

BAB II

KAJIAN PUSTAKA

A. Kajian Teori

1. Tinjauan tentang Pengukuran Hasil Belajar

a. Pengertian Tes

Menurut Zainal Arifin (2014: 118) tes merupakan teknik atau cara yang dilaksanakan untuk melakukan kegiatan pengukuran, didalamnya terdapat berbagai macam pertanyaan atau bisa juga dalam bentuk tugas yang harus dikerjakan oleh peserta didik yang berfungsi untuk mengukur aspek tingkah laku manusia. Menurut Anas Sudijono (2011: 67) tes merupakan alat atau prosedur yang digunakan dalam rangka mengukur dan menilai di bidang pendidikan yang berbentuk pemberian tugas sehingga dapat dihasilkan nilai yang menggambarkan tingkah laku, di mana nilai dapat dibandingkan dengan nilai standar tertentu. Menurut Suharsimi Arikunto (2013: 3) mengadakan evaluasi berarti meliputi dua kegiatan yaitu penilaian dan pengukuran. Penilaian dalam dunia pendidikan merupakan upaya untuk mengetahui seberapa tinggi tingkat keberhasilan kegiatan pendidikan. Penilaian juga dapat diartikan sebagai suatu proses yang digunakan untuk mengetahui taraf pengetahuan dan keterampilan peserta didik yang hasilnya digunakan untuk keperluan evaluasi atau penilaian (Bambang Subali, 2012: 1). Iriani dan Soeharto (2015: 279) menjelaskan bahwa “evaluasi atau penilain berarti tindakan untuk menentukan nilai sesuatu”. Lebih luasnya yaitu suatu proses yang dilakukan mulai dari merencanakan, memperoleh dan menyediakan

informasi yang sangat diperlukan untuk membuat alternatif-alternatif keputusan. Mengukur pada hakikatnya adalah membandingkan sesuatu dengan dasar ukuran tertentu. Berdasarkan pendapat dari beberapa ahli di atas mengenai evaluasi, dapat ditarik kesimpulan bahwa definisi dari evaluasi adalah suatu rangkaian kegiatan sistematis yang didalamnya terdapat kegiatan mengukur dan menilai hasil belajar peserta didik, di mana hasilnya dapat digunakan untuk mengambil keputusan atau membuat kebijakan. Berdasarkan pendapat ahli, dapat dinyatakan bahwa pengertian dari tes adalah suatu teknik atau cara yang digunakan untuk mengukur tingkah laku peserta didik dengan cara memberikan soal atau tugas kepadanya.

b. Macam-macam Tes

Menurut Daryanto (2012: 36) ditinjau dari segi kegunaan untuk mengukur siswa, maka tes dibedakan menjadi 3 macam yaitu:

- 1) Tes diagnostik adalah tes yang digunakan untuk mengetahui kelemahan-kelemahan peserta didik sehingga dengan mengetahui kelemahan tersebut dapat menentukan perlakuan yang tepat untuk menghadapinya.
- 2) Tes formatif bertujuan untuk mengetahui sejauh mana siswa telah terbentuk perilakunya setelah mengikuti sesuatu program tertentu. Tes formatif juga dapat dipandang seperti tes diagnostik pada akhir pelajaran. Di dalam sekolah, tes formatif dapat juga disamakan dengan ulangan harian.

- 3) Tes sumatif dilaksanakan setelah berakhirnya pemberian sekelompok program atau sebuah program yang lebih besar. Didalam sekolah, tes sumatif dapat disamakan dengan ujian akhir semester.

c. Fungsi Tes

Menurut Sudji Munadi (2009: 163-164) menjelaskan bahwa tingkat kecerdasan seseorang dapat dilihat dan diukur melalui berbagai cara salah satunya dengan tes. Hal inilah yang sering digunakan untuk menggambarkan kecerdasan seseorang. Dengan ini maka salah satu fungsi dari tes adalah untuk mengukur kecerdasan seseorang setelah mendapatkan ilmu dari berbagai hal. Fungsi tes Menurut Anas Sudijono (2012: 67) menyatakan bahwa secara umum ada dua yaitu:

- 1) Sebagai alat ukur terhadap peserta didik. Dalam hal ini tes berfungsi untuk mengukur tingkat perkembangan atau kemajuan peserta didik setelah melaksanakan proses pembelajaran dalam jangka waktu tertentu.
- 2) Sebagai alat ukur keberhasilan program pembelajaran. Dalam hal ini tes berfungsi untuk mengetahui sejauh mana keberhasilan program pembelajaran yang telah ditentukan di awal, sehingga di masa yang akan datang dapat diperbaiki lagi.

Fungsi tes yaitu sebagai motivator dalam belajar. Walaupun nilai yang diperoleh dalam tes hendaknya tidak dijadikan tujuan utama bagi siswa dalam belajar akan tetapi tes dapat digunakan sebagai sarana peningkatan motivasi untuk belajar siswa. Pengalaman menunjukkan bahwa siswa akan belajar lebih giat dan berusaha lebih keras apabila mereka mengetahui bahwa di akhir

program yang sedang ditempuh akan diadakan tes untuk mengetahui nilai dan prestasi mereka (Saifuddin Azwar, 2007: 15)

d. Prinsip Tes

Menurut Anas Sudijono (2012: 97) menyatakan bahwa ada beberapa prinsip dasar yang perlu dicermati dalam menyusun tes hasil belajar yaitu:

- 1) Tes belajar harus bisa mengukur secara jelas hasil belajar (*learning outcomes*) yang telah ditentukan sesuai dengan tujuan instruksional.
- 2) Butir-butir soal harus merupakan sampel yang representatif dari populasi bahan pelajar yang telah diajarkan di dalam kelas.
- 3) Bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat dengan beberapa variasi.
- 4) Tes hasil belajar harus didesain sesuai dengan kegunaannya untuk memperoleh hasil yang maksimal dan sesuai dengan yang diinginkan.
- 5) Tes hasil belajar harus memiliki reliabilitas yang dapat diandalkan.
- 6) Tes hasil belajar yang dijadikan untuk mengukur keberhasilan peserta didik, juga harus dapat dijadikan sebagai alat untuk mencari informasi yang berguna untuk memperbaiki cara belajar siswa dan cara mengajar guru. Sehingga setelah mengetahui hasil tes maka dapat diperbaiki lagi di masa yang akan datang.

e. Kriteria Penyusunan Instrumen yang Baik

Menurut Bradford Strand (1993) karakteristik atau kriteria dari instrumen yang baik adalah sebagai berikut:

- 1) Validitas

Instrumen evaluasi dikatakan valid atau shahih apabila tes tersebut dapat mengukur secermat-cermatnya apa yang hendak diukur.

