EVALUASI HASIL BELAJAR PENDIDIKAN AGAMA ISLAM (PAI)

M. Arfah
(Kepala MIN 2 Tanjang Jabung Timur)

Abstrak

Penelitian ini bertujuan untuk mengetahui evaluasi hasil belajar Pendidikan Agama Islam (PAI). Melalui library research, evaluasi diketahui dengan menggunakan pengukuran informasi dan informasi hasil penilaian. Hasilnya diukur dengan memberikan skor (angka). Kemudian, skor tersebut dinilai dan ditafsirkan oleh aturan tertentu untuk menentukan tingkat kemampuan pribadi.
Selain itu, hasil dari proses penilaian ini selanjutnya dievaluasi untuk menentukan tingkat pencapaian pribadi atau terprogram. Secara umum, ada dua teknik penilaian pendidikan, yaitu tes dan non-tes. Berdasarkan hasil pengkajian perpustakaan secara deskriptif kualitatif dengan memanfaatkan sumber berupa data atau dokumen, pengkajian ini menggambarkan bagaimana persiapan instrumen berdasarkan kognitif, afektif, dan psikomotor evaluasi domain. Suatu set tes dan non-tes yang baik sebagai pengukur prestasi harus memiliki kriteria; validitas, kepraktisan, kehandalan, dan ekonomi. Selanjutnya, analisis tes yang sesuai pada evaluasi pembelajaranPendidikan Agama Islam (PAI) terdiri dari tingkat kesukaran soal atau indeks kesulitan, daya pembeda, analisis pengecoh, analisis homogenitas item soal, dan efektivitas fungsi opsi.

Kata kunci: Pendidikan Agama Islam (PAI), Belajar, Evaluasi.

A. Pendahuluan

Istilah pengujian, pengukuran, penilaian, dan evaluasi kadangkadang digunakan secara bergantian, namun sebagian besar pengguna membuat perbedaan di antara empat istilah tersebut. Penilaian dan evaluasi lebih bersifat komprehensif yang meliputi pengukuran, sedangkan tes merupakan salah satu alat **(instrument)** pengukuran (Arifin, 2012: 9). Pengukuran lebih membatasi kepada gambaran yang bersifat kuantitatif (angka-angka) tentang kemajuan belajarpeserta didik **(*learning progress*),** sedangkan Penilaian dan evaluasi lebih bersifat kualitatif. Penilaian dan evaluasi pada hakikatnya juga merupakan suatu proses membuat keputusan tentang nilai suatu objek.

Keputusan penilaian (*value judgement*) tidak hanya didasarkan kepada hasil pengukuran (*quantitative description*), tetapi dapat pula didasarkan kepada hasil pengamatan dan wawancara (*qualitative description*). Sehingga secara hirarki, ruang lingkup Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI) evaluasi mencakup penilaian yang di dalamnya memuat pengukuran dan pengukuran membutuhkan alat ukur untuk pengujian.

1. Pengujian

Data kuantitatif dapat diperoleh melalui tes dan nontes (Subali, 2010: 3-4), keduanya yang dimaksud adalah pengujian. Tes merupakan metode pengukuran yang menggunakan alat ukur berbentuk satu *set* pertanyaan untuk mengukur sampel tingkah laku, dan

jawabannya dapat dikategorikan benar dan salah. Menurut Mehrens dan Lehmann (1991: 4), tes merupakan penyajian satu set standar pertanyaan yang harus dijawab. Sedangkan menurut Sax (1980: 13) bahwa "*a test may be defined as a task or series of task used to obtain sistematic observations presumed to be representative of educational or psychological traits or attributes*" (tes dapat didefinisikan sebagai tugas atau serangkaian tugas yang digunakan untuk memperoleh pengamatan-pengamatan sistematis, yang dianggap mewakili ciri atau atribut pendidikan atau psikologis). Non-tes merupakan metode pengukuran yang menggunakan alat ukur untuk mengukur sampel tingkah laku, tetapi jawabannya tidak dapat dikategorikan benar dan salah, misal positif dan negatif, setuju dan tidak setuju, suka dan tidak suka.

Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan menyatakan bahwa ulangan adalah proses yang dilakukan untuk mengukur pencapaian kompetensi peserta didik secara berkelanjutan dalam proses pembelajaran, untuk memantau kemajuan, melakukan perbaikan pembelajaran, dan menentukan keberhasilan belajar peserta didik. Ulangan harian adalah kegiatan yang dilakukan secara periodik untuk mengukur pencapaian kompetensi peserta didik setelah menyelesaikan satu Kompetensi Dasar (KD) atau lebih. Ulangan tengah semester adalah kegiatan yang dilakukan oleh pendidik untuk mengukur pencapaian kompetensi peserta didik setelah melaksanakan 8–9 minggu kegiatan pembelajaran. Cakupan ulangan meliputi seluruh indikator yang merepresentasikan seluruh KD pada periode tersebut. Ulangan akhir semester adalah kegiatan yang dilakukan oleh pendidik untuk mengukur pencapaian kompetensi peserta didik di akhir semester. Cakupan ulangan meliputi seluruh indikator yang merepresentasikan semua KD pada semester tersebut.

Ulangan kenaikan kelas adalah kegiatan yang dilakukan oleh pendidik di akhir semester genap untuk mengukur pencapaian kompetensi peserta didik di akhir semester genap pada satuan pendidikan yang menggunakan sistem paket. Cakupan ulangan meliputi seluruh indikator yang merepresentasikan KD pada semester tersebut. Ujian sekolah/madrasah adalah kegiatan pengukuran pencapaian kompetensi peserta didik yang dilakukan oleh satuan pendidikan untuk memperoleh pengakuan atas prestasi belajar dan merupakan salah satu persyaratan kelulusan dari satuan pendidikan.

2. Pengukuran

Pengukuran adalah proses pemberian bilangan atau angka pada objek-objek atau sesuatu kejadian menurut aturan tertentu (Kerlinger, 1986), pengukuran terdiri dari aturan aturan tertentu untuk memberikan angka atau bilangan kepada objek dengan cara tertentu pula sehingga angka itu dapat mempresentasikan dalam bentuk kuantitatif sifat-sifat dari objek tersebut (Purnomo dan Munadi, 2005: 265-266). Menurut Allen dan Yen (1979: 2), pengukuran didefinisikan sebagai penetapan suatu angka terhadap suatu subjek dengan cara yang sistematik. Jadi pengukuran adalah memberi bentuk kuantitatif pada subjek, objek atau kejadian dengan memperhatikan aturan-aturan tertentu sehingga bentuk kuantitatif tersebut betul-betul menunjukkan keadaaan yang sebenarnya yang diukur.

