

TEKNIK PENILAIAN HASIL PEMBELAJARAN

Zaimul Am

zaimul07@yahoo.com

(Dosen UIN Syarif Hidayatullah Jakarta Pada Fakultas Agama Islam, Universitas Muhammadiyah Tangerang)

Abstrak:

Mutu pendidikan sangat ditentukan oleh kemampuan satuan pendidikan dalam mengelola proses pembelajaran. Penilaian merupakan bagian yang sangat penting dalam pembelajaran. Berdasarkan hasil penilaian, pendidik dapat mengambil keputusan secara tepat untuk menentukan langkah yang harus dilakukan selanjutnya. Penilaian hasil pembelajaran merupakan kewajiban yang harus ditunaikan oleh setiap guru maupun dosen dengan tujuan untuk mengetahui, hasil penilaian juga dapat memberikan motivasi kepada peserta didik untuk berprestasi lebih baik. Teknik penilaian dapat dilakukan secara komplementer (saling melengkapi) sesuai dengan kompetensi yang dinilai.

Kata Kunci: Teknik, Penilaian, Belajar.

A. Pendahuluan

Salah satu bagian penting dalam pelaksanaan kegiatan pembelajaran adalah teknik penilaian (*assessment*) hasil pembelajaran. Ada serangkaian aspek yang berkaitan erat dengan penilaian hasil pembelajaran yang bersifat kongruen terhadap berbagai teori mengenai. Aspek-aspek yang berkaitan erat dengan penilaian hasil pembelajaran mencakup beberapa istilah penting yakni penilaian, pengukuran, dan evaluasi. Penilaian merupakan istilah sangat umum yang menjelaskan banyak teknik yang digunakan untuk mengukur dan menilai perilaku maupun prestasi siswa. Pengukuran merupakan penggunaan bilangan orang atau objek berdasarkan karakteristik yang dimiliki oleh orang atau objek tersebut. Sedangkan evaluasi melibatkan penggunaan ukuran untuk membuat keputusan mengenai atau menentukan kualitas orang atau objek.¹

Penilaian memiliki beberapa tipe, yakni penilaian formal dan informal, penilaian kelas, penilaian diagnostik, penilaian formatif, dan penilaian sumatif. Masing-masing tipe penilaian ini memiliki pengertian dan karakteristik tersendiri. Penilaian kelas, yang disebut pula sebagai penilaian penempatan (*placement assessment*), merupakan tes yang diadakan pada dua minggu pertama dalam semester dan dirancang untuk mengukur keterampilan dasar siswa. Penilaian diagnostik merupakan setiap bentuk penilaian yang digunakan untuk mengidentifikasi kekurangan masing-masing siswa dalam pembelajaran. Penilaian formatif merupakan setiap bentuk penilaian yang digunakan pada saat materi pelajaran sedang diajarkan dan pada dasarnya digunakan untuk memberi masukan kepada guru mengenai perkembangan kegiatan pembelajaran. Penilaian sumatif dilaksanakan pada masa akhir pembelajaran untuk menentukan

¹Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, (New York: Rutledge, 2008), 7

tingkat kompetensi siswa terhadap seluruh materi pembelajaran dan kenaikan kelas.²

Dalam penilaian hasil belajar siswa, ada beberapa langkah yang harus ditempuh. Pertama, menentukan tujuan tes. Kedua, membuat tabel spesifikasi (kisi-kisi tes). Ketiga, merumuskan soal tes berdasarkan tabel spesifikasi. Keempat, menetapkan standar penilaian.

B. Langkah-langkah Penilaian Hasil Pembelajaran

1. Menentukan Tujuan Tes

Tujuan tes sangat bergantung kepada tujuan pembelajaran. Gambaran mengenai apa yang akan diuji atau dijadikan bagian dari tes tidak akan menjadi jelas jika tidak mengacu kepada tujuan pembelajaran yang telah dirumuskan. Karena itu, merumuskan tujuan pembelajaran menjadi dasar bagi perumusan tujuan tes. Dalam kajian mengenai konsep tujuan pembelajaran, ada beberapa teori yang dapat dikemukakan. Teori taksonomi tujuan pendidikan Benjamin S. Bloom membagi tujuan pembelajaran ke dalam 3 (tiga) ranah. Pertama, ranah kognitif. Kedua, ranah psikomotorik. Ketiga, ranah afektif. Ranah kognitif berkaitan dengan kemampuan siswa dalam mengetahui, mengingat, memahami dan sebagainya. Ranah afektif berkaitan dengan perhatian dan perasaan siswa. Sedangkan ranah psikomotorik berkaitan dengan keterampilan fisik siswa dan kemampuan mereka dalam melakukan sesuatu.³

