

BAB II

KAJIAN TEORITIS

A. Landasan Teori

Untuk mendapatkan gambaran yang jelas tentang arah penulisan skripsi ini, dijelaskan lebih dahulu kata kunci yang terdapat dalam pembahasan ini sekaligus penggunaan secara operasional.

1. Pengertian Evaluasi dan Evaluasi Pendidikan

Secara harfiah kata evaluasi berasal dari bahasa Inggris *evaluation*; dalam bahasa Indonesia berarti: penilaian. Akar katanya adalah *value*; dalam bahasa Indonesia berarti; nilai. Dengan demikian, secara harfiah, evaluasi pendidikan (*education evaluation*) dapat diartikan sebagai: penilaian dalam (bidang) pendidikan atau penilaian mengenai hal-hal yang berkaitan dengan kegiatan pendidikan.¹ Secara istilah evaluasi pendidikan adalah kegiatan menilai yang terjadi dalam aktivitas pendidikan dalam pengukuran untuk mengetahui keberhasilan anak didik dalam mengikuti mata pelajaran tertentu.²

Dalam berlangsungnya proses pembelajaran dan pendidikan tidak akan terlepas dari yang namanya kegiatan evaluasi (penilaian). Melakukan kegiatan penilaian dapat dijadikan sebagai acuan atau tolak ukur sejauh mana kompetensi yang telah dikuasai oleh peserta didik.

¹ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, cet kesepuluh, (Jakarta: RajaGrafindo Persada, 2011), hal. 1.

² Hasan Basri, *Filsafat Pendidikan Islam*, cet ketiga, (Bandung: Pustaka Setia, 2009), hal. 142.

Penilaian hasil belajar peserta didik pada jenjang pendidikan dasar dan menengah dilaksanakan berdasarkan standar penilaian pendidikan yang berlaku secara nasional.

2. Fungsi Evaluasi Pendidikan

Kegiatan evaluasi yang tidak menghasilkan titik tolak untuk perbaikan adalah hampa dan tidak ada artinya sama sekali. Oleh sebab itu, secara khusus evaluasi dalam dunia pendidikan dapat dilihat dari segi psikologis, sosiologis, didaktik, dan administratif.

Secara psikologis, peserta didik selalu butuh untuk mengetahui sejauh mana kegiatan yang telah dilakukan sesuai dengan tujuan yang hendak dicapai. Dalam pembelajaran, mereka perlu mengetahui prestasi belajarnya sehingga ia merasakan kepuasan dan ketenangan.³

Secara sosiologis, evaluasi berfungsi untuk mengetahui apakah peserta didik sudah cukup mampu untuk terjun ke masyarakat. Artinya peserta didik dapat berkomunikasi dan beradaptasi terhadap seluruh lapisan masyarakat dengan segala karakteristiknya.⁴

Secara didaktik, evaluasi berfungsi untuk membantu guru dalam menempatkan peserta didik pada kelompok tertentu sesuai dengan kemampuan dan kecakapannya masing-masing serta membantu guru dalam usaha memperbaiki proses pembelajarannya.⁵

³ Zainal Arifin, *Evaluasi Pembelajaran Prinsip, Teknik, dan Prosedur*, cet kesepuluh, (Bandung: Remaja Rosdakarya, 2017), hal. 1

⁴ *Ibid.*, hal. 17.

⁵ *Ibid.*

Secara administratif, evaluasi berfungsi untuk memberikan laporan tentang kemajuan peserta didik kepada orang tua, pejabat pemerintah yang berwenang, kepala sekolah, guru-guru, dan peserta didik itu sendiri. Hasil evaluasi dapat memberikan gambaran secara umum tentang semua hasil usaha yang dilakukan oleh institusi pendidikan.⁶

3. Tujuan Evaluasi Pendidikan

Tujuan evaluasi pendidikan adalah untuk mengetahui kadar pemahaman peserta didik terhadap mata pelajaran, melatih keberanian peserta didik untuk mengingat kembali materi yang telah diberikan, untuk mengetahui siapa diantara peserta didik yang cerdas dan lemah, sehingga yang lemah diberikan perhatian khusus, dan bagi pendidik untuk mengetahui sejauh mana pendidik sungguh-sungguh dalam menjalankan tugasnya untuk mencapai tujuan pendidikan.⁷

Tujuan khusus evaluasi pendidikan terutama dalam pembelajaran evaluasi disesuaikan dengan jenis evaluasi pembelajaran itu sendiri, seperti evaluasi perencanaan dan pengembangan, evaluasi *monitoring*, evaluasi dampak, evaluasi efisiensi-ekonomik, dan evaluasi program komprehensif.⁸ Secara khusus, tujuan evaluasi pendidikan yaitu memberikan standar keberhasilan, baik dari sisi aspek kognitif, aspek afektif, dan aspek psikomotoriknya kemudian berimplikasi pada

⁶ *Ibid.*

⁷ Abdul Aziz, *Filsafat Pendidikan Islam*, cet pertama, (Yogyakarta: Penerbit TERAS, 2009), hal.173.