Pada hasil pengukuran yang berupa angka/skor, objek yang diukur berupa pengetahuan, sikap, dan keterampilan sebagai satu kesatuan yang utuh yang menunjukkan kualitas perilaku

belajar dari peserta didik. Subjek dalam hal ini menunjuk pada peserta didik, objek menunjuk kepada domain hasil belajar, dan kejadian ditunjukkan oleh kualitas perilaku belajar peserta didik.

3. Penilaian

Penilaian merupakan suatu kegiatan untuk menentukan tingkat atau derajat sesuatu objek atau kejadian yang didasarkan atas hasil pengukuran objek tersebut. Menurut Hill (1997), penilaian adalah kegiatan mengolah informasi yang diperoleh melalui pengukuran untuk menganalisis dan mempertimbangkan unjukkerja peserta didik pada tugas-tugas yang relevan. Kegiatan ini juga digunakan untuk menilai materi, program, atau kebijakan-kebijakan dengan maksud untuk menetapkan nilai kelayakan peserta didik. Nitko (1996: 4) menjelaskan "*assessment is a broad term defined as a process for obtaining information that is used for makingdecisions about students, curricula and programs, and educational policy*" (penilaian adalah suatu proses untuk memperoleh informasi yang digunakan untuk membuat keputusan tentang peserta didik, kurikulum, program, dan kebijakan pendidikan). Jadi, penilaian pada dasarnya merupakan suatu kegiatan formal untuk menentukan tingkat atau status, penafsiran dan deksripsi hasil pengukuran hasil belajar peserta didik dibandingkan dengan aturan tertentu. Penilaian (*assessment*) diartikan sebagai prosedur yang digunakan untuk mendapatkan informasi untuk mengukur taraf pengetahuan dan keterampilan subjek didik yang hasilnya akan digunakan untuk keperluan evaluasi (Subali, 2010: 3).

Penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik. Informasi adalah data yang diperoleh melalui pengukuran dan non pengukuran termasuk di dalamnya dengan melakukan observasi kelas, menggunakan tes yang standar atau tes buatan guru, proyek, dan portofolio subjek belajar. Berdasarkan Peraturan Pemerintah Nomor 19 Tahun 2005 Pasal 63 bahwa penilaian hasil belajar dilakukan oleh pendidik, satuan pendidikan, dan oleh pemerintah. Penilaian hasil belajar oleh pendidik dilakukan secara berkesinambungan untuk memantau proses, kemajuan, dan perbaikan hasil dalam bentuk ulangan harian, ulangan tengah semester, ulangan akhir semester, dan ulangan kenaikan kelas.

Penilaian pendidik digunakan untuk menilai pencapaian kompetensi peserta didik, bahan penyusunan laporan kemajuan hasil belajar,dan memperbaiki proses pembelajaran. Penilaian hasil belajar oleh satuan pendidikan bertujuan menilai pencapaian standar kompetensi lulusan untuk semua mata pelajaran, sedangkan penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi dan dilakukan dalam bentuk Ujian Nasional (Kemendiknas, 2010: 1)

4. Evaluasi

Stufflebeam, dkk (1971: vxx) menyatakan bahwa, evaluasi adalah proses menggambarkan, memperoleh, dan memberikan informasi yang berguna untuk menilai alternatif keputusan. Senada dengan pendapat Mardapi (2009: 231), evaluasi memiliki makna adanya pengumpulan informasi, penggambaran, pencarian, dan penyajian informasi guna

pengambilan keputusan tentang program yang dilaksanakan. Sax (1980: 18) juga berpendapat "*evaluationis a process through which a value judgement or decision is made from a variety of observations and from the background and training of the evaluator*" (evaluasi adalah suatu proses dimana pertimbangan atau keputusan suatu nilai dibuat dari berbagai pengamatan, latar belakang serta pelatihan dari evaluator). Evaluasi menggunakan informasi hasil pengukuran dan penilaian. Hasil pengukuran berbentuk skor (angka) yang kemudian skor ini dinilai dan ditafsirkan berdasarkan aturan untuk ditentukan tingkat kemampuan seseorang. Hasil proses penilaian ini kemudian dilakukan evaluasi untuk menentukan tingkat keberhasilan seseorang atau suatu program.

Dalam dunia pendidikan,menilai sering diartikan sama dengan melakukan evaluasi. Perbedaan antara kedua kata tersebut terletak pada pemanfaatan informasi, dimana informasi penilaian merupakan hasil pengukuran, sedangkan informasi pada evaluasi berupa nilai. Paparan di atas merupakan idealitas dari evaluasi pembelajaran. Melalui *library research*, peneliti ingin mengetahui bagaimana evaluasi hasil belajar yang sesuai untuk Pendidikan Agama Islam (PAI).

B. Pembahasan

1. Teknik Penilaian

Teknik penilaian pendidikan secara garis besar ada dua, yaitu tes bila menyangkut benar salah dan nontes bila tidak menyangkut benar salah. Berikut ini diuraikan beberapa teknik penilaian menurut BSNP (2007) dan Kanwil Kemenag Provinsi Jambi (2013), sebagai penjabaran dari teknik tes dan nontes dengan masing-masing ciri dan bentuknya diantaranya adalah:

a) Penilaian tertulis, merupakan tes yang soal dan jawaban yang diberikan kepada peserta didik dalam bentuk tulisan. Ada dua bentuk soal tes tertulis, yaitu soal dengan memilih jawaban (seperti: pilihan ganda, dua pilihan (benar-salah), menjodohkan); dan soal dengan mensuplai jawaban (seperti: isian atau melengkapi, jawaban singkat atau pendek, soal uraian.
b) Penilaian lisan, merupakan tes yang soal yang diberikankepada peserta didik dan jawaban peserta didik dalam bentuk lisan. Bentuk tesnya berupa daftar pertanyaan atau kuis dimana penilaiannya dalam rentang 0–10 atau 1–100.
c) Penilaian unjuk kerja atau praktik, merupakan penilaian yang dilakukan dengan mengamati kegiatan peserta didik dalam melakukan sesuatu, seperti praktik sholat dan praktik baca tulis al-Quran. Cara penilaian ini dianggap lebih otentik daripada tes tertulis karena apa yang dinilai lebih mencerminkan kemampuan peserta didik yang sebenarnya.Teknik penilaian berupa: daftar cek (*check-list*) dan skala penilaian (*rating scale*). Daftar cek lebih praktis digunakan mengamati subjek dalam jumlah besar, dengan cara memberi tanda cek/contreng untuk peserta didik yang kompeten atau tidak kompeten dalam kegiatan praktik. Sedangkan skala penilaian pemberian nilainya secara kontinum, misalnya: 1 = tidak baik, 2 = cukup baik, 3 = baik dan 4 =

sangat baik. Untuk memperkecil faktor subjektivitas, perlu dilakukan penilaian oleh lebih dari satu orang sehingga hasil penilaian lebih akurat.