Teori tujuan pembelajaran Robert Mager membagi tujuan pembelajaran ke dalam 3 (tiga) komponen. Pertama, kata kerja aktif (*action verb*). Menurut Mager, guru tidak perlu memerintahkan siswa untuk menghafal atau mengingat apa yang dipelajari. Tetapi, membaca atau

menuliskan apa yang dipelajari ke dalam sebuah daftar (*list*). Dalam konteks kata kerja aktif ini, demikian menurut Mager, guru memerintahkan siswa untuk melakukan sesuatu yang dapat diamati dan dengan cara ini dapat diketahui bahwa siswa telah mencapai tujuan pembelajaran. Kedua, kriteria. Dalam hal ini, guru menetapkan kriteria atau standar pencapaian tujuan pembelajaran. Misalnya, siswa harus mampu memenuhi standar minimum jawaban benar sebesar 70%-80 dari tes yang dilaksanakan untuk mencerminkan pencapaian tujuan pembelajaran. Ketiga, syarat penilaian. Dalam hal ini, guru harus menjelaskan kapan dan bagaimana siswa harus membuktikan bahwa mereka telah berhasil mencapai tujuan pembelajaran. Selain itu, guru juga harus menjelaskan ketentuan-ketentuan yang ditetapkannya dalam pelaksanaan tes hasil pembelajaran.⁴

Teori pembelajaran Norman Gronlund menganjurkan agar tujuan pembelajaran tujuan pembelajaran ditulis dalam 2 (dua) tingkatan. Pertama, guru harus mengembangkan tujuan pembelajaran yang bersifat umum. Misalnya, setelah mempelajari bab tentang demokrasi, siswa dapat memahami karakteristik tiga bagian dalam pemerintahan. Kedua, guru harus menjelaskan bagian-bagian spesifik dari tujuan pembelajaran. Misalnya, mendefinisikan peran bagian-bagian pemerintahan, menjelaskan contoh-contoh *check and balance* dan menuliskan departemen-departemen yang menjadi bagian dari pemerintahan.⁵

⁴Mager juga menyatakan bahwa tujuan umum pembelajaran merupakan sebuah deskripsi mengenai hasil yang diharapkan dalam proses pembelajaran. Lihat, Robert Mager, *Preparing Instructional Objectives* (3rd ed.), (Atlanta, GA: Center for Effective Performance, 1997), 23

⁵Norman Gronlund menyatakan bahwa tujuan umum pembelajaran merupakan taraf pertama spesifikasi unit pembelajaran yang harus

²Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 10

³Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 31

Teori hasil pembelajaran Robert Gagne membagi tujuan pembelajaran ke dalam 5 (lima) kategori. Pertama, informasi verbal yang mencakup penyebutan, penulisan, pengetahuan dan pengingatan (*recalling*). Kedua, keterampilan intelektual yang meliputi semua keterampilan kognitif versi Bloom namun disusun secara berbeda. Ketiga, kemampuan membedakan yakni kemampuan menentukan apakah dua objek itu sama atau berbeda. Keempat, konsep kongkret yang mencakup pengidentifikasian konsep-konsep sebagai bagian dari sebuah konsep kongkret tertentu (misalnya menunjukkan kepada siswa serangkaian bentuk geometrik kemudian memintanya untuk menunjukkan bentuk persegi empat dari berbagai bentuk yang ada). Kelima, konsep definitif yang mencakup berbagai konsep pada tataran yang abstrak. Contohnya, siswa harus mampu mengetahui bahwa renang, membaca atau menonton film merupakan kegiatan di waktu senggang. Kaidah ini berkaitan dengan hubungan antara berbagai konsep satu sama lain.⁶

Perumusan tujuan pembelajaran sangatlah penting dalam merencanakan tes hasil pembelajaran dan merupakan bagian tak terpisahkan dari pengukuran hasil atau prestasi pembelajaran.

