

BAB III METODE PENELITIAN

3.1 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah siswa kelas IX salah satu SMP Negeri di Kabupaten Indramayu dengan sampel dua ruang kelas, satu guru IPA dan satu operator sekolah. Teknik pengambilan sampel yang digunakan adalah *purposive sampling*. Menurut Sugiyono (2013), *purposive sampling* adalah teknik pengambilan sampel sumber data dengan pertimbangan tertentu agar nantinya data yang diperoleh dapat lebih representatif. Karena studi ini berfokus pada pengembangan dan penggunaan aplikasi SBT sebagai alat penilaian pembelajaran, pertimbangan yang digunakan dalam pemilihan sampel meliputi:

1. Peserta adalah siswa kelas 9 salah satu SMP Negeri di Kabupaten Indramayu yang menggunakan atau terbiasa menggunakan smartphone berbasis android.
2. Peserta dengan spesifikasi minimal smartphone:
 - a. Sistem Operasi: Jelly Bean
 - b. Ukuran layar ponsel: minimal 4,5 inci
 - c. RAM: 128MB

3.2 Metode dan Desain Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian dan pengembangan. metode penelitian dan pengembangan (*research and development/ R&D*) merupakan metode penelitian yang digunakan untuk meneliti sehingga menghasilkan produk baru dan selanjutnya menguji keefektifan produk tersebut.

Adapun desain penelitian yang digunakan dalam penelitian ini adalah ADDIE yaitu: *Analysis* (Analisis), *Design* (Desain), *Development & Implementation* (Pengembangan dan Penerapan), *Evaluation* (Evaluasi). Adapun diagram alir yang akan digunakan dalam penelitian ini bisa dilihat di gambar berikut ini:

Gambar 3.1 Diagram alir penelitian



3.3 Prosedur Penelitian

Berikut penjelasan dari tahap pengembangan ADDIE yang dilakukan oleh peneliti. Adapun Langkah-langkah penelitian bisa dilihat pada Gambar 3.1 dengan penjelasan sebagai berikut:

1. Analisis

- ✚ Analisis kebutuhan harus dilakukan pada awal upaya pengembangan untuk menentukan apakah aplikasi SBT berbasis Android diperlukan untuk mengimbangi kekurangan *Paper-based Test* (PBT). Pada tahap ini dilakukan identifikasi masalah terlebih dahulu terhadap jenis penilaian yang sering digunakan dan SBT berbasis android yang sudah pernah dikembangkan oleh peneliti lain. Data didapat melalui studi literatur.
- ✚ Analisis target audiens adalah langkah penting lainnya. Desain dan penyampaian aplikasi SBT berbasis android akan dipengaruhi oleh karakteristik utama peserta didik. Misalnya pengetahuan dan keterampilan mereka sebelumnya mengenai penggunaan android; konteks pembelajaran; fasilitas dan akses ke teknologi.
- ✚ Analisis tugas merupakan langkah mengidentifikasi tugas-tugas pada soal yang akan dikerjakan, apakah diperlukan penyertaan gambar, video, grafik, hingga laboratorium virtual sehingga melalui tugas-tugas tersebut diharapkan soal dapat mengungkap kemampuan siswa untuk dapat mencapai indikator soal yang telah ditentukan.

- ✚ Analisis topik dilakukan untuk mengidentifikasi dan mengklasifikasikan isi materi pada bab mata pelajaran IPA yang akan diujikan.

2. Desain

Tahap desain meliputi kegiatan berikut:

- a. Membuat *use case scenario*. Pada tahap ini, peneliti membuat gambaran interaksi antar pengguna aplikasi SBT berbasis android yang akan dikembangkan.
- b. Membuat *mockup* aplikasi. Tahap selanjutnya yaitu membuat gambaran antar muka yang akan muncul pada aplikasi SBT berbasis android yang akan dikembangkan.
- c. Membuat kisi-kisi pertanyaan untuk dimasukkan ke dalam aplikasi SBT yang akan dikembangkan