⁸ Daryanto, *Evaluasi Pendidikan*, cet kelima, (Jakarta: Rineka Cipta, 2008) hal. 11.

penanganan peserta didik oleh pendidiknya untuk bisa memberikan perhatian lebih baik secara fisik maupun secara psikis tergantung dari gejala yang dialami oleh peserta didik.⁹

4. Kegunaan Evaluasi Pendidikan

Di antara kegunaan dari kegiatan evaluasi dalam pendidikan yaitu: terbentuknya kemungkinan bagi *evaluator* guna memperoleh informasi tentang hasil-hasil yang telah dicapai dalam rangka pelaksanaan program pendidikan, terbentuknya kemungkinan untuk dapat diketahuinya relevansi antara program pendidikan yang telah dirumuskan, dengan tujuan yang hendak dicapai, dan terbentuknya kemungkinan untuk dapat dilakukannya usaha perbaikan, penyesuaian, dan penyempurnaan program pendidikan yang dipandang lebih berdaya guna dan berhasil guna, sehingga apa yang dicita-citakan, akan dapat dicapai dengan hasil yang sebaik-baiknya.¹⁰

Dalam penggunaan administratif, data evaluasi dapat digunakan sebagai kartu catatan tingkah laku peserta didik, dan untuk melengkapi laporan-laporan pendidik tentang kemajuan sekolah kepada instansi atasan yang memerlukan. Dalam bidang instruksional, data evaluasi bagi pendidik membantu dalam memperbaiki cara mengajar, metode, dan bahan pelajaran yang lebih efektif dan menarik bagi peserta didik.¹¹

⁹ Abdul Aziz, *Op. Cit.*, hal. 174.

¹⁰ Anas Sudijono, *Op. Cit.*, hal. 17

¹¹ Sulistyorini, *Evaluasi Pendidikan Dalam Meningkatkan Mutu Pendidikan*, cet pertama, (Yogyakarta: Penerbit TERAS, 2009), hal. 61.

5. Tes

a. Pengertian Tes

Istilah “tes” berasal dari bahasa Prancis, yaitu “*testum*”, berarti piring yang digunakan untuk memilah logam mulia yang nilainya sangat tinggi.¹² Pengertian dari segi istilah, menurut Anne Anastasi tes adalah alat pengukur yang mempunyai standar yang objektif sehingga dapat digunakan secara meluas, serta dapat betul-betul digunakan untuk mengukur dan membandingkan keadaan psikis atau tingkah laku individu. Menurut Lee J. Cronbach tes adalah suatu prosedur yang sistematis untuk membandingkan tingkah laku dua orang atau lebih, sedangkan menurut F.L. Goodenough, tes merupakan suatu tugas atau serangkaian tugas yang diberikan kepada individu atau sekelompok individu, dengan maksud untuk membandingkan kecakapan mereka, satu dengan yang lain.¹³

Tes adalah suatu teknik atau cara yang digunakan dalam rangka melaksanakan kegiatan pengukuran, yang di dalamnya terdapat berbagai pertanyaan, pernyataan, atau serangkaian tugas yang harus dikerjakan atau dijawab oleh peserta didik untuk mengukur aspek perilaku peserta didik.¹⁴

b. Bentuk-Bentuk Tes Hasil Belajar

¹² Eko Putro Widoyoko, *Penilaian Hasil Pembelajaran di Sekolah*, cet kedua, (Yogyakarta: Pustaka Pelajar, 2016), hal. 65.

¹³ Anas Sudiono, *Op. Cit.*, hal. 66 – 67.

¹⁴ Zainal Arifin, *Op. Cit.*, hal 118.

1) Tes Uraian

Tes uraian adalah tes yang butir-butirnya berupa suatu pertanyaan yang menghendaki jawaban yang berupa uraian-uraian yang relatif panjang.¹⁵ Bentuk-bentuk pertanyaan atau perintah itu menuntut kepada peserta tes untuk memberikan penjelasan, komentar, penafsiran, membandingkan, membedakan, dan sebagainya. jumlah butir soal terbatas berkisar antara lima sampai sepuluh butir soal. Pada umumnya butir-butir soal tes uraian itu diawali dengan kata-kata: “Jelaskan.....”, “Terangkan.....”, “Uraikan.....”, “Mengapa.....”, “Bagaimana.....” atau kata-kata yang serupa dengan itu.

Tes hasil belajar bentuk uraian, di samping memiliki keunggulan juga tidak terlepas dari kekurangan.

Di antara keunggulan yang dimiliki oleh tes uraian yaitu:

- a) Tes uraian dapat digunakan untuk menilai tingkat pemahaman peserta didik pada level tinggi.
- b) Para siswa mempunyai kebebasan untuk memilih menyiapkan, dan menyajikan gagasan di dalam kata-kata mereka sendiri.
- c) Tes uraian dapat menunjukkan kebaikan siswa dalam mengorganisir pemikiran, pendukung pandangan, dan menciptakan gagasan, metode, dan solusi.

¹⁵ Suwanto, *Pengembangan Tes Diagnostik Dalam Pembelajaran Panduan Praktis Bagi Pendidik dan Calon Pendidik*, cet pertama, (Yogyakarta: Pustaka Pelajar, 2013), hal. 47.