d) Penilaian produk, merupakan penilaian kemampuan peserta didik dalam pembuatan produk-produk teknologi seni dan hasil karya, seperti makanan, pakaian, gambar, teks pidato khutbah, gambar, peta, kliping, sinopsis, dan lain-lain. Teknik penilaian produk dapat menggunakan cara holistik atau analitik. Cara holistik berdasarkan kesan keseluruhan dari produk dengan menggunakan kriteria keindahan dan kegunaan produk tersebut pada skala skor 0–10 atau 1–100. Sedangkan cara analitik berdasarkan aspek-aspek produk, biasanya dilakukan terhadap semua kriteria yang terdapat pada semua tahap proses pengembangan, yaitu mulai dari tahap persiapan, tahap pembuatan, dan tahap penilaian, masing-masing diberi skor 0–10 atau 1–100 kemudian dihitung reratanya.
e) Penugasan, yaitu penilaian yang menuntut peserta didik melakukan kegiatan tertentu di luar kegiatan pembelajaran di kelas. Penugasan dapat diberikan dalam bentuk individual atau kelompok. Penugasan ada yang berupa pekerjaan rumah atau berupa proyek. Pekerjaan rumah adalah tugas yang harus diselesaikan peserta didik di luar kegiatan kelas, misalnya menyelesaikan soal-soal dan melakukan latihan.
f) Penilaian proyek, merupakan kegiatan penilaian terhadap suatu tugas yang harus diselesaikan dalam periode/ waktu tertentu. Dalam penilaian proyek setidaknya ada 3 (tiga) hal yang perlu dipertimbangkan, yaitu kemampuan pengelolaan (seperti: pemilihan topik, pencarian informasi dan pengelolaan waktu, pengumpulan data, dan penulisan laporan); relevansi (seperti: kesesuaian dengan tema mata pelajaran, dan pertimbangan terhadap tahap pengetahuan/ pemahaman keterampilan dalam pembelajaran); serta keaslian sebagai wujud hasil karya sendiri. Penilaian proyek dilakukan mulai dari perencanaan, proses pengerjaan, sampai hasil akhir proyek. Pelaksanaan penilaian dapat menggunakan alat/instrumen penilaian berupa daftar cek ataupun skala penilaian.
g) Penilaian portofolio, merupakan penilaian berkelanjutan yang didasarkan pada kumpulan informasi (berupa karya dari proses pembelajaran yang dianggap terbaik oleh peserta didik) yang menunjukkan perkembangan kemampuan individu peserta didik dalam satu periode tertentu. Teknik penilaian portofolio di dalam kelas memperhatikan langkah-langkah sebagai berikut: tujuan penggunaan portofolio, penentuan sampel-sampel portofolio yang akan dibuat (bias sama bias beda), pengumpulan/penyimpanan karya-karya tiap peserta didik dalam satu map atau *folder*, pemberian tanggal pembuatan, tentukan kriteria penilaian sampel portofolio dan bobotnya, meminta peserta didik menilai karyanya secara berkesinambungan dengan guru memberi keterangan tentang kelebihan dan kekurangan karya tersebut, serta bagaimana cara memperbaikinya, pemberian kesempatan untuk memperbaiki dengan jangka waktu tertentu bagi peserta didik yang tidak puas dengan hasil karyanya, penjadwalan pertemuan untuk membahas portofolio.
h) Penilaian sikap, yang dinilai dalam proses pembelajaran berupa: sikap terhadap materi pelajaran, guru/pengajar, proses pembelajaran, nilai atau norma, dan kompetensi afektif lintas kurikulum yang relevan dengan mata pelajaran. Penilaian sikap dapat

dilakukan dengan beberapa cara, seperti: observasi perilaku, pertanyaan langsung, dan laporan pribadi.

i) Jurnal, merupakan catatan pendidik selama proses pembelajaran yang berisi informasi kekuatan dan kelemahan peserta didik yang berkait dengan kinerja ataupun sikap peserta didik yang dipaparkan secara deskriptif.

j) Penilaian diri (*self assessment*), di mana peserta didiknya diminta untuk menilai dirinya sendiri berkaitan dengan status, proses, dan tingkat pencapaian kompetensi yang dipelajarinya (kompetensi kognitif, afektif, dan psikomotor. Penilaian diri dilakukan berdasarkan kriteria yang jelas dan objektif, dengan langkah-langkah sebagai berikut: menentukan kompetensi atau aspek kemampuan yang akan dinilai, membuat kriteria penilaian yang akan digunakan, merumuskan format penilaian (berupa pedoman penskoran, daftar tanda cek, atau skala penilaian), meminta peserta didik untuk melakukan penilaian diri, guru mengkaji sampel hasil penilaian secara acak untuk memotivasi peserta didik supaya senantiasamelakukan penilaian diri secara cermat dan objektif, serta menyampaikan umpan balik kepada peserta didik berdasarkanhasil kajian terhadap sampel hasil penilaian yang diambil secara acak. Strategi *self assessment* seperti temuan penelitian Marrinawati (2013) bahwa dengan strategi ini peserta didik secara berangsur bisa mengekspresikan dirinya dengan cara menilai dirinya sendiri secara objektif mampu menunjukkan karakter jujur, bertanggung jawab atas penilaiannya sendiri serta percaya diri dalam memberikan penilaian, sedangkan peranan guru fikih tetap mengontrol dan mengamati sikap peserta didik tersebut.

k) Penilaian antar teman, merupakan teknik penilaian dengan cara meminta peserta didik untuk mengemukakan kelebihan dan kekurangan temannya dalam berbagai hal. Untuk itu perlu ada pedoman penilaian antarteman yang memuat indikator prilaku yang dinilai. Satu di antara penilaiannya dengan teknik sosiometri. Sebagai catatan bahwa, tidak ada satu pun alat penilaian yang dapat mengumpulkan informasi hasil dan kemajuan belajar peserta didik secara lengkap. Penilaian tunggal tidak cukup untuk memberikan gambaran/informasi tentang kemampuan, keterampilan, pengetahuan dan sikap peserta didik. Lagi pula, interpretasi hasil tes tidak mutlak dan abadi karena anak terus berkembang sesuai dengan pengalaman belajar yang dialaminya, dan secara komprehensif manakala posisi dan peran dari hasil tes dan nontes secara paralel dan terakumulasi dalam pelaksanaan evaluasinya.

