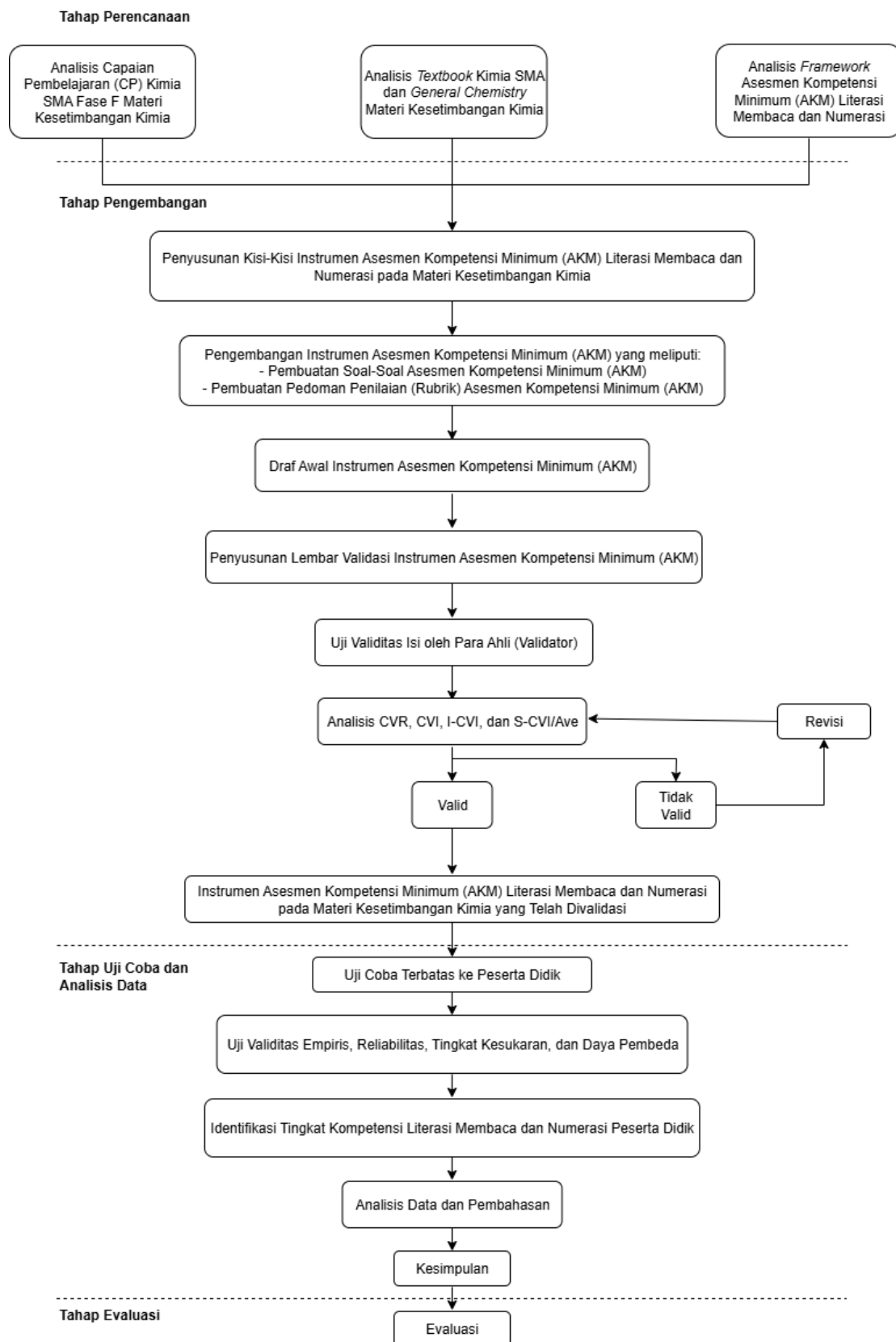
Teks ...					
Capaian Pembelajaran	Sub-Materi	Aspek–Aspek Literasi Membaca dan Numerasi (<i>framework</i> AKM)	Indikator Butir Soal	Bentuk Soal	No. Soal

3.3.3 Butir-Butir Soal AKM

Butir-butir soal AKM literasi membaca dan numerasi pada materi kesetimbangan kimia yang telah dinyatakan valid oleh validator, selanjutnya digunakan untuk menentukan nilai validitas empiris, reliabilitas, tingkat kesukaran, daya pembeda, dan tingkat kompetensi literasi membaca dan numerasi peserta didik.

3.4 Prosedur Penelitian

Prosedur penelitian merupakan langkah-langkah yang akan dilakukan selama penelitian yang dijadikan sebagai pedoman hingga diperoleh hasil yang diharapkan. Penelitian ini dilakukan melalui empat tahapan yaitu tahap perencanaan, tahap pengembangan, tahap uji coba dan analisis data, serta tahap evaluasi. Tahapan yang dilakukan dalam penelitian dapat dilihat pada alur penelitian yang disajikan pada Gambar 3.1 berikut.



