

Analisis Soal Evaluasi Melalui Pendekatan Kuantitatif

Budi Nurdiansyah

Widyaiswara Ahli Pertama, BBPPKS Banjarmasin

E-mail: jangbud@gmail.com

Abstrak

Evaluasi pembelajaran merupakan suatu proses yang sistematis dan berkesinambungan untuk mengetahui efisiensi kegiatan pembelajaran dan efektivitas pencapaian tujuan pembelajaran yang telah ditetapkan. Salah satu bentuk kegiatan evaluasi pembelajaran adalah dengan memberikan soal yang harus dijawab oleh peserta diklat. Untuk mengetahui kualitas soal apakah layak dijadikan evaluasi atau tidak, maka perlu dilakukan analisis soal. Salah satu cara analisis soal adalah melalui pendekatan kuantitatif yaitu dengan cara menghitung koefisien validitas, koefisien reliabilitas, dan indeks kesukaran. Soal dikatakan valid dan reliabel apabila soal tersebut memiliki kemampuan yang baik untuk mengukur apa yang seharusnya diukur yaitu dalam hal ini kemampuan peserta diklat memahami materi. Sementara itu, indeks kesukaran soal adalah tingkat kesulitan soal untuk dijawab oleh peserta diklat. Kriteria yang diharapkan dalam suatu evaluasi pembelajaran adalah tiap butir soal tidak ada yang validitasnya sangat jelek, koefisien reliabilitas evaluasi pembelajaran tidak boleh memiliki interpretasi sangat rendah, dan komposisi soal yang mudah sebanyak 25 persen, soal sedang 50 persen, dan sukar 25 persen.

Kata Kunci: Analisis soal, evaluasi pembelajaran, validitas, reliabilitas, dan indeks kesukaran

Abstract

Learning evaluation is a systematic and continuous process to determine the efficiency of learning activities and the effectiveness of achieving the learning objectives that have been set. One of learning evaluation activities is to provide questions that must be answered by training participants. To find out whether the quality of the questions is good or not, it is necessary to analyze the questions. One way of analyzing them is through a quantitative approach, i.e. calculating the validity coefficient, reliability coefficient, and difficulty index. Questions are said to be valid and reliable if they have a good ability to measure what should be measured, exactly the ability of the training participants to understand the training material. Meanwhile, the item difficulty index is the level of difficulty of the questions to be answered by participants. The expected criteria in learning evaluation are that each item does not have very poor validity, the reliability coefficient of learning evaluation should not have very low interpretations, and the composition of easy questions is 25 percent, moderate questions are 50 percent, and difficult questions are 25 percent.

Keywords: Questions analysis, learning evaluation, validity, reliability, and difficulty index.

A. PENDAHULUAN

Evaluasi pembelajaran merupakan suatu proses yang sistematis dan berkesinambungan untuk mengetahui sampai sejauh mana efisiensi kegiatan pembelajaran yang dilaksanakan dan efektivitas pencapaian tujuan pembelajaran yang telah ditetapkan (Suherman, 2003). Evaluasi pembelajaran tidak hanya memberikan gambaran tentang kemampuan yang dimiliki peserta pendidikan dan pelatihan (diklat), tetapi bisa pula untuk memberikan informasi mengenai keberhasilan atau kegagalan widyaiswara dalam menerapkan metode pembelajarannya. Selain itu, evaluasi pembelajaran bisa pula bermanfaat untuk menentukan arah kebijakan bagi penyelenggara diklat. Peranan evaluasi begitu hakiki dalam kegiatan pembelajaran. Evaluasi dapat mendorong peserta diklat untuk lebih giat belajar dan mendorong widyaiswara untuk lebih meningkatkan kualitas proses pembelajaran serta mendorong pengelola pendidikan untuk lebih meningkatkan fasilitas dan kualitas belajar peserta diklat (Idrus, 2019, p.921).

Salah satu bentuk kegiatan evaluasi pembelajaran adalah dengan memberikan soal-soal yang harus dijawab oleh peserta diklat baik itu berupa soal latihan, kuis, pretest, posttest, maupun ujian komprehensif. Untuk mengetahui apakah soal yang dibuat memiliki kualitas yang baik atau buruk maka diperlukan analisis soal

evaluasi. Analisis soal evaluasi bisa dilakukan secara objektif melalui pendekatan kuantitatif sehingga terbebas dari persepsi individu yang cenderung menjelekkan atau membaguskan kualitas soal. Dalam hal ini, analisis soal evaluasi melalui pendekatan kuantitatif dianggap lebih adil dalam menguji kualitas soal evaluasi. Analisis soal evaluasi pendekatan kuantitatif tersebut meliputi koefisien validitas, koefisien reliabilitas, dan indeks kesukaran. Ketiga unsur tersebut cukup untuk menentukan apakah soal memiliki kualitas baik atau buruk.