Pengembangan I 2. instrumen Penilaian Ragam teknik penilaian di atas selanjutnya dipaparkan bagaimana penyusunan instrumennya, baik pada domain kognisi, afeksi, dan psikomotorik. *Pertama* adalah penyusunan instrumen penilaian kognitif. Tes tertulis yang lazim digunakan untuk mengukur kemampuan kognitif dapat dibedakan untuk tujuan mengukur kemampuan kognitif tingkat rendah (kemampuan mengetahui, memahami, dan menerapkan), dan kemampuan kognitif tingkat tinggi (menganalisis, mengevaluasi, menyintesis, berimajinasi, dan mengkreasi). Butir-butir pada tes dalam bentuk pilihan terdiri atas soal dan kunci jawaban, sedangkan butir bentuk mengisikan jawaban singkat terdiri atas soal beserta rubrik dan atau pedoman penskoran. Butir tes bentuk uraian terbuka terdiri atas soal, rubrik, dan pedoman penskoran. Penyusun suatu butir harus memperhatikan aspek

substansi/ isi, konstruksi, dan bahasa. Persyaratan tersebut akan dapat diketahui manakala diadakan analisis secara kualitatif. Aspek isi materi dengan memperhatikan 1) Butir soal sesuai indikator, 2) batasan pertanyaan dan jawaban yang diharapkan jelas, 3) isi materi sesuai dengan tujuan pengukuran, dan 4) isi materi yang ditanyakan sesuai dengan jenjang dan jenis pendidikan. Aspek konstruksinya berupa 1) rumusan kalimat dalam bentuk kalimat tanya atau perintah yang menuntut jawaban, 2) ada petunjuk yang jelas cara mengerjakan/ menyelesaikan soal, 3) Ada pedoman penskorannya, 4) tabel, grafik, diagram, kasus, atau yang sejenisnya bermakna/berfungsi, 5) antarbutir soal tidak saling bergantung, dan 6) untuk soal pilihan, pilihan jawaban homogen.

Aspek bahasa meliputi 1) rumusan kalimat komunikatif, 2) kalimat menggunakan bahasa yang baik dan benar, sesuai dengan jenis bahasanya/EYD, 3) rumusan kalimat tidak menimbulkan penafsiran ganda atau salah pengertian, 4) menggunakan bahasa/kata yang umum (bukan bahasa lokal), 5) rumusan soal tidak mengandung kata-kata yang dapat menyinggung perasaan peserta didik (nirbias Suku, Agama, Ras dan Adat Istiadat (SARA) dan *gender*)**.** *Kedua* adalah penyusunan instrumen penilaian afektif. Instrumen penilaian afektif dapat berupa angket, daftar penilaian, dan/atau skala penilaian. Penyusunan angket harus memperhatikan skala sikap yang digunakan, berikut ini beberapa pengukuran sikap yang dapat digunakan:

a. Skala Likert, merupakan suatu skala penilaian untuk mengukur sikap peserta didik terhadap suatu kegiatan, menggunakanskala ordinal. Rentangan yang dipilih dari yang sangat positif sampai sangat negatif, dari pertanyaan atau pernyataan yang positif atau negatif. Misalnya alternatif pilihan jawaban sangat setuju (SS) atau selalu (S), setuju (S) atau hampir selalu (HS), ragu-ragu (R) atau sering (SR), tidak setuju (T) atau kadangkadang/ jarang (J), dan sangat tidak setuju (ST) atau tidak pernah (TP). Penyusunan skala Likert dapat memperhatikan langkah-langkah seperti: a) merumuskan difinisi variabel yang diukur, b) merumuskan dimensinya (apakah bersifat multidimensi) dan c) merumuskan indikatornya.
b. Skala pilihan ganda, bentuknya seperti soal bentuk pilihan ganda yaitu pernyataan yang diikuti oleh sejumlah alternatif pendapat. Contoh: Dalam melaksanakan shalat fardhu, saya merasa: (pembelajaran Fikih) 1) Senang karena dapat berdialog dengan Allah 2) mudah untuk melakukan konsentrasi 3) tidak begitu sulit untuk berkonsentrasi 4) dapat berkonsentrasi tetapi mudah terganggu 5) sulit berkonsentrasi.
c. Skala Guttman, berupa pernyataan yang dirumuskan sejumlah tiga atau empat pernyataan, di mana setiap pilihan pernyataan tersebut menunjukan tingkatan yang berurutan, apabila responden setuju pernyataan 2 diduga setuju pernyataan 1, selanjutnya setuju pernyataan 3 diduga setuju pernyataan 1 dan2, dan seterusnya. Contoh: Hormat pada orang tua: (pembelajaran Akidah Akhlak) 1) Saya permisi kepada orang tua bila bermain ke tetangga2) Saya permisi kepada orang tua bila pergi ke mana saja 3) Saya permisi kepada orang tua bila pergi kapan saja dan ke mana saja 4) Saya tidak pergi kemana saja tanpa permisi kepada orang tu.
d. Skala Perbedaan Semantik, merupakan suatu model skala dengan meletakkan suatu rentangan di antara dua kata atau ide yang berlawanan, sehingga berupa skala

perbedaan sematik. Model skala yang bipolar ini sangat baik untuk mengungkap unsur evaluasi (baik/buruk, bersih/kotor, jujur/tidak jujur, menguntungkan/merugikan dan sejenisnya), atau untuk mengungkap unsur potensi (besar/kecil, kuat/lemah, berat/ ringan, dan sejenisnya), ataupun unsur aktivitas (aktif/pasif, cepat/lambat, loyal/tak loyal, penuh perhatian/tak acuh). Pasangan adjektif tersebut harus dicari yang sesuai dengan konsep atau obyek yang akan diukur, contoh: Pembelajaran Fikih yang telah berjalan selama setengah semester I: 1) Menarik .____.____.____.____. Membosankan, 2) Mudah .____.____.____.____. Sukar 3) Ringan .____.____.____.____. Berat Menguntungkan .____.____.____.____. Merugikan 4) Bermanfaat .____.____.____.____. Merugika 5) Menantang .____.____.____.____. Tidak menantang 6) Mengasyikkan .____.____.____.____. Menjemukan Penskoran butir-butir di atas, semakin ke arah yang positif semakin besar, dan skor total merupakan penjumlahan skor setiap pasangan ajektif.