- d) Kompleksitas pertanyaan tes uraian dan kompleksitas berfikir yang diharapkan dari peserta didik dapat disesuaikan dengan umur peserta didik, kemampuan peserta didik, dan pengalaman peserta didik.
- e) Pendidik dapat memahami peserta didiknya dalam membuat jawaban dan tidak hanya memilih jawaban yang terbaik dari alternatif pilihan yang ditetapkan, seperti pada tes pilihan ganda.¹⁶

Adapun kelemahan-kelemahan yang disandang oleh tes uraian, antara lain:

- a) Tes uraian pada umumnya kurang dapat menampung atau mencakup dan mewakili isi dan luasnya materi atau bahan pelajaran yang telah diberikan kepada peserta tes, yang seharusnya diujikan dalam tes hasil belajar.
- b) Cara mengoreksi jawaban soal tes uraian cukup sulit.
- c) Dalam pemberian skor hasil tes uraian, terdapat kecenderungan bahwa peserta tes lebih banyak bersifat subjektif.
- d) Pekerjaan koreksi terhadap lembar-lembar jawaban hasil tes uraian sulit untuk diserahkan kepada orang lain, sebab pada

¹⁶ *Ibid.*, hal. 55.

tes uraian orang yang paling tahu mengenai jawaban yang sempurna adalah penyusun tes itu sendiri.¹⁷

2) Tes Jawaban Pendek (Objektif)

Tes jawaban pendek atau tes objektif adalah tes hasil belajar yang terdiri dari butir-butir soal yang dapat dijawab oleh peserta tes dengan jalan memiliki salah satu di antara beberapa kemungkinan jawaban yang telah dipasangkan pada masing-masing *items*, atau dengan jalan menuliskan (mengisikan) jawabannya berupa kata-kata atau simbol-simbol tertentu pada tempat atau ruang yang telah disediakan untuk masing-masing butir item yang bersangkutan.¹⁸

Tes objektif adalah tes yang skornya bersifat objektif, yaitu hanya dipengaruhi oleh objek jawaban atau respons yang diberikan oleh peserta tes.¹⁹ Dalam penggunaan tes objektif ini jumlah soal yang diajukan jauh lebih banyak daripada tes subjektif.

Tes jawaban pendek dapat dibedakan menjadi lima golongan yaitu: tes objektif bentuk benar-salah (*true-false test*), tes objektif bentuk menjodohkan (*matching test*), tes objektif bentuk melengkapi (*completion test*), tes objektif bentuk isian (*fill*

¹⁷ Anas Sudijono, *Op. Cit.*, hal. 103 – 104

¹⁸ *Ibid.*, hal. 107.

¹⁹ Eko Putro Widoyoko, *Op. Cit.*, hal. 71.

in test), dan tes objektif bentuk bentuk pilihan ganda (*multiple choice item test*).²⁰

Seperti halnya tes uraian, sebagai alat pengukur keberhasilan belajar peserta didik, tes objektif ini di samping memiliki keunggulan tertentu juga tidak dapat terlepas dari kekurangan. Keunggulan-keunggulan itu diantaranya:

- a. Lebih representatif mewakili isi dan luas bahan.
- b. Lebih mudah dan cepat cara memeriksanya karena dapat menggunakan kunci jawaban, bahkan dapat menggunakan alat-alat kemajuan teknologi seperti misal mesin *scanner*.
- c. Pemeriksaannya dapat diserahkan orang lain.
- d. Dalam pemeriksaan maupun penyekoran, tidak ada unsur subjektif yang memengaruhi, baik dari segi guru maupun siswa.²¹

Adapun untuk kelemahan-kelemahan dari tes jawaban pendek adalah:

- a. Membutuhkan persiapan yang lebih sulit daripada tes subjektif karena butir item tesnya banyak dan harus

²⁰ Anas Sudijono, *Op. Cit.*, hal. 107.

²¹ Eko Putro Widoyoko, *Evaluasi Program Pembelajaran Panduan Praktis Bagi Pendidik dan Calon Pendidik*, cet pertama, (Yogyakarta: Pustaka Pelajar, 2009), hal. 49.

teliti untuk menghindari kelemahan-kelemahan yang lain.

- b. Butir-butir soal cenderung hanya mengungkap ingatan dan pengenalan kembali saja dan sukar untuk mengukur kemampuan berfikir yang tinggi seperti sintesis maupun kreativitas.
- c. Banyak kesempatan bagi siswa untuk spekulasi atau untung-untungan dalam menjawab soal.
- d. Kerjasama antar siswa pada waktu mengerjakan soal tes lebih terbuka.²²

6. Prinsip-Prinsip Dasar dalam Penyusunan Tes Hasil Belajar

Ada beberapa prinsip dasar yang perlu dicermati di dalam menyusun tes hasil belajar agar tes tersebut dapat mengukur tujuan intruksional khusus untuk mata pelajaran yang telah diajarkan, atau mengukur kemampuan dan keterampilan peserta didik setelah mereka menyelesaikan suatu unit pengajaran tertentu. Beberapa prinsip-prinsip dasar tersebut diantaranya:

- a. Tes hasil belajar harus dapat mengukur secara jelas hasil belajar yang telah ditetapkan sesuai dengan tujuan intruksional.
- b. Butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi bahan pelajaran yang telah diajarkan,

²² *Ibid.*, hal. 50.

sehingga dapat dianggap mewakili seluruh *performance* yang telah diperoleh selama peserta didik mengikuti suatu unit pengajaran.