Gambar 3.1 Prosedur penelitian

Prosedur penelitian tersebut diuraikan sebagai berikut:

Tahap I: Perencanaan

- a. Analisis Capaian Pembelajaran (CP) Kimia SMA pada fase F pada Kurikulum Merdeka dalam materi kesetimbangan kimia bertujuan untuk menentukan sejauh mana topik ini perlu dikaji, baik dalam hal keluasan maupun kedalaman materi. Hasil analisis ini akan digunakan sebagai dasar pengembangan instrumen penilaian, khususnya yang berkaitan dengan literasi membaca dan numerasi. Selain itu, analisis CP juga berperan dalam merumuskan indikator-indikator soal, sehingga butir soal yang dihasilkan dapat selaras dengan capaian pembelajaran yang ditetapkan.
- b. Analisis menyeluruh terhadap buku teks kimia serta sumber-sumber relevan lainnya yang membahas materi kesetimbangan kimia. Analisis ini mencakup baik buku yang digunakan oleh peserta didik maupun buku referensi bagi pendidik, seperti *General Chemistry* dan sumber lainnya dengan tetap memperhatikan batasan sesuai kurikulum yang berlaku agar materi yang dikaji tetap relevan dan sesuai dengan standar pembelajaran.
- c. Analisis *framework* AKM literasi membaca dan numerasi berdasarkan standar yang sudah dikembangkan oleh Pusat Asesmen dan Pembelajaran (Pusmenjar), Badan Litbang dan Perbukuan, Kemendikbud pada tahun 2021 dalam mengembangkan instrumen AKM untuk mengukur kemampuan literasi dan numerasi.

Tahap II: Pengembangan

- a. Penyusunan kisi-kisi instrumen AKM literasi membaca dan numerasi dilakukan dengan mencakup aspek-aspek penting. Di dalamnya terdapat judul teks, Capaian Kompetensi (CP) kimia yang diteliti, submateri kesetimbangan kimia, aspek-aspek literasi membaca dan numerasi yang sesuai dengan *framework* AKM, indikator butir soal, bentuk soal, dan nomor soal. Dengan adanya kisi-kisi ini, proses pengembangan instrumen AKM menjadi lebih mudah dan sistematis.
- b. Dalam tahap pengembangan, instrumen AKM literasi membaca dan numerasi pada materi kesetimbangan kimia yang mencakup teks informasi dan butir soal disusun berdasarkan kisi-kisi yang telah dibuat. Bentuk soal yang

dikembangkan mencakup pilihan ganda, pilihan ganda kompleks (benar-salah dan *multiple choice*), menjodohkan, isian singkat, serta esai atau uraian. Selain itu, disusun pula rubrik penilaian sebagai pedoman dalam menilai asesmen kompetensi minimum. Rubrik ini nantinya digunakan untuk memperoleh data yang diperlukan dalam menghitung validitas empiris, reliabilitas, tingkat kesulitan, daya pembeda, serta kompetensi literasi membaca dan numerasi peserta didik.

- c. Draf awal instrumen AKM literasi membaca dan numerasi pada materi Kesetimbangan Kimia disusun dalam bentuk tabel. Tabel ini mencakup teks informasi, indikator butir soal, butir soal, serta pedoman penilaian (rubrik) yang telah dikembangkan, sehingga lebih sistematis dan mudah digunakan.
- d. Pada tahap ini juga dilakukan proses validasi isi, yang diawali dengan penyusunan format lembar validasi instrumen AKM. Lembar tersebut memuat identitas instrumen secara umum, pengantar, tujuan validasi isi, penjelasan istilah, uraian materi yang digunakan beserta jumlah dan bentuk soal, serta petunjuk penilaian. Selain itu, lembar validasi juga berisi teks informasi, butir soal lengkap dengan kolom kesesuaian aspek soal, pedoman penskoran, serta kolom saran.
- e. Selanjutnya, dilakukan penentuan validator yang akan melakukan validasi isi dan rubrik dari instrumen AKM yang dikembangkan. Soal-soal AKM yang telah dibuat kemudian divalidasi oleh para ahli kimia, yaitu dosen dari program studi pendidikan kimia dan guru kimia SMA dari sekolah yang berbeda. Validasi ini bertujuan untuk mengetahui apakah instrumen AKM yang dikembangkan sudah valid dan dapat digunakan. Proses validasi dilakukan dengan mengecek kesesuaian antara teks bacaan dengan butir soal dan kesesuaian indikator butir soal dengan butir soal. Selain itu, validator juga berperan dalam memberikan saran dan masukan untuk menyempurnakan instrumen yang dikembangkan.
- f. Hasil validasi isi dianalisis dengan menghitung nilai CVR dan I-CVI untuk setiap butir soal AKM serta CVI dan S-CVI/Ave untuk keseluruhan instrumen. Jika validitas isi instrumen tersebut dinyatakan tidak valid, maka dilakukan perbaikan atau revisi berdasarkan saran yang diberikan oleh validator. Namun,

jika sudah valid, instrumen tersebut dapat langsung diujicobakan kepada peserta didik.

