

BAB II

TINJAUAN PUSTAKA

A. Evaluasi

Evaluasi berarti penentuan sampai berapa jauh sesuatu berharga, atau bernilai. Evaluasi terhadap hasil belajar yang dicapai oleh siswa dan terhadap proses mengajar-belajar mengandung penilaian terhadap hasil belajar atau proses belajar itu, sampai berapa jauh keduanya dapat dinilai baik (Winkel, 2004: 531).

Evaluasi juga dapat diartikan suatu proses pemaknaan atau penetapan kualitas hasil belajar dengan cara membandingkannya dengan kriteria atau prosedur tertentu. Hal ini sesuai dengan pendapat Hamzah B Uno (2012:12) dalam bukunya bahwa evaluasi adalah proses pemberian makna atau ketetapan kualitas hasil pengukuran dengan cara membandingkan angka hasil pengukuran tersebut dengan kriteria.

Definisi lain yang berkaitan dengan pengukuran hasil belajar siswa, yaitu *evaluation is process of making and assessment of a student's growth*. Evaluasi merupakan proses penilaian pertumbuhan siswa dalam proses belajar mengajar (Sukardi, 2011: 2).

Evaluasi adalah suatu tindakan atau proses untuk menentukan nilai dari sesuatu. Senada dengan pendapat tersebut, Wayan Nurkencana dan Sumartana (1983) yang dikutip Pupuh Faturrohman (2010: 16), bahwa evaluasi pendidikan dapat diartikan sebagai suatu

tindakan atau suatu proses untuk menentukan nilai segala sesuatu yang ada hubungannya dengan dunia pendidikan.

Jadi pada intinya, evaluasi merupakan kegiatan mengukur dan menilai berdasarkan parameter tertentu yang tujuan akhirnya adalah untuk melihat kualitas/tingkat ketercapaian dari kegiatan yang sudah dijalankan. Apakah sudah sesuai dengan yang telah direncanakan atau belum, apakah kegiatan yang dijalankan sudah layak atau bermanfaat, apakah kegiatan yang dijalankan harus dimodifikasi atau diganti sama sekali.

Menurut Suharsimi Arikunto (2012: 38-39), ada satu prinsip umum dan penting dalam kegiatan evaluasi, yaitu adanya triangulasi atau hubungan erat tiga komponen, yaitu:

1. Hubungan antara Tujuan dengan KBM,

Kegiatan belajar mengajar yang dirancang dalam bentuk rencana mengajar disusun oleh guru dengan mengacu pada tujuan yang hendak dicapai.

2. Hubungan antara Tujuan dengan Evaluasi

Evaluasi adalah kegiatan pengumpulan data untuk mengukur sejauh mana tujuan sudah tercapai.

3. Hubungan antara KBM dengan Evaluasi

Seperti yang sudah disebutkan dalam nomor (1), KBM dirancang dan disusun dengan mengacu pada tujuan yang telah dirumuskan. Telah disebutkan pula dalam nomor (2), bahwa alat evaluasi juga

disusun dengan mengacu pada tujuan. Selain mengacu pada tujuan, evaluasi juga harus mengacu atau disesuaikan dengan KBM yang dilaksanakan.

Sedangkan menurut Sukardi (2011: 4), evaluasi memiliki beberapa prinsip antara lain:

1. Evaluasi harus masih dalam kisi-kisi kerja tujuan yang telah ditentukan.
2. Evaluasi sebaiknya ditentukan secara komprehensif.
3. Evaluasi diselenggarakan dalam proses yang kooperatif antara guru dan peserta didik.
4. Evaluasi dilaksanakan dalam proses kontinu.
5. Evaluasi harus peduli dan mempertimbangkan nilai-nilai yang berlaku.

Hal ini senada dengan pendapat Zaenal Arifin (2009: 30), bahwa untuk memperoleh hasil evaluasi yang lebih baik, maka kegiatan evaluasi harus bertitik tolak pada prinsip-prinsip umum sebagai berikut:

1. **Kontinuitas**

Evaluasi tidak boleh dilakukan secara insidental karena pembelajaran itu sendiri adalah suatu proses yang kontinu.

2. **Komprehensif**

Dalam melakukan evaluasi terhadap suatu objek, guru harus mengambil seluruh objek itu sebagai bahan evaluasi.

3. Adil dan Objektif

Dalam melaksanakan evaluasi, guru harus berlaku adil tanpa pilih kasih.

4. Kooperatif

Dalam kegiatan evaluasi guru hendaknya bekerja sama dengan semua pihak, seperti orang tua peserta didik, sesama guru, kepala sekolah, termasuk dengan peserta didik itu sendiri.

5. Praktis

Praktis mengandung arti mudah digunakan, baik oleh guru itu sendiri yang menyusun alat evaluasi maupun orang lain yang akan menggunakan alat tersebut.

Jadi, dapat dikatakan bahwa dalam melakukan kegiatan evaluasi, kita tidak boleh lepas dari prinsip-prinsip evaluasi, antara lain sesuai dengan tujuan pembelajaran, dilakukan secara kontinu, menyeluruh serta mampu membedakan (terlihat dari hasil evaluasi).

Selain prinsip-prinsip, evaluasi juga mempunyai karakteristik yang membedakannya dari metode-metode analisis kebijakan lainnya yaitu:

1. Fokus nilai. Evaluasi berbeda dengan pemantauan, dipusatkan pada penilaian menyangkut keperluan atau nilai dari sesuatu kebijakan dan program.
2. Interdependensi Fakta-Nilai. Tuntutan evaluasi tergantung baik "fakta" maupun "nilai".

3. Orientasi Masa Kini dan Masa Lampau. Tuntutan evaluatif, berbeda dengan tuntutan-tuntutan advokat, diarahkan pada hasil sekarang dan masa lalu, ketimbang hasil di masa depan.
4. Dualitas nilai. Nilai-nilai yang mendasari tuntutan evaluasi mempunyai kualitas ganda, karena mereka dipandang sebagai tujuan dan sekaligus cara. (Dunn, 2003:608-609)

Adapun menurut Sukardi (2011: 3) kegiatan evaluasi dalam proses belajar mengajar mempunyai beberapa karakteristik penting, diantaranya sebagai berikut:

1. Memiliki implikasi tidak langsung terhadap siswa yang dievaluasi. Hal ini terjadi misalnya seorang guru melakukan penilaian terhadap kemampuan yang tidak nampak dari siswa. Apa yang dilakukan adalah ia lebih banyak menafsir melalui beberapa aspek penting yang diizinkan seperti melalui penampilan, keterampilan, atau reaksi mereka terhadap suatu stimulus yang diberikan secara terencana.
2. Lebih bersifat tidak lengkap. Dikarenakan evaluasi tidak dilakukan secara kontinu maka hanya sebagian fenomena saja. Atau dengan kata lain, apa yang dievaluasi hanya sesuai dengan pertanyaan *item* yang direncanakan oleh seorang guru.
3. Mempunyai sifat kebermaknaan relatif. Ini berarti, hasil penilaian tergantung pada tolok ukur yang digunakan oleh guru. Disamping

itu, evaluasi pun tergantung dengan tingkat ketelitian alat ukur yang digunakan.

Sedangkan menurut Supriyanto (2012), evaluasi yang baik adalah evaluasi yang dapat membantu anak mencapai tujuan proses belajarnya. Oleh karena itu, evaluasi yang baik harus memperhatikan hal-hal ini, yaitu:

1. validitas: alat evaluasi harus sahih yaitu mengukur apa yang hendak diukur dengan tepat,
2. reliabilitas: alat evaluasi harus handal yaitu alat evaluasi harus menghasilkan hasil ukur yang sama jika dipakai kapanpun dan dimanapun,
3. deskriminatif: alat evaluasi harus dapat menunjukkan hasil yang berbeda pada setiap siswa, dan
4. praktis: alat evaluasi harus praktis yaitu mudah digunakan.

Hasil evaluasi dapat digunakan oleh evaluator atau guru untuk memperbaiki proses pembelajaran berikutnya guna pencapaian tujuan pembelajaran.