2) Reliabilitas

Tes dikatakan reliabel atau dapat diandalkan apabila tes tersebut diberikan dua kali berturut-turut kepada kelompok yang sama dalam kondisi yang sama pula maka menghasilkan skor yang sama atau hampir sama.

3) Objektivitas

Sebuah tes dikatakan objektif apabila dua orang atau lebih testor memberikan skor yang sama atau hampir sama kepada testi untuk satu tes. Untuk memperoleh objektivitas dalam memberikan skor perlu disepakati terlebih dahulu kriteria yang digunakan untuk mempertimbangkan kualitas unjuk kerja.

4) Ekonomis dan Praktis

Instrumen dikatakan ekonomis apabila instrumen tidak membutuhkan peralatan yang mahal dalam pembuatannya, atau tidak mahal harganya. Instrumen dikatakan praktis apabila sekali dilakukan dapat mengukur semua yang hendak diukur dan menghasilkan hasil yang baik. Instrumen yang tidak ekonomis dan praktis akan sulit digunakan meskipun instrumen tersebut sudah shahih, reliabel, dan objektif.

5) Norma Penilaian

Norma penilaian adalah satu skala yang memungkinkan konversi dari skor mentah ke skor yang dapat digunakan untuk keperluan membandingkan interpretasi.

6) *Direction* (Petunjuk Pelaksanaan)

Petunjuk pelaksanaan dari tes harus dibakukan, sehingga siapapun yang akan melaksanakannya, maka pelaksanaannya akan sama. Petunjuk pelaksanaan harus dibuat dengan kalimat yang mudah dipahami.

7) *Interest* (Menarik)

Tes diharapkan dapat memperoleh hasil yang maksimal. Tes harus dibuat dengan menarik dan memberikan tantangan bagi peserta didik. Apabila tes membosankan dan tidak menarik maka peserta didik akan enggan untuk mengerjakannya dan juga hasilnya tidak sesuai dengan apa yang hendak diukur.

f. Ciri-ciri Tes yang Baik

Menurut Suharsimi Arikunto (2013: 72) suatu tes dapat dikatakan baik apabila memenuhi lima persyaratan, yaitu:

1) Validitas

Kata valid sering diartikan dengan : tepat, benar, absah dan shahih. Jadi kata validitas ketepatan, kebenaran, keabsahan. Apabila dikaitkan dengan fungsi tes sebagai alat pengukur maka sebuah tes dikatakan valid apabila alat ukur tersebut dapat dengan tepat mengukur apa yang hendak diukur atau diungkap lewat tes tersebut. Jadi tes hasil belajar dapat dinyatakan valid (alat pengukur keberhasilan) dengan secara tepat dapat mengukur atau mengungkap hasil-hasil belajar yang telah dicapai oleh

peserta didik setelah menempuh proses belajar mengajar dalam waktu tertentu

2) Reliabilitas

Kata reliabilitas dari kata *reliability* (Inggris) yang artinya dapat dipercaya. Tes yang reliable jika memberikan hasil yang tetap (*consistent*) apabila diteskan berkali-kali. Jika kepada siswa diberikan tes yang sama yang pada waktu yang berlainan, maka setiap siswa akan tetap berada dalam urutan ranking yang sama tetap (tidak berubah-ubah) dalam kelompoknya. Validitas berhubungan dengan ketepatan sedangkan reliabilitas berhubungan dengan ketetapan atau keajekan.

Sebuah tes dikatakan reliabel apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama hasilnya tetap sama atau sifatnya stabil. Yang dimaksud Stabil disini yaitu tetap berada pada urutan kelompoknya ketika tes dilakukan berulang-ulang meskipun terjadi perubahan nilai secara keseluruhan oleh kelompoknya tetapi pada posisi urutan rankingnya tetap atau berubah tetapi perubahannya tidak berarti. Jadi penekannya bukan pada tetapnya nilai tetapi pada tetapnya posisi urutan nilai atau ranking dalam kelompoknya. Walaupun tampaknya hasil tes pada tes kedua lebih baik karena kenaikannya dialami oleh semua siswa maka tes yang digunakan dapat dikatakan memiliki reliabilitas yang tinggi. Kenaikan hasil yang kedua bisa jadi disebabkan adanya pengalaman yang diperoleh pada waktu mengerjakan tes pertama.

3) Objektivitas

Objektif berarti tidak adanya unsur pribadi yang mempengaruhinya bukan subjektif. Sebuah tes dikatakan memiliki objektivitas apabila dalam melaksanakan tes tidak ada faktor subjektif yang mempengaruhi terutama dalam sistem skornya. Apabila dikaitkan dengan reliabilitas maka objectifitas menekankan ketetapan (*consistency*) pada sistem skoring, sedangkan reliabilitas menekankan ketetapan dalam hasil tes.

Sedangkan Menurut Anas Sudijono (2011: 96) Suatu tes belajar dapat disebut tes belajar yang obyektif apabila tes tersebut disusun dan dilaksanakan menurut apa adanya. Ditinjau dari segi isi atau materinya artinya bahwa materi tes diambilkan atau bersumber dari materi atau bahan pelajaran yang telah diberikan sesuai dengan instruksional khusus yang telah ditentukan atau bahan pelajaran yang telah dipelajari oleh peserta didik yang dijadikan acuan dalam penyusunan hasil belajar tersebut.

4) Praktibilitas (*Practibility*)

Sebuah tes disebut memiliki praktibilitas yang tinggi apabila tes tersebut bersifat praktis, Tes yang praktis adalah tes yang :

- a) Mudah dilaksanakan, tidak menuntut peralatan yang banyak dan memberi kebebasan kepada siswa mengerjakan terlebih dahulu bagian yang dianggap mudah. Karena bersifat sederhana dalam arti tidak memerlukan peralatan yang sulit pengadaannya.

- b) Mudah pemeriksaannya artinya bahwa tes itu dilengkapi kunci jawaban maupun pedoman skoringnya. Dilengkapi dengan petunjuk-petunjuk yang jelas sehingga dapat diberikan atau diawali orang lain.
- 5) Dilengkapi dengan petunjuk-petunjuk yang jelas sehingga dapat diberikan atau diawasi oleh orang lain. Pelaksanaan tes tersebut tidak membutuhkan ongkos atau biaya yang mahal, tenaga yang banyak serta waktu yang lama.