2. Membuat Tabel Spesifikasi

Langkah berikutnya dalam teknik penilaian hasil pembelajaran adalah membuat tabel spesifikasi. Menurut Bloom, pembuatan tabel spesifikasi akan memungkinkan guru untuk melihat unsur-

unsur yang ada di dalam sebuah unit pembelajaran maupun hubungan antar unsur-unsur ini sebagaimana yang dikembangkan di dalam unit pembelajaran tersebut. Tabel spesifikasi juga berfungsi untuk menjelaskan unsur-unsur, perilaku dan berbagai hubungannya seperti yang dikembangkan di dalam materi pembelajaran.⁷

Tabel spesifikasi meliputi pokok-pokok materi pembelajaran yang akan diujikan, ruang lingkup materi pembelajaran berdasarkan taksonomi Bloom (kognitif, psikomotorik, afektif) dan jumlah soal tes. Dalam membuat tabel spesifikasi, perlu ditetapkan persentase dan bobot penilaian pada masing-masing materi pembelajaran yang akan diujikan. Dapat dikemukakan bahwa jumlah soal tes yang akan diujikan ditentukan berdasarkan perkalian persentase materi pembelajaran yang akan diujikan dengan jumlah keseluruhan soal tes. Persentase materi pembelajaran dan jumlah keseluruhan soal tes yang akan diujikan ditetapkan oleh guru.

Dalam membuat tabel spesifikasi, patut pula dicermati aspek-aspek yang ada pada masing-masing ranah di dalam ruang lingkup taksonomi Bloom. Dalam hal ini, Bloom membagi ranah kognitif ke dalam 6 (enam) aspek, yakni pengetahuan, pemahaman, aplikasi, analisis, sintesis dan evaluasi. Ranah psikomotorik pada dasarnya meliputi semua kemampuan siswa dalam melakukan sesuatu yang terdiri dari gerakan fisik (*physical movement*), kordinasi dan penggunaan aspek keterampilan motorik. Sedangkan ranah afektif mencakup internalisasi nilai,

dicapai siswa pada akhir pembelajaran. Untuk menjelaskan mengenai hal ini, lihat Norman Gronlund, *Writing Instructional Objectives*, 8th ed.. (Upper Saddle River, NJ: Pearson Education, Inc., 2009), 13

⁶Robert Gagne, *Learning Outcomes and Their Effects*, (Florida: American Psychologist, 1984), 378-383

⁷Benjamin S. Bloom, et.al., *Evaluation to Improve Learning*, (New York: Mc. Graw Hill, 1981), 159

penataan nilai, proses menilai (*valuing*), respon terhadap fenomena dan menerima fenomena.⁸

3. Merumuskan Soal Tes Berdasarkan Tabel Spesifikasi

Setelah tabel spesifikasi dibuat, soal-soal tes dapat dirumuskan dengan mengacu kepada tabel spesifikasi tersebut sesuai dengan materi pembelajaran dan ruang lingkup ranah (kognitif, afektif dan psikomotorik) yang berhubungan dengan materi pembelajaran. Banyak tipe soal yang dapat dipertimbangkan dan setiap tipe soal memiliki kelebihan maupun kelemahan masing-masing. Tipe soal jawaban singkat (*short answer*) mengharuskan siswa untuk menjawab dengan singkat dengan hanya menggunakan satu kata atau sebanyak-banyaknya satu kalimat. Salah satu bentuk dari tipe ini adalah mengisi, yakni sebuah kalimat yang tidak sempurna berupa bagian kosong pada akhir kalimat dan siswa diharuskan mengisinya. Jika tipe ini berbentuk pertanyaan, maka jawaban yang diberikan siswa adalah berupa jawaban singkat.⁹

Ada beberapa kelebihan pada tipe jawaban singkat ini. Pertama, tipe ini merupakan tipe yang mudah untuk dibuat. Ia memungkinkan pertanyaan untuk ditulis dengan cara yang lebih sederhana dan langsung. Guru tak membutuhkan waktu yang terlalu lama untuk mempersiapkan soal-soal tipe ini. kedua, format jawaban singkat relatif lebih mudah untuk dinilai. Ketiga, format jawaban singkat mendorong siswa untuk memberikan jawaban dengan cara memilih atau mengisi hanya satu jawaban yang benar. Dengan format pilihan ganda, misalnya, jawaban yang

ditawarkan hanya satu pilihan yang benar dan siswa cuma memilih satu jawaban yang benar saja.

Tetapi format jawaban singkat memiliki beberapa kelemahan. Pertama, format jawaban singkat ini tampaknya hanya cocok untuk soal-soal yang berkaitan dengan aspek pengetahuan dan hapalan yang sering kali berhubungan dengan ingatan tentang informasi. Kelemahan lainnya adalah bahwa format jawaban singkat tidak bisa digunakan untuk menguji tingkat keterampilan kognitif yang lebih tinggi seperti analisis, sintesis dan evaluasi.

Masih terdapat cukup banyak bentuk tes yang dapat dikemukakan, misalnya bentuk esai, pilihan ganda, bentuk pilihan salah dan benar, dan sebagainya. Bentuk tes tersebut masing-masing memiliki kelebihan maupun kekurangan. Pilihan terhadap bentuk tes tentu akan sangat bergantung kepada kebutuhan dan relevansinya dengan tujuan pembelajaran.

