

## BAB II

### LANDASAN TEORI

#### A. Penelitian Relevan

Penelitian tentang evaluasi di bidang pendidikan telah banyak dilakukan oleh beberapa penulis. Penelitian ini juga didasarkan pada penelitian yang telah dilakukan sebelumnya. Ada beberapa hasil penelitian yang mendukung dan dapat dijadikan referensi dalam penelitian ini, yaitu sebagai berikut.

**1. Penelitian yang dilakukan oleh Reni Setianingsih (IAIN Purwokerto) pada tahun 2018 yang berjudul *Analisis Kualitas Butir Soal Ujian Akhir Semester 2 Mata Pelajaran Sejarah Kebudayaan Islam Kelas VIII Di MTs Ma'arif Nu Kemranjen Kabupaten Banyumas Tahun Ajaran 2016/2017.***

Penelitian tersebut dapat diketahui hasil penelitian 50 butir soal ujian akhir semester 2 mata pelajaran Sejarah Kebudayaan Islam yang diikuti oleh 191 siswa di 3 MTs Ma'arif NU Kemranjen Kabupaten Banyumas belum baik. Berdasarkan analisis yang dilakukan dari Tingkat kesukaran, daya pembeda dan fungsi distraktor dapat diketahui: tingkat kesukaran meliputi kategori sangat sukar 2 soal, sukar 5 soal, sedang 33 soal, mudah 5 soal, dan sangat mudah 5 soal. Daya pembeda meliputi: sangat baik 4 soal, baik 22 soal, cukup 12 soal, jelek 10 soal, dan sangat jelek 2 soal. Dilihat dari fungsi pengecoh ditemukan Fungsi Distraktor dengan kriteria berfungsi sebanyak 33. Terdapat 13 butir soal yang hanya dua pengecohnya berfungsi dengan baik. Terdapat 2 butir soal yang hanya satu pengecohnya berfungsi. Terdapat 2 butir soal yang fungsi distraktornya tidak berfungsi.

Persamaan dengan penelitian yang dilakukan oleh penulis yaitu sama- sama menganalisis tentang kualitas butir soal. Sedangkan perbedaannya terletak pada teknik

pengumpulan data, skripsi yang ditulis oleh Reni Setianingsih menggunakan teknik observasi, wawancara, dan dokumentasi. Sedangkan teknik pengumpulan data yang dilakukan penulis hanya teknik dokumentasi. Soal yang dianalisis oleh Reni Setianingsih yaitu soal Ujian Akhir Semester 2 mata pelajaran Sejarah Kebudayaan Islam tingkat MTs. Sedangkan penulis menganalisis soal Penilaian Akhir Tahun semester genap mata pelajaran Bahasa Indonesia tingkat SMA serta sistem yang digunakan pada adalah sistem daring.

**2. Penelitian yang dilakukan oleh Andi Surahma Halik (UIN Alauddin Makassar) pada tahun 2017 yang berjudul *Analisis Butir Soal Ujian Akhir Sekolah (UAS) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/2016 SMP Negeri 36 Makassar.***

Hasil penelitian tersebut diantaranya yaitu : (1) Berdasarkan kriteria tingkat kesukaran soal diperoleh soal sangat mudah 2 butir (5%), sedang 24 butir (60%), sukar 12 butir (30%), dan sangat sukar 2 butir (5%), (2) Berdasarkan kriteria daya beda soal diperoleh soal dengan daya beda sangat jelek 3 butir (7,5%), jelek 12 butir (30%), cukup 14 butir (35%), dan baik 11 butir (27,5%) dan (3) Berdasarkan keefektifan pengecoh diperoleh 28 butir (70%) dalam kategori sangat baik, 9 butir (22,5%) dalam kategori baik, dan 3 butir (7,5%) dalam kategori buruk. Persamaan dengan penelitian yang dilakukan oleh penulis yaitu sama-sama menganalisis tentang kualitas butir soal. teknik pengumpulan data yaitu dokumentasi dan dianalisis dengan bantuan aplikasi Anates V4. Sedangkan perbedaannya terletak pada soal yang dianalisis, skripsi yang dilakukan oleh Andi Surahma Malik menganalisis soal mata pelajaran Matematika tingkat SMP. Sedangkan penulis menganalisis soal Penilaian Akhir Tahun semester genap mata pelajaran bahasa Indonesia tingkat SMA serta

berbasis daring. Analisis yang dilakukan oleh Andi Surahma Malik hanya dari segi tingkat kesukaran, daya beda dan efektivitas pengecoh. Sedangkan penulis menganalisis dari segi validitas, reliabilitas, daya pembeda, tingkat kesukaran dan efektivitas pengecoh.

## **B. Landasan Teori**

Landasan teori dalam penelitian ini menyajikan berbagai teori yang digunakan sebagai acuan pelaksanaan penelitian yang berisi; evaluasi pendidikan, kurikulum 2013, tes, karakteristik soal objektif, analisis butir soal, dan kerangka teoretis penelitian.

