

## **BAB III**

### **METODE PENELITIAN**

#### **A. Jenis Penelitian**

Penelitian ini merupakan penelitian pengembangan atau *research & development (R & D)* dengan menggunakan model pengembangan instrumen yang dikemukakan oleh Retnawati (2016: 3-6). Model ini terdiri dari 9 langkah, yaitu: (1) menentukan tujuan pengembangan, (2) menentukan cakupan materi, (3) menentukan indikator instrumen, (4) menyusun butir instrumen, (5) validasi isi, (6) revisi, (7) uji coba instrumen, (8) analisis hasil uji coba, dan (9) merakit instrumen.

#### **B. Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan pada hari Kamis-Jum'at, 17-18 Maret 2016 di SMA N 1 Boyolali yang beralamat di Jalan Kates No. 8 Boyolali.

#### **C. Subjek Penelitian**

Subjek dalam penelitian ini adalah siswa kelas XI SMA N 1 Boyolali kelompok matematika dan ilmu alam. Jumlah sampel yang diambil sebanyak 5 kelas yang terdiri dari 146 siswa.

#### **D. Desain Penelitian**

1. Menentukan tujuan pengembangan

Tujuan dilakukannya penelitian adalah untuk: (1) menghasilkan instrumen evaluasi matematika untuk mengukur kemampuan berpikir kritis siswa SMA kelas XI, dan (2) mendeskripsikan kualitas instrumen evaluasi matematika untuk mengukur kemampuan berpikir kritis siswa SMA kelas XI.

2. Menentukan cakupan materi

Pada penelitian ini, peneliti memilih materi aritmetika dan aljabar sederhana. Kedua materi tersebut dipilih karena lebih dapat dibuat menjadi masalah yang realistis maupun kontekstual. Hal ini ditujukan agar soal yang disusun memenuhi kriteria soal matematika yang baik menurut Fung & Roland (2004: 290), yaitu menarik, sesuai dengan kehidupan sehari-hari siswa, dan mengandung konsep matematika yang nyata sehingga dapat meningkatkan pemahaman dan pengetahuan siswa.

3. Menentukan indikator instrumen

Indikator instrumen penelitian ditentukan berdasarkan kajian terhadap buku dan penelitian terkait kemampuan berpikir kritis yang sudah ada sebelumnya. Berdasarkan kajian pustaka yang dilakukan oleh peneliti, indikator kemampuan berpikir kritis yang digunakan adalah indikator yang dapat muncul dan diamati dari hasil tes yang diperoleh. Sebagai hasilnya, ada 4 indikator yang dipilih, yaitu: (1) menginterpretasi; (2) menganalisis; (3) mengevaluasi; dan (4)

menjelaskan. Penjelasan masing-masing indikator disajikan pada Tabel 2 halaman 26. Indikator yang ada disusun ke dalam kisi-kisi soal agar penyebarannya lebih merata. Kemudian, butir soal disusun berdasarkan kisi-kisi yang sudah ada.

#### 4. Menyusun butir instrumen

Butir instrumen harus disesuaikan dengan tujuan penelitian. Pengukuran kemampuan berpikir kritis menitikberatkan pada proses. Maka dari itu, butir soal yang digunakan dalam penelitian ini berbentuk esai/uraian yang lebih mampu menampilkan proses siswa memperoleh jawaban. Selain itu, jumlah butir yang digunakan menyesuaikan alokasi waktu yang tersedia. Pada rancangan awal, peneliti menyusun 4 paket soal dengan masing-masing paket terdiri dari 10 butir soal uraian. Keempat paket tersebut selanjutnya dinamai dengan “**Paket 1**”, “**Paket 2**”, “**Paket 3**”, dan “**Paket 4**”. Selengkapnya, instrumen awal penelitian dapat dilihat pada **Lampiran 2** halaman 83-133.

#### 5. Validasi isi

Dalam melakukan validasi isi butir soal, peneliti meminta bantuan kepada 3 orang ahli/validator yang terdiri dari 3 orang dosen matematika. Ahli/validator menilai dan memberikan masukan menggunakan lembar validasi yang telah disediakan terhadap keempat paket instrumen yang telah disusun. Waktu yang diperlukan pada tahap ini adalah sekitar 1 hingga 2 pekan. Sedangkan nama validator yang dipilih dalam penelitian ini disajikan pada Tabel 4 halaman 39.