4. Menetapkan Standar Penilaian

Ada 2 (dua) macam acuan yang populer digunakan dalam penilaian hasil belajar. Pertama, penilaian acuan norma (*norm referenced interpretation*). Kedua, penilaian acuan patokan (*criterion referenced interpretation*). Dengan menggunakan penilaian acuan norma (PAN), nilai yang dicapai siswa pada sebuah tes dibandingkan dengan prestasi siswa lain yang mengikuti tes yang sama. Dalam penerapannya, penilaian acuan patokan merupakan perbandingan antara nilai yang dicapai siswa dengan nilai rata-rata yang dicapai oleh siswa lain pada kelas yang sama (nilai rata-rata kelas).¹⁰

⁸Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 31

⁹Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 30

¹⁰Shafaat Hussain, Tessema Tadesse, and Sumaiya Sajid, Norm Referenced and Criterion-Referenced Test in EFL Classroom, *International Journal of Humanities and Social Science Invention*, 4(10), October 2015, 27

Standar kelulusan siswa dalam tes dapat ditentukan dengan cara menetapkan standar bahwa nilai siswa yang berada di atas atau sama dengan nilai rata-rata kelas merupakan siswa yang lulus dalam tes. Siswa yang nilainya berada di bawah nilai rata-rata kelas dapat dinyatakan tidak lulus dalam tes.

Sedangkan dengan menggunakan penilaian acuan patokan (PAP), nilai siswa maupun standar minimum kelulusan dalam tes dibandingkan dengan kriteria atau standar yang ditetapkan sebelum tes dilaksanakan.¹¹

Kedua standar penilaian ini sering digunakan di berbagai lembaga pendidikan baik secara sendiri-sendiri maupun secara integral. Di masa lalu, rapor pendidikan lazim mencantumkan penilaian acuan patokan (PAP) dan penilaian acuan norma (PAN). Hanya saja penilaian acuan patokan (PAP) tidak menjadi basis dalam menentukan kenaikan kelas maupun kelulusan siswa dalam tes pembelajaran yang dilaksanakan.

C. Uji Validitas dan Reliabilitas Soal Tes

1. Uji Validitas Soal Tes

Kata validitas merujuk kepada derajat kebenaran tes atau instrumen lain yang mengukur apa yang ingin diukur. Definisi awal validitas memberi tekanan yang besar terhadap tes itu sendiri sebab validitas dianggap sebagai hal yang bersifat statis. Sebuah tes dianggap valid atau tidak valid melalui hubungan antara tes dan beberapa ukuran kriteria eksternal yang lain. Validitas juga didefinisikan sebagai derajat sebuah tes ketika mengukur apa yang direncanakan untuk diukur.¹²

Ada cukup banyak perspektif mengenai validitas ini. Pertama, perspektif tentang bukti validitas yang berkaitan dengan isi. Validitas tipe ini didefinisikan sebagai kesesuaian antara item tes dan materi pembelajaran yang diajarkan. Pertanyaan yang diajukan dalam konteks ini adalah apakah tes yang dilaksanakan sesuai dengan isi yang diajarkan. Kedua, perspektif tentang bukti validitas yang berkaitan dengan kriteria. Pertanyaan dalam konteks ini adalah apakah tes yang dilaksanakan merupakan cara alternatif untuk mengukur keterampilan atau bakat. Ketiga, perspektif tentang bukti validitas yang berkaitan dengan konstruk. Pertanyaan dalam konteks ini adalah apakah tes yang dilaksanakan sungguh-sungguh mengukur apa yang menurut teori memang harus diukur oleh tes.¹³

Pada 1955, Cronbach mengidentifikasi 4 (empat) tipe validitas: validitas prediktif, validitas bersamaan, validitas isi dan validitas konstruk. Validitas prediktif dibahas jika kriteria dirumuskan setelah tes dilaksanakan. Validitas bersamaan (*concurrent validity*) ketika skor tes dan skor kriteria ditentukan pada saat yang bersamaan dan dapat dibahas ketika sebuah tes diajukan sebagai pengganti tes yang lain. Validitas isi (*content validity*) mencerminkan sejauh mana item-item tes sesuai dengan domain isi tes dan ditetapkan dengan cara menunjukkan bahwa item-item tes merupakan contoh yang baik dari sebuah pokok bahasan atau materi pembelajaran. Validitas konstruk (*construct validity*) berfungsi untuk menafsirkan sebuah tes sebagai sebuah ukuran terhadap beberapa kualitas yang secara operasional belum ditentukan. Konstruk tes merupakan inti validitas dan dibahas ketika guru ingin membuktikan sebuah elemen valid dengan

