

BAB III METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini dirancang untuk mengembangkan instrumen *authentic assessment* pada kompetensi pemeliharaan berkala system sasis. Bentuk rancangan dalam penelitian ini untuk menghasilkan alat ukur penilaian yang teruji. Metode Kualitatif digunakan untuk menganalisis kebutuhan dan membuat rancangan, sedangkan untuk menguji dan mengembangkan produk menggunakan metode kuantitatif. Penelitian ini menggunakan *mixed method*, yang bertujuan merancang instrument *assessment authentic* secara valid dan reliabel.

Langkah pengembangan dalam penyusunan instrument *assessment authentic*, (Witjaksono, 2010) menjelaskan langkah utama dalam penelitian desain pengembangan adalah sebagai berikut : (1) identifikasi masalah, (2) merumuskan produk pengembangan, (3) desain dan pengembangan produk, (4) uji coba produk, (5) evaluasi, (6) mengkomunikasikan hasil.

Tesis ini menggunakan model 4D untuk penelitian pengembangannya, dimana mencoba mengembangkan *authentic assessment* dalam penelitian pengembangannya. Penelitian pengembangan model 4-D oleh Thiagarajan dan terdiri dari empat langkah yakni perencanaan (*define*), perancangan (*design*), pengembangan (*develope*), dan penyebaran (*dissemination*) (Thiagarajan, 1974), (Sugiyono., 2016b))



Gambar 3.1. Langkah-Langkah Penelitian Pengembangan 4-D

Berdasarkan penggunaan model pengembangan 4-D, maka tahapan pengembangan asesmen autentik dalam penelitian ini sebagai berikut:

1. Tahap Perencanaan pada tahap ini penulis merancang kebutuhan *authentic assessment* berdasarkan kebutuhan lapangan. Kebutuhan pengembangan *authentic assessment* di peroleh melalui kuesioner yang di sebarakan oleh peneliti kepada 4 orang guru Teknik Bisnis Sepeda Motor.
2. Tahap Perancangan pada tahap ini penulis merancang desain produk awal, dimana desain produk ini berupa alat tes evaluasi. Desain produk yang akan di rancang untuk aspek kognitif berupa soal pilihan ganda *authentic assessment*. Peneliti mengembangkan soal *authentic assessment* pada bidang perawatan berkala. Pengembangan butir soal melalui penyusunan kisi-kisi soal di sesuaikan dengan kompetensi dasar perawatan berkala sistem sasis. Desain produk untuk aspek psikomotor peneliti mengembangkan rubrik penilaian yang di sesuaikan dengan langkah kerja pada *jobsheet*. Peneliti juga menyusun buku pedoman penilaian untuk melengkapi proses penilaian. Kompetensi dasar yang akan dikembangkan untuk aspek kognitif maupun psikomotor di sesuaikan dengan kompetensi dasar perawatan berkala sistem sasis.
3. Tahap pengembangan peneliti melakukan beberapa tahapan pengembangan diantaranya:
 - a validasi konten baik aspek kognitif maupun psikomotor. Dimana masing-masing aspek divalidasi oleh 5 orang ahli. Validasi yang dilakukan ahli pada aspek kognitif berupa kesesuaian kisi-kisi soal dengan butir soal yang telah di buat berdasarkan kompetensi dasar yang disesuaikan. Validasi yang dilakukan ahli pada aspek psikomotor berupa keterbacaan rubrik nilai, *jobsheet* dan buku pedoman.
 - b Peneliti melakukan proses revisi pada produk setelah proses validasi. Revisi produk berdasarkan saran dari para ahli sehingga menghasilkan instrumen yang akan dilanjutkan pada proses uji coba.