3. Pengembangan

Pada tahap ini, aplikasi SBT berbasis android benar-benar diproduksi sesuai dengan *mockup* yang sudah dibuat. Pada tahap pengembangan ini dilakukan pula uji kualitas dengan memvalidasi setiap fungsi pada aplikasi SBT yang dikembangkan yang kemudian hasilnya digunakan untuk memperbaiki aplikasi SBT agar dapat berjalan sesuai dengan yang diinginkan. Uji coba pertama dibutuhkan dua orang siswa, yang hasilnya masih perlu beberapa perbaikan seperti *timer* ujian yang kurang tepat, terdapat soal yang dua kali muncul, hitungan waktu pengerjaan yang belum tepat, dan beberapa siswa masih terdapat kendala dalam pemutaran video di dalam soal. Uji coba kedua dibutuhkan lima orang siswa, dan semua kendala pada uji coba pertama sudah teratasi, aplikasi SBT sudah berjalan sesuai desain yang sudah dibuat.

4. Implementasi

Tahap implementasi dilaksanakan pada tanggal 8 Oktober 2021. Pada tahap ini aplikasi SBT direalisasikan kepada 2 orang guru yaitu guru mata pelajaran IPA dan operator sekolah, selain itu direalisasikan juga kepada 62 siswa kelas IX. Pada tahap ini juga termasuk mengelola dan

memfasilitasi kegiatan guru dan siswa. Adapun kegiatannya yaitu: meliputi peng-*install*-an aplikasi oleh guru dan siswa; Membuat pengaturan ujian dan meng-*input* soal ujian oleh guru; Pelaksanaan ujian dengan menggunakan aplikasi SBT berbasis android yang akan dikembangkan.

5. Evaluasi

Setelah tahap implementasi, langkah selanjutnya adalah mengevaluasi penggunaan aplikasi SBT yang dikembangkan. Termasuk didalamnya evaluasi mengenai capaian nilai siswa yang menggunakan SBT dan PBT, evaluasi mengenai kualitas aplikasi SBT berbasis android yang telah dikembangkan, serta evaluasi mengenai tanggapan siswa dan guru setelah menggunakan aplikasi SBT berbasis android yang telah dikembangkan. Hasil evaluasi ini dapat digunakan untuk perbaikan pengembangan aplikasi SBT selanjutnya.

3.4 Instrumen Penelitian

Penelitian membutuhkan alat yang digunakan untuk mengumpulkan, menyelidiki, mengolah, menganalisis, dan menyajikan data secara sistematis dan obyektif, dan alat tersebut disebut sebagai alat penelitian (Creswell, 2017. P. 253). Alat penelitian yang digunakan dalam penelitian ini adalah:

a. Instrumen aspek *Functionality*

Instrumen penelitian yang digunakan untuk menguji aspek *functionality* berupa *checklist* daftar fungsi aplikasi SBT yang telah dikembangkan. Pengujian dilakukan dengan mengisi daftar *check* yang sesuai dengan fungsi dari aplikasi tersebut, sehingga dapat diketahui apakah fungsi-fungsi pada aplikasi berjalan sesuai fungsinya atau tidak. Adapun *checklist* fungsi-fungsi yang digunakan untuk pengujian aspek *functionality* ditunjukkan oleh tabel berikut ini:

Tabel 3.1 Instrumen pengujian aspek *Functionality*

No.	Fungsi	Hasil	
		Sukses	Gagal
A. General			
1.	Penginstalan aplikasi server SBT basis android berjalan dengan baik		
2.	Penginstalan aplikasi client SBT basis android berjalan dengan baik		
3.	Aplikasi server SBT basis android bisa dijalankan dengan baik		
4.	Aplikasi client SBT basis android bisa dijalankan dengan baik		
5.	Semua navigasi berjalan dengan baik tanpa ada broken link		
B. Server Side			
6.	Melakukan Login sebagai User guru		
7.	Melakukan import data mata pelajaran		
8.	Melakukan import username siswa		
9.	Melakukan import password siswa		
10.	Melakukan import informasi ujian		
11.	Melakukan setting data bank soal		
12.	Melakukan import gambar pendukung soal		
13.	Melakukan edit informasi ujian		
14.	Melihat status peserta ujian		
15.	Melakukan reset Login peserta ujian		
16.	Melakukan analisis hasil ujian		
C. Client Side			
17.	Melakukan Login sebagai peserta ujian		
18.	Melihat konfirmasi data peserta		
19.	Melihat daftar ujian		
20.	Melihat informasi ujian yang akan berlangsung		
21.	Melihat daftar ujian yang telah berlangsung		
22.	Memulai pelaksanaan ujian		
23.	Melakukan pengecekan tombol navigasi soal		
24.	Melakukan menjawab pilihan soal		
25.	Mengakhiri sesi ujian		
26.	Melihat hasil ujian		
27.	Melihat kembali hasil ujian sebelumnya		