- c. Bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat bervariasi, sehingga betul-betul cocok untuk mengukur hasil belajar yang diinginkan sesuai dengan tujuan tes itu sendiri.
- d. Tes hasil belajar harus didesain dengan kegunaannya untuk memperoleh hasil yang diinginkan.
- e. Tes hasil belajar harus memiliki reliabilitas yang dapat diandalkan.
- f. Tes hasil belajar di samping harus dapat dijadikan alat pengukur keberhasilan belajar siswa, juga harus dapat dijadikan alat untuk mencari informasi yang berguna untuk memperbaiki cara belajar siswa dan cara mengajar guru itu sendiri.²³

7. Pengertian Tentang Penilaian Tengah Semester

Evaluasi hasil belajar dilaksanakan tidak lain adalah untuk memperoleh tujuan pembelajaran yang baik. Kegiatan belajar-mengajar tanpa evaluasi akan terasa kurang lengkap karena seorang guru tidak mampu menilai dan memantau peserta didik apakah sudah menguasai materi atau tidak, dan seorang guru juga tidak dapat melakukan introspeksi diri tanpa adanya proses evaluasi pembelajaran. Alat evaluasi yang biasa digunakan dalam pendidikan di sekolah adalah tes.

²³ Anas Sudijono, *Op.Cit.*, hal. 97 – 99.

Penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik.²⁴ Penilaian Tengah Semester (PTS) adalah kegiatan penilaian sumatif yang dilakukan untuk mengukur pencapaian kompetensi dasar mata pelajaran setelah kegiatan pembelajaran berlangsung 8-9 minggu.²⁵

Bentuk tes yang biasa digunakan diantaranya tes objektif, dan perpaduan antara tes objektif dengan tes subjektif. Tes objektif menggunakan variasi bentuk soal pilihan ganda sedangkan tes subjektif menggunakan variasi bentuk soal uraian panjang.

Data hasil Penilaian Tengah Semester (PTS), sangat memberikan manfaat bagi guru dalam rangka memperbaiki proses belajar-mengajar di sekolah dalam segi perencanaan, pelaksanaan, dan penilaian hasil belajar.

Dari uraian dia atas dapat disimpulkan bahwa Penilaian Tengah Semester (PTS) yang merupakan salah satu bentuk alat evaluasi yang dilaksanakan setelah 8-9 minggu pembelajaran akan digunakan guna mengambil keputusan. Keputusan yang diambil akan tepat mengenai sasaran apabila didukung oleh alat evaluasi yang berupa butir soal penilaian tengah semester yang memenuhi syarat validitas, reliabilitas, derajat kesukaran item, daya pembeda item, dan fungsi pengecoh.

8. Evaluasi Hasil Belajar Pendidikan Agama Islam

²⁴ Peraturan Menteri Pendidikan dan Kebudayaan Nomor 23 Tahun 2016.

²⁵ Permendikbud Nomor 66 Tahun 2013.

Evaluasi pembelajaran adalah suatu proses kegiatan evaluasi yang hanya terbatas pada ruang lingkup kelas dalam konteks proses belajar-mengajar.²⁶ Evaluasi hasil belajar Pendidikan Agama Islam adalah proses kegiatan evaluasi yang berfokus pada mata pelajaran Pendidikan Agama Islam untuk mengetahui seberapa jauh daya tangkap siswa terhadap pembelajaran PAI.

Tujuan evaluasi hasil belajar pada mata pelajaran Pendidikan Agama Islam diantaranya:

- a. Sebagai sarana untuk mengetahui tingkat pengetahuan siswa terhadap materi Pendidikan Agama Islam.
- b. Sebagai alat motivasi semangat siswa dalam mendalami ilmu agama Islam.
- c. Sebagai menilai ketercapaian tujuan, salah satunya adalah menanamkan sikap baik, dan mengasah siswa dalam mendalami ilmu Agama Islam dalam segi fiqih, akhlak, dan baca tulis al-Qur'an.²⁷

9. Analisis Butir Soal

Analisis soal bertujuan untuk mengidentifikasi soal-soal yang baik, kurang baik dan soal yang jelek. Keberhasilan, dan kemampuan suatu penilaian bergantung dari kualitas saat penilaian itu sendiri. Untuk

²⁶ Sigit Pramono, *Panduan Evaluasi Kegiatan Belajar-Mengajar*, cet pertama, (Yogyakarta: DIVA Press, 2014), hal. 13.