Menurut Daryanto (2010: 11), bahwa tujuan utama melakukan evaluasi dalam proses belajar mengajar adalah untuk mendapatkan informasi yang akurat mengenai tingkat pencapaian tujuan instruksional oleh siswa sehingga dapat diupayakan tindak lanjutnya. Tindak lanjut termaksud merupakan fungsi evaluasi dan dapat berupa:

1. Penempatan pada tempat yang tepat,
2. Pemberian umpan balik,
3. Diagnosis kesulitan belajar siswa, atau
4. Penentuan kelulusan.

Hal-hal tersebut tidak dapat ditinggalkan atau dihilangkan manakala para pendidik menyusun serta membuat suatu bentuk evaluasi dalam proses pembelajaran.

Menurut Sukardi (2011: 4), evaluasi mempunyai fungsi yang bervariasi di dalam proses belajar mengajar, yaitu sebagai berikut:

1. Untuk mengetahui aspek-aspek kelemahan peserta didik dalam melakukan kegiatan belajar.
2. Mengetahui tingkat ketercapaian siswa dalam kegiatan belajar.
3. Sebagai sarana umpan balik bagi seorang guru, yang bersumber dari siswa.
4. Sebagai alat untuk mengetahui perkembangan belajar siswa.
5. Sebagai materi utama laporan hasil belajar kepada para orang tua siswa.

Sedangkan menurut Zaenal Arifin (2009:16-17), fungsi evaluasi adalah sebagai berikut:

1. Secara psikologis, peserta didik selalu butuh untuk mengetahui sejauh mana kegiatan yang telah dilakukan sesuai dengan tujuan yang hendak dicapai.

2. Secara sosiologis, evaluasi berfungsi untuk mengetahui apakah peserta didik sudah cukup mampu untuk terjun ke masyarakat.
3. Secara deduktis-metodis, 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.
4. Evaluasi berfungsi untuk mengetahui kedudukan peserta didik dalam kelompok, apakah dia termasuk anak yang pandai, sedang, atau kurang pandai.
5. Evaluasi berfungsi untuk mengetahui taraf kesiapan peserta didik dalam menempuh program pendidikannya.
6. Evaluasi berfungsi membantu guru dalam memberikan bimbingan dan seleksi, baik dalam rangka menentukan jenis pendidikan, jurusan, maupun kenaikan kelas.
7. 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 sendiri.

Guru dituntut mampu mengolah hasil evaluasi sebagai alat atau sarana yang dapat memudahkannya untuk mengelola kegiatan pembelajaran yang lebih baik. Evaluasi berfungsi memberikan berbagai informasi kepada seluruh elemen pendidikan, yang nantinya

hasil evaluasi tersebut ditindak lanjuti oleh pihak yang berwenang dalam rangka memperbaiki serta mengoptimalkan tujuan yang ingin dicapai.

B. Tes dan Butir Soal

Tes adalah seperangkat tugas yang harus dikerjakan atau sejumlah pertanyaan yang harus dijawab oleh peserta didik untuk mengukur tingkat pemahaman dan penguasaan terhadap cakupan materi yang dipersyaratkan dan sesuai dengan tujuan pengajaran tertentu (Uno, 2012: 3).

Menurut Suaharsimi Arikunto (2012), tes adalah merupakan alat atau prosedur yang digunakan untuk mengetahui atau mengukur sesuatu dalam suasana, dengan cara dan aturan-aturan yang sudah ditentukan.

Tes merupakan salah satu alat yang di dalamnya terdapat cara atau prosedur dalam rangka penilaian atau pengukuran dalam dunia pendidikan, baik dalam bentuk pemberian tugas atau rangkaian yang diberikan kepada siswa, sehingga diperoleh hasil yang dapat digunakan sebagai tolok ukur kemampuan siswa selama dalam proses pembelajaran. Sebagaimana pendapat Darajat dalam bukunya (2005: 17), tes adalah alat, instrumen atau prosedur yang digunakan dalam rangka pengukuran dan penilaian hasil belajar. Alat tersebut dipersyaratkan memiliki kualitas yang baik atau mempunyai standar yang obyektif sehingga dapat digunakan secara meluas, serta dapat

betul-betul digunakan untuk mengukur dan membandingkan keadaan psikis atau tingkah laku individu siswa.

Tes yang bermutu adalah tes yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuannya diantaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru.

Menurut Asmawi Zainul dan Noehl Nasution (1997: 17), bahwa tes hasil belajar (*achievement test*) haruslah disusun atas butir-butir soal yang dipilih, yang secara akademik dapat dipertanggungjawabkan sebagai sampel yang representatif dari ilmu atau bidang studi yang diuji dengan perangkat tes tersebut.

Menurut Sukardi dalam bukunya (2011: 135), tes pendidikan disusun dengan menggunakan sejumlah *item*. Skor pada setiap *item* ditambahkan pada skor item lainnya untuk mencapai skor total. Tes yang digunakan untuk tujuan evaluasi sebaiknya memenuhi dua persyaratan penting, yaitu validitas dan reliabilitas.

Sedangkan menurut Asmawi Zainul dan Noehl Nasution (1997:17), tes baru akan berarti bila terdiri dari butir-butir soal yang menguji tujuan yang penting dan mewakili ranah pengetahuan, kemampuan dan keterampilan secara representatif.

Dengan demikian, dalam pembuatan tes harus memperhatikan butir-butir soal yang ada di dalamnya yang menyangkut maksud, tujuan serta mengacu pada materi yang diujikan.

Tes memiliki tipe-tipe tertentu, menurut Ebel dan Frisbie yang dikutip Asmawi Zainul dan Noehl Nasution (1997: 18), membagi tiga tipe soal: (1) esai, (2) objektif, dan (3) problem matematik. Disamping itu masih juga dikenal soal-soal lisan.

Sedangkan menurut Sukardi (2005: 93) Suatu bentuk tes dapat digolongkan menjadi dua macam, yaitu tes tertulis dan tes lisan.

1. Tes tertulis adalah sekumpulan item pertanyaan dan atau pertanyaan yang direncanakan guru maupun para evaluator secara sistematis, guna memperoleh informasi tentang siswa.
2. Tes lisan merupakan sekumpulan item pertanyaan dan atau pertanyaan yang disusun secara terencana, diberikan oleh seorang guru kepada para siswanya tanpa melalui media tulis.

Tes tertulis tidak lain seperti yang biasa kita temui dapat berupa tes tulis, tes psikologi, tes pengetahuan umum dan lain-lain. Sedangkan tes lisan yang dimaksud adalah tes yang di berikan oleh guru tanpa media tulis; berupa tes keterampilan, tes wawancara dan lain sebagainya. Pemilihan tipe tes yang akan digunakan lebih banyak ditentukan oleh kemampuan dan waktu yang tersedia pada penyusun tes dari pada peserta tes atau aspek yang ingin diatur.

Tidak kalah pentingnya dalam pembuatan soal disinggung masalah jumlah butir soal. Menurut Asmawi Zainul dan Noehl Nasution (1997: 20) untuk jumlah butir soal tidak ada ketentuan yang pasti. Tetapi yang harus diingat adalah jumlah butir soal berhubungan

langsung dengan reliabilitas tes dan representasi isi bidang studi yang dites. Jumlah butir soal disesuaikan dengan tujuan tes serta bidang studi yang bersangkutan, artinya jumlah butir soal bersifat relatif sesuai dengan kebutuhan.

C. Analisis Butir Soal (*Item Analysis*)

Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Tujuan penelaahan/penganalisisan adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum soal digunakan. Disamping itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada siswa apakah mereka sudah atau belum memahami materi yang telah diajarkan (Deptdiknas, 2008: 9).

Menurut Anas Sudijono (2006: 370), penganalisisan terhadap butir-butir item tes hasil belajar dapat dilakukan dari tiga segi, yaitu: (1) dari segi derajat kesukaran itemnya, (2) dari segi daya pembeda itemnya, (3) dari segi fungsi distraktornya.