¹¹Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 23

¹²Claudia D'Este, New Views of Validity in Language Testing, *ELLE*, 1(1), Maret 2012, 62

¹³Malcolm L. Van Blerkom, *Measurement and Statistics for Teachers*, 27

cara mengkaitkannya dengan elemen lain yang dianggap valid dan ketika konstruk yang mendasari tes lebih penting daripada tes perilaku atau skor pada kriteria.¹⁴

2. Teknik Mengukur Validitas Soal Tes

Dalam pengukuran validitas soal tes, ada 2 (dua) rumus yang lazim digunakan. Pertama, rumus korelasi *point biserial* untuk mengukur data yang bersifat diskret (misalnya tes objektif). Kedua, rumus koefisien *product moment Pearson* untuk mengukur data yang bersifat kontinu (misalnya hasil tes uraian atau skala sikap).

Teknik korelasi point biserial memiliki rumus:

$$r_{pbis} = \frac{M_p - M_t}{Sd_t} \sqrt{\frac{p}{q}}$$

r_{pbis} = koefisien korelasi point biserial

M_p = skor rata-rata hitung untuk butir soal yang dijawab benar

M_t = skor rata-rata dari skor total

Sd_t = standar deviasi dari skor total

P = proporsi siswa yang menjawab benar pada butir yang diuji validitasnya (1)

q = proporsi siswa yang menjawab salah pada butir yang diuji validitasnya (0).¹⁵

Contoh Kasus Butir Soal Objektif:

Siswa	Angka Butir Instrumen									
	1	2	3	4	5	6	7	8	9	10
A	0	1	1	1	1	0	0	1	1	1
B	1	0	1	1	1	1	0	0	1	1
C	0	0	1	1	1	1	1	1	0	0
D	1	1	0	0	0	1	0	0	1	1
E	0	1	1	1	0	0	1	1	1	0

¹⁴Claudia D'Este, New Views of Validity in Language Testing, 63

¹⁵Untuk penjelasan mengenai hal ini lihat, James Dean Brown, Point-biserial Correlation Coefficient, *Shiken: JLT Testing & Evolution SIG Newsletter*, 5(3) October, 2001, 13-17)

Persiapan Perhitungan Validitas (Korelasi Point Biserial)

Siswa	Angka Butir Instrumen										X_t	X_t^2
	1	2	3	4	5	6	7	8	9	10		
A	0	1	1	1	1	0	0	1	1	1	7	49
B	1	0	1	1	1	1	0	0	1	1	7	49
C	0	0	1	1	1	1	1	1	0	0	6	36
D	1	1	0	0	0	1	0	0	1	1	5	25
E	0	1	1	1	0	0	1	1	1	0	6	36
N	2	3	4	4	3	3	2	3	4	3	31	195
p	0.4	0.6	0.8	0.8	0.6	0.6	0.4	0.6	0.8	0.8		
q	0.6	0.4	0.2	0.2	0.4	0.4	0.6	0.4	0.2	0.2		

Dari tabel di atas sudah dapat diketahui nilai ΣX_t , yaitu 31 dan nilai ΣX_t^2 , yaitu 195.

Nilai p adalah jumlah siswa yang menjawab benar pada butir soal tertentu dibagi jumlah siswa. Dari tabel di atas, jumlah siswa yang menjawab benar pada butir 1 adalah sebanyak 2 orang. Hal ini berarti bahwa $p = 2/5 = 0.4$. Kemudian q dihitung berdasarkan ketentuan $q = 1 - p$. Pada butir 1, $q = 1 - 0.4 = 0.6$. Demikian seterusnya hingga nilai p pada tabel di atas selesai dihitung semua.

Rata-rata skor total dihitung berdasarkan ketentuan:

$$M_t = \frac{\Sigma X_t}{N} = \frac{31}{5} = 6.2$$

M_p setiap butir soal (rata-rata hitung dari skor total yang dijawab dengan benar) pada butir 1 dihitung dengan ketentuan:

Jumlah siswa yang menjawab benar pada butir 1 adalah 2 orang (siswa B dan D). Sedangkan skor total setiap siswa adalah $7 + 5 = 12$. Jadi $M_p = 12/2 = 6$.