Tahap III: Uji Coba dan Analisis Data

Soal-soal AKM yang telah dinyatakan valid dan terverifikasi selanjutnya diuji coba secara terbatas pada peserta didik kelas XI SMA. Data yang diperoleh dari uji coba kemudian dianalisis dan dihitung, mulai dari nilai validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda, serta tingkat kompetensi literasi membaca dan numerasi peserta didik. Analisis ini bertujuan untuk menjawab rumusan masalah yang telah ditetapkan sebelumnya. Hasil analisis tersebut disampaikan secara jelas sesuai dengan hasil pengolahan data dan dibahas dalam pembahasan, kemudian ditafsirkan dalam bentuk simpulan terkait kualitas instrumen AKM yang dikembangkan.

Tahap IV: Evaluasi

Tahap evaluasi merupakan tahap akhir dalam prosedur penelitian yang bertujuan untuk memastikan ketercapaian tujuan penelitian serta menegaskan kualitas instrumen yang dikembangkan. Berdasarkan metode Adams & Wieman (2011), evaluasi idealnya mencakup perakitan serta pengujian instrumen secara operasional setelah melalui validasi isi, uji coba, dan analisis butir soal. Selain itu, tahap evaluasi berfungsi untuk memastikan bahwa instrumen tidak hanya memenuhi aspek teknis pengukuran, tetapi juga sesuai dengan tujuan awal pengembangan, yakni mengidentifikasi tingkat kompetensi literasi membaca dan numerasi peserta didik pada konteks kimia, khususnya materi kesetimbangan kimia. Tahap ini dilakukan untuk mengevaluasi soal AKM secara menyeluruh, sehingga dapat diketahui kualitas instrumen yang dikembangkan setelah melalui proses uji coba dan analisis data.

3.5 Teknik Analisis Data

Tabel 3.3 berikut menyajikan uraian mengenai teknik analisis data yang digunakan dalam penelitian ini.

Tabel 3.3 Teknis Analisis Data

No.	Pertanyaan Penelitian	Teknik Pengumpulan Data	Instrumen Penelitian	Jenis Data	Sumber Data	Analisis Data
1.	Bagaimana hasil validitas instrumen Asesmen Kompetensi Minimum (AKM) literasi membaca dan numerasi yang dikembangkan pada materi kesetimbangan kimia?	Uji validitas isi dan validitas empiris	Lembar validitas isi instrumen AKM	Validitas isi dan validitas empiris instrumen AKM	Validator dan peserta didik	Menghitung <i>Content Validity Ratio</i> (CVR), <i>Content Validity Index</i> (CVI), <i>Item-Content Validity Index</i> (I-CVI), dan <i>Scale-Content Validity Index Average</i> (S-CVI/Ave) berdasarkan Lawshe (1975) dan Polit <i>et al.</i> (2007) untuk validasi isi dan menghitung <i>Coefficient Corelation Persons's Product Moment</i> menurut Arifin (2017) untuk validasi empiris.
2.	Bagaimana reliabilitas instrumen Asesmen Kompetensi Minimum (AKM) literasi membaca dan numerasi yang dikembangkan pada materi kesetimbangan kimia?	Uji Reliabilitas	Instrumen tes AKM	Reliabilitas instrumen AKM	Peserta didik	Menghitung nilai <i>Kuder Richardson 20</i> (KR-20) menurut Frankel & Wallen (dalam Firman, 2013) untuk skor maksimal satu (dikotomis) dan koefisien <i>Cronbach's Alpha</i> menurut

No.	Pertanyaan Penelitian	Teknik Pengumpulan Data	Instrumen Penelitian	Jenis Data	Sumber Data	Analisis Data
						Gliem & Gliem (2003) untuk skor lebih dari satu (politomis).
3.	Bagaimana tingkat kesukaran dan daya beda instrumen Asesmen Kompetensi Minimum (AKM) literasi membaca dan numerasi yang dikembangkan pada materi kesetimbangan kimia?	Uji tingkat kesukaran dan daya beda	Instrumen tes AKM	Tingkat kesukaran dan daya beda butir soal AKM	Peserta didik	Menghitung indeks kesukaran (P) menurut Arifin (2017) dan menghitung indeks diskriminasi (D) atau daya pembeda menurut Sudijono (2008).
4.	Bagaimana tingkat kompetensi literasi membaca dan numerasi peserta didik berdasarkan hasil pengukuran menggunakan instrumen Asesmen Kompetensi Minimum (AKM) yang dikembangkan pada materi kesetimbangan kimia?	Uji tingkat kompetensi literasi membaca dan numerasi	Instrumen tes AKM	Tingkat kompetensi literasi membaca dan numerasi	Peserta didik	Perhitungan indeks tingkat kompetensi literasi membaca dan numerasi peserta didik menurut Pusmendik (2022).