Adapun fungsi dari analisis butir soal yaitu menurut Thorndike dan Hagen (1977) yang dikutip Ngalim Purwanto (2010: 118), analisis terhadap soal-soal (*items*) tes yang telah dijawab oleh siswa mempunyai dua tujuan penting. Pertama, jawaban-jawaban soal itu merupakan informasi diagnostik untuk meneliti pelajaran dari kelas itu

dan kegagalan-kegagalan belajarnya, serta selanjutnya untuk membimbing ke arah cara belajar yang lebih baik. Kedua, jawaban-jawaban terhadap soal-soal yang terpisah dan perbaikan (*review*) soal-soal yang didasarkan atas jawaban-jawaban itu merupakan basis bagi penyiapan tes-tes yang lebih baik untuk tahun berikutnya.

Analisis butir soal dapat dibedakan menjadi dua, yaitu analisis butir soal kualitatif dan analisis butir soal kuantitatif. Analisis butir soal kualitatif adalah analisis butir soal yang dilaksanakan berdasarkan kaidah penulisan soal (tes tertulis, perbuatan, dan sikap). Penelaahan/penganalisisan ini biasanya dilakukan sebelum soal digunakan atau diujikan. Menurut Depdiknas (2008: 4), Aspek yang diperhatikan di dalam penelaahan secara kualitatif ini adalah setiap soal ditelaah dari segi materi, konstruksi, bahasa/budaya, dan kunci jawaban/pedoman penskorannya. Dalam melakukan penelaahan setiap butir soal, penelaah perlu mempersiapkan bahan-bahan penunjang seperti: (1) kisi-kisi tes, (2) kurikulum yang digunakan, (3) buku sumber, dan (4) kamus bahasa Indonesia.

Analisis butir soal yang kedua adalah analisis butir soal kuantitatif yaitu penelaahan/penganalisisan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empirik ini diperoleh dari soal yang telah diujikan (Depdiknas, 2008: 11).

Dalam penelitian ini peneliti menggunakan analisis butir soal secara kuantitatif yaitu peneliti menganalisis butir soal ujian akhir

sekolah yang telah diujikan pada tahun pelajaran 2011/2012. Ada dua pendekatan dalam analisis butir soal secara kuantitatif, yaitu pendekatan secara klasik dan modern.

Analisis butir soal dengan pendekatan klasik adalah proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori tes klasik.

Menurut Millman dan Greene (1993) yang dikutip Depdiknas (2008: 4), kelebihan analisis butir soal secara klasik adalah murah, dapat dilaksanakan sehari-hari dengan cepat menggunakan komputer, sederhana, familier dan dapat menggunakan data dari beberapa peserta didik atau sampel kecil.

Aspek yang perlu diperhatikan dalam analisis butir soal secara klasik adalah setiap butir soal ditelaah dari segi: tingkat kesukaran butir, daya pembeda butir, dan penyebaran pilihan jawaban (untuk soal bentuk obyektif) atau frekuensi jawaban pada setiap pilihan jawaban.

1. Validitas

Alat ukur dikatakan valid apabila alat ukur itu dapat dengan tepat mengukur apa yang hendak diukur. Dengan kata lain, validitas berkaitan erat dengan “ketepatan” dengan alat ukur. Tes sebagai salah satu alat ukur hasil belajar dapat dikatakan valid

apabila tes dapat mengukur hasil belajar yang hendak diukur (Widoyoko, 2009: 98).

Hal ini senada dengan Pendapat Suharsimi Arikunto (2012: 73), bahwa sebuah tes dikatakan valid apabila tersebut dapat tepat mengukur apa yang hendak diukur.

Sedangkan menurut Waridjan (1991: 342) sesuatu tes hasil belajar sebagai alat ukur dinyatakan valid apabila tes hasil belajar itu hanya mengukur secara murni variabel yang memang seharusnya diukur, tidak mengukur variabel-variabel lain yang memang seharusnya tidak diukur.

Maksudnya sebuah tes tidak bisa dipergunakan untuk mengukur sesuatu yang hendak diukur bisa dikatakan tes tersebut tidaklah valid atau dengan kata lain tes tersebut tidak layak untuk digunakan.

Berbeda dengan pendapat Zaenal Arifin (2009: 247), bahwa validitas suatu tes erat hubungannya dengan tujuan penggunaan tes tersebut. Namun, tidak ada validitas yang berlaku secara umum. Artinya, jika suatu tes dapat memberikan informasi yang sesuai dan dapat digunakan untuk mencapai tujuan tertentu, maka tes itu valid untuk tujuan tersebut.

Senada dengan pendapat *Cronbach* yang dikutip Ngalim Purwanto dalam bukunya (2010: 138): “*How well a test or evaluative technique does the job that it is employed to do.*”

Validitas bukanlah ciri atau sifat mitlak dari suatu teknik evaluasi; ia merupakan suatu ciri yang relatif terhadap tujuan yang hendak dicapai oleh pembuat tes.

Validitas sebuah tes tidak dapat dipandang dari satu sisi saja melainkan validitas tersebut juga dapat dilihat dari tujuan dilakukannya tes tersebut.

Menurut Kerlinger (1986) yang dikutip Zaenal Arifin (2009: 248) mengemukakan, “validitas instrumen tidak cukup ditentukan oleh derajat ketepatan instrumen untuk mengukur apa yang seharusnya diukur, tetapi perlu juga dilihat dari tiga kriteria yang lain, yaitu *appropriateness*, *meaningfulness*, dan *usefulness*.” *Appropriateness* menunjukkan kelayakan dari tes sebagai suatu instrumen, yaitu seberapa jauh instrumen dapat menjangkau keragaman aspek perilaku peserta didik. *Meaningfulness* menunjukkan kemampuan instrumen dalam memberikan keseimbangan soal-soal pengukurannya berdasarkan tingkat kepentingan dari setiap fenomena. *Usefulness* menunjukkan sensitif tidaknya instrumen dalam menangkap fenomena perilaku dan tingkat ketelitian yang ditunjukkan dalam membuat kesimpulan.

Menurut Waridjan (1991: 342), hal-hal yang menjadikan suatu tes hasil belajar kurang valid ialah faktor-faktor internal tes itu sendiri, yang pertama rumusan petunjuk mengerjakan soal

yang kurang jelas dan rumusan soal yang membingungkan dalam mengemukakan masalah. Faktor eksternal tes yang mengurangi kesahihan atau kevalidan tes hasil belajar ialah penetapan kunci jawaban yang salah, pemberian skor yang kurang adil melalui penjatahan skor yang sama bagi soal-soal yang mudah, yang sedang dan yang sukar dan lebih dari sebuah jawaban, penyediaan waktu tes yang kurang memadai dan kurang terkontrolnya peserta tes untuk melakukan tindakan-tindakan tidak jujur dalam mengerjakan soal-soal.

Jadi, tingkat kevalidan atau validitas soal pada ujian akhir sekolah tersebut yang tergolong rendah, bisa dikarenakan belum memperhatikan tiga kriteria tersebut atau hanya terfokus pada validitas dalam arti kemampuan instrumen mengukur hasil yang diperoleh siswa.

Berkenaan dengan koefisien yang menunjukkan nilai validitas sebuah tes, menurut Ngalim Purwanto (2010: 139), validitas suatu tes dinyatakan dengan angka koefisien korelasi (r). adapun koefisien korelasi adalah sebagai berikut:

- a. 0,00 – 0,20 sangat rendah (hampir tidak ada korelasi)
- b. 0,20 – 0,40 korelasi rendah
- c. 0,40 – 0,70 korelasi cukup
- d. 0,70 – 0,90 korelasi tinggi
- e. 0,90 – 1,00 korelasi sangat tinggi (sempurna)

Nilai koefisien korelasi suatu tes menunjukkan derajat atau tingkatan validitas sebuah tes, makin tinggi nilai koefisiennya (mendekati angka 1) maka semakin valid tes tersebut.

2. Reliabilitas

Kata reliabilitas dalam bahasa Indonesia diambil dari kata *reliability* dalam bahasa Inggris, berasal dari kata asal *reliable* yang artinya dapat dipercaya. Seorang dikatakan dapat dipercaya jika orang tersebut selalu bicara ajek (konsisten), tidak berubah-ubah pembicaraannya dari waktu ke waktu. Demikian juga halnya dengan sebuah tes. Tes tersebut dikatakan dapat dipercaya (*reliable*) jika memberikan hasil yang tetap atau ajek (*consistent*) apabila diteskan berulang-ulang (Widoyoko, 2009: 99).

