

BAB III

METODOLOGI PENELITIAN

A. Metode Penelitian

Menurut Firman (2000) pengembangan tes adalah suatu proses perancangan dan perbaikan alat ukur (tes) agar menjadi alat ukur yang berkualitas baik. Metode penelitian yang digunakan untuk membuat dan mengembangkan suatu instrumen penilaian adalah metode penelitian dan pengembangan (*Research and Development*). Metode penelitian pengembangan merupakan dasar untuk mengembangkan produk yang akan dihasilkan. (Sugiyono, 2006)

Menurut Sukmadinata (2009) ada beberapa metode yang digunakan dalam pelaksanaan penelitian dan pengembangan, yaitu metode: deskriptif, evaluatif, dan eksperimental. Metode penelitian deskriptif, digunakan dalam penelitian awal untuk menghimpun data tentang kondisi yang ada. Kondisi yang ada mencakup: (1) kondisi produk-produk yang sudah ada sebagai bahan perbandingan atau bahan dasar untuk produk yang akan dikembangkan, (2) kondisi pihak pengguna, seperti sekolah, guru, kepala sekolah, siswa, beserta pengguna lainnya, (3) kondisi faktor-faktor pendukung dan penghambat pengembangan dan penggunaan dari produk yang akan dihasilkan, mencakup unsur manusia, saran-prasarana, biaya, pengelolaan, dan lingkungan.

Metode evaluatif digunakan untuk mengevaluasi proses uji coba pengembangan suatu produk. Produk dikembangkan yang kemudian diuji dengan serangkaian uji coba, dan untuk setiap kegiatan uji coba diadakan

evaluasi, baik evaluasi hasil maupun evaluasi proses. Berdasarkan temuan-temuan hasil uji coba diadakan penyempurnaan-penyempurnaan.

Metode eksperimen digunakan untuk menguji kemampuan dari produk yang dihasilkan. Walaupun dalam tahap uji coba telah ada evaluasi (pengukuran), tetapi pengukuran tersebut masih dalam rangka pengembangan produk, belum ada kelompok pembanding. Pengukuran tidak hanya dilakukan pada kelompok eksperimen tetapi juga juga pada kelompok pembanding atau kelompok kontrol. Perbandingan hasil eksperimen pada kedua kelompok tersebut dapat menunjukkan tingkat kemampuan dari produk yang dihasilkan.

Metode penelitian yang digunakan adalah metode penelitian evaluatif. Gambar di bawah ini merepresentasikan langkah penelitian *R & D* (Sugiyono, 2009):



Gambar 3.1 Langkah Penelitian R & D

Hal yang dikembangkan dalam penelitian ini adalah mengenai instrumen penilaian berbasis kelas yang salah satunya dapat mengukur tes keterampilan berpikir kritis. Kualitas instrumen yang dikembangkan ditentukan melalui pembuatan atau penyusunan tes, validitas dan reliabilitas tes, serta tingkat kesukaran dan daya pembeda soal (Arikunto,2007). Adapun alur penelitiannya dapat dilihat pada Gambar 3.2

Langkah-langkah penelitian tersebut dapat dijabarkan sebagai berikut:

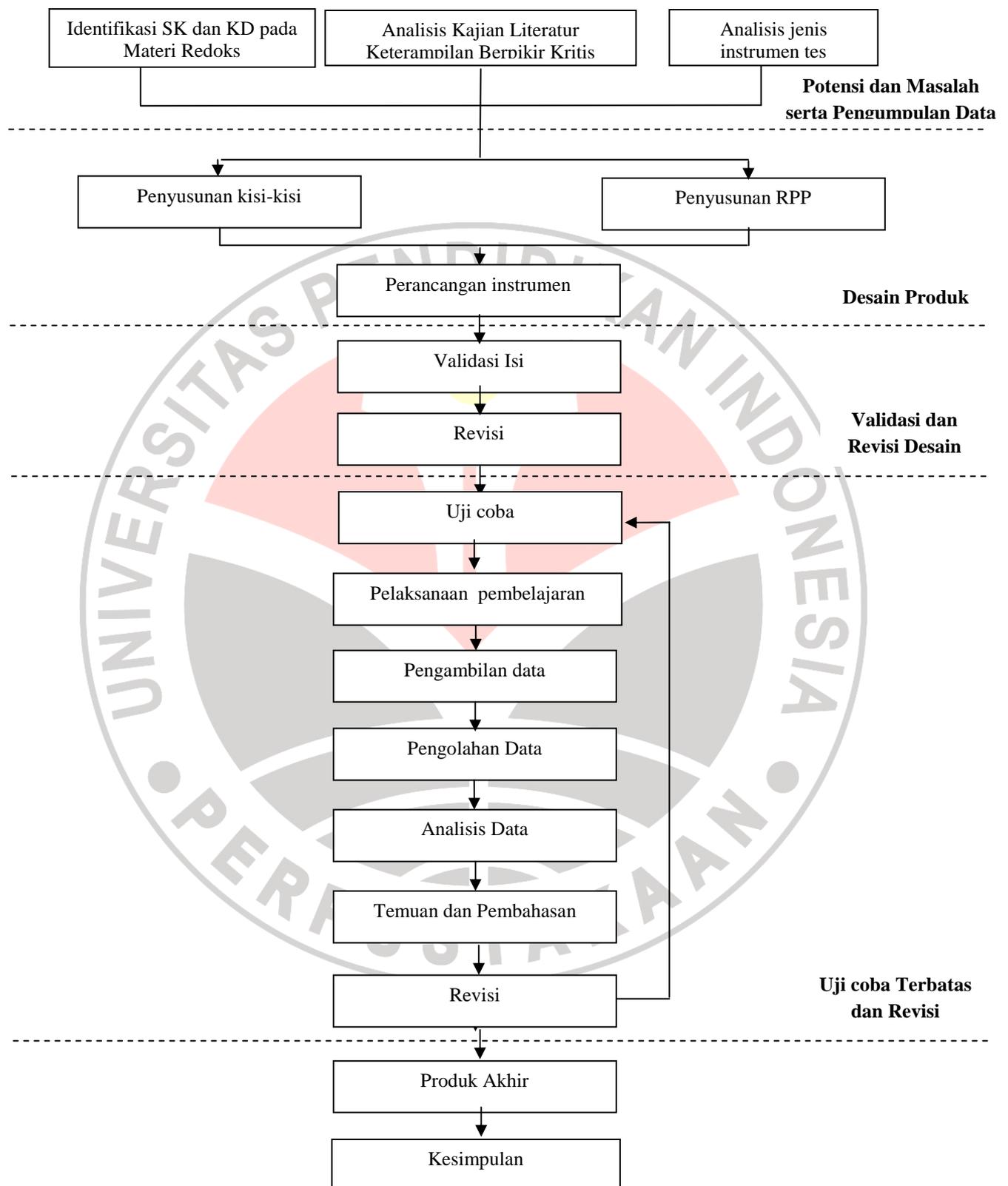
1. Potensi dan Masalah

Penelitian ini berangkat dari adanya potensi atau masalah. Potensi adalah segala sesuatu yang apabila didayagunakan akan memiliki suatu nilai tambah pada produk yang diteliti. Masalah adalah sesuatu yang bisa dicari solusi pemecahannya. Potensi dan masalah pada penelitian ini telah dibahas di latar belakang penelitian.

2. Pengumpulan Data

Setelah potensi dan masalah dapat ditunjukkan secara faktual, selanjutnya adalah mengumpulkan informasi dan melakukan studi literatur. Langkah ini ditujukan untuk menemukan konsep-konsep atau landasan teori yang bisa menjadi masukan bagi pengembangan produk.