e. Skala Thurstone, dengan memperhatikan tahapan berikut: 1) pengembangan daftar pernyataan yang ditawarkan pada panelis yakni dengan menyusun minimal 50 pernyataan dari yang sangat positif sampai sangat negatif yang berkait dengan pembelajaran Quran Hadis, 2) menyiapkan anggota panelis, misalnya dengan memilih sekurang-kurangnya 80 guru Quran Hadis dan/atau peserta didik, dan 3) meminta panelis untuk memberikan skor terhadap setiap pernyataan yang ditawarkan, dengan kisaran skor 1 (sangat negatif) sampai 11 (sangat positif). f. Lembar Observasi/Lembar Penilaian Antar teman (*Peer Assessment*), dapat digunakan untuk melihat sikap peserta didik selama kegiatan berlangsung. Observer/penilai selain guru juga teman sekelas peserta didik yang dinilai dalam kegiatan pembelajaran di kelas baik secara individu maupun interaksinya dalam suatu kelompok.

Ketiga adalah pengembangan instrumen penilaian psikomotorik. Penilaian psikomotorik atau kinerja adalah penilaian yang memfokuskan aspek keterampilan yang berkait dengan ranah psikomotor yang dapat didemonstrasikan/dipraktikkan/dikerjakan oleh peserta didik, yang di dalamnya juga mencakup ranah kognitif.

Demonstrasi/praktik/kinerja dapat digradasi dari paling rendah sampai yang paling tinggi. Dari taksonomi ranah psikomotor dapat diidentifikasikan bahwa ada aspek dari ranah psikomotor yang murni sebagai gerak fisik tubuh dan ada pula gerak dari bagian tubuh yang berkait dengan pemakaian alat dan bahan. Sebagai contoh pada pembelajaran Fikih materi merawat jenazah, praktik menyucikan dan mengkafani jenazah melibatkan gerak fisik tubuh dengan menggunakan alat dan bahan yang tersedia, berbeda dengan praktik shalatnya yang hanya melibatkan gerak fisik tubuh saja.

3. Validitas dan Reliabilitas Instrumen

Seperangkat tes yang baik sebagai alat pengukur menurut Arikunto (2008: 59) harus memenuhi persyaratan tes, yaitu memiliki; validitas, praktibilitas, reliabilitas, dan ekonomis. Hal senada juga, alat ukur yang baik adalah alat ukur yang memiliki bukti kesahihan dan keandalan (Widayati, 2009: 185). Menurut Azwar (2000: 5-6), validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam

melakukan fungsi ukurnya. Senada dengan Subali (2010: 41) dan Arifin (2012: 314), suatu alat ukur (dalam hal ini adalah tes) dinyatakan sahih (valid), jika alat ukur tersebut benar-benar mampu memberikan informasi empirik sesuai dengan apa yang diukur.

Bukti kesahihan atau validitas alat ukur dilihat pada kesesuaian antara definisi operasional dari konsep yang akan diukur dengan materi pertanyaan pada alat ukur. Bukti kesahihan alat ukur meliputi kesahihan isi, konstruk, dan kriteria. Gronlund (1985: 79- 81) mengemukan ada tiga faktor yang mempengaruhi validitas hasil tes, yaitu: 1) Faktor instrumen evaluasi, terkait prosedur penyusunan instrumen, seperti silabus, kisi-kisi soal, petunjuk mengerjakan soal dan pengisian lembar jawaban, kunci jawaban, penggunaan kalimat efektif, bentuk alternatif jawaban, tingkat kesukaran, daya pembeda, dan distraktor, 2) Faktor administrasi evaluasi dan penskoran, seperti alokasi waktu untuk pengerjaan soal yang tidak proporsional, memberikan bantuan kepada peserta didik dengan berbagai cara, peserta didik saling menyontek ketika ujian, kesalahan penskoran, termasuk kondisi fisik dan psikis peserta didik yang kurang menguntungkan, dan 3) faktor jawaban dari peserta didik, seperti kecenderungan peserta didik untuk menjawab secara cepat tetapi tidak tepat, keinginan melakukan coba-coba, dan penggunaan gaya bahasa tertentu dalam menjawab soal bentuk uraian. Berikut ini dipaparkan beberapa ragam validitas yang sering digunakan dalam pengujian instrumen, dalam hal ini tes, yaitu:

a. Validitas isi, validitas kurikuler, validitas perumusan, validitas rasional, atau validitas logis, berkenaan dengan keterkaitan tes dengan kurikulum atau isi materi pembelajaran yang sedang berlaku. Validitas kurikuler ini dapat dilakukan dengan beberapa cara, antara lain mencocokkan materi tes dengan silabus dan kisi-kisi, melakukan diskusi dengan pakar/ahli, praktisi dan/atau sesama guru, atau mencermati kembali substansi dari konsep yang akan diukur. Menurut Purwanto (2011), pengujian validitas isi dapat dilakukan dengan meminta pertimbangan para ahli (*experts judgement*), di mana butir-butir yang telah direspon oleh para ahli (dua orang) dalam bentuk skor kemudian dikorelasikan dengan teknik statistika dan jika signifikan berarti disepakati para ahli, sehingga butir-butir dikatakan valid.
b. Validitas empiris atau validitas statistik, yaitu mencari hubungan antara skor tes dengan suatu kriteria tertentu yang merupakan suatu tolok ukur di luar tes yang bersangkutan. Ada dua macam validitas empiris, yaitu: 1) validitas prediktif, yaitu kemampuan suatu tes dapat memprakirakan perilaku peserta didik pada masa yang akan datang, dan 2) Validitas konkuren ialah jika kriteria eksternal sudah ada saat pengujian, misalnya, skor tes dari perangkat tes buatan guru dengan buatan hasil MGMP. Pengujian validitas empiris menurut Purwanto (2011: 121-122) dapat dilakukan dengan mencari koefisien korelasi antar hasil pengukuran dari perangkat perangkat tes.
c. Validitas konstruk. Konstruk adalah konsep yang dapat diobservasi (*observable*) dan dapat diukur (*measurable*). Validitas konstruk adalah kemampuan tes dalam mengobservasi dan mengukur fungsi psikologis yang merupakan deskripsi perilaku peserta didik. Pengujian validitas konstruk menurut Purwanto (2011: 126-127) dapat dilakukan dengan cara: a) telaah butir, 2) *expert judement*, 3) konvergensi dan diskriminabilitas, 4) *multitrait-multimethod*, dan 5) analisis faktor.