²⁷ Abdul Rahman, *Pendidikan Agama Islam dan Pendidikan Islam*, Riset, Maret 2012, hal. 2057.

mengetahui kualitas dari alat penilaian tersebut, dapat dilakukan analisis butir soal secara kuantitatif meliputi beberapa hal berikut:

a. Validitas

Validitas adalah pertimbangan yang paling pokok di dalam mengembangkan dan mengevaluasi tes. Proses pengvaliditan melibatkan pengumpulan bukti untuk menyediakan penjelasan ilmiah penafsiran skor yang diusulkan.²⁸ Validitas tes menyangkut pada ketepatan inferensi yang dibuat guru berdasarkan data yang dikumpulkan menggunakan tes tersebut. Ada beberapa jenis ukuran validitas yang umum digunakan, antara lain validitas rasional dan validitas empirik.²⁹

1) Validitas Rasional

Validitas rasional adalah validitas yang diperoleh atas dasar pemikiran, validitas yang diperoleh dengan berpikir secara logis. Suatu tes hasil belajar dapat dikatakan melewati validitas rasional, apabila setelah dilakukan penganalisisan secara rasional ternyata bahwa tes hasil belajar itu memang secara rasional tepat dapat mengukur apa yang seharusnya diukur.³⁰

Untuk menentukan apakah tes hasil belajar sudah memiliki validitas rasional ataukah belum, dapat dilakukan penelusuran dari

²⁸ Suwanto, *Op. Cit.*, hal. 94 – 95.

²⁹ Anas Sudijono, *Op. Cit.*, hal. 182.

³⁰ *Ibid.*, hal 64.

dua segi, yaitu dari segi isinya (*content*) dan dari segi susunan atau konstruksinya (*construct*).

2) Validitas Empirik

Validitas empirik adalah ketepatan mengukur yang didasarkan pada hasil analisis yang bersifat empirik, yang bersumber pada pengamatan di lapangan.³¹

Tes hasil belajar dapat dikatakan telah memiliki validitas empirik apabila berdasarkan hasil analisis yang dilakukan terhadap data hasil pengamatan di lapangan, terbukti bahwa tes hasil belajar itu dengan secara tepat telah dapat mengukur hasil belajar yang seharusnya diungkap atau diukur lewat tes hasil belajar tersebut. Sebuah *instrument* dapat dikatakan memiliki validitas empirik, apabila sudah diuji dari pengalaman. Untuk menentukan validitas butir soal pilihan ganda dan soal uraian menggunakan korelasi skor butir dengan skor total/ *point biserial*.

Indeks korelasi *Point Biserial* diberi lambang r_{pbi} , dapat diperoleh dengan rumus:

$$r_{pbi} = \frac{Mp - Mt}{SDt} \sqrt{\frac{p}{q}}$$

³¹ *Ibid.*, hal. 167.

Keterangan:

- r_{pbi} : Koefisien korelasi *point biserial* yang melambangkan kekuatan korelasi antara variabel I dengan variabel II, yang dalam hal ini dianggap sebagai Koefisien Validitas Item.
- M_p : Skor rata-rata hitung yang dimiliki oleh peserta tes, yang untuk butir butir item yang bersangkutan telah dijawab dengan betul.
- M_t : Skor rata-rata dari skor total.
- SD_t : Deviasi standar dari skor total.
- p : proporsi peserta tes yang menjawab betul terhadap butir item yang sedang diuji validitas itemnya.
- q : proporsi peserta tes yang menjawab salah terhadap butir item yang sedang diuji validitas itemnya.

b. Reliabilitas

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata asal *reliable* yang artinya dapat dipercaya.³² Instrumen tes dapat dikatakan dapat dipercaya jika memberikan hasil yang tepat atau ajek (konsisten) apabila diteskan berkali-kali.³³ Untuk menentukan reliabilitas tes hasil belajar dibedakan menjadi dua golongan yaitu tes hasil belajar objektif (pilihan ganda) dan bentuk uraian. Untuk menentukan

³² Eko Putro Widyoko, *Op. Cit.*, hal. 252.

³³ Anas Sudijono, *Op. Cit.*, hal. 207

reliabilitas soal bentuk objektif dapat dilakukan melalui salah satu dari pendekatan, yaitu pendekatan serba satu (*single test-single trial*), pendekatan tes ulang (*tes-retest*), dan pendekatan tes sejajar (*alternate-forms*).

Analisis reliabilitas tes bentuk uraian umumnya menggunakan sebuah rumus Alpha. Ada beberapa teknik yang dapat digunakan untuk mengukur reliabilitas suatu tes, tergantung dari skala yang digunakan. Untuk data yang bersifat dikotomi menggunakan rumus Alpha dengan asumsi bahwa tingkat kesukaran tes homogen, adapun rumusnya sebagai berikut:

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

Keterangan:

r_{11} : Koefisien reliabilitas tes.

n : banyaknya butir item yang dikeluarkan dalam tes.

1 : bilangan konstan.

$\sum S_i^2$: jumlah varian skor dari tiap-tiap butir item.

S_t^2 : Varian total.