Melalui studi literatur dan pengumpulan informasi dikaji pula ruang lingkup suatu produk, keluasan penggunaan, kondisi-kondisi pendukung agar produk dapat digunakan atau diimplementasikan secara optimal. Tahap ini juga dilakukan untuk mengetahui langkah-langkah yang paling tepat dalam pengembangan produk tersebut.



Gambar 3.2 Alur penelitian

3. Desain Produk

Langkah ini meliputi pengembangan produk mengacu pada rancangan yang telah ditentukan sebelumnya. Pengembangan desain produk disertai dengan landasan pemikiran dan keilmuan berdasarkan hasil pengumpulan informasi dan studi literatur.

Penyusunan instrumen PBK diawali dengan pembuatan kisi-kisi instrumen. Karena PBK terintegrasi dan tidak dapat dipisahkan dalam proses pembelajaran (Earl dan Katz, 2006). Maka dari itu pengembangan PBK yang mengukur keterampilan berpikir kritis akan lebih baik dengan disusunnya perangkat pembelajaran.

Tidak semua metode pembelajaran dapat meningkatkan keterampilan berpikir kritis siswa, untuk itu diperlukan suatu metode pembelajaran dengan gaya belajar-mengajar aktif (Filsaime, 2010) dengan pendekatan *student-centered* (Stiggins, 1994). Salah satu model pembelajaran yang dapat meningkatkan keterampilan berpikir kritis siswa adalah model inkuiri (Purwaningsih, 2005; Rahayu, 2006 dan Risnawati 2010) dan metode pembelajaran untuk meningkatkan dan memperbaiki daya berpikir kritis siswa adalah metode diskusi dan praktikum (Filsaime, 2010).

4. Validasi Desain (*Desk Evaluation*)

Validasi desain merupakan kegiatan untuk menilai apakah rancangan produk efektif dalam mengatasi masalah yang ada. Validasi di sini masih bersifat penilaian berdasarkan pemikiran rasional, belum mencapai fakta di lapangan. Validasi desain produk dapat dilakukan dengan cara menghadirkan beberapa

pakar atau tenaga ahli yang sudah berpengalaman untuk menilai produk tersebut (Validasi isi).

Menurut Firman (2000) cara menilai atau menyelidiki validitas isi suatu alat ukur ialah dengan mengundang ‘*judgment*’ (timbangan) kelompok ahli dalam bidang yang diukur. Kemudian, timbangan para pakar dihitung dengan indeks *Content Validity Ratio* (CVR) (Lawshe, 1975). Para pakar yang memvalidasi instrumen disajikan dalam Tabel 3.1 berikut:

Tabel 3.1 Pakar yang memvalidasi instrumen

No	Panelis pakar	Status	Pendidikan terakhir
1.	Bapak Dr. Nahadi, M. Pd., M.Si.	Dosen mata kuliah Evaluasi, Jurusan pendidikan Kimia UPI	S3 Pendidikan, UPI
2.	Ibu Tuszie Widiyanti, M. Pd.	Dosen Jurusan pendidikan Kimia UPI	S2 Pendidikan Ilmu Alam, UPI
3.	Bapak Drs. Harry Firman, M.Pd.	Dosen evaluasi pendidikan, Jurusan Pendidikan Kimia UPI	S2 Pendidikan Ilmu Alam, UPI
4.	Ibu Dra. Lilis Yulianingsih	Dosen kimia di Sekolah Tinggi Farmasi, Bandung.	S1 Kimia, UNPAD
5.	Ibu Rahmah, M.Si.	Dosen Kimia analitik, YAPIKA, YPUP Makasar	S1 Kimia, Universitas Hasanudin.
6.	Ibu Wido Studianah, S.Pd.	Guru kimia SMAN 11 Bandung	S1 Pendidikan Kimia UPI
7.	Ibu Hidayah S.Pd.,	Guru kimia SMAN 11 Bandung	S1 Pendidikan Kimia UPI
8.	Ibu Dra. Sri Juliah S.Pd.,	Guru kimia SMAN 11 Bandung	S1 Pendidikan Kimia UPI
9.	Bapak Budi Santosa S.Pd.	Guru kimia SMAN 1 Ampah, Kalimantan Tengah	S1 Kimia Universitas Airlangga
10.	Bapak Khoirul Anam S.Si.	Kepala sekolah dan guru kimia di SMAN I Ampah, Kalimantan tengah	S1 jurusan Kimia di universitas Airlangga

Hasil dari validitas isi dapat dilihat pada Lampiran C1.