b. Instrumen aspek *Usability*

Kuesioner yang dibuat oleh Lund (2001), yaitu USE (*usability, satisfaction, ease of use*), digunakan untuk menguji aspek *usability*. Kuesioner USE menggunakan skala Likert berupa *checklist*. Kuesioner USE dibagi menjadi lima pilihan jawaban: setuju (SS), setuju (S), ragu-ragu (RG), tidak setuju (TS), dan sangat tidak setuju (STS). Format kuisisioner adalah sebagai berikut:

Tabel 3.2 Instrumen pengujian aspek *Usability*

Indikator	Pertanyaan	
Usefulness	1.	Aplikasi SBT ini membantu saya menjadi lebih efektif dalam mengakses ulangan harian
	2.	Aplikasi SBT ini bermanfaat untuk pelaksanaan ulangan harian
	3.	Aplikasi SBT ini membantu saya untuk lebih mengontrol aktivitas-aktivitas yang berkaitan dengan ulangan harian
	4.	Aplikasi SBT ini mempermudah saya mencapai hal-hal yang berkaitan dengan Ujian/Tes yang ada di sekolah
	5.	Aplikasi SBT ini menghemat waktu saya dalam Pelaksanaan ulangan harian
	6.	Aplikasi SBT ini memenuhi kebutuhan Evaluasi Pembelajaran saya.
	7.	Aplikasi SBT ini bekerja sesuai dengan apa yang saya harapkan.
Ease of Learning	8.	Aplikasi SBT ini mudah untuk digunakan
	9.	Aplikasi SBT ini praktis untuk digunakan
	10.	Aplikasi SBT ini user friendly / mudah dioperasikan
	11.	Langkah-langkah penggunaan Aplikasi SBT ini tidak sulit
	12.	Aplikasi SBT ini fleksibel / dapat digunakan dengan kebutuhan
	13.	Penggunaan Aplikasi SBT ini mudah/ tidak perlu bersusah payah.
	14.	Saya dapat menggunakan Aplikasi SBT ini tanpa harus membaca panduan tertulis.
	15.	Saya tidak menemukan ketidakkonsistenan dalam Aplikasi SBT ini.
	16.	Saya dapat mengatasi kendala dengan mudah dan cepat.
	17.	Saya dapat menggunakan Aplikasi SBT ini

Indikator	Pertanyaan	
		dengan lancar setiap saat.
Satisfaction	18.	Saya belajar menggunakan Aplikasi SBT ini dengan cepat.
	19.	Saya bisa mengingat bagaimana cara menggunakan Aplikasi SBT ini dengan mudah.
	20.	Aplikasi SBT ini mudah dipelajari cara penggunaannya.
	21.	Saya dapat menguasai cara menggunakan Aplikasi SBT ini dengan cepat.
	22.	Saya merasa puas dengan Aplikasi SBT ini.
	23.	Saya akan merekomendasikan Aplikasi SBT ini kepada teman-teman
	24.	Aplikasi SBT ini menyenangkan untuk digunakan.
	25.	Aplikasi SBT ini bekerja sesuai dengan keinginan saya
	26.	Aplikasi SBT ini terlihat bagus
	27.	Saya merasa memerlukan Aplikasi SBT ini
28.	Aplikasi SBT ini nyaman untuk digunakan.	

c. Instrumen skala sikap

Persepsi siswa dan guru terhadap penggunaan aplikasi pembelajaran berbasis komputer (SBT) berbasis Android yang dikembangkan diperoleh dengan memberikan lembar skala sikap yang berisi pernyataan tentang penggunaan aplikasi SBT berbasis Android beserta skala Likert. Skala Likert yang digunakan adalah skala empat langkah: sangat setuju, setuju, tidak setuju, dan sangat tidak setuju (Sugiyono, 2013).