- c Peneliti melakukan proses validasi konstruk dimana setelah proses uji coba lapangan. Peneliti melakukan validasi konstruk 35 butir soal kepada peserta didik, dan 27 langkah kerja. Validasi konstruk yang dilakukan peneliti pada aspek kognitif berupa menganalisis tingkat kesukaran, indeks pengecoh, daya beda dan analisis butir soal. Validasi konstruk untuk aspek psikomotor tersendiri berupa keterbacaan siswa terhadap *jobsheet* dan keterbacaan guru pada rubrik penilaian dan buku pedoman.
 - d Tahap selanjutnya peneliti melakukan uji coba terbatas terhadap instrumen yang telah di revisi berdasarkan hasil dari validitas konstruk. Tahap uji coba terbatas pada masing-masing aspek di lakukan kepada 15 orang peserta didik berdasarkan tingkat kemampuan yang berbeda-beda,
4. Tahap penyebaran pada tahap ini, peneliti mengimplementasikan produk akhir yang selesai di revisi kepada 52 peserta didik, dimana instrumen yang akan di hasilkan dapat mengetahui tingkat pemahaman kemampuan peserta didik.

3.2. Subjek Penelitian

Populasi dalam penelitian ini adalah siswa kelas XI TBSM yang berjumlah 180 orang siswa. Penelitian ini menggunakan sample dengan teknik *purposive random sampling*. Teknik untuk menentukan sampel penelitian dengan beberapa pertimbangan tertentu yang bertujuan agar data yang diperoleh nantinya bisa lebih representatif. (Sugiyono, 2013). Sample pada penelitian ini kelas XI TBSM 3 dengan jumlah 36 orang siswa.

3.3. Instrumen dan Teknik Pengumpulan Data Penelitian

Instrument penelitian ini berupa serangkaian instrument *authentic assessment* yang mengukur berpikir kritis. pada ranah kognitif. Teknik pengujian validitas instrumen pada penelitian ini menggunakan validitas isi (*content validity*) dan validitas konstruksi (*construct validity*). (Sudijono, 2009) mengemukakan bahwa “dalam praktik, validitas isi dari suatu tes hasil belajar dapat diketahui dengan jalan membandingkan antara isi yang terkandung dalam

tes hasil belajar dengan tujuan instruksional khusus yang telah ditentukan untuk masing-masing mata pelajaran”. Tes hasil belajar dapat dikatakan memiliki validitas konstruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir (kognitif, afektif, dan psikomotor) sebagaimana telah ditentukan dalam tujuan instruksional khusus (Sudijono, 2009).

Instrument yang akan digunakan terlebih dahulu diuji validitasnya oleh validator. Validasi instrument penting karena dapat mengetahui instrument layak atau tidak untuk mengambil data. Validasi instrument menggunakan validasi konstruk dan isi oleh validator.

Pengujian validitas pada penelitian ini menggunakan cara *expert judgement*, yaitu penilaian validitas instrumen dengan cara meminta pendapat dan rekomendasi dari para pakar yang dipandang memiliki keahlian yang ada hubungannya dengan instrumen penelitian

Penelitian ini penulis menggunakan teknik pengumpulan data antara lain:

1. Wawancara

Wawancara dalam penelitian ini dilakukan untuk mendukung data yang telah dikumpulkan. Isi dari wawancara ini berisi mengenai tanggapan guru terhadap penilaian autentik. Data yang dihasilkan pada wawancara ini untuk mendukung data dari penelitian.

2. Tes

Tes pada penelitian ini digunakan untuk menguji instrument yang telah penulis buat, pengujian pada penelitian ini di bagi 2 yaitu: 1) uji terbatas yang akan dilakukan untuk mengetahui validitas instrumen yang akan dijadikan produk akhir, 2) uji lapangan mengujikan kembali produk instrumen yang telah di revisi hingga produk instrumen akan di implementasikan..

3.4. Analisis Data

Teknik analisis pada penelitian pengembangan *authentic assessment* perawatan berkala sistem sasis sepeda motor dimana pada penelitian ini

menggunakan wawancara untuk mengetahui kebutuhan guru terhadap *authentic assessment*, dan tes untuk mengetahui validitas suatu instrument. Teknik analisa data yang di gunakan dalam penelitian pengembangan ini: uji validitas konten, uji validitas konstruk, uji reliabilitas, dan dissimination.