Tabel 4. Daftar Nama Validator

No.	Nama Validator	Keterangan
1.	Dr. Ali Mahmudi	Dosen Pendidikan Matematika UNY
2.	Musthofa, M. Sc.	Dosen Pendidikan Matematika UNY
3.	Ilham Rizkianto, M. Sc.	Dosen Pendidikan Matematika UNY

6. Revisi berdasarkan masukan validator

Instrumen yang telah disusun beserta lembar validasi diserahkan kepada ahli/validator untuk dinilai dan diberikan masukan. Kemudian hasilnya dijadikan acuan untuk melakukan revisi agar instrumen siap diujicobakan. Jumlah soal dan soal mana saja yang akan diujicobakan dipilih berdasarkan masukan ahli/validator.

Skor yang terdapat pada lembar validasi menggunakan skala 1 sampai 5, dengan keterangan sebagai berikut: 1 = sangat tidak sesuai ; 2 = kurang sesuai ; 3 = cukup sesuai ; 4 = sesuai ; dan 5 = sangat sesuai. Kemudian, dari skor tersebut dihitung indeks Aiken-nya untuk menentukan validitas isi instrumen. Indikator yang digunakan dalam lembar validasi ahli diadaptasi dari Puspendik (2007: 13-14) dan Indrayana (2010) serta disajikan pada Tabel 5 halaman 40.

Instrumen awal penelitian yang sebelumnya terdiri dari 4 paket dengan 10 soal uraian setiap paketnya direvisi menjadi 4 paket dengan 8 soal uraian setiap paketnya. Dengan kata lain, ada beberapa soal yang dibuang. Hal tersebut dilakukan karena mempertimbangkan waktu tes yang tersedia. Selain itu, pertimbangan yang digunakan adalah terpenuhinya 4 indikator kemampuan berpikir kritis di masing-masing paket soal. Hal ini memungkinkan untuk mengelompokkan kembali soal

dari keempat paket yang ada sehingga setiap paket memuat keempat indikator kemampuan berpikir kritis. Kemudian, instrumen hasil revisi disebut “**Paket A**”, “**Paket B**”, “**Paket C**”, dan “**Paket D**”.

Tabel 5. Indikator Penilaian Instrumen Soal

No.	Aspek yang Dinilai
<b>A Aspek Materi</b>	
1	Materi soal sesuai dengan tujuan pengukuran
2	Butir soal sesuai dengan indikator
3	Materi soal sesuai dengan tingkat/jenjang pendidikan siswa
4	Hanya ada satu jawaban yang benar
5	Masalah yang disajikan menarik bagi siswa
<b>B Aspek Konstruksi</b>	
6	Setiap butir soal tidak bergantung satu sama lain
7	Penyelesaian soal lebih dari satu langkah
8	Merangsang kemampuan berpikir kritis siswa
9	Merangsang rasa ingin tahu siswa
10	Rumusan pertanyaan soal menggunakan kata tanya atau perintah yang menuntut siswa menguraikan jawaban
11	Penyajian gambar, grafik, tabel, atau sejenisnya disajikan secara jelas dan dapat dipahami
12	Petunjuk mengerjakan soal jelas dan dapat dipahami
13	Terdapat pedoman penskoran
<b>C Aspek Bahasa</b>	
14	Penggunaan bahasa sesuai dengan EYD
15	Bahasa yang digunakan komunikatif
16	Tidak menggunakan bahasa yang berlaku setempat
17	Rumusan soal tidak menimbulkan penafsiran ganda
18	Rumusan soal tidak mengandung kata-kata yang dapat menyinggung perasaan siswa

7. Uji coba instrumen

Instrumen yang lolos revisi dan telah dinyatakan layak serta dapat diujicobakan dipilih untuk digunakan pada tahap uji coba. Pada tahap ini, instrumen evaluasi yang diujicobakan adalah **Paket A**.

8. Analisis hasil uji coba

Skor jawaban siswa dari soal yang diujicobakan, kemudian dianalisis untuk mengetahui kualitas instrumen/produk. Pada tahap ini, terdapat 4

analisis data yang dilakukan, diantaranya: (1) pembuktian validitas konstruk, (2) analisis butir soal, yang terdiri dari tingkat kesukaran, daya beda, dan reliabilitas butir soal.

#### 9. Merakit instrumen

Apabila terdapat soal yang tidak memenuhi kriteria sebagai butir soal yang baik, maka peneliti dapat merekomendasikan butir soal yang terdapat pada paket lain untuk dirakit dengan butir yang lolos seleksi. Selengkapnya, instrumen akhir penelitian dapat dilihat pada **Lampiran 4** halaman 146-186.

### **E. Instrumen Penelitian**

Pada penelitian ini, instrumen yang digunakan adalah lembar kisi-kisi soal, lembar soal, lembar kunci jawaban dan panduan penskoran, serta lembar validasi ahli. Secara lengkap, instrumen awal penelitian disajikan pada **Lampiran 2** halaman 83-133.