Reliabilitas berkait dengan konsistensi, keandalan, keajegan, ataupun stabilitas. Suatu alat ukur dikatakan reliabel bila memberikan hasil yang sama pada berkali-kali pengulangan pengukuran (Subali, 2010: 43-44) dan reliabilitas berlaku pada tingkat suatu perangkat tes, bukan untuk masing-masing butir tes. Kerlinger (1986: 443) mengemukakan reliabilitas dapat diukur dari tiga kriteria, yaitu *stability, dependability*, dan *predictability*. *Stability* menunjukkan keajegan suatu tes dalam mengukur gejala yang sama pada waktu yang berbeda. *Dependability* menunjukkan kemantapan/ keandalan suatu tes. *Predictability* menunjukkan kemampuan tes untuk meramalkan hasil pada pengukuran gejala selanjutnya. Bukti keandalan atau reliabilitas suatu alat ukur dapat dilihat pada besarnya indeks keandalan. Selanjutnya, Gronlund (1985: 100) mengemukakan ada empat faktor yang dapat mempengaruhi reliabilitas, yaitu 1). Panjang tes (*length of test*), semakin panjang suatu tes (banyak butir) akan lebih tinggi tingkat reliabilitas suatu tes, 2). Sebaran skor (*spread of scores*), semakin besar sebaran skor akan membuat tingkat reliabilitas menjadi lebih tinggi, 3). tingkat kesukaran (*difficulty index*), di mana tingkat kesukaran soal yang ideal untuk meningkatkan koefisien reliabilitas adalah soal yang menghasilkan sebaran skor berbentuk genta atau kurva normal, dan 4). objektifitas (*objectivity*), di mana objektifitas prosedur tes yang tinggi akan memperoleh reliabilitas hasil tes yang tidak dipengaruhi oleh prosedur penskoran. Setelah memperhatikan faktor-faktor tersebut di atas, seperangkat tes dapat diuji dengan menggunakan beberapa klasifikasi metode pengujian seperti berikut: 1) koefisien stabilitas eksternal, seperti metode tes ulang (*test-retest*) dan metode paralel, 2) koefisien konsistensi internal, meliputi metode belah dua (*split half*), metode Flanagan, metode Rulon, metode Kuder-Richardson, *anova* Hoyt, dan Cronbach *alpha* (Purwanto, 2011: 156-176), dan 3) reliabilitas *interrater*, seperti pendekatan Hoyt, pengukuran observasi, dan teori generaliabilas (Mardapi, 2012: 86-92).

4. Analisis Butir

Sebelum penyelenggaraan tes, pedoman pemberian skor juga harus tersusun, bahkan sebaiknya sudah dirancang strategi pemberian skornya sejak merumuskan kalimat pada setiap butir soal. Pedoman penskoran sangat penting disiapkan bentuk soal subjektif, afektif dan psikomotor, yang dimaksudkan untuk meminimalisir subjektifitas penilaian. Formulasi penskoran yang digunakan bergantung kepada bentuk soalnya, sedangkan bobot bergantung kepada tingkat kesukaran soal seperti sukar, sedang ataukah mudah. Menurut Arifin (2011) dalam mengolah data hasil tes, ada empat langkah pokok yang harus tempuh. *Pertama*, menskor, yaitu memberi skor pada hasil tes yang dapat dicapai oleh peserta didik dan untuk memperoleh skor mentah diperlukan tiga jenis alat bantu, yaitu: kunci jawaban, kunci skoring, dan pedoman konversi. *Kedua*, mengubah skor mentah menjadi skor standar sesuai dengan norma tertentu. *Ketiga*, mengkonversikan skor standar ke dalam nilai, baik berupa huruf atau angka. *Keempat*, melakukan analisis soal (jikadiperlukan) untuk mengetahui derajat validitas dan reliabilitas soal, tingkat kesukaran soal (*difficulty index*), dan daya pembeda.

Butir dalam bentuk uraian biasanya skor mentah dicari dengan menggunakan sistem bobot. Sistem bobot ada dua cara, yaitu: 1) bobot butir dinyatakan dalam skor maksimum sesuai dengan tingkat kesukarannya, dan 2) bobot dinyatakan dalam bilangan-bilangan tertentu sesuai dengan tingkat kesukaran soal. Untuk memudahkan pemberian skor, ada baiknya digunakan sistem yang kedua. Sistem bobot diberikan kepada soal bentuk uraian dengan maksud untuk memberikan skor secara adil kepada peserta didik berdasarkan kemampuannya masing-masing dalam menjawab soal-soal yang berbeda tingkat kesukarannya. Sedangkan penskoran pada soal tes bentuk objektif ini dapat menggunakan dua cara, yaitu: 1) tanpa rumus tebakan (*non-guessing formula*), caranya ialah menghitung jumlah jawaban yang betul saja, dan 2). Dengan rumus tebakan (*guessing formula*), digunakan apabila soal-soal tes itu sudah pernah diujicobakan dan dilaksanakan, serta sangat memungkinkan peserta didik untuk menebak. Skor total adalah jumlah skor yang diperoleh dari seluruh bentuk soal setelah diolahdengan rumus tebakan. Skor ini selanjutnya disebut skor mentah (*raw score*). Konversi skor adalah proses transformasi skor mentah yang dicapai peserta didik ke dalam skor terjabar atau skor standar untuk menetapkan nilai hasil belajar yang diperoleh.