1. Uji Validitas Konten

1.1 Validitas Konten Pilihan Ganda

Validitas mengacu pada aspek ketepatan dan kecermatan hasil pengukuran. Pengukuran sendiri dilakukan untuk mengetahui seberapa banyak (dalam arti kuantitatif), suatu aspek psikologis terdapat dalam diri seseorang, yang dinyatakan oleh skornya pada instrument pengukur yang bersangkutan Azwar (2015).

a. Perhitungan rasio validitas isi dari C H Lawshe

Dalam konteks penelitian ini, rumus CVR digunakan untuk melihat kesesuaian antara butir soal dengan aspek materi pada instrumen telaah soal. Rumus yang digunakan sebagai berikut.

$$CVR = \frac{Ne - N/2}{N/2}$$

$$CVI = \frac{\sum CVR}{\text{Jumlah sub pertanyaan}}$$

Sumber: (Gilbert, 2016)

Keterangan:

CVR : Rasio validitas konten

CVI : *Content validity index*

Ne : Jumlah validator yang menyatakan setuju

N : Total validator

Hasil perhitungan CVR akan dibandingkan dengan nilai kritis CVR Lawshe pada tabel berikut dengan ketentuan:

Bila $CVR_{hitung} > CVR_{tabel}$, maka butir soal valid

Bila $CVR_{hitung} < CVR_{tabel}$, maka butir soal tidak valid.

Kriteria	Keterangan
0-0,33	Tidak Valid
0,34-0,67	Valid
0,68-1	Sangat valid

Tabel kategori CVR dan CVI

Sumber: (Wilson, 2012)

1.2 Uji Validitas Konstruk Butir soal Pilihan ganda Pilihan Ganda

Baik buruknya tes tergantung pada butir-butir soal yang ada di dalamnya. Oleh sebab itu, untuk mendapatkan tes yang baik perlu dipilih butir-butir yang baik, sedangkan butir yang buruk harus dibuang dan yang kurang baik perlu direvisi. Untuk mengetahui kualitas tiap butir soal perlu dianalisis satu persatu. Analisis meliputi perhitungan daya beda, tingkat kesukaran, homogenitas tes serta analisis distraktor/pengecoh pada tes.

a. Daya Beda (DB)

Daya pembeda soal adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pintar (berkemampuan rendah).

Daya beda (*item discrimination*) adalah indeks yang digunakan untuk membedakan peserta tes berkemampuan tinggi (kelompok atas) dan berkemampuan rendah (kelompok bawah) (Toksöz, 2017). Langkah pertama dalam menghitung Daya Beda yaitu menentukan kelompok atas dan kelompok bawah. Rata pembagian kelompok atas 27 % dan kelompok bawah 33%. Perhitungan yang paling baik rata-rata pembagian harus sama yaitu 27%, sisanya 46% dapat di sisihkan (Surapranata, 2006); (Mukherjee, 2015). Rumus dalam menghitung DB masing-masing *item*:

$$DP = \frac{Ba}{na} - \frac{Bb}{nb}$$

Sumber: (Toksöz, 2017); (Surapranata, 2006).

Keterangan :

DP : Indeks daya pembeda

Ba : Jumlah peserta didik yang menjawab soal dengan benar pada kelompok atas

Bb : Jumlah peserta didik yang menjawab soal dengan benar pada kelompok bawah

Na : Jumlah peserta didik kelompok atas

Nb : Jumlah peserta didik kelompok bawah

Proses interpretasi daya pembeda seperti yang disajikan pada tabel 9.

Tabel 1. Kriteria Daya Beda

Interval	Kualitas <i>Item</i>
< 0,20	Buruk
0,20 – 0,40	Cukup
>0,40	Baik

Sumber: (Basuki, 2016)

Kegiatan analisis kualitas tes dan butir soal terdapat manfaat daya pembeda butir soal sebagai berikut ini. 1) Untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak. 2) Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru (Natar, 2011).

Butir soal yang tidak dapat membedakan kedua kemampuan siswa itu, maka butir soal itu dapat dicurigai “kemungkinannya” seperti berikut ini: (a) kunci jawaban butir soal itu tidak tepat; (b) butir soal itu memiliki 2 (dua) atau lebih kunci jawaban yang benar; (c) kompetensi yang diukur tidak jelas; (d) pengecoh tidak berfungsi; (e) materi yang ditanyakan terlalu sulit, sehingga banyak siswa

yang menebak; (f) sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya (Natar, Karjono 2011: 12).