### **F. Teknik Pengumpulan Data**

Pada penelitian ini, terdapat beberapa teknik pengumpulan data. Pada tahap pengumpulan informasi awal mengenai instrumen penilaian yang biasa digunakan oleh guru di kelas, peneliti melakukan wawancara. Pada tahap validasi ahli, peneliti mengumpulkan data menggunakan angket (lembar validasi ahli). Sedangkan pada tahap uji coba awal dan uji lapangan, peneliti menggunakan teknik tes. Tes dilakukan dalam waktu 2 jam pelajaran atau

sekitar 80 menit. Selanjutnya, hasil yang diperoleh dianalisis sebagai dasar untuk mengevaluasi dan memperbaiki soal agar menjadi produk yang layak digunakan.

## G. Teknik Analisis Data

### a. Validitas Isi

Validitas isi merupakan validitas yang diperoleh dari pengujian terhadap kelayakan atau kesesuaian isi tes melalui analisis rasional oleh panel yang berkompeten atau *expert judgement* (penilaian ahli). Untuk instrumen evaluasi berbentuk tes, pembuktian validitas isi dapat dilakukan dengan membandingkan antara isi instrumen dengan materi pelajaran yang telah diajarkan (Sugiyono, 2013: 353).

Dalam membuktikan validitas isi butir instrumen, peneliti menggunakan rumus indeks Aiken ( $V$ ) yang dikemukakan oleh Retnawati (2016: 18), yaitu:

$$V = \frac{\sum S}{N(C - 1)}$$

keterangan:

$V$  = indeks kesepakatan ahli mengenai validitas butir

$$S = R - L_0$$

$L_0$  = angka penilaian validitas terendah

$C$  = angka penilaian validitas tertinggi

$N$  = banyaknya ahli/validator

$R$  = angka yang diberikan oleh ahli.

Nilai yang diperoleh kemudian diklasifikasikan validitasnya.

Pengklasifikasian validitas isi instrumen didasarkan Tabel 6 berikut:

Tabel 6. Klasifikasi Validitas Isi Instrumen

No.	Indeks Aiken (V)	Validitas
1.	$0 \leq V \leq 0,4$	Kurang valid (rendah)
2.	$0,4 < V \leq 0,8$	Cukup valid (sedang)
3.	$0,8 < V \leq 1,0$	Sangat valid (tinggi)

(Retnawati, Heri. 2016: 19)

#### b. Validitas Konstruk

Sebuah instrumen evaluasi dikatakan memiliki validitas konstruk jika butir soal yang membangun instrumen tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam tujuan evaluasi (Arikunto, 2013: 83). Pembuktian validitas konstruk dapat dilakukan dengan dua cara, yaitu *exploratory factor analysis (EFA)* dan *confirmatory factor analysis (CFA)*.

Penelitian ini menggunakan metode *EFA* untuk membuktikan validitas konstruk instrumen dan dibantu program SPSS 16.0. Beberapa langkah yang dilakukan dalam metode *EFA* adalah sebagai berikut:

##### 1) Pengujian *KMO and Bartlett's Test*

Pada langkah ini, analisis faktor layak dilakukan apabila nilai *Kaiser-Meyer-Olkin measure of sampling adequacy (KMO)* lebih dari 0,5. Selain itu, harus dipenuhi syarat bahwa nilai signifikansi kurang dari 0,01. Dengan demikian, hasil tersebut menunjukkan bahwa sampel yang digunakan pada analisis faktor telah cukup (Retnawati, 2016: 47).

## 2) Pengujian *Total Variance Explained*

Pada tahap ini, pengujian difokuskan pada kolom *Initial Eigenvalues*. Nilai minimal *Eigenvalues* pembentuk faktor adalah 1. Hal ini berarti bahwa jika *Eigenvalues* kurang dari 1, maka tidak terdapat variabel pembentuk faktor. Sedangkan nilai *Cummulative %* menunjukkan berapa persen (%) instrumen mampu mengukur kemampuan berpikir kritis siswa.

## 3) Pengujian *Scree Plot*

Selain dari tabel *Total Variance Explained*, untuk menentukan jumlah faktor yang terbentuk, peneliti menggunakan grafik *Scree Plot*. Caranya adalah dengan memperhatikan garis penghubung antarnilai Eigen. Jika setelah suatu faktor mengalami penurunan tajam, maka faktor yang valid hanya sampai faktor tersebut. Selain itu, dapat pula dilihat dari nilai Eigen-nya. Apabila titik-titik tersebut berada di atas nilai Eigen 1 (satu), maka dapat disimpulkan variabel tersebut membentuk suatu faktor.