### **1. Evaluasi Pendidikan**

Evaluasi berasal dari bahasa Inggris *evaluation*, yang berarti penilaian. Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional Bab I Pasal 1 Ayat 21 menyatakan bahwa evaluasi pendidikan adalah kegiatan pengendalian, penjaminan, dan penetapan mutu pendidikan terhadap komponen pendidikan pada setiap jalur, jenjang, dan jenis pendidikan sebagai bentuk pertanggungjawaban penyelenggaraan pendidikan. Selanjutnya dalam Bab XVI Pasal 57 Ayat 1, dinyatakan bahwa evaluasi dilakukan dalam rangka pengendalian mutu pendidikan secara nasional sebagai bentuk akuntabilitas penyelenggara pendidikan pada pihak-pihak yang berkepentingan. Kemudian dalam Bab XVI Pasal 58 Ayat 1, menyatakan bahwa evaluasi hasil belajar peserta didik dilakukan oleh pendidik untuk

memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan.

Menurut Arikunto (2013 : 3) evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagian mana tujuan pendidikan sudah tercapai. Jika belum, bagaimana yang belum dan apa sebabnya. Terdapat tiga istilah dalam evaluasi pendidikan, yaitu; evaluasi, pengukuran, dan penilaian. Hal serupa juga dikemukakan oleh Solichin (2017 : 193) bahwa kegiatan evaluasi merupakan suatu komponen yang bersifat integral dalam program pembelajaran disamping kurikulum, tujuan, bentuk, metode, dan media pembelajaran dengan tujuan untuk mendapatkan informasi yang akurat mengenai tingkat pencapaian tujuan pembelajaran oleh peserta didik agar dapat diupayakan tindak lanjutnya. Senada dengan pendapat Arifin (2012 : 8) evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan arti) daripada sesuatu, berdasarkan pertimbangan dan kriteria tertentu untuk membuat suatu keputusan.

Berdasarkan pendapat yang telah dikemukakan, maka dapat disimpulkan bahwa evaluasi merupakan suatu proses yang dilakukan untuk menentukan sejauh manakah tujuan pendidikan, sudah tercapai atau belum, dan untuk menentukan tindak lanjutnya. Oleh karena itu, perencanaan evaluasi harus ditetapkan pada waktu menyusun satuan pembelajaran, agar dapat disesuaikan dengan tujuan intruksionalnya dan hasilnya dapat sesuai dengan yang diharapkan. Kaitannya dengan evaluasi, terdapat tiga istilah dalam evaluasi pembelajaran diantaranya:

### a. Penilaian

Istilah penilaian merupakan alih bahasa dari istilah *assessment*, bukan dari istilah *evaluation*. Dalam proses pembelajaran, penilaian sering dilakukan guru untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai oleh peserta didik. Artinya, penilaian tidak hanya ditujukan pada penguasaan salah satu bidang tertentu saja, tetapi bersifat menyeluruh yang mencakup aspek pengetahuan, keterampilan, sikap dan nilai-nilai.

Penilaian berhubungan dengan pengambilan keputusan. Pengambilan keputusan didasarkan pada nilai skor pengukuran hasil belajar menjadi bermakna dan dapat digunakan untuk mengambil keputusan setelah diubah menjadi nilai. Skor hanyalah bilangan yang secara arbitrer diterapkan kepada objek pada sifat tertentu yang diukur menggunakan sistem aturan tertentu. Bilangan skor hasil belajar itu belum mempunyai makna apapun dalam pengambilan keputusan pendidikan sebelum diubah menjadi nilai (Purwanto, 2018 : 204).

Arikunto (2013 : 18-19) mengemukakan empat tujuan dan fungsi penilaian, yaitu: (1) Penilaian berfungsi selektif, dilakukan dengan mengadakan seleksi terhadap peserta didik. Misalnya untuk memilih peserta didik yang dapat diterima di sekolah tertentu. (2) Penilaian bersifat diagnostik, apabila alat yang digunakan dalam penilaian cukup memenuhi persyaratan, sehingga guru dapat melihat kelemahan peserta didik berdasarkan hasil yang diperoleh. (3) Penilaian berfungsi sebagai penempatan, misalnya untuk dapat menentukan dengan pasti di kelompok mana peserta didik harus ditempatkan. (4) Penilaian berfungsi sebagai pengukur keberhasilan, untuk mengetahui sejauh mana suatu program berhasil diterapkan.

### 1) Tujuan Penilaian

Tujuan penilaian hasil belajar peserta didik menurut Kunandar (2013:70) adalah sebagai berikut:

#### a) Melacak kemajuan peserta didik.

Melakukan penilaian akan menghasilkan perkembangan hasil peserta didik. Perkembangan hasil peserta didik dapat diidentifikasi yaitu meningkat atau menurun. Guru bisa membuat data kemajuan peserta didik yang berisi pencapaian hasil belajar peserta didik secara periodik.

#### b) Mengecek ketercapaian kompetensi peserta didik.

Melakukan penilaian akan dapat mengetahui apakah peserta didik telah menguasai kompetensi tersebut atau belum. Apabila masih ada peserta didik yang belum menguasai kompetensi tertentu maka akan dicarikan tindakan seperti remedial. Hal ini bertujuan untuk membantu peserta didik mencapai kompetensi tertentu.

#### c) Mendeteksi kompetensi yang belum dikuasai oleh peserta didik.

Adanya penilaian maka dapat mendeteksi kompetensi-kompetensi yang belum dikuasai oleh peserta didik. Hal ini ditujukan untuk memperoleh data tentang kompetensi yang belum dikuasai oleh peserta didik. Sehingga guru dapat mengulang kembali kompetensi tertentu yang belum dikuasai peserta didik.