Menurut Nana Sudjana (2010: 16) reliabilitas alat penilaian adalah ketetapan atau keajegan alat tersebut dalam menilai apa yang dinilainya. Artinya, kapanpun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama.

Sebagaimana pendapat dari Sukardi dalam bukunya (2010: 127) bahwa reliabilitas sama dengan konsistensi atau keajegan. Suatu instrumen penelitian dikatakan mempunyai nilai reliabilitas yang tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur.

Menurut Depdiknas (2008), ada 3 cara yang dapat dilakukan untuk menentukan reliabilitas skor tes, yaitu :

- a. Keajegan pengukuran ulang: kesesuaian antara hasil pengukuran pertama dan kedua dari sesuatu alat ukur terhadap kelompok yang sama.
- b. Keajegan pengukuran setara: kesesuaian hasil pengukuran dan dua atau lebih alat ukur berdasarkan kompetensi kisi-kisi yang lama.
- c. Keajegan belah dua: kesesuaian antara hasil pengukuran belahan pertama dan belahan kedua dari alat ukur yang sama.

Suatu alat atau instrumen penilaian dapat dikatakan memiliki tingkat reliabilitas yang tinggi apabila ia mampu mempertahankan keajekan atau konsistensinya dalam mengukur sesuatu yang hendak diukur.

Menurut Kaplan dan Sacuzzo (1978) yang dikutip Mawardi Lubis (2008: 65) bahwa besarnya nilai reliabilitas minimum pada sebuah soal sebesar 0,7.

Sedangkan menurut Anas Sudijono (2008), Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan keajegan (*consistency*) skor tes. Indeks reliabilitas berkisar antara 0 - 1. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), makin tinggi pula keajegan/ketepatannya.

Dengan demikian, semakin tinggi nilai reliabilitas suatu tes yaitu mendekati 1, maka bisa disimpulkan bahwa tersebut memiliki reliabilitas (ketepatan/ keajegan)

Menurut Waridjan (1991: 362), faktor-faktor yang mengurangi reliabilitas suatu tes hasil belajar itu, karena tes hasil belajar itu tidak mengukur apa yang seharusnya diukur (*invalid*), hasil pengukuran yang berkali-kali dengan tes hasil belajar itu akan berubah-ubah.

Disamping itu, menurut Wood (1961: 12-14) yang dikutip oleh Waridjan (1991: 362), kurang reliabelnya suatu tes hasil belajar disebabkan oleh subjektivitas pemberi skor, kurang representatifnya soal-soal tes dalam mewakili bahan belajar-mengajar dan masih bersifat sementara tindakan yang diukur sebagai hasil belajar. Sesuai dengan pendapat Suharsimi Arikunto (2012: 76) bahwa semua hal yang mempengaruhi hasil tes akan mempengaruhi reliabilitas soal.

3. Tingkat Kesukaran

Perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proposional), maka dapat dikatakan bahwa soal tersebut baik (Arifin, 2011: 267)

Tingkat kesulitan *item* atau disebut juga indeks kesulitan *item* adalah angka yang menunjukkan tingkat kesulitan proporsi

siswa yang menjawab betul dalam suatu soal yang dilakukan dengan menggunakan tes objektif (Sukardi, 2011: 136).

Menurut Suharsimi Arikunto (2012), soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya.

Tingkat kesukaran dapat dinyatakan melalui beberapa cara diantaranya (1) proporsi menjawab benar, (2) skala kesukaran linear, (3) indeks Davis, (4) skala bivariat (Supranata, 2006: 12) .

Tingkat kesukaran mampu menggambarkan kualitas soal dilihat dari segi kemampuan siswa/testee dalam mengerjakan atau menyelesaikan tes/ujian.

Dalam hubungannya dengan distribusi tingkat kesukaran harus diperhatikan bahwa tes yang memiliki tingkat kesukaran yang rendah sebaiknya diletakkan di awal perangkat tes dan yang tinggi pada akhir perangkat tes (Zainul, 1997: 21).

Menurut Nitko (1996) yang dikutip Depdiknas (2008: 12) semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki $TK = 0,00$ artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki $TK = 1,00$ artinya bahwa siswa menjawab

benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal. Pada prinsipnya, skor rata-rata yang diperoleh peserta didik pada butir soal yang bersangkutan dinamakan tingkat kesukaran butir soal itu.

Sedangkan menurut Daryanto (2010: 180), bahwa di dalam istilah evaluasi tingkat kesukaran diberi symbol P, singkatan dari "*proporsi*". Adapun kriteria untuk tingkat kesukaran *item* adalah sebagai berikut.

- a. Item dengan $P = 0,00$ sampai $0,30$ tergolong sukar
- b. Item dengan $P = 0,31$ sampai $0,70$ tergolong sedang
- c. Item dengan $P = 0,71$ sampai $1,00$ tergolong mudah.

Tingkat kesukaran butir soal memiliki dua kegunaan, yaitu kegunaan bagi guru dan kegunaan bagi pengujian dan pengajaran. Kegunaannya bagi guru adalah: (1) sebagai pengenalan konsep terhadap pembelajaran ulang dan memberi masukan kepada siswa tentang hasil belajar mereka, (2) memperoleh informasi tentang penekanan kurikulum atau mencurigai terhadap butir soal yang bias. Adapun kegunaannya bagi pengujian dan pengajaran adalah: (a) pengenalan konsep yang diperlukan untuk diajarkan ulang, (b) tanda-tanda terhadap kelebihan dan kelemahan pada kurikulum sekolah, (c) memberi masukan kepada siswa, (d) tanda-tanda kemungkinan adanya butir soal yang bias, (e) merakit tes yang memiliki ketepatan data soal.

Kesukaran soal-soal hasil belajar dapat disusun mulai dari soal-soal yang mudah, soal-soal yang sedang sampai dengan soal-soal yang sukar. Menurut Waridjan (1991: 383), soal yang susah menurut pembuat soal bisa jadi mudah menurut siswa atau testee, mudahnya soal itu mungkin berasal dari kerja sama peserta tes yang berupa contek-mencontek, *ngawur* yang kebetulan benar, pembocoran isi tes dan penghafalan kunci jawaban tes.

Dengan demikian, dalam hal penyusunan soal perlu diperhatikan urutan serta komposisi soal yaitu dari yang mudah, sedang kemudian yang terakhir yang sulit. Jadi, siswa sedikit terbantu dengan susunan tersebut yaitu disamping mudah di pahami juga waktu yang digunakan untuk menjawab soal menjadi lebih efektif dan efisien.

4. Daya Pembeda

Menurut Zaenal Arifin (2011: 273), 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.

Menurut Barnard (1999) yang dikutip Sukardi (2011: 138), daya pembeda adalah koefisien yang memberikan informasi tentang pembeda secara individual, termasuk membedakan antara

siswa yang pencapaiannya tinggi dengan siswa yang pencapaiannya rendah dalam suatu tes pencapaian hasil belajar.

Menurut Hamzah B Uno dan Satria Koni (2012: 177), analisis daya pembeda adalah pengkajian butir-butir soal yang dimaksudkan untuk mengetahui kesanggupan siswa untuk membedakan siswa yang tergolong mampu dengan siswa yang tergolong tidak mampu (Uno, 2012: 177).

Sedangkan menurut Sumarna Supranata (2006: 23) bahwa indeks daya pembeda dihitung atas dasar pembagian kelompok menjadi dua bagian, yaitu kelompok atas yang merupakan kelompok peserta tes yang berkemampuan tinggi dengan kelompok bawah yaitu kelompok peserta tes yang berkemampuan rendah.

Senada dengan pendapat Depdiknas (2008), Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga belajar/siswa yang tidak/kurang/belum menguasai materi yang ditanyakan.

Daya pembeda dimaksudkan untuk mengklasifikasikan testee/peserta tes antara yang *high class* dan *low class* dilihat dari sisi kemampuan untuk menjawab butir soal dalam tes. Daya pembedapun baik secara langsung atau tidak sangat membantu guru terutama dalam hal *placement*/penempatan siswa misalnya.

Dengan demikian, daya pembeda adalah suatu angka atau koefisien yang menunjukkan seberapa jauh sebuah soal mampu membedakan siswa yang pandai dan siswa yang kurang pandai.