5. Revisi Desain

Setelah desain produk divalidasi melalui diskusi dengan pakar, beberapa hal yang harus diperbaiki kemudian ditindaklanjuti melalui tahap revisi. Revisi ini dilakukan bersama dengan para pakar agar revisi lebih terarah dan tepat sasaran.

Berdasarkan masukan dari kelompok ahli, maka dilakukan revisi terhadap instrumen yang dikembangkan. Perbaikan butir stem pernyataan dan pertanyaan instrumen meliputi perbaikan penulisan yang kurang tepat, isi materi reaksi redoks, kesesuaian indikator dengan butir soal, serta kesesuaian antara butir soal dengan keterampilan yang akan diukur.

6. Uji Coba Produk Terbatas (*Preliminary Field Testing*)

Uji coba lapangan bertujuan untuk mendapatkan kelayakan secara mikro. Selama pelaksanaan uji coba di lapangan, peneliti mengadakan pengamatan secara intensif dan mengumpulkan data yang akan dijadikan bahan untuk penyempurnaan produk awal tersebut.

Uji coba tes dilakukan sebanyak tiga kali pada suatu SMA kelas X untuk menghasilkan instrumen tes yang berkualitas (Sugiyono, 2008). Uji coba I dilakukan dengan responden sebanyak 31 siswa, uji coba II dilakukan dengan responden sebanyak 40 siswa dan uji coba III dilakukan dengan responden sebanyak 45 siswa. Tujuan dari uji coba adalah untuk mengetahui reliabilitas, validitas dan karakteristik psikometri tes berupa daya beda, tingkat kesukaran dan keefektifan distraktor.

Pada Pelaksanaan uji coba dilakukan pelaksanaan pembelajaran yang telah direncanakan dalam RPP. Kemudian, untuk mengukur ketercapaian indikator pembelajaran, dilaksanakan suatu tes tertulis dan sikap siswa terhadap mata pelajaran kimia. Jangka waktu pelaksanaan tes tertulis tidak begitu jauh dan terlalu dekat dari proses pembelajaran.

Setelah dilakukan penelitian, maka data yang diperoleh diolah dengan terlebih dahulu diskor. Kemudian dianalisis validitas dan reliabilitasnya serta karakteristik psikometris tesnya. Dari hasil analisis hasil tes, maka akan ditafsirkan dan menjadi temuan dalam penelitian. Temuan tersebut dibahas secara rinci.

7. Revisi Produk

Peneliti merevisi produk berdasarkan masukan yang didapat dari hasil uji coba lapangan. Revisi dilakukan untuk memperbaiki bagian dari produk yang dirasakan oleh pengguna produk masih kurang maksimal. Hasil akhir dari penelitian ini adalah menghasilkan produk berupa instrumen penilaian berbasis kelas yang berkualitas.

B. Responden Penelitian

Responden pada penelitian ini adalah siswa kelas X semester 2 di suatu SMA Negeri di kota Bandung. Jumlah siswa yang dijadikan responden penelitian sebanyak 116 siswa yang terbagi menjadi tiga kelas.