#### 4) Menjadi umpan balik untuk perbaikan bagi peserta didik.

Adanya penilaian maka guru dapat memperoleh bahan acuan untuk memperbaiki hasil belajar peserta didik yang masih dibawah kriteria ketuntasan minimum (KKM).

## 2) Prinsip dan Prosedur Penilaian

Pentingnya penilaian hasil belajar peserta didik dalam menentukan kualitas pendidikan, maka upaya merencanakan dan melaksanakan penilaian hendaknya memperhatikan prinsip dan prosedur penilaian. Sudjana (2012:8) mengemukakan prinsip dan prosedur penilaian hasil belajar sebagai berikut:

- a) Menilai hasil belajar hendaknya dirancang sedemikian rupa, sehingga jelas abilitas yang harus dinilai, materi penilaian, alat penilaian, interpretasi penilaian. Patokan atau rambu-rambu dalam merancang penilaian hasil belajar adalah kurikulum yang berlaku dan buku pelajaran yang digunakan. Kurikulum dan buku yang digunakan hendaknya dipelajari tujuannya, pokok bahasan yang diberikan, ruang lingkup dan urutan penyajian, serta pedoman pelaksanaannya.
- b) Penilaian hasil belajar hendaknya menjadi bagian integral dari proses belajar mengajar. Artinya, penilaian senantiasa dilaksanakan pada setiap saat proses belajar mengajar sehingga pelaksanaannya berkesinambungan.
- c) Penilaian harus menggunakan berbagai alat penilaian dan sifatnya komprehensif. Sifat komprehensif dimaksudkan segi atau abilitas yang dinilai tidak hanya aspek kognitif, afektif dan aspek psikomotor, melainkan semua aspek digunakan dalam penilaian. Hal ini dilakukan untuk menilai peserta didik secara objektif.
- d) Penilaian hasil belajar hendaknya diikuti dengan tindakan lanjutan. Setelah melakukan penilaian guru sudah memiliki hasil belajar peserta didik. Oleh karena itu guru harus melakukan kegiatan lanjutan yaitu mengelompokkan peserta didik yang sudah mencapai kompetensi dan peserta didik yang belum mencapai kompetensi. Peserta didik yang sudah mencapai kompetensi bisa diberi tugas

pengayaan untuk memperdalam materi dan peserta didik yang belum mencapai kompetensi dapat mengikuti remedial.

Dari pemaparan prinsip dan prosedur penilaian hasil belajar tersebut dapat disimpulkan bahwa merencanakan dan melaksanakan penilaian harus berpedoman pada kurikulum yang berlaku dan dilakukan secara berkesinambungan. Penilaian juga harus bersifat objektif kepada peserta didik serta guru harus memiliki tindakan lanjutan setelah melakukan penilaian hasil belajar.

### 3) Instrumen Penilaian

Trianto (2014:129) mengungkapkan bahwa instrumen merupakan alat yang dapat digunakan untuk mengukur tingkat ketercapaian kompetensi. Bentuk instrumen merupakan alat yang digunakan dalam melakukan penilaian/pengukuran/evaluasi terhadap pencapaian kompetensi peserta didik dalam bentuk tes maupun non tes. Penilaian menurut Trianto (2014:123) yaitu serangkaian kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang hasil belajar peserta didik yang dilakukan secara sistematis dan berkesinambungan sehingga menjadi informasi yang bermakna dalam pengambilan keputusan. Sedangkan Abidin (2014:66) mengungkapkan bahwa penilaian merupakan bagian dari kegiatan evaluasi yang berfokus pada dimensi pembelajaran yang di dalamnya terkandung juga istilah tes dan pengukuran. Dari beberapa pendapat tersebut dapat disimpulkan instrumen penilaian yaitu alat evaluasi yang digunakan untuk mengukur tingkat ketercapaian kompetensi peserta didik dalam bentuk tes maupun non tes yang dilakukan secara sistematis dan berkesinambungan.

## **b. Pengukuran**

Pengukuran adalah suatu proses atau kegiatan untuk menentukan kuantitas daripada sesuatu. Kata “sesuatu” bisa berarti peserta didik, guru, gedung sekolah, meja belajar, dan sebagainya. Dalam proses pengukuran, tentu guru harus menggunakan alat ukur (tes atau non-tes). Alat ukur tersebut harus standar, yaitu memiliki derajat validitas dan reliabilitas yang tinggi.

Pengukuran dilakukan untuk mendapatkan data yang objektif. Objektivitas dapat dicapai karena pengumpul data mengambil jarak dengan objek yang diukur dan menyerahkan wewenang pengukuran kepada alat ukur. Penderahan kewenangan pengukuran kepada alat ukur menyebabkan pengumpul data tidak lagi menyertakan subjektivitasnya ke dalam hasil ukur dan diperoleh data yang objektif (Purwanto, 2018 : 2). Dalam pengumpulan data pendidikan, pengukuran juga dilakukan untuk memperoleh data yang objektif. Dalam pengumpulan data hasil belajar misalnya, pengukuran dilakukan atas peserta didik menggunakan tes hasil belajar sebagai alat ukur.

## **c. Tes**

Tes merupakan alat ukur yang paling sering digunakan oleh guru untuk mengukur hasil belajar peserta didik. Tes pada hakikatnya adalah suatu alat yang berisi serangkaian tugas yang harus dikerjakan atau soal-soal yang harus dijawab oleh peserta didik untuk mengukur suatu aspek perilaku tertentu. Artinya, fungsi tes adalah sebagai alat ukur. Dalam tes prestasi belajar aspek perilaku yang hendak diukur adalah tingkat kemampuan peserta didik dalam menguasai materi pelajaran yang telah

disampaikan. Guru dapat menentukan sejauh mana tujuan pembelajaran telah tercapai dari hasil tes yang telah didapatkan.