Adapun manfaat daya pembeda butir soal adalah seperti berikut ini.

- a. Untuk meningkatkan mutu setiap butir soal melalui data empiriknya.
- b. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak. mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru.

Menurut Arifin (2009: 274), untuk menginterpretasikan koefisien daya pembeda, dapat menggunakan kriteria yang dikembangkan Ebel sebagai berikut:

- a. 0,40 ke atas : sangat baik
- b. 0,30 - 0,39 : baik
- c. 0,20 – 0,29 : cukup
- d. $\leq 0,19$: jelek

Dengan demikian, suatu soal dikatakan memiliki daya pembeda baik ketika ia mampu membedakan antara siswa yang pandai dengan siswa yang kurang pandai secara tepat.

Butir soal yang tidak memiliki daya pembeda diduga terlalu mudah atau terlalu sukar sehingga perlu diperbaiki atau diganti dengan pertanyaan lain. Idealnya semua butir soal memiliki daya pembeda dan tingkat kesukaran (Sudjana, 2011: 144).

Pendekatan analisis butir soal selanjutnya ialah dengan pendekatan modern yaitu analisis butir soal dengan moderen yaitu penelaahan/penganalisan butir soal dengan menggunakan *Item Response Theory* (IRT) atau teori jawaban butir soal. Teori ini merupakan suatu teori yang menggunakan fungsi matematika untuk menghubungkan antara peluang menjawab benar suatu soal dengan kemampuan siswa. Nama lain IRT adalah *latent trait theory* (LTT), atau *characteristics curve theory* (ICC).

Menurut Mc. Donald (1999) yang dikutip Depdiknas (2008), bahwa asal mula IRT adalah kombinasi suatu versi hukum phi-gamma dengan suatu analisis faktor butir soal (*item factor analisis*) kemudian bernama Teori Trait Latent (*Latent Trait Theory*), kemudian sekarang secara umum dikenal menjadi teori jawaban butir soal (*Item Response Theory*).

D. Ujian Akhir Sekolah Berstandar Nasional (UASBN)

Ujian Sekolah Berstandar Nasional (UASBN) adalah ujian nasional yang dilaksanakan secara terintegrasi dengan pelaksanaan ujian sekolah/madrasah. Menurut Arniati Persedyawati Herkusumo

(2011: 455), Ujian Akhir Sekolah Berstandar Nasional (UASBN) diselenggarakan dalam rangka memenuhi Undang-Undang Nomor 20 Tahun 2003 tentang sistem pendidikan nasional. Dengan kata lain, UASBN adalah salah satu ujian yang dilaksanakan secara terstruktur oleh sekolah untuk mengetahui kompetensi siswa serta tingkat ketercapaian tujuan pembelajaran yang bertaraf nasional sebagai salah satu wujud pelaksanaan UU No 20 tahun 2003 tentang sistem pendidikan nasional.

Ujian Akhir Sekolah Berstandar Nasional (UASBN) mulai diberlakukan pada tahun pelajaran 2007/2008, di mana propinsi diberi kewenangan untuk menyusun soal-soal ujian berdasarkan kisi-kisi yang sama yang dikeluarkan oleh BNSP (Syamsudin, 2009: 17). Jadi, propinsilah yang bertanggung jawab penuh atas soal-soal ujian, sedangkan sekolah hanya sebagai pihak pelaksana Ujian Sekolah Berstandar Nasional (UASBN), seperti halnya UASBN Pendidikan Agama Islam yang dilaksanakan di SMP Negeri 8 Purwokerto, yang mana sekolah tersebut hanya sebagai penyelenggara dan bukan pembuat soal. SMP Negeri 8 Purwokerto hanya sebagai tempat peneliti melakukan penelitian.

Adapun Tujuan dilaksanakannya UASBN adalah antara lain untuk menilai pencapaian kompetensi lulusan secara nasional, keterbandingan antar sekolah, antar daerah dan antar tahun pada mata pelajaran yang diujikan (Herkusumo, 2011: 246).

Melalui UASBN semua orang akan mampu mengetahui mutu sebuah pendidikan (sekolah), dalam hal ini mutu secara akademis dari hasil ujian tersebut. Hal ini pula yang menjadi alasan peneliti tertarik untuk menganalisis butir soal UASBN namun peneliti membatasi hanya pada mata pelajaran Pendidikan Agama Islam.

E. Program *Iteman*

Analisis butir soal selain menggunakan cara manual yaitu dengan rumus juga dapat dilakukan dengan komputer maksudnya adalah penelaahan butir soal secara kuantitatif yang penghitungannya menggunakan bantuan program komputer. Analisis data dengan menggunakan program komputer sangat tepat. Karena tingkat keakuratan hitungan dengan menggunakan program komputer lebih tinggi bila dibandingkan dengan diolah secara manual. Program komputer yang digunakan untuk menganalisis data, modelnya bermacam-macam tergantung tujuan dan maksud analisis yang diperlukan.

Program komputer yang sudah dikenal secara umum adalah EXCEL, SPSS (*Statistical Program for Social Science*), atau program khusus seperti ITEMAN (analisis secara klasik), RASCAL, ASCAL, BILOG (analisis secara *item* respon teori atau IRT), FACETS (analisis model Rasch untuk data kualitatif). Contoh program analisis data dengan menggunakan komputer, antara lain: program ITEMAN, RASCAL, ASCAL, BIGSTEP, QUEST

(Depdiknas, 2008: 28). Dalam penelitian ini peneliti menggunakan program *IteMan*.

Item And Test Analysis (IteMan) merupakan perangkat lunak (*software*) yang dibuat melalui bahasa pemrograman komputer dan dibuat khusus untuk analisis butir soal dan tes.

Hasil analisis meliputi: tingkat kesukaran butir soal, daya pembeda soal, statistik sebaran jawaban, kehandalan/reliabilitas tes, kesalahan pengukuran (*standar error*), dan distribusi skor serta skor setiap peserta tes (Hidayati, 2012: 1).

IteMan merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik. Program ini termasuk satu paket program dalam MicroCAT[®] yang dikembangkan oleh *Assessment Systems Corporation* mulai tahun 1982 dan mengalami revisi pada tahun 1984, 1986, 1988, dan 1993; mulai dari versi 2.00 sampai dengan versi 3.50. Alamatnya adalah *Assessment Systems Corporation*, 2233 University Avenue, Suite 400, St Paul, Minnesota 55114, United States of America (Salirawati, 2012: 1).

Sedang menurut *Assessment Systems Corporation* (2006: 1), "*IteMan is one of the analysis programs that comprise Assessment Systems Corporation's Item and Test Analysis. "iteMan adalah salah satu program analisis yang terdiri dari sistem penilaian dan analisis data. IteMan menganalisis item data dengan cara konvensional (misalnya, proporsi/persentase mendukung dan item-total korelasi)*

untuk setiap *item*, untuk membantu dalam menentukan sejauh mana *item* berkontribusi dalam kevalidan tes, sedangkan alternatif respon berfungsi dengan baik untuk setiap item.

Mengolah atau memasukan data (*input*) kedalam program *iteman* harus menggunakan format tertentu. Menurut *Assessment Systems Corporation* (2006), format yang tepat untuk file input adalah serupa untuk semua *Item* dan program Analisis Uji Paket. Satu-satunya perbedaan dalam format mencerminkan kemampuan *Iteman* untuk menganalisis lebih dari satu skala pada satu waktu dan untuk menganalisis data untuk *multipoint* atau rating skala-jenis item.

Tahap awal mengoperasikan program *Iteman* adalah membuat "file data" (*control tile*) yang berisi 5 komponen utama.

1. Baris pertama adalah baris pengontrol yang mendeskripsikan data.
2. Baris kedua adalah daftar kunci jawaban setiap butir soal.
3. Baris ketiga adalah daftar jumlah *option* untuk setiap butir soal.
4. Baris keempat adalah daftar butir soal yang hendak dianalisis (jika butir yang akan dianalisis diberi tanda Y (yes), jika tidak diikuti dalam analisis diberi tanda N (no)).
5. Baris kelima dan seterusnya adalah data siswa dan pilihan jawaban siswa.