Soal dikatakan mempunyai reliabilitas yang tinggi apabila koefisien reliabilitas tes yang sedang diuji reliabilitasnya lebih dari 0,70. Apabila lebih kecil dari 0,70 berarti tes yang sedang diuji

reliabilitasnya dinyatakan belum memiliki reliabilitas yang tinggi (*unreliable*).³⁴

c. Derajat Kesukaran Item

Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir yang baik, apabila butir-butir item tersebut tidak terlalu sukar dan tidak terlalu mudah dengan kata lain derajat kesukaran item itu sedang atau cukup. Witherington mengatakan bahwa sudah atau belum memadainya derajat kesukaran item tes hasil belajar dapat diketahui dari besar kecilnya angka yang melambangkan tingkat kesulitan dari item tersebut.³⁵ Angka yang dapat memberikan petunjuk mengenai tingkat kesulitan itu dikenal dengan istilah *difficulty index* (angka indek kesukaran item) yang dilambangkan dengan huruf P, singkatan dari *proporitin* (proporsi=proporsa).

Angka indeks kesukaran dapat diperoleh dengan menggunakan rumus yang dikemukakan oleh Du Bois, yaitu:

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

Keterangan:

P : Angka indeks kesukaran item.

N_p : Banyaknya testee yang dapat menjawab betul terhadap butir item yang bersangkutan.

³⁴ *Ibid.*, hal. 209

³⁵ *Ibid.*, hal.. 371.

N : Jumlah testee yang mengikuti tes hasil belajar.

Menurut Witherington dalam Sudijono (2011) mengemukakan bahwa apabila angka indeks diskriminasi item kurang dari 0,25 maka butir soal tergolong sukar, indeks diskriminasi item antara 0,25 – 0,75 dianggap sedang atau cukup, dan lebih dari 0,75 dianggap terlalu mudah.³⁶

d. Daya Pembeda Item

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara peserta tes yang berkemampuan tinggi, dengan peserta tes yang kemampuannya rendah demikian rupa sehingga sebagian peserta tes yang memiliki kemampuan tinggi untuk menjawab butir item tersebut lebih banyak menjawab betul, sementara peserta tes yang kemampuannya rendah untuk menjawab butir item tersebut sebagian besar tidak dapat menjawab item tersebut dengan betul.³⁷

Daya pembeda item dapat diketahui melalui atau dengan melihat besar kecilnya angka indeks diskriminasi item. Angka indeks diskriminasi item adalah sebuah angka atau bilangan yang menunjukkan besar kecilnya daya pembeda (*discriminatory power*) yang dimiliki oleh sebutir item. Daya pembeda pada dasarnya dihitung atas dasar pembagian peserta tes ke dalam dua kelompok, yaitu kelompok atas (*the higher group*) yang dianggap berkemampuan

³⁶ *Ibid.*, hal. 372.

³⁷ *Ibid.*, hal. 385 – 386.

tinggi dan kelompok bawah (*the lower group*) yang dianggap berkemampuan rendah.

Indeks diskriminasi item umumnya diberi lambang dengan huruf *D* (*discriminatory power*), indeks diskriminasi item besarnya berkisar antara 0 (nol) sampai dengan 1,00. Jika sebutir item angka indeks diskriminasinya = 0,00 (nihil), maka hal ini menunjukkan bahwa butir item bersangkutan tidak memiliki daya pembeda sama sekali. Apabila angka indeks diskriminasi item dari sebutir item bertanda negatif (minus), maka hal itu berarti bahwa butir item yang bersangkutan lebih banyak dijawab betul oleh peserta tes kelompok bawah ketimbang kelompok atas. Untuk mengetahui besar kecilnya angka indeks diskriminasi item dapat dipergunakan rumus sebagai berikut:

Rumus pertama

$$D = P_H - P_L$$

Keterangan:

D : Angka indeks diskriminasi item

P_H : Proporsi kelompok atas

P_L : Proporsi kelompok bawah

Rumus Kedua

$$\phi = \frac{P_H - P_L}{2 \sqrt{(p)(q)}}$$

Keterangan:

\emptyset : Angka indeks diskriminasi item.

P_H : Proporsi kelompok atas.

P_L : Proporsi kelompok bawah.

2 : Bilangan konstan.

p : Proporsi seluruh testee yang menjawab Betul.

q : Proporsi seluruh testee yang jawabannya Salah, dimana $q = (1 - p)$

Adapun patokan yang pada umumnya dipegangi adalah sebagai berikut:

Tabel I
Kriteria Angka Indeks Diskriminasi Item

Besarnya Angka Indeks Diskriminasi Item (D)	Klasifikasi	Interpretasi
Kurang dari 0,20	<i>Poor</i>	Butir item yang bersangkutan daya pembedanya lemah sekali (jelek), dianggap tidak memiliki daya pembeda yang baik.
0,20 – 0,40	<i>Satisfactory</i>	Butir item yang bersangkutan telah memiliki daya pembeda yang cukup (sedang).
0,40 – 0,70	<i>Good</i>	Butir item yang bersangkutan telah memiliki daya pembeda yang baik.
0,70 – 1,00	<i>Excellent</i>	Butir item yang bersangkutan telah memiliki daya pembeda yang baik sekali.
Bertanda negatif	-	Butir item yang bersangkutan daya pembedanya negatif