Tabel 3.3 Instrumen skala sikap untuk siswa

No.	Pernyataan	Respon			
		SS	S	TS	STS
1.	Saya dapat mengakses aplikasi SBT di android tanpa kendala				
2.	Saya dapat menjalankan aplikasi SBT dengan mudah				
3.	Saya dapat memahami soal-soal yang ditampilkan dalam aplikasi SBT				
4.	Saya dapat melihat gambar di dalam soal dengan jelas				
5.	Saya dapat mengerjakan soal dengan				

	cepat				
6.	Saya lebih termotivasi mengerjakan soal dengan aplikasi SBT				
7.	Saya lebih mandiri mengerjakan soal dengan menggunakan aplikasi SBT				
8.	Tampilan soal di dalam aplikasi SBT lebih menarik				
9.	Saya lebih cepat mengetahui hasil ulangan harian dengan menggunakan aplikasi SBT				
10.	Saya senang menggunakan aplikasi SBT untuk ulangan harian				

Tabel 3.4 Instrumen skala sikap untuk guru

No.	Pernyataan	Respon			
		SS	S	TS	STS
1.	Saya dapat mengakses aplikasi SBT di android tanpa kendala				
2.	Saya dapat menjalankan aplikasi SBT dengan mudah				
3.	Saya dapat dengan mudah mengatur jadwal ulangan harian dengan menggunakan aplikasi SBT				
4.	Saya dapat menginput dan mengedit soal dengan mudah menggunakan aplikasi SBT				
5.	Saya dapat mengatur durasi waktu pengerjaan soal tes dengan mudah menggunakan aplikasi SBT				
6.	Saya dapat mengurangi kemungkinan kecurangan siswa dalam mengerjakan tes dengan menggunakan aplikasi SBT				
7.	Saya dapat lebih menghemat anggaran produksi soal ulangan harian dengan menggunakan SBT				
8.	Saya dapat lebih menghemat waktu mempersiapkan pelaksanaan ulangan harian dengan menggunakan SBT				
9.	Saya lebih cepat mendapatkan hasil ulangan harian siswa menggunakan aplikasi SBT				
10.	Saya dapat menggunakan hasil analisis ulangan harian siswa untuk melakukan remedial				
11.	Saya dapat menggunakan hasil				

	analisis ulangan harian siswa untuk melakukan pengayaan				
12.	Saya dapat menggunakan hasil analisis ulangan harian siswa untuk proses bimbingan dengan orang tua siswa				

d. Lembar Penilaian Soal

Lembar penilaian soal ini digunakan untuk mengetahui kelayakan setiap butir soal dan saran perbaikan jika diperlukan yang dikeluarkan oleh ahli.

3.5 Tehnik Analisis data

- Analisis data aspek *Functionality*

Analisis uji fungsional dilakukan dengan teknik analisis deskriptif, yaitu dengan menganalisis hasil pengujian untuk masing-masing fungsi perangkat lunak yang dikembangkan. Skala yang digunakan dalam tes adalah skala Guttman dua poin: "ya" atau "tidak." Rumus analisis data yang digunakan adalah sebagai berikut.:

$$X = 1 - \frac{A}{B} \dots\dots\dots 3.1$$

Keterangan :

X = *functionality*

A = Jumlah total fungsi yang tidak valid

B = Jumlah seluruh fungsi

Rumus pengujian fungsional menyatakan bahwa sistem yang dikembangkan memenuhi standar atau nilai x adalah mendekati 1 ($0 \leq x \leq 1$).

- Analisis Data *Usability*

Pengujian usability menggunakan kuesioner USE dengan skala Likert sebagai ukuran pengujiannya. Tanggapan untuk setiap item dalam alat yang menggunakan skala Likert memiliki rentang dari sangat setuju hingga sangat tidak setuju. Jawaban-jawaban tersebut kemudian dinilai untuk dianalisis oleh peneliti. Berikut ini

adalah contoh skala Likert yang dikemukakan oleh Sugiyono (2009):

SS = Sangat Setuju (skor 5)

S = Setuju (skor 4)

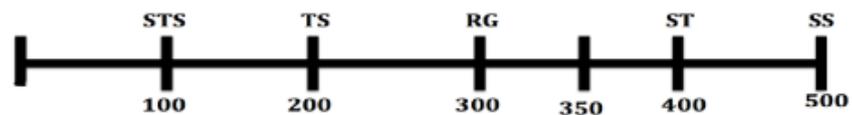
RR = Ragu-ragu (skor 3)