## 4) Pengujian *Rotated Component Matrix*

Tahap analisis data pada analisis faktor selanjutnya adalah memperhatikan *output* pada tabel *Rotated Component Matrix*. Tabel ini menunjukkan korelasi tiap butir dengan tiap komponen (faktor). Dengan kata lain, pada tahap ini dapat ditentukan butir yang belum jelas akan menjadi komponen (faktor) yang mana. Setelah itu, faktor

yang muncul dapat dinamai berdasarkan indikator kemampuan berpikir kritis yang dominan.

**c. Taraf Kesukaran ( $p$ )**

Soal yang baik adalah soal yang tidak terlalu sulit atau terlalu mudah. Soal yang terlalu mudah kurang dapat merangsang kemampuan berpikir kritis siswa. Sedangkan soal yang terlalu sulit cenderung membuat siswa putus asa dalam mengerjakannya. Rumus untuk mencari taraf kesukaran ( $p$ ) adalah:

$$p_i = \frac{J_i}{JT}$$

keterangan:

$p_i$  = indeks kesukaran butir soal ke- $i$

$J_i$  = jumlah skor yang diperoleh seluruh siswa pada soal ke- $i$

$JT$  = skor maksimal  $\times$  jumlah siswa.

Nilai yang diperoleh kemudian diklasifikasikan tingkat kesukarannya. Pengklasifikasian tersebut didasarkan Tabel 7 berikut:

Tabel 7. Klasifikasi Tingkat Kesukaran Butir Soal

No.	Indeks Kesukaran ( $p$ )	Keterangan
1	$0 \leq p \leq 0,3$	Sulit
2	$0,3 < p \leq 0,7$	Sedang
3	$0,7 < p \leq 1,0$	Mudah

(Arikunto, S. 2013: 225)

**d. Daya Beda Butir Soal**

Daya beda butir soal adalah kemampuan suatu soal untuk membedakan antara siswa yang berkemampuan tinggi dengan yang

berkemampuan rendah. Butir soal yang baik adalah soal yang memiliki indeks daya beda (diskriminasi) 0,3 sampai 0,7.

Ada beberapa cara untuk mengetahui daya beda butir soal. Dalam penelitian ini, peneliti menggunakan koefisien korelasi *point biserial* ( $r_{pbis}$ ) dalam mengukur daya beda butir soal. Untuk menentukan kriteria daya beda butir soal penelitian, koefisien korelasi *point biserial* ( $r_{pbis}$ ) yang telah ditemukan dikonsultasikan kepada Tabel 1 halaman 21. Rumus menghitung koefisien korelasi *point biserial* ( $r_{pbis}$ ):

$$r_{pbis} = \frac{M_p - M_q}{S_t} \sqrt{pq}$$

keterangan:

$r_{pbis}$  = indeks daya beda butir soal

$M_p$  = *mean* skor pada tes dari peserta pada kelompok atas

$M_q$  = *mean* skor pada tes dari peserta pada kelompok bawah

$S_t$  = standar deviasi seluruh skor tes

$p$  = proporsi peserta tes kelompok atas

$q = 1 - p$ .

#### e. **Reliabilitas**

Untuk menentukan reliabilitas soal, peneliti menggunakan perbandingan antara  $r_{hitung}$  dengan  $r_{tabel}$  pada taraf signifikan 5%. Jika  $r_{hitung} \geq r_{tabel}$ , maka tes dikatakan reliabel. Akan tetapi, jika  $r_{hitung} < r_{tabel}$ , maka tes tersebut dikatakan tidak reliabel. Nilai  $r_{hitung}$  dapat juga diinterpretasikan dengan kriteria reliabilitas pada Tabel 8 halaman 47.

Sedangkan rumus yang digunakan untuk menghitung besarnya koefisien reliabilitas butir soal uraian adalah rumus *Alpha Cronbach* (Arikunto, 2013: 122):

$$r_{11} = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

keterangan:

$r_{11}$  = reliabilitas tes secara keseluruhan

$\sum \sigma_i^2$  = jumlah varians skor tiap-tiap item

$\sigma_t^2$  = varians total

$n$  = banyak item (soal).

Tabel 8. Klasifikasi Reliabilitas Butir Soal

No.	Koefisien Reliabilitas ( $r_{11}$ )	Keterangan
1.	$0,0 \leq r_{11} \leq 0,2$	Sangat rendah
2.	$0,2 < r_{11} \leq 0,4$	Rendah
3.	$0,4 < r_{11} \leq 0,7$	Cukup
4.	$0,7 < r_{11} \leq 0,9$	Tinggi
5.	$0,9 < r_{11} \leq 1,0$	Sangat tinggi

Guilford (Sudijono, 2015: 193)