Menurut Arikunto (2019 : 37), tes merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan cara dan aturan-aturan yang sudah ditentukan. Istilah tes diambil dari kata *testum* (berasal dari bahasa Prancis kuno) yang berarti piring untuk menyisihkan logam-logam mulia. Artinya, piring tersebut dapat digunakan sebagai alat untuk memperoleh sesuatu yang bernilai tinggi. Pada umumnya bentuk tes berupa pilihan ganda dan uraian. Tes pilihan ganda merupakan tes dimana setiap butir soalnya menyediakan pilihan jawaban lebih dari satu dan salah satu opsinya merupakan jawaban yang benar, sedangkan opsi lainnya berfungsi sebagai pengecoh. Tes uraian adalah tes tertulis dengan pertanyaan yang menuntut peserta didik untuk menjawab dalam bentuk uraian.

Sudijono (2012 : 68-73) mengemukakan bahwa sebagai alat pengukur perkembangan belajar peserta didik, tes dibedakan menjadi enam yaitu : (1) Tes seleksi, digunakan untuk memilih calon peserta didik baru yang tergolong paling baik dari sekian banyak calon yang mengikuti tes; (2) Tes awal, digunakan untuk menentukan sejauh mana peserta didik menguasai materi atau bahan pelajaran yang telah diajarkan dan dilaksanakan sebelum bahan pelajaran diberikan kepada peserta didik; (3) Tes akhir, digunakan untuk menentukan sejauh mana materi pelajaran dikuasai oleh peserta didik sebaik-baiknya; (4) Tes diagnostik, digunakan untuk mengidentifikasi peserta didik yang mengalami masalah dan menelusuri jenis masalah yang dihadapi; (5) Tes formatik, digunakan untuk menentukan sejauh mana peserta didik “telah terbentuk” (sesuai dengan tujuan pengajaran yang telah ditentukan)

setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu dan biasanya dilakukan ditengah-tengah perjalanan program pengajaran; (6) Tes sumatif, digunakan untuk menentukan penguasaan peserta didik terhadap semua materi pembelajaran yang telah disampaikan dan untuk menentukan nilai peserta didik setelah menempuh proses pembelajaran dalam satu semester.

Berdasarkan uraian tersebut jika dilihat dari fungsinya, maka tes sumatif memiliki fungsi yang signifikan dalam pembelajaran. Tes sumatif di sekolah dikenal dengan istilah Penilaian Akhir Tahun (PAT). Hasil PAT digunakan untuk mengisi nilai rapot. Pada umumnya PAT berisi pertanyaan tentang materi pelajaran yang telah diperoleh selama satu semester. Tujuan utama PAT adalah untuk menentukan nilai yang melambangkan keberhasilan peserta didik setelah menempuh proses pembelajaran selama satu semester. Nilai PAT sangat menentukan peserta didik dalam mengikuti program pembelajaran berikutnya. Berdasarkan kajian tersebut dapat disimpulkan bahwa tes bentuk Penilaian Akhir Tahun (PAT) merupakan sebuah alat yang digunakan untuk mengetahui kompetensi lulusan peserta didik.

Menurut Arikunto (2013: 72) sebuah tes dikatakan baik sebagai alat pengukuran apabila memenuhi persyaratan sebagai berikut:

#### 1) Validitas

Sebuah data dapat dikatakan valid apabila sesuai dengan keadaan senyatanya. Data yang dihasilkan oleh instrumen yang benar, valid dan sesuai dengan kenyataannya maka instrumen yang digunakan juga valid. Sebuah tes dikatakan valid atau tepat apabila tes tersebut dapat mengukur apa yang hendak diukur.

#### 2) Reliabilitas

Tes dikatakan reliabel apabila memberikan hasil yang tetap walaupun diteskan berkali-kali. Dengan kata lain, yang dimaksud dari reliabilitas adalah ketetapan dari sebuah instrumen.

### 3) Objektivitas

Sebuah tes memiliki objektivitas apabila dalam melaksanakan tes tersebut tidak ada faktor subjektif yang memengaruhi.

### 4) Praktikabilitas

Sebuah tes dikatakan memiliki praktikabilitas yang tinggi apabila tes tersebut bersifat praktis. Tes yang praktis adalah tes yang mudah dalam pelaksanaannya, mudah pemeriksaannya, mudah dalam pengadministrasiannya, dan dilengkapi petunjuk yang jelas sehingga dapat diberikan oleh orang lain.

### 5) Ekonomis

Sebuah alat atau instrumen dikatakan ekonomis apabila dalam pelaksanaannya tidak memerlukan biaya yang mahal, waktu dan tenaga yang banyak.

## **2. Penilaian Akhir Tahun (PAT) Berbasis Daring**

### **a. Pengertian Penilaian Akhir Tahun (PAT) Berbasis Daring**

Penilaian Akhir Tahun adalah penilaian yang dilaksanakan pada akhir semester genap dengan materi semua Kompetensi Dasar (KD) dalam satu semester. Istilah Penilaian Akhir Tahun (PAT) merupakan pengganti dari Ulangan Kenaikan Kelas (UKK) yang dilaksanakan pada akhir semester genap. Perubahan istilah tersebut karena adanya Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 3 Tahun 2013 tentang Standar Penilaian Pendidikan, bahwa penilaian hasil

belajar oleh satuan pendidikan yang dilakukan dalam bentuk Penilaian Akhir Semester (PAS), Penilaian Akhir Tahun (PAT) dan Ujian Sekolah (US).