3.6 Analisis Data

Teknik analisis agar mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dari data penelitian yang diperoleh, sebagai berikut:

3.6.1 Uji Validitas

Validitas berasal dari kata *validity*, yang berarti tingkat ketepatan dan kecermatan suatu instrumen pengukur (tes) dalam menjalankan fungsinya (Ramadhan *et al.*, 2024). Validitas suatu instrumen menentukan akurasi pengukuran terhadap variabel yang ingin diukur. Validitas instrumen menunjukkan sejauh mana alat ukur mampu mengungkapkan data secara tepat tanpa menyimpang dari kondisi sebenarnya (Yusup, 2018). Validitas juga dapat diartikan sebagai keabsahan yang mencerminkan ketepatan instrumen dalam melakukan pengukuran terhadap variabel yang memang seharusnya diukur (Azwar, 2012).

Selain itu, validitas juga merujuk pada evaluasi komprehensif mengenai sejauh mana data empiris di lapangan serta alasan teoritis mendukung kesimpulan dan tindakan yang diambil berdasarkan hasil tes atau metode penilaian lainnya (Reynold, 2010). Dengan demikian, pemahaman tentang validitas tidak hanya terbatas pada kesesuaian isi instrumen, tetapi juga mencakup dukungan bukti yang relevan. Oleh karena itu, alat ukur yang baik harus memiliki validitas tinggi untuk memastikan bahwa instrumen tersebut dapat mengukur aspek yang sesuai dengan tujuannya secara akurat (Firman, 2013; Syahlani & Setyorini, 2023).

1. Validitas Isi

Validitas isi merujuk pada kesesuaian rancangan instrumen dengan materi pelajaran yang dievaluasi. Suatu tes dikatakan memiliki validitas isi jika mampu mengukur tujuan tertentu yang sesuai dengan materi pembelajaran (Arikunto, 2016). Validitas isi suatu alat ukur dilihat dari sejauh mana “isi” (*content*) bahan pelajaran yang tercakup dalam instrumen tersebut mencerminkan keseluruhan materi yang diajarkan. Tes dengan validitas isi yang tinggi akan memastikan bahwa materi yang diujikan bersifat representatif, baik dari segi cakupan materi maupun proses penalaran yang diperlukan. Penilaian terhadap validitas isi biasanya dilakukan melalui *judgement* oleh kelompok ahli di bidang yang relevan (Firman, 2013). Dalam

penelitian ini, proses validasi isi dilakukan oleh lima orang ahli (validator) yang memiliki kompetensi pada bidangnya.

Validitas isi juga menentukan sejauh mana butir-butir soal mampu mewakili secara menyeluruh dan proporsional perilaku yang ingin diukur dari peserta didik. Tes dinyatakan valid apabila setiap butir soal mencerminkan konten atau materi pembelajaran secara adil dan proporsional (Ramadhan *et al.*, 2024). Menurut Lawshe (1975), terdapat tahapan yang digunakan untuk memperoleh data hasil validasi isi, yaitu sebagai berikut: Instrumen AKM dilakukan validasi oleh validator. Nilai validitas isi dihitung menggunakan CVR (*content validity ratio*) dengan persamaan Lawshe berikut:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

Keterangan:

n_e = jumlah validator yang menyatakan esensial

N = jumlah keseluruhan validator

Hasil CVR tersebut kemudian dibandingkan dengan nilai CVR minimum CVR untuk 5 orang validator. Instrumen dinyatakan valid jika hasil nilai CVR pada penelitian ini lebih besar dari 0,99 yang ditunjukkan dalam Tabel 3.4 berikut:

Tabel 3.4 Nilai Minimum CVR

Jumlah Validator	Nilai Minimum CVR
5	0,99
6	0,99
7	0,99
8	0,75
9	0,78
10	0,62
11	0,59
12	0,56
13	0,54
14	0,51
15	0,49
20	0,42
25	0,37

Jumlah Validator	Nilai Minimum CVR
30	0,33
35	0,31
40	0,29

(Lawshe, 1975)

CVR berfungsi untuk menentukan apakah suatu item soal dipertahankan atau ditolak. Setelah item soal diidentifikasi untuk dimasukkan ke dalam versi akhir, langkah berikutnya adalah menghitung indeks validitas konten (CVI) untuk menilai validitas keseluruhan instrumen. CVI diperoleh dari rata-rata nilai CVR dari butir-butir soal yang telah dipertahankan menurut Lawshe. Dengan demikian, CVI dapat dihitung menggunakan rumus berikut:

$$CVI = \frac{CVR_{total}}{jumlah\ item}$$

Lawshe (1975) tidak menetapkan ambang batas tertentu untuk CVI sebagai indikator validitas sempurna, dan tidak ada interpretasi nilai CVI yang baku menurutnya. CVI hanya memberikan gambaran sejauh mana kesepakatan antara para ahli (validator) mengenai validitas isi secara keseluruhan. Untuk memperoleh nilai CVI yang lebih mendalam, termasuk I-CVI (*Item-Content Validity Index*) dan S-CVI/Ave (*Scale-Content Validity Index Average*). Pada penelitian ini, tahapan yang digunakan adalah berdasarkan Polit *et al.* (2007), yaitu sebagai berikut:

- a. Hasil validasi dari para validator dikategorikan ke dalam skala 1-4, dengan rincian sebagai berikut: kriteria “valid” diberi skor 4, “valid dengan perbaikan” diberi skor 3, sedangkan “tidak valid” diberikan skor 1 atau 2. Untuk setiap item, nilai I-CVI diperoleh dengan menghitung jumlah validator yang memberikan skor 3 atau 4, kemudian membaginya dengan total jumlah validator. Rumus perhitungan I-CVI adalah sebagai berikut:

$$I - CVI = \frac{jumlah\ validator\ yang\ memberi\ skor\ 3\ dan\ 4}{jumlah\ total\ validator}$$

- b. Untuk memperoleh nilai S-CVI/Ave, total nilai I-CVI dari seluruh item dihitung, kemudian dibagi dengan jumlah keseluruhan item yang ada. Rumus perhitungan S-CVI/Ave adalah sebagai berikut:

$$S - CVI/Ave = \frac{\Sigma I - CVI}{jumlah\ item}$$

Nilai I-CVI yang mencapai $\geq 0,78$ dan nilai S-CVI/Ave yang bernilai minimal $\geq 0,90$ menunjukkan bahwa keseluruhan (skala) item dalam instrumen tersebut memiliki validitas isi yang sangat baik (valid) (Polit *et al.*, 2007).

2. Validitas Empiris

Validitas empiris, yang juga dikenal sebagai validitas kriteria (*criterion-related validity*), berkaitan dengan penggunaan alat ukur lain sebagai acuan dalam menilai tingkat validitas suatu instrumen. Validitas ini ditentukan berdasarkan korelasi antara hasil tes yang diuji dengan hasil dari tes lain yang dijadikan sebagai kriteria; korelasi yang tinggi menunjukkan bahwa instrumen memiliki validitas yang baik, sedangkan korelasi yang rendah menunjukkan validitas yang kurang memadai (Firman, 2013). Validitas kriteria penting untuk memastikan bahwa suatu instrumen benar-benar mampu mencerminkan atau memprediksi performa berdasarkan ukuran eksternal yang relevan.

Sejalan dengan hal tersebut, penerapan validitas empiris memerlukan analisis statistik kuantitatif guna menguji ketepatan hasil pengukuran yang dilakukan (Alwi, 2015). Istilah “empiris” mengacu pada pengalaman atau data hasil observasi, sehingga suatu instrumen dikatakan memiliki validitas empiris apabila telah melalui uji coba dan memberikan hasil yang sesuai dengan tujuan pengukuran (Riyani *et al.*, 2017). Dalam prosesnya, validitas empiris diperoleh melalui pelaksanaan uji coba soal, yang kemudian dianalisis untuk menilai sejauh mana data yang dihasilkan mencerminkan objek pengukuran secara akurat (Selaras *et al.*, 2019).

Pengelolaan nilai validitas empiris dilakukan dengan metode *Coefficient Corelation Persons's Product Moment* menggunakan perangkat *Microsoft Excel* dengan rumus:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{((\sum X^2) - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

Keterangan:

r_{xy} = koefisien korelasi antara variabel X dan variabel Y

n = jumlah sampel (banyaknya peserta tes)

Berikut adalah tabel interpretasi koefisien korelasi *Pearson's Product Moment* berdasarkan klasifikasi validitas, sebagaimana disajikan pada Tabel 3.5:

Tabel 3.5 Interpretasi Koefisien Korelasi *Pearson's Product Moment*

Koefisien Korelasi	Interpretasi
$0,80 < r_{xy} \leq 1,00$	Sangat tinggi
$0,60 < r_{xy} \leq 0,80$	Tinggi
$0,40 < r_{xy} \leq 0,60$	Cukup
$0,20 < r_{xy} \leq 0,40$	Rendah
$0,00 < r_{xy} \leq 0,20$	Sangat rendah

(Arifin, 2017)

3.6.2 Uji Reliabilitas

Reliabilitas diperoleh dari istilah *reliability* yang merujuk pada kestabilan, konsistensi, dan keterandalan suatu alat ukur dalam menghasilkan data yang akurat (Azwar, 2012). Konsep utama dari reliabilitas berkaitan dengan tingkat kepercayaan terhadap hasil pengukuran (Adriweri *et al.*, 2022). Suatu instrumen dinyatakan memiliki reliabilitas tinggi apabila pengukuran berulang terhadap subjek yang sama, dalam kondisi yang serupa, menghasilkan hasil yang konsisten atau mendekati sama. Oleh karena itu, reliabilitas sering kali dipahami sebagai derajat konsistensi atau keajegan hasil pengukuran (Firman, 2013).