Setiap pilihan jawaban siswa (untuk soal bentuk pilihan ganda) diketik dengan menggunakan huruf, misal ABCD atau angka

1234 untuk 4 pilihan jawaban atau ABCDE atau 12345 untuk 5 pilihan jawaban (Depdiknas, 2008: 29).

Cara termudah adalah menggunakan program Windows yaitu dengan mengetik data di tempat Notepad. Caranya adalah klik Start-Programs- Accessories-Notepad.

Contoh pengetikan data untuk soal bentuk pilihan ganda

```
25 0 N 10
ABDCEBCEDAABEDCCBDBAEDCAB Kunci
55555555555555555555555555555555 Pilihan
YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
IWAN ABDCEBCEDAABEDCEADBAEEECB
TIKA ACCEBCDBAABEECBDBAEEAAB
YENNY ABDDBCEDAABCACCBDDBCDCAB
WIJI ACBCEBCEDDCEEDCCAADAEDBBB
HENNY ABDCECBDDAABDEACBDBBBECAB
UJANG CDDCEBCEDCDCEDCCBBCADDCAE
NIKEN CDDCEBACDAABEBBCBDBAADAAB
MIMIK ABDDBCEDAABCACCBDDBCDCAB
NUR ABDBCDCEDAABBBCDCBDBAAACAB
RURI AEDEEBCEDBBDEDCCBDCBDCAB
RYSABCDEBCEDAABCACCBDBDEBCAB
ANDRI ACDCEBCECBCBEDCADABAEBBCB
JOKO AAAABBBCCCDDEEAABBCCDDEEA
LUKMAN ACDBEBCECDBBEDCCBBAEDCBB
OTAH DBBCEBAECAABDCBCBDBAEEAAB
AKHMAD ADDCEBCEDCBCDDCCBDBEEDCAB
DENY ABCDABCEDABCBDCCBDEAEDCAB
DEWI ACCBEBCDCBABEDBCEDBDCBCAC
ISMAIL ABDBCDCEDAABBBCDCBDBAACAB
JEMI ACCEBCDBAABEECBDBAEEAAB
```

Langkah kedua data yang telah diketik disimpan, misal disimpan pada file: Tes1.txt. Selanjutnya untuk menggunakan program Iteman yaitu dengan mengklik *icon Iteman*. Kemudian isilah

pertanyaan- pertanyaan yang muncul di layar computer seperti berikut.

```

Enter the name of the input file: Tesl.txt <enter>
Enter the name of the output file: haltesl.txt
<enter>
Do you want the scores written to a file? (Y/N)
Y <enter> Enter the name of the score file:
scrtesl.txt <enter>
**ITEMAN ANALYSIS IS COMPLETE**
  
```

Langkah ketiga adalah membaca hasil, yaitu dengan mengklik *icon hsltes1*. Hasilnya adalah seperti pada contoh berikut.

```

MicroCAT (tm) Testing System
Copyright (c) 1982, 1984, 1986, 1988 by Assessment
t Systems Corporation
Item and Test Analysis Program -- ITEMAN (tm) Ver
sion 3.00
Item analysis for data from file tesl.txt
Page 1
Item Statistics Alternative Statis
tics
-----
  
```

Seq.	Scale	Prop.	Point		Prop.	Point
1	0-1	0.850	-0.018	-0.012	A	0.850 -0.018-0.012*
					B	0.000 -9.000-9.000
					C	0.100 0.047 0.028?
					D	0.050 -0.040-0.019
					E	0.000 -9.000-9.000
					Other	0.000 -9.000-9.000
2	0-2	0.450	0.534	0.425	A	0.050 -1.000-0.856
					B	0.450 0.534 0.425*
					C	0.300 -0.262-0.199
					D	0.150 0.231 0.15
					E	0.050 0.121 0.05
					Other	0.000 -9.000-9.000

3	0-3	0.600	0.515	0.406	A	0.050	-1.000	-0.856
					B	0.100	-0.142	-0.083
					C	0.250	0.039	0.029
					D	0.600	0.515	0.406*
					E	0.000	-9.000	-9.000
					Other	0.000	-9.000	-9.000
4	0-4	0.400	0.172	0.135	A	0.050	-1.000	-0.856
					B	0.200	-0.059	-0.041
					C	0.400	0.172	0.135
					D	0.200	0.474	0.332?
					E	0.150	0.018	0.012
					Other	0.000	-9.000	-9.000
CHECK THE KEY					C	0.400	0.172	0.135
C was specified, D works bet					D	0.200	0.474	0.332?
5	0-5	0.700	0.215	0.163	A	0.050	0.281	0.13
					B	0.050	-1.000	-0.856
					C	0.100	0.142	0.08
					D	0.100	0.331	0.19 ?
					E	0.700	0.215	0.16 *
					Other	0.000	-9.000	-9.000
CHECK THE KEY					C	0.100	0.142	0.08
E was specified, D works bet					D	0.100	0.331	0.19 ?
6	0-6	0.850	-0.089	-0.058	A	0.000	-9.000	-9.000
					B	0.850	-0.089	-0.058*
					C	0.050	-0.040	-0.019
					D	0.100	0.142	0.083?
					E	0.000	-9.000	-9.000
					Other	0.000	-9.000	-9.000
CHECK THE KEY					C	0.050	-0.040	-0.019
B was specified, D works bett					D	0.100	0.142	0.083?
K					E	0.000	-9.000	-9.000
					Other	0.000	-9.000	-9.000

Menurut Supranata dalam bukunya (2006:126), parameter statistik tes dan soal dengan menggunakan program *Iteman* yaitu berupa tingkat kesukaran atau proporsi menjawab benar terlihat pada *prop correct*, validitas soal dan validitas alternatif jawaban terlihat pada *point biserial* dan *biserial* sedangkan distribusi jawaban terlihat pada *prop endorsing*.

Item analysis for data from file tes1.txt

There were 20 examinees in
the data file. Scale
Statistics

Scale: 0

N of Items	25
N of Examinees	20
Mean	16.250
Variance	9.087
Std. Dev.	3.015
Skew	-2.463
Kurtosis	6.976
Minimum	5.000
Maximum	20.000
Median	17.000
Alpha	0.437
SEM	2.261
Mean P	0.650
Mean Item-Tot.	0.266
Mean Biserial	0.352

Pada tampilan data hasil analisis butir soal dengan menggunakan program *Iteman* di atas, diantaranya dapat diketahui jumlah soal, jumlah peserta ujian/siswa, skor maksimum dan minimum hasil ujian serta reliabilitas soal dapat dilihat pada *Alpha*.

Program *iteman* memiliki beberapa kegunaan, diantaranya program ini dapat digunakan untuk: (1) menganalisis data file (format ASCII) jawaban butir soal yang dihasilkan melalui *manual entry* data atau dari mesin scanner; (2) menskor dan menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 siswa dan 250 butir soal; (3) menganalisis sebuah tes yang terdiri dari 10 skala (subtes) dan memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option), reliabilitas

(KR-20/Alpha), *standar error of measurement, mean, variance, standar deviasi, skew, kurtosis* untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor. (Sulirawati, 2012: 1)

F. Pendidikan Agama Islam

Menurut A. Tafsir (2012: 1), untuk melaksanakan pendidikan agama Islam yang berhasil perlu dilakukan pendidikan agama yang terpadu. Keterpaduan yang dimaksud adalah keterpaduan tujuan, keterpaduan materi, dan keterpaduan proses.

1. Keterpaduan Tujuan, berarti pencapaian tujuan pendidikan merupakan tanggung jawab pemangku kepentingan (*Stakeholders*) pendidikan, yaitu pemerintah, kepala sekolah, guru, orang tua siswa dan masyarakat.
2. Keterpaduan Materi, ialah keterpaduan isi kurikulum yang digunakan atau materi pelajaran. Semua materi pelajaran yang dipelajari siswa hendaknya saling memiliki keterkaitan antara satu mata pelajaran dengan mata pelajaran lainnya.
3. Keterpaduan Proses, berarti para pendidik hendaknya menyadari bahwa semua kegiatan pendidikan sekurang-kurangnya tidak berlawanan dengan tujuan pendidikan keimanan dan ketakwaan, bahkan dikehendaki semua kegiatan pendidikan membantu tercapainya siswa yang beriman dan bertakwa.