TS = Tidak Setuju (skor 2)

STS = Sangat Tidak Setuju (skor 1)

Analisis data hasil pengujian usability dilakukan dengan menghitung jumlah rata-rata jawaban berdasarkan skor. Jumlah skor ideal untuk seluruh item = $5 \times 100 = 500$. Jadi tingkat persetujuannya berdasarkan data tersebut = $(350 : 500) \times 100\% = 70\%$ dari yang diharapkan. Berikut ini adalah gambar skala kontinum dapat dilihat pada Gambar 3.2 berikut ini:

Gambar 3.2 Contoh Tingkat Persetujuan Respon pada Software



(Nurhairi, 2019)

Berdasarkan data yang diperoleh dari 100 responden maka skor 350 terletak pada daerah setuju.

- Ukuran dampak (*effect size*) aplikasi SBT

Penentuan efektifitas aplikasi *Smartphone Based Test* (SBT) berbasis Android dilakukan dengan mengukur *effect size*. Ukuran efek adalah ukuran hubungan yang tidak bergantung pada besar kecilnya pengaruh suatu variabel terhadap variabel lain, besarnya perbedaan, dan pengaruh ukuran sampel (Olejnik dan Algina, 2003). Variabel terkait biasanya berupa variabel respon, disebut juga variabel bebas dan variabel hasil, atau disebut juga variabel terikat. Ukuran ini diperlukan karena signifikansi statistik tidak memberikan informasi yang berarti tentang besarnya perbedaan. Signifikansi statistik hanya memberi tahu kita bahwa peningkatan

rata-rata pada kelompok eksperimen dan kontrol berbeda, bukan seberapa kuat perbedaan peningkatan tersebut (Olejnik dan Algina, 2000). Dalam hal ini, perhitungan *effect size* bertujuan untuk mengetahui besarnya pengaruh aplikasi SBT berbasis Android yang dikembangkan terhadap prestasi belajar siswa..

Dalam penelitian ini, ukuran dampak (*effect size*) diperoleh dengan menghitung ukuran dampak (D) dari buku teks. Dalam kasus omnidirectionality, metode paling sederhana dan paling langsung untuk menghitung ukuran dampak (D) adalah sebagai berikut (Cohen, 1998):

$$D = \frac{|M_E - M_K|}{SD_{pooled}}; \dots\dots\dots 3.2$$

$$SD_{pooled} = \sqrt{\frac{(n_E - 1)S_E^2 + (n_K - 1)S_K^2}{N_E + N_K - 2}} \dots\dots\dots 3.3$$

Keterangan:

- D : Ukuran dampak
- M_E : Rata-rata kelas eksperimen
- M_K : Rata-rata kelas kontrol
- SD_{pooled} : Standar deviasi sampel-sampel yang digabungkan
- N_E : Jumlah siswa kelas eksperimen
- N_K : Jumlah siswa kelas kontrol
- S_E² : Varians kelas eksperimen
- S_K² : Varians kelas kontrol

Hasil perhitungan tersebut sesuai dengan kriteria Cohen (1998) mengenai ukuran efek (*effect size*) suatu variabel relatif terhadap variabel lainnya, yang kemudian dikembangkan oleh Sawilowsky (2009):

Tabel 3.5 Kriteria ukuran efek (*Effect Size*)

Mean yang distandardisasi	Kriteria	Referensi
$0,01 \leq D < 0,2$	Efek sangat kecil (<i>very small effect</i>)	Sawilowsky (2009)
$0,2 \leq D < 0,5$	Efek kecil (<i>small effect</i>)	Cohen (1998)
$0,5 \leq D < 0,8$	Efek sedang (<i>medium effect</i>)	Cohen (1998)
$0,8 \leq D < 1,2$	Efek besar (<i>large effect</i>)	Cohen (1998)
$1,2 \leq D < 2$	Efek sangat besar (<i>very large effect</i>)	Sawilowsky (2009)
$D \geq 2$	Efek besar sekali (<i>huge effect</i>)	Sawilowsky (2009)