2. Tinjauan tentang Analisis Butir Soal

a. Pengertian Analisis Butir Soal

Analisis soal merupakan kegiatan untuk mengkaji soal pada setiap item atau butirnya guna mengetahui kualitas dari setiap soal tersebut. Menurut Suharsimi Arikunto (2013: 220) analisis butir soal adalah suatu prosedur yang sistematis yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun. Sedangkan menurut Daryanto (2012: 179) analisis soal antara lain bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Menurut Nana Sudjana (2013: 135) analisis butir soal adalah pengkajian pertanyaan-pertanyaan tes agar diperoleh perangkat pertanyaan yang memiliki kualitas memadai.

Berdasarkan paparan beberapa ahli mengenai analisis butir soal, kesimpulannya bahwa analisis butir soal merupakan suatu kegiatan mengkaji dan mengidentifikasi setiap butir soal guna mengetahui kualitas setiap butir soal tersebut. Hasil dari proses mengkaji dan mengidentifikasi soal dapat

digunakan untuk melakukan perbaikan dan/atau penyempurnaan pada setiap butir soal.

b. Manfaat Analisis Butir Soal

Kegiatan analisis pada setiap butir soal tentunya memberikan manfaat bagi penyusun soal. Menurut Kusaeri dan Suprananto (2012 :164) manfaat tersebut antara lain:

- 1) Menentukan soal-soal yang cacat atau tidak berfungsi dengan baik, Meningkatkan butir soal melalui tiga komponen analisis yaitu, tingkat kesukaran, daya pembeda dan pengecoh soal.
- 2) Merevisi soal yang tidak relevan dengan materi yang diajarkan, ditandai dengan banyaknya anak yang tidak dapat menjawab butir soal tertentu.

c. Teknik Analisis Butir Soal

Analisis Kuantitatif

1) Validitas

Validitas berkenaan dengan ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai (Nana Sudjana, 2013: 12). Menurut Zainal Arifin (2014: 247) menyebutkan ada dua unsur penting dalam validitas yaitu validitas menunjukkan suatu derajat dan validitas selalu dihubungkan dengan suatu putusan atau tujuan yang spesifik. Suatu instrumen pengukur dapat dikatakan mempunyai validitas yang tinggi apabila instrumen tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang sesuai dengan maksud dilakukannya pengukuran tersebut.

Menurut Anas Sudijono (2011: 163) terdapat dua macam validitas yaitu validitas tes dan validitas item.

a) Validitas Tes

Validitas tes merupakan pengukuran yang digunakan untuk soal yang akan digunakan secara keseluruhan. Pengukuran validitas tes dapat dilakukan secara rasional dan secara empirik.

(1) Validitas rasional

Anas Sudijono (2011: 164) memaparkan bahwa “validitas rasional adalah validitas yang diperoleh atas dasar hasil pemikiran, validitas yang diperoleh dengan berpikir secara logis”. Validitas rasional untuk sebuah instrumen evaluasi menunjuk pada kondisi sebuah instrumen yang memenuhi persyaratan valid berdasarkan penalaran. Jadi suatu tes dikatakan memiliki validitas rasional apabila setelah dilakukan analisis secara rasional menunjukkan bahwa tes hasil belajar tersebut memang dapat mengukur apa yang seharusnya diukur. Ada dua macam validitas rasional yang dapat dicapai sebuah instrumen yaitu validitas isi dan validitas konstruksi.

(a) Validitas isi

Validitas isi menunjuk pada suatu kondisi sebuah instrumen yang disusun berdasarkan isi materi pelajaran yang dievaluasi. Validitas isi merupakan validitas yang dilihat dari segi isi tes sebagai alat pengukur hasil belajar (Anas Sudijono, 2011: 164). Suatu tes dikatakan memiliki validitas isi apabila tes tersebut dapat mewakili secara representatif bagi seluruh materi pelajaran yang akan diujikan.

(b) Validitas konstruksi

Validitas konstruk/bentuk kerangka yang akan menunjuk pada suatu kondisi sebuah instrumen yang disusun berdasarkan konstruk aspek-aspek kejiwaan yang seharusnya dievaluasi. Konstruk merupakan sesuatu yang berkaitan dengan fenomena dan objek yang abstrak, tetapi gejalanya dapat diamati dan diukur (Sumarna Surapranata, 2005: 53). Validitas tes dikatakan memiliki validitas konstruk apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir. Dengan kata lain jika butir-butir soal mengukur aspek berpikir tersebut sudah sesuai dengan aspek berpikir yang menjadi tujuan instruksional. Suatu instrumen dikatakan memiliki validitas konstruk apabila butir-butir soal yang membangun tes tersebut mengukur di setiap aspek berpikir yang ada di dalam Tujuan Instruksional Khusus (TIK). Sekarang ini Tujuan Instruksional Khusus dikenal sebagai indikator tujuan pembelajaran.

(2) Validitas Empirik

“Validitas empirik adalah validitas yang bersumber pada atau diperoleh atas dasar pengamatan dilapangan” (Anas Sudijono, 2011: 167). Sebuah instrumen dikatakan memiliki validitas empirik apabila sudah diuji dari pengalaman atau pengamatan di lapangan, dan terbukti bahwa tes tersebut dapat mengukur apa yang seharusnya diukur. Validitas empirik dapat dilihat dari dua segi yaitu validitas ramalan dan validitas bandingan.

(a) Validitas ramalan

Validitas prediksi menunjuk pada hubungan antara tes skor yang diperoleh peserta tes dengan keadaan yang akan terjadi diwaktu mendatang (Sumarna Surapranata, 2005: 54). Sebuah tes dikatakan memiliki validitas prediksi apabila mempunyai kemampuan untuk meramalkan suatu soal yang akan terjadi pada masa mendatang. Untuk mengetahui bahwa suatu tes memiliki validitas ramalan dapat dilakukan dengan mencari korelasi antara tes hasil belajar yang sedang diuji dengan hasil belajar setelah peserta tes mengikuti pembelajaran.