Contoh berikutnya adalah butir 10. Jumlah siswa yang menjawab dengan benar adalah 3 orang (Siswa A, B dan D). Sedangkan skor total setiap siswa adalah $7 + 7 + 5 = 19$. Jadi $M_p = 19/3 = 6.3$

Standar deviasi total dihitung dengan menggunakan rumus:

$$Sd_t = \sqrt{\frac{\Sigma X_t^2}{N} - \left(\frac{\Sigma X_t}{N}\right)^2}$$

$$Sd_t = \sqrt{\frac{195}{5} - \left(\frac{31}{5}\right)^2}$$

$$Sd_t = \sqrt{39 - 38.44} = 0.56$$

Sedangkan validitas butir soal nomor 1 berdasarkan korelasi point biserial dihitung dengan ketentuan:

$$r_{pbis} = \sqrt{\frac{M_p - M_t}{Sd_t} \frac{p}{q}}$$

$$r_{pbis} = \sqrt{\frac{6 - 6.2}{0.56} \frac{0.4}{0.6}}$$

$$r_{pbis} = 0.60$$

Makin tinggi koefisien korelasi yang dimiliki makin valid butir instrumen tersebut. Secara umum, jika koefisien korelasi sudah lebih besar dari 0,3 maka butir instrumen tersebut sudah dikategorikan valid.

Uji validitas terhadap soal uraian (esai) dapat dilakukan dengan menggunakan rumus koefisien korelasi *product moment*, sebagai berikut:

$$r = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N \Sigma X^2 - (\Sigma X)^2\}} \sqrt{\{N \Sigma Y^2 - (\Sigma Y)^2\}}}$$

r = koefisien korelasi Pearson

ΣXY = Jumlah hasil kali skor X dan Y

ΣX = Jumlah skor X

ΣY = Jumlah skor Y

ΣX^2 = Jumlah kuadrat skor X

ΣY^2 = Jumlah kuadrat skor Y

N = Jumlah peserta tes

Contoh persiapan perhitungan soal uraian:

Siswa	Angka Butir Instrumen					Y_t	Y_t^2
	1	2	3	4	5		
A	3	1	4	2	4	14	196
B	4	3	3	4	3	17	289
C	2	4	3	4	3	16	256
D	3	2	2	3	3	13	169
E	4	3	3	2	4	16	256
5	16	13	15	15	14	76	1166

Persiapan perhitungan validitas butir instrumen

No	X_1	X_1^2	Y	Y^2	$X_1 Y$
A	3	9	14	196	42
B	4	16	17	289	68
C	2	4	16	256	32
D	3	9	13	169	39
E	4	16	16	256	64
N=5	$\Sigma X_1=16$	$\Sigma X_1^2=54$	$\Sigma Y=76$	$\Sigma Y^2=1166$	$\Sigma X_1 Y=245$

$$r = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N \Sigma X^2 - (\Sigma X)^2\}} \sqrt{\{N \Sigma Y^2 - (\Sigma Y)^2\}}}$$

$$r = \frac{5(245) - (16)(76)}{\sqrt{\{N \Sigma X^2 - (\Sigma X)^2\}} \sqrt{\{N \Sigma Y^2 - (\Sigma Y)^2\}}}$$

$$r = \frac{1225 - 1216}{\sqrt{\{270 - 256\}} \sqrt{\{5830 - 5776\}}}$$

$$r = \frac{9}{\sqrt{(3.74)(7.34)}}$$

$$r = 0.32 \quad r = 0.32$$

Semakin tinggi koefisien korelasi semakin valid butir instrumen. Secara umum, jika koefisien korelasi lebih besar dari 0,3, butir instrumen tersebut sudah dapat dikategorikan sebagai valid.¹⁶

3. Teknik Mengukur Reliabilitas Soal Tes

Menurut teori reliabilitas klasik, setiap skor yang diperoleh terdiri dari skor jawaban yang benar (yang tidak diketahui) dan skor jawaban yang salah dalam proses pengukuran.¹⁷ Uji reliabilitas digunakan untuk mengevaluasi kestabilan pengukuran yang dilaksanakan pada peserta tes yang sama atau menggunakan standar yang sama namun pada waktu yang berbeda (reliabilitas tes-tes). Uji reliabilitas juga digunakan untuk mengevaluasi kesetaraan

¹⁶W. Wiersma and S. G. Jurs, *Educational Measurement and Testing*, (2ndEd.), (Boston, MA: Allyn and Bacon, 1990), 145

¹⁷L. Crocker, J. Algina, *Introduction to Classical and Modern Test Theory*, (Orlando, FL: Harcourt Brace Jovanovich; 1986), 49