C. Teknik Pengumpulan Data

Dalam penelitian ini, teknik pengumpulan data meliputi pengumpulan data tes tertulis, skala sikap, observasi sikap dan observasi tes kinerja.

a. Pengumpulan Data Tes Tertulis dan Skala Sikap

Pengumpulan data untuk tes tertulis dilakukan dengan mengajukan pertanyaan-pertanyaan tertulis sebanyak 16 butir, yang terdiri dari 14 soal

pilihan ganda dan 2 soal uraian. Sementara pengumpulan data skala sikap dilakukan dengan mengajukan pernyataan-pernyataan tertulis sebanyak 46 butir. Pengumpulan data tes tertulis dan skala sikap dilakukan setelah responden menyelesaikan pembelajaran reaksi redoks.

b. Pengumpulan Data Observasi Sikap dan Tes Kinerja

Observasi sikap dilakukan pada saat dilaksanakannya pembelajaran. Untuk mengefektifkan observasi, peneliti dibantu 8 observer, masing-masing mengamati satu kelompok belajar yang merupakan kelompok praktikum juga. Observasi yang dilakukan merupakan observasi sistematis, yaitu observasi yang dilakukan dengan menggunakan format/blangko pengamatan sebagai instrumen. Format yang disusun berbentuk *check list* berisi item-item kejadian/tingkah laku yang digambarkan akan terjadi. Observer hanya tinggal mencek nilai keterampilan siswa yang berdasarkan standar penilaian yang diberikan.

Observer dilatih terlebih dahulu sebelum melakukan pengamatan melalui dua tahap, Tahap pertama yaitu mendiskusikan format observasi, menjelaskan dengan contoh kejadian atau gerak untuk setiap item, memahami apa yang harus diamati bagaimana cara member keputusan skor/cek terhadap hasil observasi. Tahap kedua adalah tahap simulasi, dimana salah seorang observer menjadi model sedangkan yang lain menjadi pengamat dan mengisi format (Brualdi (1998) dan Moskal (2003)).

Setelah dilaksanakan observasi, observer memberikan saran mengenai efektivitas instrumen observasi yang telah digunakan berikut saran perbaikan instrumen.

D. Teknik Pengolahan Data

Data yang diperoleh dari pelaksanaan penelitian kemudian diolah dan dianalisis. Teknik pengolahan data yang dilakukan pada penelitian ini adalah:

a. Validitas Tes

Validitas adalah suatu ukuran yang menunjukkan tingkat kevalidan atau kesahihan suatu instrumen. Dalam penelitian ini digunakan validitas isi dan validitas berdasarkan kriteria.

Validitas isi berkenaan dengan kevalidan suatu alat ukur dipandang dari segi isi (*content*) materi pelajaran yang dicakup oleh alat ukur tersebut. Untuk keperluan perlu melibatkan penelaah (*reviewer*) dalam menilai apakah instrumen yang dibuat sudah memenuhi syarat validitas.

Indeks untuk menyatakan keshahihan berdasarkan validitas isi secara kuantitatif yaitu *Content Validity Ratio* (CVR). Butir dianggap tidak baik atau tidak memenuhi syarat jika dianggap tidak penting atau tidak cocok, oleh mayoritas pakar.

Lawshe (1975) membuat rumus rasio validitas isi tentang rasio penilaian penting atau tidak penting (valid tidaknya butir) dari para pakar yang memvalidasi.

Rumus Lawshe:

$$CVR = \frac{n_e - N/2}{N/2} = \frac{2n_e}{N} - 1$$

n_e = banyaknya pakar yang sepakat

N = banyaknya pakar yang memvalidasi

Nilai CVR hitung kemudian dibandingkan dengan CVR tabel (lampiran C1), jika $CVR_h > CVR_t$ maka butir tersebut dinyatakan signifikan / valid.