(b) Validitas bandingan

Validitas bandingan juga dapat disebut sebagai validitas ada sekarang (Anas Sudijono, 2011:177). Validitas ada sekarang menunjuk pada hubungan antara tes skor yang dicapai dengan keadaan sekarang. Pada keadaan ini, tes dipasangkan dengan hasil pengalaman. Pengalaman selalu mengenai kejadian yang lampau sehingga data pengalaman tersedut sekarang sudah ada. Tes dikatakan memiliki validitas bandingan apabila tes tersebut dalam kurun waktu yang sama secara tepat telah mampu menunjukkan adanya hubungan yang searah antara tes pertama dengan tes yang dilakukan berikutnya.

b) Validitas item

Validitas item merupakan suatu ketepatan mengukur yang dimiliki oleh instrumen tes pada tiap butirnya yang menjadi bagian tak terpisahkan dari tes secara keseluruhan, dalam mengukur apa yang seharusnya diukur oleh tes tersebut (Anas Sudijono, 2011: 182). Rumus yang dapat digunakan

untuk menghitung validitas item bentuk pilihan ganda dapat dihitung dengan menggunakan rumus korelasi point biserial sebagai berikut:

$$y_{pbi} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}}$$

Keterangan :

y_{pbi} = korelasi *point* biserial

Mp = rerata skor dari subjek yang menjawab betul bagi item yang dicarivaliditasnya

Mt = rerata skor total

St = standar deviasi dari skor total

p = proporsi siswa yang menjawab benar

q = proporsi siswa yang menjawab salah

(Suharsimi Arikunto, 2018: 196)

Validitas item untuk soal uraian dapat menggunakan rumus *product moment* 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 :

r_{xy} = Koefisien korelasi antara X dan Y

N = Jumlah peserta tes (*testee*)

Σxy = Total perkalian skor item dan total

Σx = Jumlah skor butir soal

Σy = Jumlah skor total

Σx^2 = Jumlah kuadrat skor butir soal

Σy^2 = Jumlah kuadrat skor total

(Suharsimi Arikunto, 2018: 190)

Mencari validitas item juga bisa menggunakan Microsoft Exel dengan menggunakan rumus correlasi. Software ini sengaja dibuat khusus untuk melakukan perhitungan uji validitas dan reliabilitas dengan uji statistik

biserial point dan *KR-20*.

Menurut Rahayu dan Dzajari (2016: 91) tindak lanjut terhadap hasil analisis validitas butir soal sebagai berikut:

- a) Butir soal yang valid dapat digunakan kembali pada tes berikutnya atau disimpan pada bank soal.
- b) Butir soal yang tidak valid diperbaiki disesuaikan dengan indikator pencapaian maupun disesuaikan dengan teknik penyusunan butir soal.

2) Reliabilitas

Menurut Suharsimi Arikunto (2013: 100) pengertian reliabilitas tes, berhubungan dengan masalah keajegan hasil tes. Didukung dengan pendapat Zainal Arifin (2014: 258) bahwa reliabilitas merupakan tingkat atau derajat konsistensi dari suatu alat pengukur atau instrumen. Berdasarkan kedua pendapat tersebut, maka reliabilitas dapat diartikan sebagai tingkat keajegan hasil pengukuran dengan tes yang sama pada waktu yang berbeda. Sebuah instrumen dikatakan memiliki tingkat reliabilitas yang tinggi apabila dapat dengan tetap memberikan data yang sesuai dengan kenyataan.

Nilai dari reliabilitas diberi lambang “*r*” yang dapat dicari besarnya dengan menggunakan rumus *KR 20* atau *Spearman Brown*. Menurut Anas Sudijono (2011: 209) interpretasi terhadap koefisien reliabilitas tes (*r₁₁*) pada umumnya digunakan patokan sebagai berikut:

- a) Apabila nilai *r₁₁* sama dengan atau lebih besar daripada 0,70 berarti tes hasil belajar yang sedang diuji reliabilitasnya dinyatakan memiliki reliabilitas yang tinggi (=reliable).

- b) Apabila nilai r_{11} lebih kecil daripada 0,70 berarti tes hasil belajar yang sedang diuji reliabilitasnya belum memiliki reliabilitas yang tinggi (=unreliable).

Menurut Suharsimi Arikunto (2018:204-207) memaparkan bahwa hal-hal yang mempengaruhi hasil tes dapat dikelompokkan menjadi 3 sebagai berikut.

- a) Hal yang berhubungan dengan tes itu sendiri, yaitu panjang tes dan kualitas butir-butir soalnya. Tes dengan jumlah yang banyak butirnya tentu saja lebih valid dibandingkan dengan tes yang hanya terdiri dari beberapa butir soal. Semakin panjang tes maka reliabilitasnya semakin tinggi.
- b) Hal yang berhubungan dengan terdoba (*testee*). Suatu tes yang dicobakan kepada peserta didik yang jumlahnya banyak akan mencerminkan keragaman hasil yang menggambarkan besar atau kecilnya reliabilitas tes.
- c) Hal yang berhubungan dengan penyelenggaraan tes. Penyelenggaraan tes yang bersifat administratif sangat menentukan hasil tes. Hasil tes tersebut akan mempengaruhi reliabilitas soal tes.

Menurut Suharsimi Arikunto (2018:207-210) ada tiga macam metode dalam menghitung reliabilitas, yaitu:

- a) Metode Bentuk Paralel (*Equivalent*)

Tes paralel merupakan dua buah tes yang memiliki kesamaan tujuan, tingkat kesukaran dan susunan yang sama, tetapi butir-butir soalnya berbeda. Kelemahan dari metode ini adalah memerlukan waktu yang cukup lama karena harus menyusun dua buah tes dan dicoba sebanyak dua kali juga.

Kelemahan dari metode ini adalah bahwa pengetes pekerjaannya berat karena harus menyusun dua seri tes, dan juga harus tersedia waktu yang lama untuk mengujikan dua kali tes.

b) Metode Tes Ulang

Pada metode ini, tes diujikan sebanyak dua kali pada kelompok peserta didik yang sama dalam waktu yang berbeda dan hanya terdiri dari satu seri tes. Hasil dari kedua tes tersebut dihitung korelasinya untuk mendapatkan nilai reliabilitas.

Pada umumnya, hasil tes yang kedua cenderung lebih baik dari hasil tes pertama. Hal ini tidak mengapa karena pengetes harus sadar akan adanya *practice effect* dan *carry over effect*. Yang penting adalah adanya kesejajaran hasil atau ketetapan hasil yang ditunjukkan oleh koefisien korelasi yang tinggi.

c) Belah Dua

Kelemahan penggunaan metode dua-tes dua kali percobaan dan satu-tes dua kali percobaan diatasi dengan metode ketiga ini, yaitu metode belah dua. Dalam menggunakan metode ini pengetes hanya menggunakan sebuah tes dan dicobakan satu kali.