Ada dua cara mengubah atau mengolah skor menjadi nilai, yaitu :

1. Pertama, Penilaian Acuan Patokan (PAP) yaitu meneliti apa yang dapat dikerjakan oleh peserta didik dengan suatu kriteria atau patokan yang spesifik, dengan kata lain membandingkan skor mentah dengan kriteria. Kriteria yang dimaksud adalah suatu tingkat pengalaman belajar yang diharapkan tercapai sesudah selesai kegiatan belajar atau sejumlah kompetensi dasar yang telah ditetapkan terlebih dahulu sebelum kegiatan belajar berlangsung. Misalnya, kriteria yang digunakan 75%. Bagi peserta didik yang kemampuannya di bawah kriteria yang telah ditetapkan dinyatakan tidak berhasil dan harus mendapatkan remedial.
2. Kedua, penilaian acuan norma (PAN) yaitu penilaian akhir terhadap seorang peserta didik, dibandingkan dengan prestasi seluruh peserta didik dengan kaidah distribusi normal, dengan cara ini akan dapat dilihat kedudukan peserta didik di dalam kelompoknya. Guna meningkatkan derajat validitas dan reliabilitas, dapat dilakukan analisis butir soal. Berikut ini setidaknya analisis butir yang dilakukan pada evaluasi pembelajaran PAI:

   1) Tingkat kesukaran soal (*difficulty index*), yaitu pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional dalam sebaran tingkat kesukaran), maka dapat dikatakan bahwa soal tersebut baik. Menurut Allen dan Yen (1979: 121), Gregory (2007: 153), dan Mardapi (2002: 116) tingkat kesulitan butir soal sebaiknya terletak pada interval 0,3 sampai 0,8 karena pada interval ini informasi tentang kemampuan peserta didik akan diperoleh secara maksimal. Untuk menghitung tingkat kesukaran soal bentuk objektif dapat dilakukan dengan menggunakan rumus tingkat kesukaran, kemudian menggunakan tabel batas tingkat kesukaran. Dalam analisis soal secara klasikal, tingkat kesukaran dapat diperoleh dengan beberapa cara, antara lain: skala kesukaran linier, skala bivariat, indeks davis, dan proporsi menjawab benar. Cara menghitung tingkat kesukaran untuk soal bentuk uraian

adalah menghitung berapa persen peserta didik yang gagal menjawab benar atau ada di bawah batas lulus (*passing grade*) untuk setiap butir soal.

2) Daya Pembeda (*discriminating power*), yaitu kemampuan butir soal dalam membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/ kurang menguasai kompetensi berdasarkan kriteria tertentu. Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi. Teknik yang digunakan untuk menghitung daya pembeda soal bentuk uraian adalah menghitung perbedaan dua rata-rata (*mean*), yaitu antara rata-rata dari kelompok atas dengan rata-rata dari kelompok bawah untuk tiap-tiap soal. Menurut teori klasik, ada hubungan antara tingkat kesukaran dan daya pembeda butir (Kartowagiran, 2008: 189), seperti gambar berikut: *Gambar 1*. Tingkat kesukaran dan daya pembeda butir (Kartowagiran, 2008: 189).

3) Analisis Pengecoh. Butir soal pilihan ganda dikatakan baik jika pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang, pengecohnya tidak dipilih atau dipilih secara tidak merata. Menurut pendapat Fernandes (1984) menjelaskan bahwa distraktor dikatakan baik apabila paling tidak dipilih oleh 2% dari seluruh peserta. Sementara itu, Nitko (1996) mengatakan distraktor dikatakan berfungsi manakala paling tidak dipilih oleh seorang peserta tes dari kelompok rendah. Pemilih dari kelompok rendah harus lebih banyak daripada kelompok atas (Kartowagiran, 2008: 190). Distraktor juga dapat dikatakan berfungsi manakala peserta tes dari kelompok atas dapat membedakan antara distraktor dan kunci jawaban sehingga yang memilih kunci jawaban lebih banyak dari pada yang memilih distraktor.

4) Analisis Homogenitas Soal, dilakukan dengan menghitung koefisien korelasi antara skor tiap butir soal dengan skor total, di mana skor setiap butir soal adalah 1 atau 0, sedang skor total tiap peserta didik akan bervariasi. Teknik korelasi yang dapat digunakan adalah korelasi Pearson/*product moment* atau korelasi *point biserial*. Butir soal dikatakan homogen, apabila koefisien korelasinya sama atau lebih besat dari harga kritik korelasi pada tabel, atau sebaliknya. Butir soal yang tidak homogen kemungkinan besar mengukur aspek lain di luar materi/bahan yang diajarkan, karena tidak sesuai dengan kompetensi yang telah ditetapkan, sehingga butir soal yang demikian sebaiknya direvisi atau dibuang.

5) Efektifitas Fungsi Opsi. Analisis butir perlu juga dicari apakah suatu opsi (alternatif jawaban) dari setiap soal berfungsi secara efektif atau tidak, dengan langkah-langkah sebagai berikut : a). menentukan jumlah peserta didik (N), b). menentukan jumlah sampel (n), baik untuk kelompok atas maupun kelompok bawah, yaitu 27 % x N, c). membuat tabel pengujian efektifitas opsi, d. menghitung jumlah alternatif jawaban yang dipilih peserta didik, baik untuk kelompok atas maupun kelompok bawah, dan e). Menentukan efektifitas fungsi opsi dengan kriteria tertentu.

5. Feedback Penilaian: Remediasi dan Pengayaan

Guru dalam mengatasi perilaku *under achievement* (berprestasi kurang) pada peserta didik seharusnya menyadari bahwa anak memiliki kekuatan dan kelemahan baik berkenaan dengan kebutuhan sosial, emosional, maupun intelektual (Wahab, 2005: 8-9). Dengan strategi

remedial, peserta didik diberikan kesempatan untuk mempercepat dalam bidang-bidang yang menjadi kekuatannya dan minat melalui penganyaan, sementara itu kesempatan diberikan untuk bidang-bidang spesifik yang dirasakan ada kesulitan belajar bagi peserta remedial. Remediasi khususnya dilakukan dalam suatu lingkungan yang aman dan kondusif, yaitu suatu lingkungan yang kesalahan-kesalahan terjadi dianggap menjadi bagian dari belajar setiap orang dapat dihindari, termasuk guru.