- Menganalisis persepsi siswa dan guru tentang penggunaan aplikasi SBT Pada bagian ini, Skala Sikap digunakan untuk mengetahui persepsi siswa dan guru terhadap penggunaan aplikasi SBT yang dikembangkan. Data yang diperoleh melalui skala sikap diubah dari skala kualitatif menjadi skala kuantitatif. Langkah-langkah untuk menganalisis skala ini adalah:

a. Memberikan skor jawaban dengan kriteria:

SS = Sangat setuju dengan bobot 4

S = Setuju dengan bobot 3

TS = Tidak setuju dengan bobot 2

STS = Sangat tidak setuju dengan bobot 1

b. tentukan skor tertinggi

c. Tentukan skor total untuk setiap komponen, kemudian jumlahkan skor total untuk semua komponen.

d. Tingkat persetujuan yang dirasakan untuk aplikasi SBT dapat dihitung menggunakan persamaan berikut::

$$\% \text{ persetujuan} = \frac{\text{jumlah skor yang diperoleh tiap item}}{\text{jumlah skor ideal untuk seluruh item}} \times 100\%$$

e. Skor yang diperoleh dinyatakan berdasarkan indikator pernyataan seperti yang diinterpretasikan berikut ini. Tabel 3.6 berikut ini:

Tabel 3.6 Kriteria Tanggapan Siswa dan Guru
(Sugiyono, 2013)

Interval Presentase Tanggapan Responden (%)	Kriteria
$80 \leq K \leq 100$	Sangat baik
$60 \leq K < 80$	Baik
$40 \leq K < 60$	Ragu-ragu
$20 \leq K < 40$	Kurang baik
$0 \leq K < 20$	Sangat tidak baik

- Validitas dan Reliabilitas Soal

✚ Validitas

Scarvia B. Anderson (1975) (Arikunto, 2007, hlm. 65) menyatakan bahwa suatu tes dikatakan valid jika mengukur apa yang hendak diukur. Sedangkan menurut American Educational Research Association, American Psychological Association, dan National Council on Measurement in Education (dalam Heri Retnawati, 2016, hlm. 16), validitas adalah skor tes Tingkat fakta dan teori yang mendukung interpretasi pengujian, dan merupakan pertimbangan terpenting dalam pengembangan pengujian. Validitas dapat diklasifikasikan menjadi tiga jenis: validitas kriteria (criteria-related), validitas isi, dan validitas konstruk (Heri Retnawati, 2016, p.16). Validitas yang digunakan dalam penelitian ini adalah validitas isi.

Menurut Fernandes (1984) (dalam Herawati, 2000, hlm.16), validitas isi adalah sejauh mana suatu item dalam suatu alat mewakili suatu komponen dalam domain isi keseluruhan dari objek yang diukur, dan item tersebut mencerminkan karakteristik perilaku itu mencoba untuk mengukur. Sedangkan menurut Lawrence (1994) (dalam Heri Retnawati, 2016, p.16), validitas isi adalah ungkapan suatu pertanyaan tentang suatu kemampuan khusus yang perlu diukur. Berdasarkan hal tersebut dapat disimpulkan bahwa validitas

isi berkaitan dengan analisis rasional terhadap daerah yang akan diukur untuk menentukan ekspresi suatu alat dengan kemampuan pengukuran (Herawati, 2000, hlm. 18). Validitas indeks validitas ini dapat ditentukan, termasuk penggunaan indeks yang dikemukakan oleh Aiken (1980) (dalam Heri Retnawati, 2016, hal.18).

$$V = \frac{\sum s}{n(c-1)} \dots\dots\dots 3.4$$

V adalah indeks persetujuan penilai untuk validitas item, s adalah skor yang diberikan oleh setiap penilai dikurangi skor terendah dalam kategori yang digunakan (s= r - lo di mana r adalah skor untuk kategori pilihan penilai dan lo adalah skor terendah). dalam kategori penilaian); n jumlah penilai; c Jumlah kategori yang dapat dipilih oleh penilai.

Nilai indeks V berkisar dari nol (0) hingga satu (1). Pada hasil perhitungan dapat diklasifikasikan berdasarkan nilai metrik, yang ditunjukkan pada Tabel 3.7 di bawah ini.