Berbeda dengan metode pertama dan kedua yang setelah ditemukannya koefisien korelasi langsung ditafsirkan itulah koefisien reliabilitas, dengan metode ketiga ini tidak dapat demikian. Pada waktu membelah dua dan mengkorelasikan dua belahan, baru diketahui reliabilitas separo tes. Untuk mengetahui realibilitas seluruh tes menggunakan rumus sebagai berikut,

(1) Rumus Spearman-Brown

$$r_{11} = \frac{2r_{1/2}^{1/2}}{(1 + r_{1/2}^{1/2})}$$

Keterangan:

r_{11} = koefisien eabilitas yang sudah disesuaikan

$r_{1/2}^{1/2}$ = korelasi antara skor-skor setiap belahan tes

(Suharsimi Arikunto, 2018: 210)

(2) Rumus Flanagan

$$r_{11} = 2 \left(1 - \frac{s_1^2 + s_2^2}{s_t^2} \right)$$

Keterangan:

r_{11} = reliabilitas tes

s_1^2 = varians belahan pertama (varian skor item ganjil)

s_2^2 = varians belahan kedua (varian skor item genap)

s_t^2 = varians skor total

(Suharsimi Arikunto, 2018: 214)

(3) Rumus Rulon

$$r_{11} = 1 - \frac{s_d^2}{s_t^2}$$

Keterangan:

s_d^2 = varians beda (variens *difference*)

D = *difference* yaitu perbedaan antara skor belahan pertama (awal) dengan skor belahan kedua (akhir)

s_t^2 = varians total yaitu varians skor total

(Suharsimi Arikunto, 2018: 216)

(4) Rumus KR 20

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

Keterangan :

r_{11} = reliabilitas tes secara keseluruhan

$\sum pq$ = jumlah hasil perkalian antara p dan q

n = banyaknya item

p = proporsi subjek yang menjawab item dengan benar

q = proporsi subjek yang menjawab item dengan salah ($q = 1 - p$)

S = standard deviasi dari tes (akar varians)

(Suharsimi Arikunto, 2018: 217)

(5) Rumus Hoyt

$$r_{11} = 1 - \frac{V_s}{V_r}$$

Keterangan :

r_{11} = reliabilitas seluruh soal

V_r = varians responden

V_s = varians sisa

(Suharsimi Arikunto, 2018: 220)

3) Tingkat Kesukaran

Menurut Suharsimi Arikunto (2013: 222) soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit. Soal yang terlalu mudah tidak dapat merangsang siswa untuk mengembangkan kemampuannya dalam memecahkan soal tersebut. Soal yang terlalu sulit akan membuat siswa menjadi putus asa untuk mencoba lagi karena di luar kemampuan siswa.

Menurut Zainal Arifin (2017: 266) Tingkat kesukaran soal merupakan pengukuran seberapa besar derajat kesukaran soal. Suatu soal dikatakan baik, apabila memiliki tingkat kesukaran soal yang seimbang (proporsional) dalam artian soal tersebut tidak terlalu mudah atau terlalu sukar. Rumus yang digunakan untuk mencari tingkat kesukaran adalah sebagai berikut:

$$TK = \frac{(WL+WH)}{(nL+nH)} \times 100\%$$

Keterangan :

WL = jumlah peserta didik yang menjawab salah dari kelompok bawah

WH = jumlah peserta didik yang menjawab salah dari kelompok atas

nL = jumlah kelompok bawah

Nh = jumlah kelompok atas

(Zainal Arifin, 2017: 266)

Bilangan yang menunjukkan sulit dan mudahnya suatu soal dinamakan indeks kesukaran. Besarnya indeks kesukaran berkisar antara 0,00 sampai 1,00. Butir soal dengan indeks kesukaran yang mendekati angka 1,00

berarti soal tersebut semakin mudah. Indeks kesukaran dapat diklasifikasikan sebagai berikut:

Soal dengan $P = 0,00$ sampai $0,30$ adalah soal sukar.

Soal dengan $P = 0,31$ sampai $0,70$ adalah soal sedang.

Soal dengan $P = 0,71$ sampai $1,00$ adalah soal mudah.

(Suharsimi Arikunto, 2018: 235)

Kriteria untuk mengklasifikasikan tingkat/kualitas kesukaran soal dalam bentuk pilihan ganda adalah apabila nilai Tingkat Kesukaran antara $0,00-0,30$ dikategorikan terlalu sukar dengan keputusan ditolak/direvisi, $0,30-0,70$ dikategorikan cukup dengan keputusan diteima, dan $0,70-1,0$ dikategorikan terlalu mudah dengan keputusan ditolak/direvisi (Anas Sudijono, 2011: 372).

Menurut Anas Sudijono (2011: 376) tindak lanjut yang perlu dilakukan setelah analisis item tingkat kesukaran adalah sebagai berikut:

- a) Butir item tingkat kesukaran berdasarkan hasil analisis dalam kategori baik (tingkat kesukaran cukup atau sedang), sebaiknya butir item tersebut disimpan dalam bank soal. Selanjutnya butir soal tersebut dapat digunakan lagi pada tes berikutnya.
- b) Butir item tingkat kesukaran berdasarkan analisis dalam kategori terlalu sukar ada 3 kemungkinan tindak lanjut yaitu:
 - (1) Butir item tersebut dibuang dan tidak akan dikeluarkan lagi dalam tes selanjutnya.

(2) Diteliti ulang dan ditelusuri faktor yang menyebabkan butir item soal sulit dijawab oleh peserta didik. Setelah mengetahui faktornya maka dapat dilakukan perbaikan dengan cara menyederhanakan lagi supaya mudah dipahami oleh peserta didik. Setelah diperbaiki maka soal dapat digunakan lagi di tes selanjutnya.

(3) Butir item masih dapat diambil manfaatnya dalam tes-tes yang sifatnya ketat. Dalam arti sebagian besar dari testee tidak akan diluluskan dalam seleksi tersebut.

c) Butir item tingkat kesukaran berdasarkan hasil analisis dalam kategori terlalu mudah, ada 3 kemungkinan tindak lanjut yaitu:

(1) Butir item soal tersebut dibuang dan tidak akan digunakan lagi di tes selanjutnya.