(ekuivalensi) kumpulan item dari tes yang sama (konsistensi internal) atau mengevaluasi berbagai tes dengan menggunakan instrumen yang sama (*inter-rater reliability*). Koefisien reliabilitas merentang dari 0.00 hingga 1.00 dan koefisien yang lebih tinggi menunjukkan tingkat reliabilitas yang lebih tinggi.¹⁸

Koefisien konsistensi internal memberikan sebuah perhitungan mengenai pengukuran reliabilitas dan didasarkan atas asumsi bahwa pengukuran item-item dengan konstruk yang sama pasti berkorelasi. Metode yang mungkin paling sering digunakan untuk mengukur reliabilitas konsistensi internal adalah Cronbach Alpha.¹⁹

Reliabilitas antar penilai (*inter rater reliability*) yang juga disebut sebagai kesepakatan antar penguji (*agreement interobserver*) menetapkan kesetaraan penilaian yang diperoleh sebuah instrumen yang digunakan oleh penguji yang berbeda. Reliabilitas antar penilai membutuhkan penilaian yang sangat independen terhadap sebuah tes yang sama yang dilaksanakan oleh penilai yang berbeda. Dalam hal ini, reliabilitas ditentukan oleh korelasi skor dari dua atau lebih penilai yang independen atau koefisien kesepakatan penilaian para penilai. Rumus Cohen Kappa lazim digunakan untuk menentukan koefisien kesepakatan tersebut.

Reliabilitas soal tes pilihan ganda dapat diukur dengan menggunakan rumus KR 20 sebagai berikut:

$$r_1 = \frac{k}{(k-1)} \left(\frac{S_1^2 - \sum p_1 q_1}{S_1^2} \right)$$

¹⁸Carole L. Kimberlin and Almut G. Winterstein, *Validity And Reliability Of Measurement Instruments Used In Research*, Am J Health-Syst Pharm—Vol 65 Dec 1, 2008, 2277

¹⁹Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*. 1951, 16, 297-334.

k = Jumlah item soal tes

p₁ = Banyaknya peserta tes yang menjawab benar pada butir 1 soal tes

q₁ = 1 – p₁

S₁² = varian total

Persiapan perhitungan realibilitas (KR 20)

Siswa	Angka Butir Instrumen										X _i	X _i ²
	1	2	3	4	5	6	7	8	9	10		
A	0	1	1	1	1	0	0	1	1	1	7	49
B	1	0	1	1	1	1	0	0	1	1	7	49
C	0	0	1	1	1	1	1	1	0	0	6	36
D	1	1	0	0	0	1	0	0	1	1	5	25
E	0	1	1	1	0	0	1	1	1	0	6	36
N	2	3	4	4	3	3	2	3	4	3	31	195
p	0.4	0.6	0.8	0.8	0.6	0.6	0.4	0.6	0.8	0.8	Σp=	6.4
q	0.6	0.4	0.2	0.2	0.4	0.4	0.6	0.4	0.2	0.2		
pq	0.24	0.24	0.16	0.16	0.24	0.24	0.24	0.24	0.16	0.16	Σpq=	2.08

Sebelum angka-angka ini dimasukkan ke dalam rumus, varian total harus dihitung terlebih dahulu dengan menggunakan rumus:

$$S_1^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{(N)}$$

$$S_1^2 = \frac{195 - \frac{(31)^2}{5}}{5}$$

$$S_1^2 = \frac{195 - 192.2}{5}$$

$$S_1^2 = 0.56$$

Hasil perhitungan ini kemudian dimasukkan ke dalam rumus KR 20 sebagai berikut:

$$r_1 = \frac{k}{(k-1)} \left(\frac{S_1^2 - \sum p_1 q_1}{S_1^2} \right)$$

$$r_1 = \frac{10}{(10-1)} \left(\frac{0.56 - 2.08}{0.56} \right)$$

$$= 1.1 (-2.7)$$

$$= -2.97$$

Sedangkan uji reliabilitas tes berbentuk esai menggunakan rumus Alfa Cronbach sebagai berikut:

$$r_1 = \frac{k}{(k-1)} \left(1 - \frac{\sum S_1^2}{S_1^2} \right)$$

k = mean kuadrat antara subjek

$\sum S_1^2$ = mean kuadrat kesalahan

S_1^2 = varian total

Rumus untuk varians total dan varians item:

$$S_1^2 = \frac{\sum X_1^2}{n} - \frac{(\sum X_1)^2}{n^2}$$

$$S_1^2 = \frac{JK_i}{n} - \frac{JK_s}{n^2}$$

JK_i = jumlah kuadrat seluruh skor item

JK_s = jumlah kuadrat subjek

Contoh persiapan perhitungan soal uraian:

Siswa	Angka Butir Instrumen					X_i	X_i^2
	1	2	3	4	5		
A	3	1	4	2	4	14	196
B	4	3	3	4	3	27	729
C	2	4	3	4	3	16	256
D	3	2	2	3	3	13	169
E	4	3	3	2	4	16	256
5	16	13	15	15	17	88	1606
	54	39	47	47	59	7744	

Dengan harga-harga sebagaimana tercantum di dalam tabel di atas, varian total dan varian item sudah dapat ditentukan.