Soal dapat dikatakan mempunyai daya pembeda jika soal tersebut dapat dijawab oleh siswa berkemampuan tinggi dan tidak dapat dijawab oleh siswa berkemampuan rendah. Jika suatu soal dapat dijawab oleh siswa pintar maupun kurang, berarti soal tersebut tidak mempunyai daya beda, demikian juga jika soal tersebut tidak dapat dijawab oleh siswa pintar dan siswa kurang, berarti soal tersebut tidak baik sebab tidak mempunyai daya pembeda Purwanto (2011: 102)

b. Tingkat Kesukaran (TK)

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya. Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (*difficulty index*). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah Arikunto (1996:207). Senada dengan surapranata menyatakan tingkat kesukaran (*Difficulty Index*) suatu butir soal ditentukan kedalaman soal, kompleksitas, atau hal-hal lain yang berkaitan dengan kemampuan yang diukur oleh soal (Surapranata, 2006).

Rumus yang cocok untuk menentukan tingkat kesukaran yakni merujuk banyaknya peserta tes yang dapat menjawab pernyataan yang benar pada suatu soal. Jawaban yang benar diberi nilai 1, sedangkan jawaban yang salah diberi nilai 0. Rumus yang digunakan untuk menghitung tingkat kesukaran yaitu:

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

Sumber: (Surapranata, 2006); (Toksöz, 2017)

Keterangan:

P : tingkat kesukaran perangkat tes

B : Banyaknya peserta tes yang menjawab soal dengan benar

JS : Jumlah seluruh peserta tes

Proses interpretasi tingkat kesukaran seperti yang disajikan pada tabel 10.

Tabel 2. Kriteria Tingkat Kesukaran

Interval	Kualitas Item
< 0,30	Sukar
0,30 - 0,70	Sedang
> 0,70	Mudah

Sumber: (Rao, 2016); (Kolte, 2015)

c. Indeks Pengecoh (*Distractor Efficiency*)

Pola jawaban soal adalah distribusi testee (responden yang sedang mengerjakan tes) dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal diperoleh dengan menghitung banyaknya testee yang memilih pilihan jawaban a, b, c, atau d atau yang tidak memilih pilihan manapun (blangko). Istilah evaluasi disebut Omit (tidak menjawab), disingkat.

Berikut beberapa pertimbangan terhadap analisis pengecoh: 1) diterima, karena sudah baik; 2) ditolak, karena tidak baik; 3) ditulis kembali, karena kurang baik; 4) Sebuah pengecoh dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% pengikut tes. Tes pilihan ganda dengan 5 alternatif jawaban dan $P = 0,8$, dilihat dari segi omitted (O), sebuah butir soal dikatakan baik jika persentase O-nya $\leq 10\%$.²⁹ Pengecoh non fungsional (*Non-Functional Distractors*) adalah opsi yang jarang dipilih (<5%) oleh peserta tes dan pengecoh fungsional (efektif) adalah pilihan yang dipilih oleh 5% atau lebih peserta tes (Mukherjee, 2015). Pengecoh non fungsional seharusnya direvisi, dihapus atau diganti dengan opsi yang lebih masuk akal. Efisiensi pengecoh ditentukan untuk setiap item berdasarkan jumlah NFD di dalamnya dan berkisar dari 0 hingga 100%.

$$pxi = \frac{fxi}{M} \times 100\%$$

Keterangan:

p_{xi} :proporsi masing-masing pilihan jawaban suatu butir tes

f_{xi} :frekuensi masing-masing pilihan jawaban suatu butir tes

M : jumlah responden

Proses interpretasi indeks pengecoh seperti yang disajikan pada tabel berikut.