(2) Diteliti ulang dan ditelusuri faktor apa saja yang membuat peserta didik terlalu mudah untuk menjawabnya. Setelah diketahui faktornya, maka dapat dilakukan revisi terhadap butir item soal sehingga dapat digunakan lagi di tes berikutnya.

(3) Butir item yang termasuk dalam kategori mudah ini dapat dimanfaatkan pada tes yang sifatnya longgar. Bisa dikatakan hanya sebagai formalitas saja.

4) Daya Pembeda

Daya pembeda adalah pengukuran sejauhmana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi berdasarkan kriteria

tertentu. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi. Seluruh peserta tes dikelompokkan menjadi 2 yaitu kelompok atas (*upper*) dan kelompok bawah (*lower*). Daya pembeda merupakan kemampuan pada setiap butir soal untuk membedakan antara siswa yang mempunyai kemampuan tinggi dan kemampuan rendah (Daryanto, 2008: 183). Didukung dengan pendapat Zainal Arifin (2014: 273) yang memaparkan bahwa “Perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi berdasarkan kriteria tertentu”. Perhitungan daya pembeda dibedakan antara kelompok kecil dan kelompok besar. Kelompok kecil merupakan kelompok yang terdiri kurang dari 100 (seratus) orang, sebaliknya kelompok besar adalah kelompok yang terdiri lebih dari 100 (seratus) orang. Angka yang menunjukkan besarnya daya pembeda adalah disebut ”*Discriminating Power*” yang diberi lambang D. Besarnya daya pembeda berkisar antara dibawah 0,19 sampai 0,40 keatas. Rumus yang dapat digunakan untuk menghitung daya pembeda pada soal pilihan ganda adalah sebagai berikut.

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Keterangan :

D = daya pembeda

J_A = banyaknya peserta kelompok atas

J_B = banyaknya peserta kelompok bawah

B_A = banyaknya peserta kelompok atas yang jawab benar

B_B = banyaknya peserta kelompok bawah yang jawab salah

P_A = proporsi peserta kelompok atas yang jawab benar

P_B = proporsi peserta kelompok bawah yang jawab benar

(Suharsimi Arikunto, 2018: 238-239)

Rumus dan klasifikasi daya pembeda soal adalah sebagai berikut:

$$DP = \frac{(WL-WH)}{n}$$

Keterangan :

DP = daya pembeda

WL = jumlah peserta didik yang gagal dari kelompok bawah

WH = jumlah peserta yang gagal dari kelompok atas

$n = 27\% \times N$

(Zainal Arifin ,2017: 273)

D = dibawah 0,19 daya pembeda soal adalah jelek

D = 0,20 – 0,29 daya pembeda soal adalah cukup

D = 0,30 – 0,39 daya pembeda soal adalah baik

D = 0,40 keatas daya pembeda soal adalah baik sekali

D = Negatif daya pembeda soal adalah sangat jelek.

(Zainal Arifin ,2017: 273)

5) Efektivitas Pengecoh/Distraktor

Pengecoh merupakan option atau pilihan jawaban yang lain dari jawaban yang benar. Suatu option disebut efektif jika memenuhi fungsinya atau tujuan disajikannya option tersebut tercapai. Hal ini berarti bahwa setiap option yang disajikan masing- masing mempunyai kemungkinan yang sama

untuk dipilih, jika tes menjawab soal itu dengan menerka-nerka (spekulasi). Pengecoh dikatakan dapat menjalankan fungsinya dengan baik apabila pengecoh tersebut dipilih sekurang-kurangnya 5% dari seluruh peserta tes. Pengecoh yang telah menjalankan fungsinya dengan baik dapat digunakan kembali pada tes yang akan datang.

Distraktor yaitu suatu pola yang dapat menggambarkan bagaimana teste menentukan pilihan jawabannya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir item (Anas Sudijono: 2011: 387). Distraktor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih 5% dari seluruh peserta tes.

Menurut Rahmah Zulaiha (2008: 18) tindak lanjut setelah dilakukan analisis keefektifan distraktor sebagai berikut ini:

- a) Distraktor diterima karena sudah baik, artinya semua distraktor pada soal sudah dipilih 2,5% dari peserta tes.
- b) Distraktor ditulis kembali karena kurang baik, artinya distraktor belum menjalankan fungsinya dengan baik (distraktor dipilih kurang dari 2,5%).
- c) Ditolak karena tidak baik, artinya distraktor sama sekali tidak dipilih peserta tes (0%).

Efektivitas pengecoh dapat dihitung dengan rumus sebagai berikut.

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

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 (opsi)

1 = bilangan tetap

(Zainal Arifin, 2014: 279)

Interpretasi terhadap hasil perhitungan setiap pengecoh pada suatu butir soal dapat menggunakan kriteria sebagai berikut.

Sangat baik IP = 76% - 125%

Baik IP = 51% - 75%

Kurang Baik IP = 26% - 50%

Jelek IP = 0% - 25%

(Zainal Arifin, 2014: 280)

B. Penelitian yang Relevan

1. Penelitian yang dilakukan oleh Ruspidu (2014: 60) yang berjudul “Analisis Butir Soal Ujian Akhir Semester Gasal Ekonomi Akuntansi Kelas XI IPS SMA Negeri 11 Yogyakarta Tahun Ajaran 2013/2014”. Dalam penelitian tersebut dapat diketahui hasil penelitian tersebut yaitu: Berdasarkan hasil analisis dari segi validitas, soal yang valid 18 butir (45%) dan soal yang tidak valid 22 butir (55%). Berdasarkan uji reliabilitas, termasuk soal yang memiliki reliabilitas rendah yaitu 0,477. Berdasarkan analisis daya pembeda, yang termasuk soal yang daya pembedanya sangat rendah berjumlah 6 butir (15%), rendah berjumlah 5 butir (12,5%), cukup berjumlah 16 butir (40%), baik berjumlah 11 butir

(27,5%), sangat baik berjumlah 2 butir (5%). Berdasarkan analisis tingkat kesukaran yang termasuk soal yang sukar berjumlah 7 butir (17,5%), sedang berjumlah 13 butir (32,5%) dan mudah berjumlah 20 butir (50%). Berdasarkan analisis pola sebaran jawaban yang termasuk kategori soal yang memiliki pengecoh berfungsi sangat baik berjumlah 4 butir (10%), baik berjumlah 8 butir (20%), cukup berjumlah 10 butir (25%), jelek berjumlah 8 butir (20%), sangat jelek berjumlah 10 butir (25%).