Tabel 3.7 Kategori indeks validitas

Indeks validitas	Kategori
≤ 0,40	Kurang
0,41 – 0,80	Sedang
>0,80	Tinggi

(Heri Retnawati, 2016, hlm.19)

Reliabilitas

Reliabilitas adalah derajat invariansi antara skor dua ukuran untuk subjek yang sama, bahkan ketika menggunakan alat ukur yang berbeda dan skala yang berbeda (Heri Retnawati, 2016, hlm. 84). Suatu tes dapat dikatakan reliabel hanya jika hasil pengukurannya mendekati kondisi sebenarnya dari si pengambil tes (Heri Retnawati, 2016, hlm. 84).

Keandalan suatu tes biasanya dinyatakan secara numerik dalam bentuk koefisien antara -1,00 dan +1,00. Koefisien yang tinggi menunjukkan keandalan yang tinggi. Sebaliknya, ketika koefisien nilai tes rendah, reliabilitas tes juga rendah (Heri Retnawati, 2016, hlm. 85).

Reliabilitas tes dalam penelitian ini menggunakan rumus Alpha Cronbach sebagai berikut:

$$r = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right) \dots\dots\dots 3.5$$

Keterangan:

- r : koefisien reliabilitas
- k : banyaknya butir pertanyaan
- $\sum \sigma_i^2$: jumlah varians butir
- σ_t^2 : varians skor total

Serupa dengan indeks validitas V, koefisien reliabilitas r memiliki nilai antara nol (0) dan satu (1). Dari hasil perhitungan dapat diklasifikasikan berdasarkan koefisien reliabilitas yang disajikan pada Tabel 3.8.

Tabel 3.8 Kategori koefisien reliabilitas

Indeks validitas	Kategori
0,0 – 0,2	Sangat rendah
0,2 – 0,4	Rendah
0,4 – 0,6	Cukup
0,6 – 0,8	Tinggi
0,8 – 1,0	Sangat tinggi

(Arikunto, 2007, hlm.75)

3.6 Hasil Validitas dan Reliabilitas Soal

Dalam penelitian, alat penelitian harus diuji validitas dan reliabilitasnya. Validitas yang digunakan dalam penelitian ini adalah validitas isi. Validitas isi berkaitan dengan analisis rasional terhadap wilayah yang akan diukur. Anda dapat menggunakan indeks validitas untuk menentukan validitas ini. Keyakinan itu sendiri diukur agar mengetahui jika hasil pengukuran mendekati kondisi sebenarnya dari peserta tes.

Hasil perhitungan didapat bahwa terdapat tiga soal termasuk ke dalam kategori dengan validitas sedang yaitu nomor 2, 6, dan 25. Selanjutnya, untuk 27 soal lainnya termasuk ke dalam kategori dengan validitas yang tinggi yaitu nomor 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, dan nomor 30. Perhitungan validitas dan perhitungan reliabilitas bisa dilihat di dalam lampiran.

Validitas ini terkait dengan analisis rasional terhadap domain yang hendak diukur. Dalam hal ini adalah validitas soal yang digunakan untuk mengukur ketercapaian hasil belajar siswa melalui ulangan harian. Berdasarkan tabel 3.9 diketahui bahwa indeks validasi soal pada penelitian ini adalah 0,883. Jika kita bandingkan dengan tabel 3.7, menurut Heri Retnawati (2016) indeks validitas 0,883 termasuk ke dalam kategori tinggi. Sehingga dapat kita katakan bahwa instrumen soal yang digunakan dalam penelitian ini dapat mengukur hasil belajar siswa melalui ulangan harian secara valid. Selain itu, butir-butir soal di dalam instrumen penelitian ini dapat mencerminkan hasil belajar siswa yang telah diukur secara valid.

Reliabilitas soal tentunya tidak kalah penting dengan validitas soal. Maka soal yang digunakan pada penelitian ini juga dihitung koefisien reliabilitasnya. Berdasarkan hasil perhitungan (terlampir), didapatkan koefisien reliabilitas soal pada penelitian ini adalah 0,62. Menurut Arikunto (2007) koefisien reliabilitas soal sebesar 0,62 ini termasuk dalam kategori tinggi. Sehingga dapat disebutkan bahwa hasil pengukuran dengan menggunakan instrumen soal dalam penelitian ini adalah mendekati keadaan peserta tes yang sebenarnya.