Remidiasi dan pengayaan merupakan program pembelajaran bagi peserta didik baik secara individu maupun kolektif yang pelaksanaannya relatif paralel atau bersamaan, di mana remidiasi dan pengayaan dirancang untuk kebutuhan pembelajaran kelompok peserta didik yang mengalami kegagalan atau keberhasilan dalam penguasaan Kompetensi Dasar (KD) melalui *feed back* dari penilaian formatif (Subali, 2010: 69), dan berikut ini alur remidiasi dan pengayaan dalam proses pembelajaran: Penilaian/asesmen formatif sebagai dasar untuk evaluasi formatif. Keputusan bagi peserta didik untuk remedial bukan sekedar mengulang proses pembelajaran, namun juga disesuaikan dengan penyebab kegagalannya. Demikian pula keputusan bagi peserta didik untuk program pengayaan adalah program yang diberikan yang telah berhasil menguasai hasil belajar sesuai yang ditargetkan.

Untuk mengidentifikasi peserta didik yang mengalami remediasi ataupun pengayaan dapat dilakukan dengan melihat dalam proses pembelajaran, meliputi: a) Aspek prestasi akademis (seperti interaksi dalam kelas saat memberikan jawaban, otentisitas dan kualitas tugas rumah, dan kemampuan dalam pengerjaan tes); b) aspek perilaku (seperti minat belajar, dan kehadiran di kelas). Jenis-jenis kegagalan ataupun keberhasilan dapat didiagnosis dengan karakteristik kompetensi yang dikembangkan, yaitu terkait dengan: 1) penguasaan konsep atau rumus, 2) keterampilan motorik dan penggunaan alat/bahan, 3) penguasaan prosedur pemecahan masalah, dan 4) Kemampuan mengaplikasikan konsep pada situasi baru (situasi selain yang pernah dibahas guru). Beberapa hal yang perlu dipertimbangkan untuk melakukan remediasi adalah: 1) peserta remediasi dalam satu kelompok sejumlah 5 – 10 peserta didik 2) memberi perhatian secara individual kepada peserta remedial 3) hindari mencemooh peserta didik yang mengalami kegagalan 4) proses belajar dapat dilakukan dengan cara *drill* (berulangulang dalam membaca, menulis, latihan soal) 5) memberikan asesmen secara khusus secara bertingkat dan berangsur mulai asesmen dengan butir yang berkategori sangat mudah sampai dengan kompleks 6) model belajar kelompok atau belajar melalui kelompok belajar akan sangat membantu 7) berlatih membuat catatan-cacatan kecil untuk mempermudah belajar 8) difokuskan pada target dalam silabus.

C. Kesimpulan

Berdasarkan analisis dan pembahasan di atas, peneliti dapat menyimpulkan bahwa evaluasi hasil belajar Pendidikan Agama Islam (PAI) meliputi beberapa hal, yaitu: (1) tingkat kesukaran soal (*difficulty index*); (2) daya pembeda (*discriminating power*); (3) analisis pengecoh; (4) analisis homogenitas soal; dan (5) efektifitas fungsi opsi.

Daftar Pustaka

Allen, M.J. dan Yen, W.M. 1979. *Introduction to Measurement Theory*. Monterey: Wardsworth, Inc., 1979.

Arifin, Zainal, *Evaluasi Pembelajaran: Prinsip-Teknik-Prosedur.* Bandung: PT.Remaja Rosdakarya, 2011.

Anonim, *Evaluasi Pembelajaran.* Jakarta : Subdit Dirjen Pendis Kemenag RI, 2012.

Anonim, *Evaluasi Penerapan Ujian Akhir Sekolah Dasar berbasis Standar Nasional*. Jurnal Penelitian dan Evaluasi Pendidikan T.13 no. 2, 2009.

Anonim, *Panduan Penilaian Kelompok Mata Pelajaran Ilmu Pengetahuan dan Teknologi.* Jakarta : BSNP, 2007.

Anonim, *Pengukuran, Penilaian dan Evaluasi Pendidikan.* Yogyakarta : Nuha Medika, 2012.

Arikunto Suharsimi, *Dasar-Dasar Evaluasi Pendidikan.* Jakarta : Bumi Aksara, 2001.

Azwar Saifuddin, *Reliabitas dan Validitas.* Yogyakarta : Sigma Alpha, 2000.

Gregory, N.E, *Measurencet and Evaluation in Principles and Application.* New York : Pearson Education.Inc, 2007.

Gronlund, N.E. *Measurencet and Evaluation in Teaching.* New York : Mc.Millan Publishing Co.,Inc, 1985.

Hill, R.B, *The Design or an Instrument to Assess Problem Solving Activities in Technology Education*. Jurnal of Tekonlogy Education.Vol.9, no. 1, 1997.

Kerlinger, F.N, *Evaluation of Behavior Research. Halt-Rinechart and Winston,Inc, 1986.rtikel tes keterampilan Olah Raga Judo bagi mahapeserta didik*. Jurnal Kependidikan no.1, 2002.

Marrinawati Rina, *Penerepan Strategi Self Assissment dalam Pembentukan Karakter Peserta Didi pada Pembelajaran Fiqg di Kleas XI IPA MAN Yogyakarta III.* Yogyakarta : UIN Suka, 2013.

Mehrens, W.A and Lehmann, I.J, *Meausurment and Evaluation in Education and Psychology*. California : Wadsworth/Thomson Learning, 1991.

Nitko, A. J.. *Educational Assessment of Students*, New Jersey : Englewood Cliffs, 1996.

Purwanto, *Evaluasi Hasil Belajar.* Yogyakarta : Pustaka Pelajar, 2011.

Sax, G, *Principles of Educational and Psychological Measurement and Evaluation*. Belmont California : Wads Worth Pub.Co, 1980.

Stufflebeam,D.L., et al, *Evaluational Evaluation and Decision Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI)* Vol 9, no. 2., Making, Bloominton,IN: Phi Delta Kappa, 1971.

Subali Bambang, *Penilaian, Evaluasi dan Remediasi Pembelajaran Biologi*. Yogyakarta : Fak. MIPA, 2010.

Wahab Rochmat, *Anak Berbakat Berprestasi Kurang (the Underchiving Gifted) dan Strategi Penanganannya.* Jakarta : Dit.PLB Dirjen Dikdasmen, 2005.

Widayati Catharina Sri Wahyu, *Komparasi beberapa Metode Estimasi Kesalahan pengukuran*. Jurnal Penelitian dan Evaluasi pendidikan T.13 no.2, 2009.