2. Penelitian yang dilakukan oleh Nugrahanti (2013: 42) dengan berjudul “Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Akuntansi Keuangan Kelas XI Kompetensi keahlian Akuntansi SMK Negeri 1 Yogyakarta Tahun Ajaran 2012/2013”. Hasil penelitian tersebut sebagai berikut: Berdasarkan validitasnya, pada soal pilihan ganda terdapat soal yang valid sebesar 70% dan soal yang tidak valid berjumlah 30%. Dalam bentuk soal uraian, semua soal dinyatakan valid. Berdasarkan hasil analisis dari segi reliabilitas, soal tersebut memiliki koefisien reliabilitas yang rendah, pada bentuk soal pilihan ganda 0,610, sedangkan bentuk soal uraian 0,49. Berdasarkan tingkat kesukaran, pada bentuk soal pilihan ganda terdapat soal yang sukar sebesar 10%, soal yang sedang sebesar 53,33%, dan soal yang mudah sebesar 36,67%. Sedangkan, pada bentuk soal uraian yang termasuk dalam tingkat kesukaran berkategori sukar sebesar 50%, kategori sedang 25%, dan kategori mudah sebesar 25%. Berdasarkan daya pembeda, pada bentuk soal pilihan ganda yang termasuk soal dengan daya pembeda jelek sebesar 20%, soal dengan daya pembeda

cukup sebesar 10%, daya pembedanya baik sebesar 10%, dan daya pembeda yang baik sekali sebesar 60%. Bentuk soal uraian yang memiliki daya pembeda yang jelek adalah 75% dan daya pembeda yang cukup adalah 25%. Berdasarkan hasil analisis efektivitas pengecoh, pada soal pilihan ganda yang termasuk soal dengan pengecoh yang berkualitas sangat baik sebesar 33,33%, berkualitas baik sebesar 23,33%, berkualitas cukup baik sebesar 20%, berkualitas kurang baik sebesar 16,67%, dan berkualitas tidak baik sebesar 6,67%.

3. Penelitian yang dilakukan oleh Oktanin dan Sukirno (2014: 35) yang berjudul “Analisis Butir Soal Ujian Akhir Semester Genap Mata Pelajaran Ekonomi Akuntansi Kelas XI IPS SMA N 1 Kalasan Tahun Ajaran 2013/2014”. Hasil penelitian menunjukkan bahwa: Dilihat dari segi validitas, terdapat 26 butir soal (52%) dinyatakan valid, 24 butir soal (48%) tidak valid. Dilihat dari segi reliabilitas, keseluruhan butir soal memiliki kategori reliabilitas yang tinggi dengan indeks 0,727. Dilihat dari segi daya pembeda, butir soal dengan daya pembeda jelek berjumlah 33 butir (66%), cukup berjumlah 11 butir (22%), baik berjumlah 3 butir (6%), tidak baik berjumlah 3 butir (6%). Dilihat dari tingkat kesukarannya, butir soal yang termasuk dalam kategori sukar berjumlah 5 butir (10%), sedang berjumlah 15 butir (30%), dan mudah berjumlah 30 butir (60%). Dilihat dari segi efektivitas pengecoh, soal dengan pengecoh yang berfungsi sangat baik berjumlah 1 butir (2%), pengecoh yang berfungsi baik berjumlah 7 butir (14%), berfungsi cukup berjumlah 15 butir (30%),

berfungsi kurang baik berjumlah 14 butir (28%), dan tidak berfungsi berjumlah 13 butir (26%). Hasil analisis soal secara keseluruhan terdapat 3 butir (6%) termasuk soal berkualitas sangat baik, 9 butir (18%) termasuk soal berkualitas baik, 9 butir (18%) termasuk soal berkualitas sedang, 21 butir (42%) termasuk soal berkualitas tidak baik, dan 8 butir (16%) termasuk soal berkualitas sangat tidak baik.

Berdasarkan penelitian yang telah dilakukan maka terdapat persamaan dan perbedaan dengan penelitian yang akan dilakukan. Persamaan penelitian yang dilakukan oleh ketiga peneliti tersebut dengan penelitian yang akan dilakukan adalah sama-sama penelitian deskriptif kuantitatif dan sama-sama meneliti tentang analisis butir soal. Tempat penelitian yaitu di SMK Muhammadiyah Gamping, waktu penelitian, subjek penelitian, dan program yang digunakan untuk mengolah data.

C. Kerangka Berpikir

Evaluasi adalah kegiatan yang dilakukan untuk mengetahui ketercapaian program yang sudah dilaksanakan. Terkait dengan pembelajaran, evaluasi bertujuan untuk faktor-faktor apa saja yang mempengaruhi keberhasilan dan kegagalan dalam proses pembelajaran. Selain itu, evaluasi juga digunakan untuk mengetahui seberapa jauh kemampuan peserta didik dalam menguasai materi pelajaran yang telah dilaksanakan. Salah satu cara untuk mengevaluasi peserta didik adalah dengan melakukan tes. Tes merupakan alat yang digunakan untuk mengukur kemampuan peserta didik.

Penilaian Akhir Tahun (PAT) penting dilakukan karena untuk melihat sejauh mana siswa sudah memahami materi yang telah diberikan oleh guru. Pentingnya PAT ini menuntut seorang guru untuk membuat tes dengan kualitas soal yang baik. Kualitas soal dapat diketahui dengan cara menganalisis soal tersebut. Analisis kualitas soal dilakukan untuk mengkaji tingkat kualitas dari soal tersebut apakah sudah memenuhi persyaratan sebagai tes yang baik. Analisis kualitas soal PAT dilakukan setelah PAT selesai dilaksanakan. Analisis yang dilakukan yaitu analisis kuantitatif. Analisis soal yang dilakukan secara kuantitatif meliputi validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan distraktor. Didalam melakukan analisis soal pada penelitian yang dilakukan menggunakan program komputer Microsoft Exel dan ITEMAN 4.3.