Daring merupakan akronim dari kata “dalam jaringan”. Dalam Kamus Besar Bahasa Indonesia (KBBI) istilah daring memiliki arti terhubung melalui jejaring komputer, internet dan sebagainya. Istilah tersebut memiliki arti yang sama dengan *online*, aktivitas daring terjadi ketika perangkat elektronik seperti komputer, laptop, maupun gawai terhubung ke internet. Dengan internet, informa dapat diterima dan diakses dalam format dari seluruh penjuru dunia. Kehadiran internet juga dapat memberikan kemudahan dalam dunia pendidikan, hal ini terlihat dengan begitu banyaknya situs web yang menyediakan media pembelajaran yang semakin interaktif serta mudah untuk dipelajari.

Maka dapat disimpulkan bahwa Penilaian akhir Tahun berbasis daring merupakan penilaian yang dilaksanakan pada akhir semester genap dengan materi selama satu semester yang memanfaatkan jaringan internet sebagai media pembelajarannya.

#### **b. Karakteristik Soal Penilaian Akhir Tahun (PAT)**

Soal merupakan suatu alat yang dibuat untuk mengukur kemampuan ranah koognitif peserta didik. Soal Penilaian Akhir Tahun pada tahun ajaran 2019/2020 hanya terdiri dari soal pilihan ganda yang berjumlah 30 butir soal. Materi soal pilihan ganda dalam Soal Penilaian Akhir Tahun adalah materi yang telah diajarkan pada semester genap, tetapi dimungkinkan mencakup materi semester sebelumnya atau semester ganjil. Soal pilihan ganda merupakan soal dengan bentuk pertanyaan yang

memiliki satu jawaban benar. Soal-soal bentuk pilihan ganda sering digunakan dalam menilai hasil belajar ranah koognitif karena luasnya bahan pelajaran yang harus dicakup dalam tes. Selain itu, soal pilihan ganda merupakan bentuk tes yang pemeriksaannya dapat dilakukan secara objektif, sehingga memudahkan pemeriksa dalam menilai jawaban yang diberikan. Arikunto (2013 : 183) mengatakan bahwa soal pilihan ganda merupakan jenis tes objektif yang paling banyak digunakan, karena banyak sekali materi yang dapat dicakup. Tes pilihan ganda terdiri dari dua bagian, yaitu keterangan (*stem*) dan alternatif jawaban (*option*). Ketika seseorang hendak membuat soal bentuk pilihan ganda, perlu memerhatikan variasi soal bentuk pilihan ganda, agar tidak terlihat monoton. Kaidah penulisan soal bentuk pilihan ganda (Kemendikbud, 2015 : 26) sebagai berikut.

- 1) Substansi/materi
  - a) Soal sesuai dengan indikator (menuntut tes bentuk pilihan ganda).
  - b) Materi yang diukur sesuai dengan kompetensi.
  - c) Pilihan jawaban homogen dan logis.
  - d) Hanya ada satu kunci jawaban yang tepat.
- 2) Konstruksi soal
  - a) Pokok soal dirumuskan secara singkat, jelas, dan tegas.
  - b) Rumusan pokok soal dan pilihan jawaban mengandung pernyataan yang diperlukan saja.
  - c) Pokok soal tidak menggunakan pernyataan negatif ganda.
  - d) Gambar/grafik/table/diagram dan sebagainya jelas dan berfungsi.
  - e) Panjang rumusan pilihan jawaban relatif sama.

- f) Pilihan jawaban tidak menggunakan “semua jawaban benar” atau “semua jawaban salah”.
  - g) Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan besar kecilnya angka atau kronologis kejadian.
  - h) Butir soal tidak bergantung pada jawaban soal sebelumnya
- 3) Bahasa
- a) Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia.
  - b) Menggunakan bahasa yang komunikatif.
  - c) Pilihan jawaban tidak mengulang kata atau kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian.
  - d) Tidak menggunakan bahasa yang berlaku setempat/tabu.

Soal objektif merupakan soal dengan bentuk pertanyaan yang memiliki satu jawaban yang benar. Soal-soal bentuk objektif sering digunakan dalam menilai hasil belajar ranah kognitif, karena luasnya bahan pelajaran yang harus dicakup dalam tes. Selain itu, soal objektif merupakan bentuk tes yang pemeriksaannya dapat dilakukan secara objektif, sehingga memudahkan pemeriksa dalam menilai jawaban yang diberikan. Arikunto (2013, h. 179) menjelaskan, “Tes objektif merupakan tes yang dalam pemeriksaannya dapat dilakukan secara objektif. Hal ini memang dimaksudkan untuk mengatasi kelemahan-kelemahan dari tes bentuk esai”. Sudijono (2012, h. 107) menyatakan, “Tes objektif dibedakan menjadi lima macam, yaitu: tes bentuk benar-salah, tes bentuk menjodohkan, tes bentuk melengkapi, tes bentuk isian, dan tes bentuk pilihan ganda”.