Lebih lanjut, reliabilitas juga dapat dimaknai sebagai hubungan antara dua nilai yang diperoleh dari instrumen penelitian yang diberikan secara paralel kepada responden, dan dinilai berdasarkan koefisien korelasi (Haq, 2022). Instrumen yang memiliki reliabilitas yang baik akan tetap menghasilkan data yang konsisten meskipun digunakan dalam waktu dan situasi yang berbeda (Ndiung & Jediut, 2020). Dengan demikian, uji reliabilitas menjadi penting untuk memastikan bahwa data yang dihasilkan dapat dipercaya dalam menggambarkan objek yang diukur.

Data hasil uji reliabilitas ini diolah dengan menggunakan *Microsoft Excel*. Reliabilitas soal pilihan ganda dan isian singkat dengan skor maksimal satu (dikotomis) dalam penelitian ini ditentukan menggunakan teknik *Kuder-Richardson 20* (KR-20) yang dinyatakan dengan rumus sebagai berikut:

$$\rho_{KR20} = \frac{k}{k-1} \left(1 - \frac{\sum pq}{\sigma^2} \right)$$

Untuk menentukan varian skor tes (σ^2), dapat dihitung menggunakan rumus:

$$\sigma^2 = \frac{N\Sigma X^2 - (\Sigma X)^2}{N^2}$$

Keterangan:

ρ_{KR20} = koefisien reliabilitas

p = proporsi jawaban benar

k = jumlah butir soal

σ^2 = varian skor tes

q = proporsi jawaban benar

X = total skor setiap responden

N = jumlah responden

Suatu tes yang baku minimal mempunyai koefisien reliabilitas sebesar 0,70 agar tes ini memenuhi syarat reliabilitas (Frankel & Wallen dalam Firman, 2013). Berikut merupakan tabel interpretasi koefisien reliabilitas *Kuder-Richardson 20* (KR-20) yang diuraikan pada Tabel 3.6 sebagai berikut:

Tabel 3.6 Interpretasi Koefisien Reliabilitas *Kuder-Richardson 20* (KR-20)

Koefisien Korelasi	Interpretasi
0,96 – 0,99	Sangat Tinggi
0,91 – 0,95	Tinggi
0,81 – 0,90	Cukup
0,71 – 0,80	Sedang
<0,70	Rendah

(Djaali & Muljono, 2008)

Reliabilitas soal pilihan ganda kompleks dan urain dengan skor maksimal lebih dari satu (politomis) dalam penelitian ini ditentukan menggunakan koefisien *Cronbach's Alpha*, yang dinyatakan dengan rumus sebagai berikut:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

Keterangan:

α = reliabilitas butir soal

n = jumlah butir soal

σ_i^2 = varian skor setiap butir soal

σ_t^2 = varian skor soal

Adapun:

$$\sigma_i^2 = \frac{\sum(x_i^2) - \frac{(\sum x_i)^2}{N}}{N} \quad \text{dan} \quad \sigma_t^2 = \frac{\sum(X^2) - \frac{(\sum X)^2}{N}}{N}$$

Keterangan:

X_i = skor setiap soal

X = jumlah skor setiap responden

N = jumlah responden

Adapun tabel interpretasi koefisien *Cronbach's Alpha* menurut Gliem & Gliem (2003) untuk menentukan reliabilitas internal soal dengan maksimal lebih dari satu (politomis) diuraikan pada Tabel 3.7 sebagai berikut:

Tabel 3.7 Interpretasi Reliabilitas Internal (Koefisien *Cronbach's Alpha*)

Kriteria Reliabilitas Internal	Interpretasi
$\alpha \geq 0,9$	Sangat Tinggi
$0,8 \leq \alpha < 0,9$	Tinggi
$0,7 \leq \alpha < 0,8$	Dapat diterima
$0,6 \leq \alpha < 0,7$	Diragukan
$0,5 \leq \alpha < 0,6$	Jelek
$\alpha < 0,5$	Tidak dapat diterima

(Gliem & Gliem, 2003)