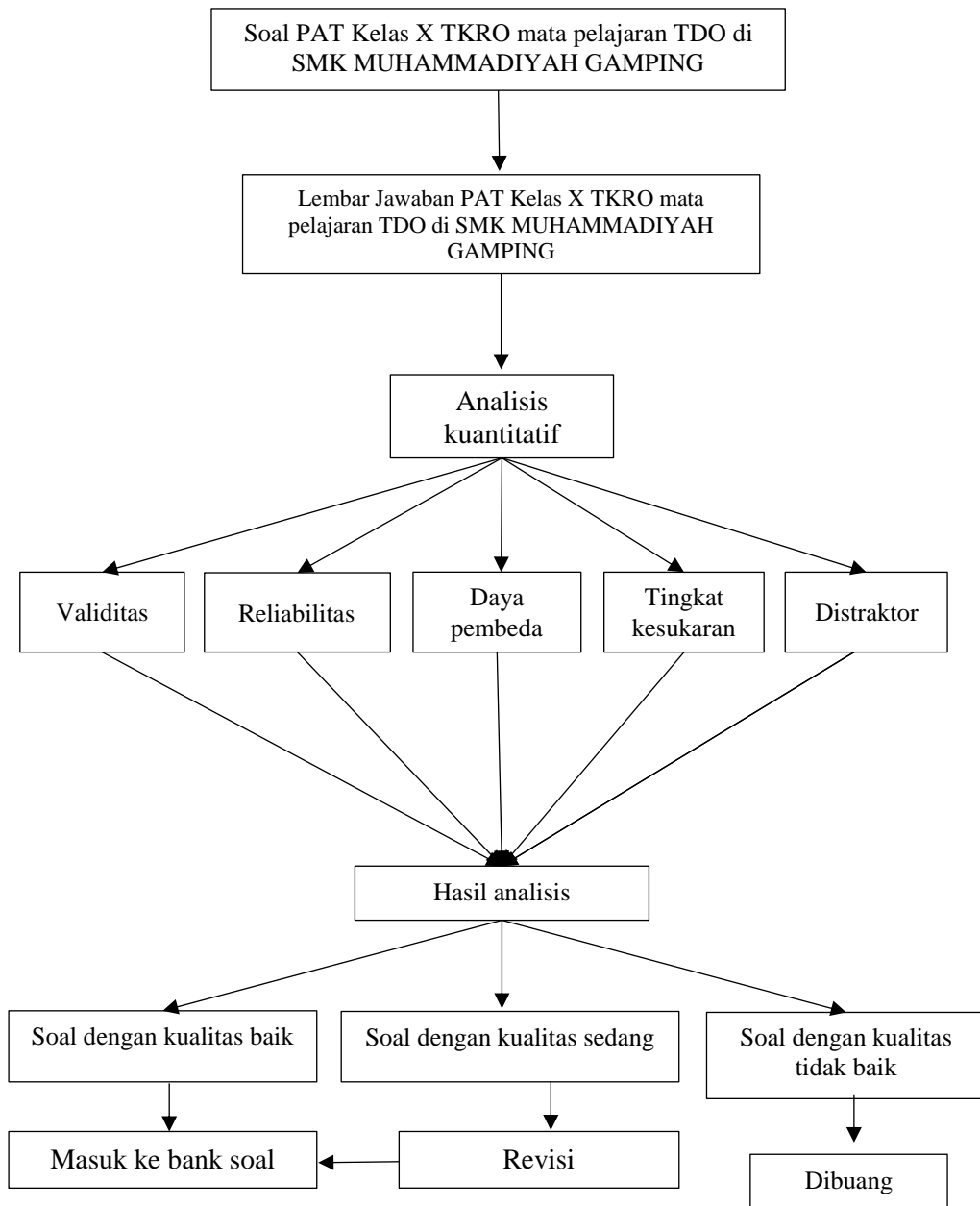
Validitas merupakan tingkat ketepatan atau keshahihan dari suatu tes. Sebuah tes dikatakan valid apabila dapat mengukur apa yang hendak diukur. Reliabilitas adalah keajegan dari tes tersebut. Sebuah tes dikatakan reliabel apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama hasilnya tetap sama atau sifatnya stabil. Daya pembeda merupakan pengukuran yang digunakan untuk mengetahui siswa yang sudah menguasai materi dan yang belum menguasai materi.

Tingkat kesukaran adalah analisis yang dilakukan untuk mengetahui golongan dari soal tersebut, apakah termasuk kedalam golongan yang sulit, sedang, atau mudah. Distraktor atau pengecoh merupakan opsi jawaban lain

yang hampir sama dengan jawaban yang benar. Pengecoh banyak ditemukan pada salah satu bentuk tes yaitu pilihan ganda. Jawaban dari peserta didik dapat digunakan sebagai bahan acuan untuk mengetahui apakah distraktor sudah berfungsi dengan baik atau belum.

Hasil dari penelitian ini adalah akan didapatkan soal-soal yang dibagi menjadi tiga kategori yaitu: baik, sedang, dan tidak baik. soal dengan kualitas baik maka dapat dimasukkan kedalam bank soal, sehingga dapat digunakan lagi di masa yang akan datang.

D. Paradigma Penelitian



Gambar 1. Skema Kerangka berpikir

E. Pertanyaan Penelitian

1. Bagaimanakah kualitas soal Penilaian Akhir Tahun pada Mata Pelajaran Teknologi Dasar Otomotif Kelas X TKRO di SMK Muhammadiyah Gamping Tahun Ajaran 2018/2019 dilihat dari segi validitas?
2. Bagaimanakah kualitas soal Penilaian Akhir Tahun pada Mata Pelajaran Teknologi Dasar Otomotif Kelas X TKRO di SMK Muhammadiyah Gamping Tahun Ajaran 2018/2019 dilihat dari segi reliabilitas?
3. Bagaimanakah kualitas soal Penilaian Akhir Tahun pada Mata Pelajaran Teknologi Dasar Otomotif Kelas X TKRO di SMK Muhammadiyah Gamping Tahun Ajaran 2018/2019 dilihat dari segi daya pembeda?
4. Bagaimanakah kualitas soal Penilaian Akhir Tahun pada Mata Pelajaran Teknologi Dasar Otomotif Kelas X TKRO di SMK Muhammadiyah Gamping Tahun Ajaran 2018/2019 dilihat dari segi tingkat kesukaran?
5. Bagaimanakah kualitas soal Penilaian Akhir Tahun pada Mata Pelajaran Teknologi Dasar Otomotif Kelas X TKRO di SMK Muhammadiyah Gamping Tahun Ajaran 2018/2019 dilihat dari segi efektivitas pengecoh?