Pendidikan agama Islam merupakan suatu bentuk usaha yang dilakukan guru atau pihak sekolah dalam menyiapkan siswa untuk mengenal, memahami, menghayati, hingga mengimani, ajaran agama Islam dengan tetap mengarah pada tujuan.

Mata pelajaran pendidikan agama Islam di sekolah, secara keseluruhannya mencakup Al-Qur'an dan al-hadis, keimanan, akhlak, fiqh/ibadah, dan sejarah. Dapat dikatakan bahwa ruang lingkup pendidikan agama Islam mencakup perwujudan keserasian, keselarasan, dan keseimbangan hubungan manusia dengan Allah SWT, diri sendiri, sesama manusia, makhluk lainnya maupun lingkungannya.

Menurut Hasbi Ash-Shidiqi yang dikutip Majid (2005: 138), mengatakan bahwa ruang lingkup pendidikan agama Islam antara lain:

1. *Tarbiyah Jismiyah*, yaitu segala rupa pendidikan yang wujudnya menyuburkan dan menyehatkan tubuh serta menegakannya, supaya dapat merintangangi kesukaran yang dihadapi dalam pengalamannya.
2. *Tarbiyah Aqliyah*, yaitu sebagaimana rupa pendidikan dan pelajaran yang akibatnya mencerdaskan akal menajamkan otak semisal ilmu berhitung.
3. *Tarbiyah Adabiyah*, yaitu segala rupa praktek maupun berupa teori yang wujudnya meningkatkan budi dan meningkatkan perangai.

Bisa dikatakan lingkup pendidikan agama Islam secara umum meliputi, Qur'an, hadis/sunnah, akhlak, muamalah yang nantinya akan diaplikasikan serta disinergikan dengan kehidupan sehari-hari.

Adapun tujuan pendidikan agama Islam adalah untuk meningkatkan pemahaman, keterampilan melakukan, dan pengamalan ajaran Islam dalam kehidupan sehari-hari. Tujuan utama pendidikan agama Islam di sekolah ialah keberagamaan, yaitu menjadi muslim yang sebenarnya (Tafsir, 2012: 1).

Masih menjadi problema ialah bagaimana melaksanakannya seefektif mungkin. Sehingga program pendidikan agama Islam tersebut dapat memperoleh sukses. Kenyataan di lapangan, pelaksanaan pendidikan agama Islam masih banyak mengalami kendala, diantaranya jam pelajaran pendidikan agama Islam yang relatif sedikit yaitu hanya 2 jam setiap minggunya, keterbatasan media pembelajaran serta kemampuan guru dalam menciptakan pembelajaran yang inovatif.

Menurut Arifin dalam bukunya (2011: 197) ada faktor-faktor yang menjadi penghambat perkembangan pendidikan agama Islam di sekolah, antara lain tujuan/cita-cita; alat-alat yang membantu pelaksanaan baik moril maupun materiil pendidikannya. Bagaimanakah mutu keilmiahannya dan personaliti kemampuannya dalam menguasai metode-metode yang mungkin dan yang ada.

Pendidikan agama Islam di sekolah dapat terlaksana dengan baik atau tepat sasaran manakala didukung oleh faktor-faktor terkait, antara lain sekolah berupa penyediaan sarana dan prasarana pendukung, dari pihak guru yaitu menguasai metode, media serta perencanaan pembelajaran yang sesuai dengan tujuan, serta dari siswa sendiri yaitu kesadaran serta ketertarikan dalam mengikuti kegiatan pembelajaran.

Pendidikan agama Islam di sekolah bisa dikatakan sebagai pelajaran penting, karena pendidikan agama Islam sebagai pembentuk pola dasar kepribadian siswa yang nantinya akan sangat berpengaruh terhadap pola hidup siswa saat sekolah maupun kelak setelah mereka terjun di masyarakat.

Sebagaimana Pendapat Abdul Majid dan Dian Andayani (2005: 138), bahwa pendidikan agama Islam merupakan ikhtiar manusia dengan jalan bimbingan dan pimpinan untuk membantu dan mengarahkan fitrah agama si anak didik menuju terbentuknya kepribadian utama sesuai dengan ajaran agama.

Pendidikan agama Islam sangatlah penting sebab dengan pendidikan agama Islam, orang tua atau guru berusaha secara sadar memimpin dan mendidik anak diarahkan kepada bentuk perkembangan jasmani dan rohani sehingga mampu membentuk kepribadian yang utama yang sesuai dengan ajaran Islam. Hal ini senada dengan pendapat Zakiah Darajat (tt:48) bahwa: “pada

umumnya agama seseorang ditentukan oleh pendidikan, pengalaman dan latihan yang dilalui sejak kecil”.

Jadi, perkembangan agama pada seseorang sangat ditentukan oleh pendidikan dan pengalaman hidup sejak kecil; baik dalam keluarga, sekolah maupun dalam lingkungan masyarakat terutama pada masa pertumbuhan dan perkembangannya.

G. Penelitian Terdahulu

Korih Wasilatus Saadah (UMP: 2005) dalam skripsinya yang berjudul “Analisis Butir Soal Ulangan Umum Akhir Semester II Mata Pelajaran Biologi Kelas 1 Tahun Pelajaran 2003/2004 di SMA PGRI 03 Randudongkal Kabupaten Pemalang,” menyimpulkan bahwa soal Ulangan Umum Akhir Semester II Mata Pelajaran Biologi Kelas 1 Tahun Pelajaran 2003/2004 di SMA PGRI 03 Randudongkal Kabupaten Pemalang, tidak layak diujikan dilihat dari validitas, reliabilitas, indeks kesukaran dan daya pembeda serta tipe butir soalnya.

Dalam skripsi tersebut metode pengumpulan data yang digunakan adalah survey. Analisis data yang digunakan yaitu proposionalitas materi dengan kurikulum menggunakan analisis deskriptif kualitatif (untuk memaparkan hasil analisis data) dan analisis validitas, reliabilitas, tingkat kesukaran dan daya pembeda menggunakan rumus manual, salah satunya rumus yang digunakan adalah *Product Moment*. Sedangkan metode penelitian yang

digunakan oleh peneliti dalam penelitian ini adalah observasi dan dokumentasi, analisis data deskriptif kualitatif, sedangkan untuk mencari nilai validitas, reliabilitas, tingkat kesukaran dan daya pembeda, peneliti menggunakan salah satu program komputer yaitu Program *Iteman*.

Dwi Purwanti (UMP: 2005) dalam skripsinya yang berjudul “Upaya Evaluasi Pelaksanaan Kurikulum 1994 Melalui Analisis Butir Soal Ujian Akhir Semester Genap Bidang Studi Biologi Kelas 1 Tahun Pelajaran 2003/2004 Pada SMA Veteran dan SMA Jenderal Soedirman Purwokerto.” Menyimpulkan bahwa ada ketidaksesuaian materi pembelajaran (kurikulum 1994) dengan soal ujian, dan soal ujian akhir semester tidak sesuai dengan prosedur (validitas, reliabilitas, indeks kesukaran dan daya pembeda).

Dalam skripsi tersebut metode pengumpulan data yang digunakan adalah survey. Analisis data yang digunakan yaitu proposionalitas materi dengan kurikulum menggunakan analisis deskriptif kualitatif (untuk mengetahui kesesuaian materi dengan kurikulum 1994) dan analisis validitas, reliabilitas, tingkat kesukaran dan daya pembeda menggunakan rumus manual, salah satunya rumus yang digunakan adalah *Product Moment*. Sedangkan metode penelitian yang digunakan oleh peneliti dalam penelitian ini adalah observasi dan dokumentasi, analisis data deskriptif kualitatif (untuk memaparkan hasil analisis data), sedangkan untuk mencari nilai

validitas, reliabilitas, tingkat kesukaran dan daya pembeda, peneliti menggunakan salah satu program komputer yaitu Program *Iteman*.