Tabel 3. Kriteria Indeks Pengecoh

Interval	Tingkat Hubungan
>0,05	Distraktor berfungsi/efektif
<0,05 atau 5%	Distraktor tidak berfungsi

Sumber: (Rao, 2016)

Option biasanya berjumlah tiga atau lima buah, dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir soal itu salah satunya adalah jawaban betul (kunci jawaban) sedangkan sisanya merupakan jawaban salah. Jawaban salah itulah yang biasa dikenal dengan istilah “distractor” (distractor: pengecoh).

Tujuan pemasangan distraktor pada setiap butir item adalah agar dari sekian banyak siswa mengikuti tes ada yang tertarik memilihnya, distraktor yang mereka pilih merupakan jawaban betul. Semakin banyak siswa terkecoh, maka distraktor makin dapat menjalankan fungsinya sebaikbaiknya. Sebaliknya, jika distraktor yang dipasang tidak ada yang memilih, maka distraktor tidak dapat menjalankan fungsinya dengan baik. distraktor dinyatakan telah berfungsi dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih 5% dari seluruh peserta tes.(Sudijono, 2011: 411).

1.3 Validitas Konten Psikomotor

Validitas mengacu pada aspek ketepatan dan kecermatan hasil pengukuran. Pengukuran sendiri dilakukan untuk mengetahui seberapa banyak (dalam arti

kuantitatif), suatu aspek psikologis terdapat dalam diri seseorang, yang dinyatakan oleh skornya pada instrument pengukur yang bersangkutan Azwar (2015).

b. Perhitungan rasio validitas isi dari C H Lawshe

Dalam konteks penelitian ini, rumus CVR digunakan untuk melihat kesesuaian antara butir soal dengan aspek materi pada instrumen telaah soal
Rumus yang digunakan sebagai berikut:

$$CVR = \frac{Ne - N/2}{N/2}$$

$$CVI = \frac{\sum CVR}{\text{Jumlah sub pertanyaan}}$$

Sumber: (Gilbert, 2016)

Keterangan:

CVR : Rasio validitas konten

CVI : *Content validity index*

Ne : Jumlah validator yang menyatakan setuju

N : Total validator

Hasil perhitungan CVR akan dibandingkan dengan nilai kritis CVR Lawshe pada tabel berikut dengan ketentuan:

Bila $CVR_{hitung} > CVR_{tabel}$, maka butir soal valid

Bila $CVR_{hitung} < CVR_{tabel}$, maka butir soal tidak valid.

Tabel kategori CVR dan CVI

Kriteria	Keterangan
0-0,33	Tidak Valid
0,34-0,67	Valid
0,68-1	Sangat valid

Sumber: (Wilson, 2012)

2. Uji Reliabilitas Instrumen

Realibilitas memberikan pengertian bahwa suatu tes dapat dikatakan realibel apabila test tersebut mampu memberikan hasil yang konsisten meskipun tes di berikan, dan di skor oleh penilai yang berbeda, atau diberikan pada waktu yang berlainan atau menggunakan bentuk paralel dari tes tersebut (Anastasi dan Urbina, 1997). Realibilitas atau keandalan adalah kemantapan, konsistensi, prediktabilitas/keteramalan (Kerlinger, 2006).

Uji reliabilitas dimaksudkan untuk mengetahui tingkat konsistensi terhadap instrumen-instrumen yang mengukur konsep. Reliabilitas merupakan syarat untuk tercapainya validitas suatu kuesioner dngan tujuan tertentu. Untuk menguji reliabilitas digunakan Cronbach Alpha dengan bantuan SPSS FOR WINDOWS.

Hair (1998) menyatakan bahwa nilai Cronbach Alpha dapat dikatakan reliable (andal) apabila nilainya $> 0,60$.

Uji reliabilitas dilakukan dengan melihat kondisi internal berdasarkan koefisien Alpha Cronbach's Indikator pengukuran reliabilitas menurut Sekaran (2000) yang membagi tingkatan reliabilitas dengan kriteria sebagai berikut: Jika alpha atau r hitung:

- a. $0,8-1,0$ = Reliabilitas baik
- b. $0,6-0,799$ = Reliabilitas diterima
- c. $< 0,6$ = Reliabilitas kurang baik.