### 3. Analisis Butir Soal

Nitko (2008 : 1) mengatakan bahwa kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban peserta didik untuk membuat keputusan tentang setiap penilaian”. Senada dengan pendapat tersebut, Departemen Pendidikan Nasional (2008 : 1) mengatakan bahwa analisis butir soal bertujuan untuk mengidentifikasi kekurangan-kekurangan yang ada dalam tes atau dalam pembelajaran. Analisis butir soal merupakan kegiatan mengkaji pertanyaan-pertanyaan tes, agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai (Sudjana, 2009 : 135). Berdasarkan tujuan ini, manfaat kegiatan analisis butir soal yaitu; membantu para pengguna tes dalam evaluasi atas tes yang digunakan, sangat relevan bagi penyusunan tes informal dan lokal, mendukung penulisan butir soal yang efektif, dan secara materi dapat memperbaiki tes di kelas, meningkatkan validitas soal dan reliabilitas. Tidak hanya itu, Departemen Pendidikan Nasional (2008 : 2) juga mengatakan bahwa pelaksanaan kegiatan analisis butir soal didesain untuk menjawab pertanyaan-pertanyaan berikut ini: Apakah fungsi soal sudah tepat?; Apakah soal ini memiliki tingkat kesukaran yang tepat?; Apakah soal bebas dari hal-hal yang tidak relevan?; Apakah pilihan jawabannya efektif?. Dalam pelaksanaannya, kegiatan analisis butir soal dapat dilakukan secara kuantitatif.

Departemen Pendidikan Nasional (2008 : 8) menjelaskan bahwa analisis butir soal secara kuantitatif merupakan penelaahan butir soal didasarkan pada data empiris dari butir soal yang telah diujikan. Ada dua pendekatan dalam analisis butir soal secara kuantitatif, yaitu pendekatan secara klasik dan modern. Penelitian ini

membahas tentang analisis butir soal secara klasik, sesuai dengan fokus penelitian. Analisis butir soal secara klasik yaitu proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori tes klasik. Aspek yang perlu diperhatikan dalam analisis butir soal secara klasik yaitu: validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh.

#### a. Validitas

Validitas tes yaitu ketepatan pengukuran yang dimiliki oleh butir soal dalam mengukur apa yang seharusnya diukur lewat butir soal tersebut. Sudijono (2012 : 184) mengemukakan, “Butir soal dapat dikatakan valid apabila skor soal tersebut terbukti memiliki korelasi positif yang signifikan dengan skor totalnya. Skor total berkedudukan sebagai variabel terikat (dependent variable), sedangkan skor butir soal berkedudukan sebagai variabel bebas (independent variable)”.

Sudijono (2012 : 185) menjelaskan bahwa setiap butir soal yang dijawab benar diberi skor satu (1), sedangkan butir soal yang dijawab salah diberi skor nol (0). Jenis data tersebut dalam ilmu statistik dikenal dengan nama data diskret murni atau dikotomik. Skor total merupakan hasil penjumlahan dari setiap skor butir soal (misalnya:  $0 + 1 + 1 + 0 + 1 + 0 + 1 + 1 + 0 + 0 + 1 = 6$ ) merupakan data kontinu. Apabila variabel I berupa data diskret murni atau dikotomik, dan variabel II berupa data kontinu, maka teknik korelasi yang tepat digunakan adalah teknik korelasi point biserial yang diberi lambang  $r_{pb}$ . Senada dengan pendapat tersebut, Purwanto (2016 :

115-128) mengelompokkan metode pengujian validitas menjadi tiga macam, yaitu; (1) Validitas isi (*content validity*) adalah pengujian validitas yang dilakukan terhadap isi butir soal untuk memastikan apakah butir soal tersebut dapat mengukur secara tepat keadaan yang hendak diukur, (2) Validitas kriteria (*criterion related validity*) adalah pengujian validitas yang dilakukan dengan membandingkan tes hasil belajar dengan kriteria tertentu di luar tes hasil belajar, dan (3) Validitas konstruk (*construct validity*) adalah pengujian validitas yang dilakukan dengan melihat kesesuaian konstruksi butir soal yang ditulis dengan kisi-kisinya. Kriteria korelasi koefisien validitas suatu tes menurut Purwanto (2013 : 139) disajikan dalam Tabel 2.1.

**Tabel 2.1 Tingkat Validitas**

No.	Koefisien Validitas	Tingkat Validitas
1.	0,00 – 0,19	Sangat rendah (hampir tidak ada korelasi)
2.	0,20 – 0,39	Korelasi rendah
3.	0,40 – 0,69	Korelasi cukup
4.	0,70 – 0,89	Korelasi tinggi
5.	0,90 – 1,00	Korelasi sangat tinggi (sempurna)

Sumber: Purwanto (2013 : 139)

#### **b. Reliabilitas**

Kata reliabilitas berasal dari kata *reliable* yang artinya dapat dipercaya. Reliabilitas mengacu pada ketepatan hasil yang diperoleh dari suatu pengukuran. Tes dikatakan memiliki reliabilitas tinggi jika memberikan hasil pengukuran yang tetap (*consistent*) apabila diteskan berkali-kali (Widoyoko 2016 : 252). Arikunto (2013 : 100) mengatakan bahwa suatu tes dapat dikatakan mempunyai taraf kepercayaan yang tinggi jika tes tersebut dapat memberikan hasil yang tetap. Hal senada juga dikemukakan oleh Azwar (2013 : 7) bahwa.