Utami Budiastuti (UMP: 2006) dalam skripsinya yang berjudul “Studi Komparasi Hasil Analisis Butir Soal Ujian Semester Genap Mata Pelajaran Biologi Kelas X Tahun Pelajaran 2004/2005 Pada SMA N Jatilawang, SMA N Patikraja dan SMA N Sokaraja,” menyimpulkan bahwa SMA N Jatilawang mempunyai kualitas soal yang lebih baik dibandingkan SMA N Patikraja dan SMA N Sokaraja (dilihat dari kesesuaian materi pembelajaran dengan soal yang diujikan, validitas, reliabilitas, indeks kesukaran dan daya beda) atau bisa dikatakan bahwa pelaksanaan kurikulum 2004/KBK di SMA N Jatilawang berhasil.

Dalam skripsi tersebut metode pengumpulan data yang di gunakan adalah survey. Analisis data yang digunakan yaitu proposionalitas materi dengan kurikulum menggunakan analisis deskriptif kualitatif (untuk mengetahui kesesuaian materi dengan kurikulum 2004/KBK) dan analisis validitas, reliabilitas, tingkat kesukaran dan daya beda menggunakan rumus manual, salah satunya rumus yang digunakan adalah *Product Moment*. Sedangkan metode penelitian yang digunakan oleh peneliti dalam penelitian ini adalah observasi dan dokumentasi, analisis data deskriptif kualitatif (untuk memaparkan hasil analisis data), sedangkan untuk mencari nilai

validitas, reliabilitas, tingkat kesukaran dan daya pembeda, peneliti menggunakan salah satu program komputer yaitu Program *Iteman*.

Dian Fitria Hani (UMP: 2006) dalam skripsinya yang berjudul “Analisis Butir Soal Akhir Semester Genap Kelas X Bidang Studi Biologi Kurikulum 2004 Tahun Pelajaran 2004/2005 di SMA Negeri Sumpiuh Kabupaten Banyumas,” menyimpulkan bahwa secara umum soal yang digunakan di SMA Negeri Sumpiuh pelaksanaan kurikulum 2004/KBK belum terlaksana dengan baik, hal ini dapat terlihat dari hasil analisis soal dimana kualitas soal tergolong kurang bagus, sehingga untuk tahun berikutnya soal tersebut tidak digunakan lagi atau diganti dengan soal yang baru.

Dalam skripsi tersebut metode pengumpulan data yang digunakan adalah survey. Analisis data yang digunakan yaitu proposionalitas materi dengan kurikulum menggunakan analisis deskriptif kualitatif (untuk mengetahui kesesuaian materi dengan kurikulum 2004/KBK) dan analisis validitas, reliabilitas, tingkat kesukaran dan daya beda menggunakan rumus manual, salah satunya rumus yang digunakan adalah *Product Moment*. Sedangkan metode penelitian yang digunakan oleh peneliti dalam penelitian ini adalah observasi dan dokumentasi, analisis data deskriptif kualitatif (untuk memaparkan hasil analisis data), sedangkan untuk mencari nilai validitas, reliabilitas, tingkat kesukaran dan daya pembeda, peneliti menggunakan salah satu program komputer yaitu Program *Iteman*.

Siti Nasiroh (UMP: 2009) dalam skripsinya yang berjudul “Analisis Butir Soal Ulangan Akhir Semester Ganjil Tahun Ajaran 2008/2009 Mata Pelajaran Matematika Kelas XI di SMA Negeri 1 Jeruklegi Kabupaten Cilacap,” menyimpulkan bahwa soal ulangan akhir semester ganjil tahun ajaran 2008/2009 di SMA Negeri 1 memiliki kualitas soal yang kurang baik, sehingga dari hasil tersebut bisa dikatakan bahwa kemampuan guru di SMA Negeri 1 Jeruklegi dalam membuat instrumen evaluasi perlu ditingkatkan.

Dalam skripsi tersebut metode pengumpulan data tidak dicantumkan. Analisis data yang digunakan yaitu proporsionalitas materi dengan kurikulum menggunakan analisis deskriptif kualitatif (untuk mengetahui kesesuaian materi dengan kurikulum 2004/KBK) dan analisis validitas, reliabilitas, tingkat kesukaran dan daya beda menggunakan rumus manual, salah satunya rumus yang digunakan adalah *Product Moment*. Sedangkan metode penelitian yang digunakan oleh peneliti dalam penelitian ini adalah observasi dan dokumentasi, analisis data deskriptif kualitatif (untuk memaparkan hasil analisis data), sedangkan untuk mencari nilai validitas, reliabilitas, tingkat kesukaran dan daya pembeda, peneliti menggunakan salah satu program komputer yaitu Program *Iteman*.

Suwarsito, S.Pi., M.Si. dan Hidayati Mustafidah, S.Si., M.Kom (UMP: 2009), dalam laporan program penerapan Iptek yang berjudul “Pelatihan Analisis Butir Soal Tes Menggunakan Open *Software*”

ANATES ONLINE” Bagi Guru-guru di Kabupaten Banyumas. Dalam laporan tersebut menjelaskan/menginformasikan bahwa pengabdian ini bertujuan untuk memberikan pelatihan kepada tenaga pengajar (guru) di kabupaten Banyumas tentang pemanfaatan teknologi *software* komputer berupa “ANATES ONLINE” untuk membantu guru dalam melakukan analisis butir soal yang meliputi: validitas, reliabilitas, tingkat kesukaran dan daya pembeda, selain lebih cepat dibanding dengan cara perhitungan manual, dengan menggunakan peralatan komputer hasil perhitungan yang didapatkan juga lebih teliti.

Dalam laporan tersebut, dapat diketahui bahwa kegiatan tersebut berupa pengenalan serta pelatihan pemanfaatan komputer dalam kegiatan analisis butir soal yaitu dengan menggunakan *Software* yang dinamakan “ANATES ONLINE”. Sedang peneliti mengenalkan dan penerapan langsung program komputer yang dinamakan “ITEMAN” untuk menganalisis butir soal ujian akhir sekolah.

Berdasarkan, penelitian-penelitian terdahulu yang telah peneliti uraikan di atas, terdapat beberapa hal yang membedakan penelitian ini dengan penelitian-penelitian terdahulu tersebut, diantaranya mata pelajaran yang dianalisis soal ujiannya; semua penelitian terdahulu menganalisis soal ujian untuk mata pelajaran Biologi dengan maksud melakukan konfirmasi apakah soal tersebut sesuai dengan kurikulum yang berlaku atau tidak, sedangkan peneliti menganalisis soal ujian

mata pelajaran Pendidikan Agama Islam dengan maksud mengetahui apakah soal tersebut sudah sesuai dengan prosedur pembuatan soal (validitas, reliabilitas, tingkat kesukaran dan daya pembeda) mengingat nilai hasil ujian untuk mata pelajaran Pendidikan Agama Islam selalu memiliki *rating* tertatas (nilai tertinggi) dibanding mata pelajaran lainnya. Tempat penelitian; untuk penelitian-penelitian terdahulu tempat penelitiannya antara lain SMA PGRI 03 Randudongkal, SMA Veteran Purwokerto, SMA Jenderal Soedirman Purwokerto, SMA N Jatilawang, SMA N Patikraja, SMA N Sokaraja, SMA N Sumpiuh dan SMA Negeri 1 Jeruklegi, sedangkan peneliti melakukan penelitian ini di SMP Negeri 8 Purwokerto. Metode pengumpulan data; peneliti tidak menggunakan metode survey sebagaimana dijelaskan dalam penelitian terdahulu namun peneliti menggunakan metode observasi dan dokumentasi. Metode analisis data; penelitian terdahulu menggunakan analisis deskriptif kualitatif untuk menguji proposionalitas materi ujian dengan kurikulum, sedangkan peneliti menggunakan analisis deskriptif kualitatif untuk mengungkap serta mnguraikan hasil proses analisis dengan menggunakan program *Iteman*. Kegiatan analisis; penelitian terdahulu menggunakan rumus manual salah satunya rumus *Produck Moment* sedangkan peneliti menggunakan program komputer (*Iteman*). Program komputer yang digunakan; dalam salah satu penelitian terdahulu juga menggunakan komputer dalam menganalisis butir soal

yaitu “Anates Online” sedangkan peneliti menggunakan program
Itaman dalam penelitian ini.