Varian total:

$$S_1^2 = \frac{(\sum X_1^2)}{n} - \frac{(\sum X_1)^2}{n^2}$$

$$\begin{aligned} S_1^2 &= \frac{1606}{5} - \frac{(88)^2}{25} \\ &= 321.2 - 309.76 \\ &= 11.4 \end{aligned}$$

Varian Item:

$$S_1^2 = \frac{JK_i}{n} - \frac{JK_s}{n^2}$$

$$\begin{aligned} S_1^2 &= \frac{2916}{5} - \frac{7744}{25} \\ &= 583.2 - 309.76 \\ &= 273.4 \end{aligned}$$

Jika dimasukkan ke dalam rumus Alfa Cronbach, maka hasilnya adalah sebagai berikut:

$$r_1 = \frac{k}{(k-1)} \left(1 - \frac{\sum S_1^2}{S_1^2} \right)$$

k = mean kuadrat antara subjek

$\sum S_1^2$ = mean kuadrat kesalahan

S_1^2 = varian total

$$\begin{aligned} r_1 &= \frac{5}{(5-1)} \left(1 - \frac{11.4}{273.4} \right) \\ &= 1.25 (1 - 0.04) \\ &= 1.25 (0.96) \\ &= 1.2 \end{aligned}$$

Jadi, koefisien reliabilitas soal tes adalah 1.2.²⁰

D. Kesimpulan

Dari uraian di atas dapat disimpulkan bahwa teknik penilaian hasil pembelajaran mencakup:

1. Analisis yang cermat terhadap materi pembelajaran, tujuan pembelajaran dan penyusunan tabel spesifikasi baik menurut teori Bloom maupun teori Gagne.
2. Tabel spesifikasi juga berfungsi untuk merumuskan soal tes formatif maupun sumatif dengan bentuk soal pilihan ganda maupun uraian (esai).
3. Uji validitas soal tes memastikan apakah soal tes yang dirumuskan valid ataukah tidak.
4. Uji reliabilitas soal tes memastikan apakah soal tes yang dirumuskan reliabel ataukah tidak.

²⁰Untuk penjelasan mengenai hal ini, lihat Sugiyono, *Statistika untuk Penelitian*, (Bandung: Alfabeta, 2010), 365-366

DAFTAR PUSTAKA

- Bloom, Benjamin S., et.al., *Evaluation to Improve Learning*, New York: Mc. Graw Hill, 1981
- Brown, James Dean, Point-biserial Correlation Coefficient, *Shiken: JLT Testing & Evolution SIG Newsletter*, 5(3) October, 2001
- Crocker, L., Algina J., *Introduction to Classical and Modern Test Theory*, Orlando, FL: Harcourt Brace Jovanovich: 1986
- D'Este, Claudia, New Views of Validity in Language Testing, *EL.LE*, 1(1), Maret 2012
- Gagne, Robert, *Learning Outcomes and Their Effects*, Florida: American Psychologist, 1984
- Gronlund, Norman, *Writing Instructional Objectives*, 8th ed., Upper Saddle River, NJ: Pearson Education, Inc., 2009
- Hussain, Shafaat, Tadesse, Tessema, and Sajid, Sumaiya, Norm Referenced and Criterion-Referenced Test in EFL Classroom, *International Journal of Humanities and Social Science Invention*, 4(10), October 2015
- Kimberlin, Carole L. and Winterstein Almut G., *Validity And Reliability Of Measurement Instruments Used In Research*, Am J Health-Syst Pharm—Vol 65 Dec 1, 2008
- LJ., Cronbach, Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*. 16, 1951
- Mager, Robert, *Preparing Instructional Objectives* (3rd ed.), Atlanta, GA: Center for Effective Performance, 1997
- Sugiyono, *Statistika untuk Penelitian*, Bandung: Alfabeta, 2010
- Van Blerkom, Malcolm L., *Measurement and Statistics for Teachers*, New York: Routledge, 2008
- Wiersma, W. and Jurs S. G., *Educational Measurement and Testing*, (2ndEd.), Boston, MA: Allyn and Bacon, 1990