Hasil suatu pengukuran akan dapat dipercaya hanya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, selama aspek yang diukur dalam diri subjek memang belum berubah. Dalam hal ini, relatif sama berarti tetap adanya toleransi terhadap perbedaan-perbedaan kecil yang biasanya terjadi di antara hasil beberapa pengukuran. Bila perbedaan yang terjadi sangat besar dari waktu ke waktu maka hasil pengukuran tersebut tidak dapat dipercaya dan dikatakan sebagai tidak reliabel. Pengukuran yang hasilnya tidak reliabel tentu tidak dikatakan akurat karena konsistensi menjadi syarat bagi akurasi.

Tujuan utama menghitung reliabilitas skor tes adalah untuk menentukan

tingkat ketepatan dan keajegan skor tes. Indeks reliabilitas berkisar antara 0-1.

Semakin tinggi koefisien reliabilitas (mendekati 1), semakin tinggi pula keajegan/ketetapan (Departemen Pendidikan Nasional 2008 : 16). Koefisien reliabilitas menurut Guilford (Ruseffendi, 199 : 191) dalam Jihad & Haris (2013 : 181) disajikan dalam Tabel 2.2.

**Tabel 2.2 Tingkat Reliabilitas**

No.	Koefisien Reliabilitas	Tingkat Reliabilitas
1.	$r_{11} \leq 0,20$	Reliabilitas : Sangat rendah
2.	$0,20 < r_{11} 0,40$	Reliabilitas : rendah
3.	$0,40 < r_{11} 0,70$	Reliabilitas : sedang
4.	$0,70 < r_{11} 0,90$	Reliabilitas : tinggi
5.	$0,90 < r_{11} 1,00$	Reliabilitas : sangat tinggi

Sumber: Jihad, A. & Haris, A. (2013 : 181)

### c. Tingkat Kesukaran

Arikunto (2013 : 222) menyatakan bahwa soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Perlu diketahui bahwa soal-soal yang terlalu mudah atau terlalu sukar bukan berarti soal tersebut tidak boleh digunakan, hal ini bergantung pada penggunaannya. Soal yang sukar akan menambah semangat

belajar bagi peserta didik yang pandai, sedangkan soal yang terlalu mudah akan membangkitkan semangat peserta didik dengan tingkat kecerdasan yang kurang.

Hal serupa juga dikemukakan oleh Sudijono (2012 : 370), bermutu tidaknya butir soal tes hasil belajar dapat ditentukan dari derajat kesukaran atau tingkat kesukaran butir soal tersebut. Bilangan yang menunjukkan sukar-mudahnya suatu soal disebut indeks kesukaran. Besarnya antara 0,00 sampai 1,0. Soal dengan indeks kesukaran 0,00, menunjukkan soal itu terlalu sukar, sebaliknya indeks 1,0, menunjukkan bahwa soalnya terlalu mudah (Daryanto 2010 : 180).



Indeks kesukaran dalam istilah evaluasi diberi simbol P (p besar), singkatan dari kata “proporsi”. Dengan demikian, soal dengan  $P = 0,70$  lebih mudah daripada soal dengan  $P = 0,20$ . Sebaliknya, soal dengan  $P = 0,30$  lebih sukar daripada soal dengan  $P = 0,80$ . Penelitian ini menggunakan ketentuan indeks kesukaran menurut Daryanto (2010 : 182) yang disajikan dalam Tabel 2.3.

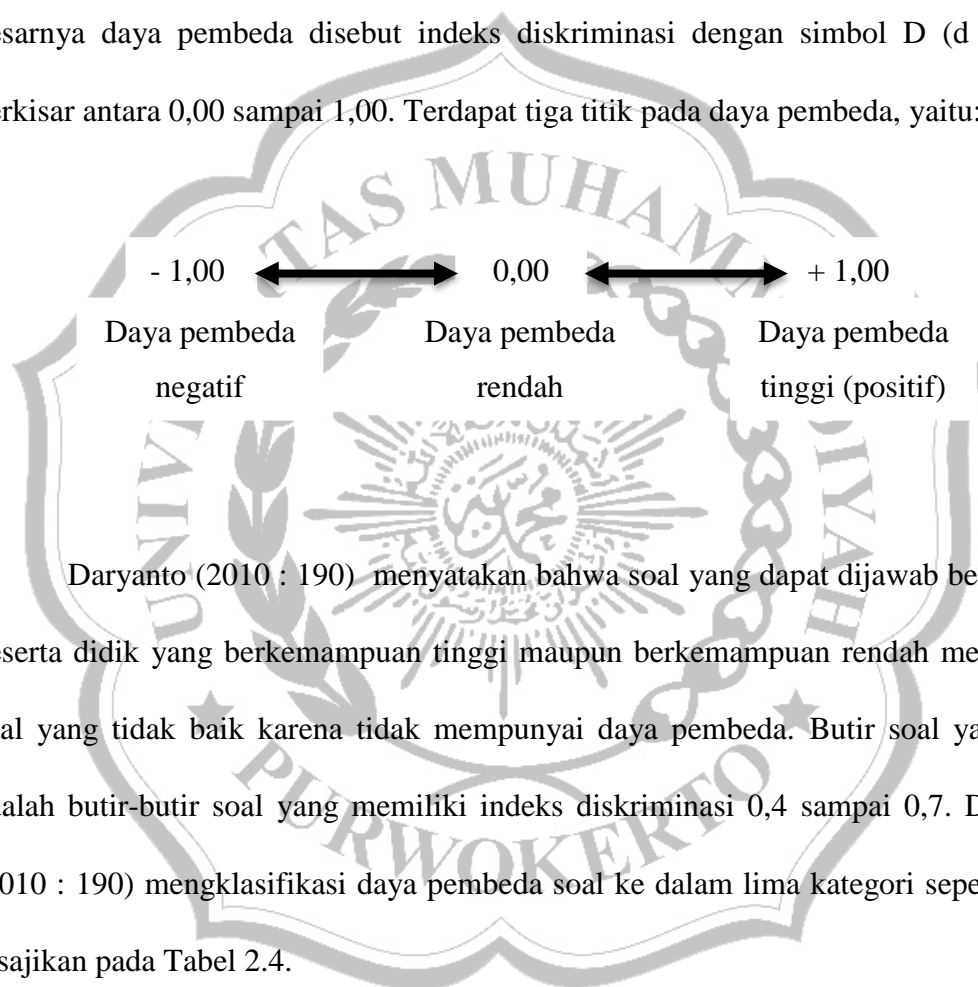
**Tabel 2.3 Kategori Indeks Kesukaran Soal**