		(jelek sekali).
--	--	-----------------

e. Fungsi Distraktor (Pengecoh)

Instrumen evaluasi yang berbentuk tes objektif, harus mempunyai distraktor yang efektif. Distraktor atau pengecoh merupakan opsi-opsi yang bukan merupakan kunci jawaban (jawaban benar). Suatu distraktor disebut efektif jika mampu menarik minat testee untuk memilihnya, semakin banyak pemilihnya, semakin efektif fungsi distraktor tersebut.³⁸ Tujuan utama dari pemasangan distraktor pada setiap butir item adalah agar dari sekian banyak peserta tes yang mengikuti tes hasil belajar ada yang tertarik atau terangsang untuk memilihnya, sebab mereka menyangka bahwa pengecoh yang mereka pilih itu merupakan jawaban betul.³⁹ Makin banyak peserta tes yang terkecoh maka kita dapat menyimpulkan bahwa distraktor itu dapat menjalankan fungsinya dengan sebaik-baiknya. Namun sebaliknya, apabila pengecoh yang dipasang di setiap item “tidak laku” dapat disimpulkan distraktor tersebut tidak menjalankan fungsinya dengan baik.

³⁸ Shodiq Abdullah, *Evaluasi Pembelajaran Konsep Dasar, Teori, dan Aplikasi*, cet pertama, (Semarang: Pustaka Rizki Putra, 2012), hal. 111.

³⁹ Anas Sudijono, *Op. Cit.*, hal. 410.

Untuk mengetahui distraktor (pengecoh) telah menjalankan fungsinya dengan baik, yaitu apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5% dari seluruh peserta tes.⁴⁰

10. Analisis Butir Soal Dalam Penelitian

Analisis butir soal bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik, dan soal yang jelek. Melakukan analisis terhadap butir soal akan mendapat informasi tentang kejelekan sebuah soal dan “petunjuk” untuk mengadakan perbaikan. Analisis butir soal dapat dilakukan dengan menganalisis dari segi validitas, reliabilitas, taraf kesukaraan, daya pembeda, dan fungsi distraktor (pengecoh) butir-butir soal.

Analisis butir soal dalam penelitian ini mengkaji dan mengidentifikasi instrumen penilaian terutama pada Penilaian Tengah Semester (PTS). Instrumen penilaian berupa tes hasil belajar yang dilaksanakan pada separuh pembelajaran setelah 8 – 9 minggu dilaksanakan proses belajar mengajar.

Peneliti menganalisis instrumen penilaian berupa tes Penilaian Tengah Semester (PTS), sehingga peneliti dapat menggolongkan butir soal yang baik, kurang baik, dan jelek. Melalui analisis butir soal dalam penelitian dapat menjadi “Petunjuk” untuk mengadakan perbaikan terhadap butir-butir soal yang kurang baik, dan jelek. Perbaikan dilakukan

⁴⁰ *Ibid.*, hal. 411.

sesuai dengan kaidah dalam penyusunan tes hasil belajar. Analisis butir soal dalam penelitian juga dapat membantu pendidik untuk memperbaiki instrumen evaluasi hasil belajar terutama pada aspek penilaian. Instrumen penilaian yang baik akan menentukan sejauh mana peserta didik dapat memahami materi yang sudah diajarkan dan hasil tes yang dilakukan dapat menentukan kualitas pendidik dalam menyampaikan materi pembelajaran.

B. Hasil Penelitian Terdahulu

Untuk mendukung penelaahan penelitian yang lebih komprehensif, peneliti berusaha melakukan kajian terhadap beberapa penelitian terdahulu yang mempunyai relevansi dengan topik yang ingin diteliti. Beberapa penelitian terdahulu yang relevan dengan penelitian ini adalah:

1. Skripsi berjudul “Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Pendidikan Agama Islam (PAI) Kelas VIII Semester Genap Sekolah Menengah Pertama (SMP) Kota Semarang Tahun Pelajaran 2016/2017, yang ditulis oleh Azis Muslim, mahasiswa jurusan Pendidikan Agama Islam, Fakultas Ilmu Tarbiyah dan Keguruan Universitas Agama Islam Negeri Walisongo Semarang 2017. Skripsi ini membahas mengenai kualitas butir soal ulangan akhir semester genap soal pilihan ganda. Analisis data dilakukan secara kuantitatif dengan mempergunakan statistik inferensial, dengan dua metode analisis yaitu dengan analisis secara teoritik yang meliputi keseusian isi soal dengan Standar Kmpetensi dan Kompetensi Dasar yang berlaku dan kaidah

penulisan soal yang benar, dan analisis secara empiris meliputi analisis tingkat kesukaran, daya pembeda, validitas, reliabilitas, dan efektifitas fungsi pengecoh pada soal pilihan ganda.⁴¹