Validitas berdasarkan kriteria/validitas empiris, yaitu dengan menggunakan teknik statistik yang dilakukan yaitu dengan menghitung koefisien validitas dengan menggunakan rumus korelasi *product moment* menggunakan angka kasar. Rumusnya adalah sebagai berikut:

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

Keterangan:

X = skor suatu butir soal

Y = skor total

r_{xy} = Koefisien validitas

Hasil dari r_{hitung} akan dibandingkan dengan r_{tabel} . Jika $r_{hitung} > r_{tabel}$ maka soal tersebut dikatakan signifikan/ valid.

b. Realibilitas

Reliabilitas adalah suatu ukuran yang menunjukkan sejauh mana suatu alat ukur memberikan gambaran yang benar-benar dapat dipercaya mengenai kemampuan seseorang. Pada penelitian ini reliabilitas diukur dengan menggunakan rumus KR#20 untuk soal pilihan ganda dengan

mempertimbangkan jumlah butir item yang sedikit dan kesetaraan tingkat kesukarannya (Firman, 2000).

Rumus KR#20:

$$r = \frac{n}{n-1} \left(1 - \frac{\sum pq}{s^2} \right)$$

r = reliabilitas tes

n = jumlah butir soal

p = proporsi jawaban benar

q = proporsi jawaban salah

s^2 = variansi

Kemudian untuk soal uraian dan skala sikap digunakan rumus koefisien alpha, dengan pertimbangan soal yang digunakan berbentuk uraian/essay yang skornya bukan 1 dan 0 (Arikunto, 2007).

Rumus koefisien alpha:

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

Keterangan:

r_{11} = reliabilitas instrumen

p = proporsi respon soal benar pada suatu soal

q = proporsi respon soal salah pada suatu soal

$\sum \delta^2$ = varian skor-skor tiap butir item

δ^2 = varian total

n = jumlah butir soal

Tabel 3.2 Klasifikasi Reliabilitas (Arikunto, 2010)

Reabilitas	Kriteria
0,8 -1,00	sangat tinggi
0,6 - 0,79	Tinggi
0,4 -0,59	Sedang
0,2 - 0,39	Rendah
0 < 0,2	sangat rendah

c. Daya Pembeda

Daya pembeda soal (D) merupakan salah satu karakteristik psikometrik tes, untuk menunjukkan kemampuan suatu soal dalam membedakan antara siswa yang berkemampuan tinggi dan siswa yang berkemampuan rendah. Sehingga, analisis dengan daya pembeda soal pada instrumen penelitian yang dikembangkan hanyalah tes tertulis.

Untuk soal pilihan ganda, indeks D dapat dihitung dengan menggunakan rumus:

$$D = \frac{n_t - n_r}{N_t}$$

Keterangan:

D = daya pembeda

n_t = jumlah siswa dari kelompok tinggi yang menjawab benar

n_r = jumlah siswa dari kelompok rendah yang menjawab benar

N_t = jumlah siswa kelompok tinggi

Tabel 3.3 Klasifikasi Daya Pembeda Soal (Arikunto, 2010)

Daya Pembeda	Kriteria
Negatif	Sangat jelek
0,00-0,20	Jelek
0,20- 0,40	Cukup
0,40-0,70	Baik
0,70-1,00	Baik sekali

Sementara untuk soal uraian daya pembeda soal dapat menggunakan uji t (Arifin, 2010), rumusnya sebagai berikut:

$$t = \frac{\bar{X}_a - \bar{X}_b}{\sqrt{\frac{\Sigma(X_1 - \bar{X}_a)^2 + \Sigma(X_2 - \bar{X}_b)^2}{n(n-1)}}$$

Keterangan:

\bar{X}_a = rata-rata kelompok atas

\bar{X}_b = rata-rata kelompok bawah

X_1 = skor kelompok atas

X_2 = skor kelompok bawah

$n = 27\% \times N$ (kelompok atas/kelompok bawah)

Kemudian untuk menentukan signifikansi butir soal, dihitung terlebih dahulu *degree of freedom* (df). $df = (n_1 - 1) + (n_2 - 1)$. Jika t hitung > t tabel, maka soal tersebut signifikan. Artinya dapat membedakan kemampuan siswa yang menguasai materi dengan yang tidak mampu menguasai materi.