3.6.3 Tingkat Kesukaran

Tingkat kesukaran, atau indeks kesukaran, menunjukkan proporsi peserta didik yang menjawab suatu soal dengan benar dalam tes objektif (Sukardi, 2009). Soal yang baik tidak boleh terlalu mudah maupun terlalu sulit, karena soal yang terlalu mudah tidak mendorong peserta didik untuk meningkatkan usahanya, sedangkan soal yang terlalu sulit dapat menyebabkan keputusasaan dan penurunan motivasi belajar (Daryanto, 2012; Fatimah & Alfath, 2019; Saputri *et al.*, 2023). Oleh karena itu, analisis tingkat kesukaran penting dilakukan untuk mengelompokkan soal ke dalam kategori mudah, sedang, atau sulit. Proporsi penyusunan soal dapat mengikuti pola 3-4-3 (30% soal mudah, 40% soal sedang, dan 30% soal sulit) atau 3-5-2 (30% soal mudah, 50% soal sedang, dan 20% soal sulit) agar tercipta keseimbangan dalam tingkat kesukaran suatu tes (Magdalena *et al.*, 2021; Novalia & Syazali, 2014; Sudjana, 2009). Selain itu, Arifin (2017) juga

mengatur proporsi penyebaran tingkat kesukaran soal secara normal, yaitu soal sukar 25%, soal sedang 50%, soal mudah 25%, atau soal sukar 20%, soal sedang 60%, soal mudah 20%, atau soal sukar 15%, soal sedang 70%, soal mudah 15%.

Indeks kesukaran diperoleh berdasarkan data hasil pengerjaan soal oleh peserta didik, bukan dari analisis subjektif guru saat menyusun soal (Bagiyono, 2017; Sulistyorini, 2009). Semakin tinggi persentase peserta yang menjawab benar, maka soal tersebut dianggap semakin mudah; sebaliknya, semakin rendah persentasenya, maka soal tersebut dikategorikan semakin sulit (Setiyawan & Wijayanti, 2020). Indeks kesukaran dinyatakan dalam bentuk angka yang menunjukkan derajat kesulitan suatu butir soal (Lestari & Yudhanegara, 2015). Nilai indeks berkisar antara 0,00 hingga 1,00, di mana nilai 0,00 menunjukkan bahwa tidak ada peserta yang menjawab benar (soal sangat sulit), dan nilai 1,00 berarti semua peserta menjawab benar (soal sangat mudah). Indeks ini menjadi acuan dalam mengevaluasi sejauh mana suatu soal sesuai dengan tujuan pengukuran pembelajaran (Komarudin & Sarkadi, 2017).

Pengolahan data tingkat kesukaran ini dilakukan dengan program *Microsoft Excel* yang rumusnya disesuaikan dengan bentuk pertanyaan dan evaluasi (skor maksimal satu (dikotomis) dan skor maksimal lebih dari satu (politis). Rumus penentuan tingkat kesukaran soal pilihan ganda dan isian singkat dengan skor maksimal satu adalah sebagai berikut:

$$P = \frac{B}{J}$$

Keterangan:

P = indeks kesukaran

B = banyaknya peserta didik yang menjawab soal dengan benar

J = jumlah seluruh peserta didik peserta tes

Rumus yang digunakan untuk soal pilihan ganda kompleks dan uraian dengan skor maksimal lebih dari satu (politis) yaitu:

$$\text{mean} = \frac{\text{jumlah skor siswa pada soal tertentu}}{\text{jumlah siswa yang mengikuti tes}}$$

$$P = \frac{\text{mean}}{\text{skor maksimum suatu soal}}$$

Penafsiran tingkat kesukaran butir tes diuraikan pada Tabel 3.8 berikut:

Tabel 3.8 Interpretasi Tingkat Kesukaran Butir Tes

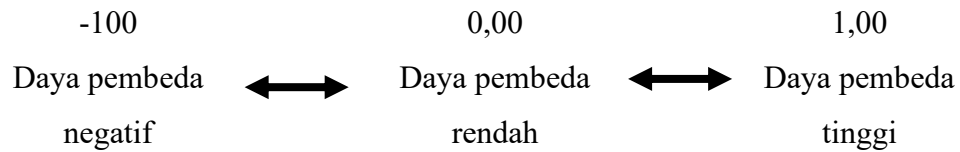
Indeks Kesukaran (P)	Interpretasi
$P > 0,30$	Soal sukar
$0,30 \leq P \leq 0,70$	Soal sedang
$0,70 < P$	Soal mudah

Arifin (2017)

3.6.4 Daya Pembeda

Daya pembeda soal adalah kemampuan suatu butir soal untuk membedakan antara peserta didik yang memiliki kemampuan tinggi dan peserta didik yang berkemampuan rendah (Daryanto, 2012). Selain itu, daya pembeda juga diartikan sebagai indikator sejauh mana suatu butir soal mampu mengukur kemampuan kognitif peserta didik secara akurat (Yani *et al.*, 2014). Fungsi utama dari daya pembeda adalah untuk menilai sejauh mana peserta didik yang telah menguasai materi dapat dibedakan dari peserta didik yang belum memahami materi tersebut (Ratnawulan & Rusdiana, 2015). Oleh karena itu, mengetahui nilai daya pembeda dari suatu soal menjadi hal yang penting, karena soal yang baik seharusnya mampu merepresentasikan variasi kemampuan peserta didik secara jelas (Sudijono, 2008).