2. Skripsi berjudul “Analisis Butir Soal Pada Penilaian Akhir Semester Ganjil Mata Pelajaran PAI kelas XI di SMA Negeri 10 Malang Tahun ajaran 2017/2018”, yang ditulis oleh Eka Fitrianiingsih, Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Maulana Malik Ibrahim Malang 2018. Skripsi ini membahas mengenai kualitas butir soal penilaian akhir semester. Analisis data yang dilakukan dengan menggunakan perhitungan komputer yaitu menggunakan program SPSS dan ANATES. Dalam analisis kualitas butir soal ini ditinjau secara kuantitatif dengan meneliti tingkat validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal.⁴²
3. Jurnal berjudul “Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam di SMP Bintang Persada Tabanan-Bali”, yang ditulis oleh Rusmayani 2019. Jurnal ini membahas mengenai kualitas butir soal penilaian akhir semester genap bentuk pilihan ganda. Analisis data yang digunakan adalah analisis campuran yaitu secara kuantitatif dan kualitatif dengan menyajikan data validitas,

⁴¹ Aziz Muslim, *Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Pendidikan Agama Islam (PAI) Kelas VIII Semester Genap Sekolah Menengah Pertama (SMP) Kota Semarang Tahun Pelajaran 2016/2017*, (Semarang: Jurusan Pendidikan Agama Islam, Fakultas Ilmu Tarbiyah dan Keguruan, UIN Walisongo, 2018)

⁴² Eka Fitrianiingsih, *Analisis Butir Soal Pada Penilaian Akhir Semester Ganjil Mata Pelajaran PAI Kelas XI di SMA Negeri 10 Malang Tahun Ajaran 2017/2018*, (Malang: Jurusan Pendidikan Agama Islam, Fakultas Tarbiyah dan Keguruan, UIN Maulana Malik Ibrahim Malang, 2018)

reliabilitas, daya pembeda, dan taraf kesukaran dalam bentuk numerik yang selanjutnya diinterpretasikan ke dalam kriteria yang telah ditetapkan.⁴³

Persamaan skripsi yang penulis angkat dengan penelitian di atas, yaitu sama-sama meneliti kualitas butir soal. Sedangkan perbedaannya adalah penulis melakukan penelitian dengan soal, tempat, dan subjek yang berbeda serta metode analisis data yang berbeda. Skripsi yang disusun ini, tepat penelitiannya di SDN 1 Bumiharjo, Kecamatan Klirong, Kabupaten Kebumen dengan objek penelitian mengenai analisis butir soal Penilaian Tengah Semester dalam mata pelajaran Pendidikan Agama Islam (PAI) semester genap kelas V secara kuantitatif. Sedangkan subjeknya adalah, siswa kelas V SDN 1 Bumiharjo Tahun Ajaran 2021/2022. Untuk metode analisis datanya secara kuantitatif dengan menggunakan program Anates versi 4.0.

C. Hipotesis Penelitian

Hipotesis adalah dugaan atau jawaban sementara terhadap permasalahan yang akan diteliti.⁴⁴ Hipotesis atau kesimpulan dari penelitian ini adalah hasil penilaian butir soal Penilaian Tengah Semester (PTS) mata pelajaran Pendidikan Agama Islam (PAI) semester genap kelas V SDN 1 Bumiharjo tahun pelajaran 2021/2022 memiliki hasil sesuai dengan standar

⁴³ Rusmayani, *Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam di SMP Bintang Persada Tabanan-Bali*, (Bali: Sekolah Tinggi Agama Islam Denpasar Bali, 2019) hal. 5.

⁴⁴ Syahrur, Salim, *Metode Penelitian Kuantitatif*, cet pertama (Bandung: Citapustaka Media, 2014), hal. 98.

dalam menyusun soal dari segi validitas, reliabilitas, taraf kesukaran, daya pembeda, dan fungsi pengecoh pada soal pilihan ganda.

D. Instrumen Penelitian

Penelitian kuantitatif adalah penelitian yang menggunakan data berupa angka sebagai alat untuk menemukan keterangan mengenai apa yang diteliti.⁴⁵ Instrumen penelitian adalah alat yang digunakan untuk melakukan kegiatan penelitian terutama sebagai pengukuran dan pengumpulan data. Kegunaan instrumen penelitian adalah untuk memperoleh data yang diperlukan ketika peneliti sudah menginjak pada langkah pengumpulan informasi di lapangan.⁴⁶ Instrumen penelitian kuantitatif berkaitan dengan validitas dan reliabilitas instrumen. Bentuk instrumen dalam penelitian kuantitatif biasanya berupa tes, pedoman wawancara, pedoman observasi, dan kuisioner. Dalam penelitian ini, menggunakan tes Penilaian Tengah Semester Gasal sebagai instrumen yang mana soal dan jawaban dikuantifikasi serta dapat diolah dan dianalisis secara statistik.

⁴⁵ Samsu, *Metode Penelitian Teori Aplikasi Penelitian Kualitatif, Kuantitatif, Mixed Method, serta Research & Development*, cet pertama, (Jambi: Pustaka Jambi, 2017), hal. 125.

⁴⁶ Sukardi, *Metode Penelitian Pendidikan Kompetensi dan Praktiknya*, cet pertama, (Jakarta: Bumi Aksara, 2003), hal. 75