d. Tingkat Kesukaran

Bilangan yang menunjukkan sukar dan mudahnya suatu soal disebut dengan indeks kesukaran. Indeks kesukaran menunjukkan taraf kesukaran soal. Untuk soal pilihan ganda, tingkat kesukaran dapat dicari dengan rumus:

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

Keterangan:

P = indeks kesukaran

B = banyaknya siswa yang menjawab benar

N = Jumlah seluruh peserta tes

Sementara untuk soal uraian tingkat kesukaran soal dapat dicari dengan menghitung proporsi menjawab benar yaitu jumlah peserta tes yang menjawab benar pada butir soal yang dianalisis dibandingkan dengan jumlah peserta tes seluruhnya merupakan tingkat kesukaran yang paling umum digunakan. Persamaan yang digunakan untuk menentukan tingkat kesukaran dengan proporsi menjawab benar adalah (Surapranata, 2004):

$$p = \frac{\sum x}{Sm \cdot N}$$

Keterangan:

p = Proporsi menjawab benar atau tingkat kesukaran

$\sum x$ = Banyaknya peserta tes yang menjawab benar

Sm = Skor maksimum

N = Jumlah peserta tes

Tabel 3.4 Kategori Tingkat Kesukaran (Arikunto, 2007)

Tingkat Kesukaran	Tafsiran
$p < 0,3$	Sukar
$0,3 < p \leq 0,7$	Sedang
$P > 0,7$	Mudah

e. Keefektifan Distraktor

Pada soal bentuk pilihan ganda, ada alternatif jawaban (opsi) yang merupakan pengecoh. Butir soal yang baik, pengecohnya akan dipilih secara

merata oleh peserta didik yang menjawab salah. Sebaliknya butir soal yang kurang baik pengecohnya akan dipilih secara tidak merata. Pengecoh dianggap baik, apabila jumlah peserta didik yang memilih pengecoh itu sama atau mendekati jumlah ideal. Indeks pengecoh dihitung dengan menggunakan rumus (Arifin, 2010):

$$IP = \frac{P}{(N - B)/(n - 1)} \times 100\%$$

Keterangan:

IP = indeks pengecoh

P = Jumlah peserta didik yang memilih pengecoh

N = jumlah peserta didik yang ikut tes

B = jumlah peserta didik yang menjawab benar pada setiap soal

n = jumlah alternatif jawaban

Kualitas pengecoh berdasarkan indeks pengecoh adalah ditafsirkan dalam tabel berikut:

Tabel 3.5 Klasifikasi Indeks Pengecoh (Arifin, 2010)

Indeks Pengecoh (%)	Kriteria
76-125	Sangat baik
51-75 atau 126-150	Baik
26-50 atau 151-175	Cukup
0-25 atau 176-200	Buruk
> 200	Sangat buruk

Selanjutnya, perlu diketahui pula, apakah suatu opsi (alternatif jawaban) berfungsi secara efektif atau tidak. Maka dari itu, efektivitas fungsi opsi dapat ditentukan dengan kriteria sebagai berikut:

1. Untuk opsi kunci:

a. Jumlah pemilih kelompok atas dan bawah tidak kurang dari 25%-75%

Menentukan jumlah pemilih dengan rumus $= \frac{\Sigma PKA + \Sigma PKB}{n_a + n_b} \times 100\%$

ΣPKA = jumlah pemilih kelompok atas

ΣPKB = jumlah pemilih kelompok bawah

n_a = jumlah sampel kelompok atas

n_b = jumlah sampel kelompok bawah

b. Jumlah pemilih kelompok atas > jumlah pemilih kelompok bawah

2. Untuk opsi pengecoh:

a. Jumlah Jumlah pemilih kelompok atas dan bawah tidak kurang dari

$$25\% \times \frac{1}{2(\Sigma d)} \times (Ka + Kb)$$

d = jumlah opsi pengecoh

Ka = kelompok atas

Kb = kelompok bawah

b. Jumlah pemilih kelompok bawah > jumlah pemilih kelompok atas