Analisis daya pembeda dilakukan untuk menilai efektivitas soal dalam memilah peserta didik ke dalam kelompok dengan kemampuan tinggi dan rendah (Magdalena *et al.*, 2021; Setiyawan & Wijayanti, 2020). Tujuannya adalah untuk mengidentifikasi butir soal yang mampu membedakan secara konsisten antara peserta yang memahami materi dengan baik dan peserta yang masih kesulitan memahami materi (Bagiyono, 2017). Dalam praktiknya, perhitungan daya pembeda dilakukan dengan membagi peserta ke dalam dua kelompok, yaitu kelompok atas (berkemampuan tinggi) dan kelompok bawah (berkemampuan rendah). Nilai daya pembeda dinyatakan dalam indeks diskriminasi (D) dengan rentang antara 0,00 hingga 1,00. Jika indeks yang diperoleh bernilai negatif, maka hal ini menunjukkan bahwa butir soal tidak bekerja sebagaimana mestinya, karena peserta dengan kemampuan tinggi justru cenderung menjawab salah, dan peserta dengan kemampuan rendah menjawab benar. Dengan demikian terdapat 3 titik daya pembeda yaitu:



(Fatimah & Alfath, 2019)

Data hasil uji daya pembeda diolah dengan *Microsoft Excel* yang rumusnya disesuaikan berdasarkan skor maksimal butir soal. Rumus untuk menentukan indeks diskriminasi (D) menurut Sudijono (2008) adalah:

$$D = \frac{A_B}{A} - \frac{B_B}{B} = P_A - P_B$$

Keterangan:

D = indeks diskriminasi

A = jumlah peserta kelompok atas

A_B = peserta kelompok atas yang menjawab benar

B = jumlah peserta kelompok bawah

B_B = peserta kelompok bawah yang menjawab benar

P_A = proporsi peserta kelompok atas yang menjawab benar

P_B = proporsi peserta kelompok bawah yang menjawab benar

Klasifikasi daya pembeda ditunjukkan pada Tabel 3.9 berikut:

Tabel 3.9 Kriteria Daya Pembeda Butir Soal

Daya Pembeda (D)	Interpretasi
$0,70 < D \leq 1,00$	Baik sekali
$0,40 < D \leq 0,70$	Baik
$0,20 < D \leq 0,40$	Sedang
$0,00 < D \leq 0,20$	Jelek
$D \leq 0,00$	Jelek sekali

(Sudijono, 2008)

3.6.5 Tingkat Kompetensi Literasi Membaca dan Numerasi

Tingkat kompetensi peserta didik ditentukan dengan standar deviasi (SD) untuk menentukan batas dari tingkat kemampuan peserta didik setelah memperoleh nilai tes peserta didik. Tingkat kompetensi minimum dikategorikan berdasarkan nilai standar deviasi dan nilai rata-rata peserta didik sesuai Tabel 3.10.

Tabel 3.10 Interval Tingkat Kompetensi Literasi Membaca dan Numerasi

Interval	Kategori
$>M + SD$	Mahir

Interval	Kategori
$M \leq X \leq (M + SD)$	Cakap
$(M - SD) \leq (M + SD)$	Dasar
$< (M - SD)$	Perlu Intervensi Khusus

Keterangan: (Haryani *et al.*, 2023)

SD = Standar Deviasi

M = rata-rata jumlah skor yang diperoleh peserta didik (mean)

Tingkat kompetensi literasi membaca dan numerasi dapat dilihat dari nilai indeks yang ditentukan oleh persentase peserta didik di setiap tingkat kompetensi dan bobot. Berikut disajikan Tabel 3.11 untuk kriteria indeks literasi dan numerasi.

Tabel 3.11 Nilai Indeks Tingkat Kompetensi Peserta Didik Literasi Membaca dan Numerasi

Tingkat Kompetensi Peserta Didik	Jumlah	Persentase	Bobot	Indeks	Jumlah Indeks
(a)		(b)	(c)	(b*c)	Sum(b*c)
Mahir	...	%	3	...	...
Cakap	...	%	2	...	
Dasar	...	%	1,5	...	
Perlu Intervensi Khusus	...	%	1	...	

Tingkat kompetensi literasi membaca dan numerasi dikelompokkan menjadi empat yaitu mahir, cakap, dasar, dan perlu intervensi khusus. Berikut disajikan pada Tabel 3.12 mengenai tingkat kompetensi literasi membaca dan numerasi peserta didik.

Tabel 3.12 Tingkat Kompetensi Literasi Membaca dan Numerasi Berdasarkan Indeks

Tingkat Kompetensi	Rentang Nilai (Jumlah Indeks)
Mahir	2,10 s.d. 3,00
Cakap	1,80 s.d. 2,09
Dasar	1,40 s.d. 1,79
Perlu Intervensi Khusus	1,00 s.d. 1,39

(Pusmendik, 2022)