No.	Indeks Taraf Kesukaran	Kategori Soal
1.	0,00 – 0,30	Sukar
2.	0,30 – 0,70	Sedang
3.	0,70 – 1,00	Mudah

Sumber: Daryanto (2010, h. 182).

#### d. Daya Pembeda

Daryanto (2010 : 183) menyatakan bahwa daya pembeda soal adalah kemampuan sesuatu untuk membedakan antara peserta didik yang berkemampuan tinggi dan peserta didik yang berkemampuan rendah. Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi dengan simbol  $D$  ( $d$  kapital) berkisar antara 0,00 sampai 1,00. Terdapat tiga titik pada daya pembeda, yaitu:



Daryanto (2010 : 190) menyatakan bahwa soal yang dapat dijawab benar oleh peserta didik yang berkemampuan tinggi maupun berkemampuan rendah merupakan soal yang tidak baik karena tidak mempunyai daya pembeda. Butir soal yang baik adalah butir-butir soal yang memiliki indeks diskriminasi 0,4 sampai 0,7. Daryanto (2010 : 190) mengklasifikasi daya pembeda soal ke dalam lima kategori seperti yang disajikan pada Tabel 2.4.

**Tabel 2.4 Kategori Daya Pembeda Soal**

No.	Indeks Daya Beda	Kategori Soal
1.	0,00 – 0,20	Jelek
2.	0,21 – 0,40	Cukup
3.	0,41 – 0,70	Baik
4.	0,71 – 1,00	Baik Sekali
5.	Negatif (-)	Semuanya tidak baik

Sumber: Daryanto (2010 : 190)

### e. Efektivitas Pengecoh

Penyebaran pilihan jawaban dijadikan dasar penelaahan butir soal untuk menentukan berfungsi atau tidaknya jawaban yang tersedia. Pengecoh dikatakan berfungsi apabila pengecoh paling tidak dipilih oleh 5% peserta tes/peserta didik, dan lebih banyak dipilih oleh kelompok peserta didik yang belum memahami materi (Departemen Pendidikan Nasional, 2008 : 14). Berdasarkan pengertian tersebut, pengecoh merupakan pilihan jawaban yang bukan kunci jawaban.

Analisis butir soal secara kuantitatif juga dapat dilakukan menggunakan program komputer. Menurut Departemen Pendidikan Nasional (2008 : 29), analisis data menggunakan program komputer sangat tepat dilakukan, karena tingkat keakuratan hitungan lebih tinggi daripada secara manual. Program komputer yang digunakan dalam penelitian ini yaitu program Anates V4.

Analisis butir soal merupakan kegiatan mengkaji pertanyaan-pertanyaan tes, agar diperoleh perangkat pertanyaan yang memiliki kualitas yang memadai (Sudjana, 2009 : 135). Kegiatan analisis butir soal dapat dilakukan secara kualitatif dan/atau secara kuantitatif.

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

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 (opsi)

1 = bilangan tetap

(Arifin, 2014 : 279)

### C. Kerangka Teoretis Penelitian

Evaluasi pembelajaran adalah proses mengukur untuk menentukan apakah tujuan pembelajaran telah tercapai sesuai yang diharapkan, sehingga berguna dalam pembuatan keputusan dalam dunia pendidikan. Evaluasi terbagi atas penilaian, pengukuran dan juga tes. Salah satu alat yang digunakan untuk melakukan evaluasi adalah tes. Tes pada umumnya meliputi tes pilihan ganda maupun tes uraian. Berdasarkan judul penelitian, peneliti hanya memfokuskan pada tes pilihan ganda yang dilaksanakan dalam Penilaian Akhir Tahun semester genap untuk menentukan apakah butir soal dalam tes layak untuk diteskan, diperlukan analisis butir soal terlebih dahulu. Aspek analisis butir soal dalam penelitian ini diantaranya yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh. Tujuan kegiatan analisis butir soal yaitu untuk mengetahui kualitas butir soal, sehingga dapat ditentukan butir soal yang baik, tidak baik dan yang masih perlu direvisi. Kerangka teoretis penelitian ini disajikan pada Gambar 2.1.

**Gambar 2.1 Bagan Kerangka Teoretis**