Analisis soal dilakukan setelah diperoleh data skor peserta diklat. Data tersebut diolah dengan perhitungan matematika sehingga diperoleh nilai koefisien validitas, koefisien reliabilitas, dan indeks kesukaran. Suatu instrumen penelitian dikatakan berkualitas dan dapat dipertanggungjawabkan jika sudah dapat terbukti validitas dan reliabilitasnya (Priatna, 2008). Setelah diperoleh nilai-nilai tersebut, soal-soal dipilah yang memenuhi kriteria baik dipertahankan sedangkan yang buruk bisa direvisi atau diganti dengan yang baru.

B. PEMBAHASAN

1. Validitas

Suatu soal dikatakan valid apabila soal tersebut mampu mengukur apa yang seharusnya diukur (Ary, 2010). Sebagai analogi istilah valid adalah sebuah pisau

silet sangat valid digunakan untuk mencukur jenggot dibandingkan dengan gergaji meskipun gergajinya tajam. Pisau dapur yang tajam boleh dikatakan valid untuk mencukur jenggot tetapi lebih valid lagi untuk memotong sayuran. Begitu juga dengan soal yang sudah dibuat, apakah sudah benar-benar dapat mengukur keberhasilan peserta diklat dalam menguasai materi pada modul tertentu. Meskipun peserta diklat banyak yang menjawab benar, belum tentu benarnya itu menggambarkan bahwa peserta diklat menguasai materi dalam modul.

Cara cepat untuk menentukan soal evaluasi ini valid atau tidak valid bisa dengan pendekatan kualitatif yaitu menanyakan kepada yang ahli dalam menguasai modul setidaknya kepada tiga

orang ahli. Tetapi terkadang hal itu tidak bisa dilakukan karena beberapa hal seperti tidak tersedianya waktu orang yang dianggap ahli untuk menentukan validitas soal atau masih tabu untuk mengatakan bahwa soal yang dibuat tidak valid atau jelek. Dengan demikian jalan alternatif yang lebih “halus” untuk menentukan validitas adalah dengan pendekatan kuantitatif yaitu menentukan koefisien validitas.

Penentuan koefisien validitas soal dilakukan dengan cara menghitung koefisien korelasi antara Skor tiap butir soal dan skor totalnya. Rumus yang digunakan untuk menghitung korelasi ditunjukkan Gambar 1, dan klasifikasi koefisien validitas tampak pada Tabel 1.

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}$$

Keterangan:

r_{xy}	= koefisien validitas/korelasi
x_i	= skor butir soal
y_i	= Skor total
n	= banyaknya peserta diklat

Gambar 1. Rumus penghitungan korelasi (Suherman, 2003)

Tabel 1. Klasifikasi koefisien validitas

Besarnya r_{xy}	Interpretasi
$0,90 < r_{xy} \leq 1,00$	Validitas sangat tinggi (sangat baik)
$0,70 < r_{xy} \leq 0,90$	Validitas tinggi (baik)
$0,40 < r_{xy} \leq 0,70$	Validitas sedang (cukup)
$0,20 < r_{xy} \leq 0,40$	Validitas rendah (jelek)
$0,00 < r_{xy} \leq 0,20$	Validitas sangat rendah (sangat jelek)
$r_{xy} = 0,00$	Tidak valid

Sumber: Guilford (1956, p. 145)

Apabila ada butir soal yang koefisien validitasnya tidak valid dan sangat rendah, maka sebaiknya soal tersebut tidak digunakan lagi. Apabila koefisiennya validitasnya rendah, maka bisa menjadi pertimbangan apakah masih mau digunakan atau tidak tergantung komponen lainnya seperti reliabilitas dan indeks kesukaran. Untuk yang validitasnya sedang, tinggi, dan sangat tinggi, soal tersebut masih bisa digunakan.

Beberapa yang menyebabkan suatu evaluasi menjadi tidak valid adalah petunjuk tidak jelas, kalimat yang sukar dimengerti, soal tes yang terlalu sukar /mudah, materi tes tidak sesuai dengan yang diujikan.

2. Reliabilitas Soal

Suatu mistar terbuat dari besi digunakan untuk mengukur pensil menunjukkan angka 20 cm. Mistar tersebut digunakan oleh orang lain untuk mengukur pensil yang sama maka

hasilnya juga akan tetap 20 cm. Lain halnya apabila mistar tersebut terbuat dari karet, maka belum tentu hasilnya sama. Kondisi tersebut menggambarkan reliabilitas suatu mistar digunakan untuk mengukur panjang. Tentunya mistar yang terbuat dari besi memiliki reliabilitas yang lebih bagus dibandingkan dengan yang terbuat dari karet.

Suatu soal dikatakan memiliki reliabilitas yang baik apabila soal tersebut memiliki hasil pengukuran yang sama terhadap subjek yang sama (subjek sama bukan berarti orang sama) dan tidak berpengaruh terhadap pelaku tes, situasi, dan kondisi. Untuk menentukan reliabilitas soal, bisa dihitung nilai koefisien reliabilitasnya. Untuk menghitung koefisien reliabilitas soal *essay*, maka rumus yang digunakan ditunjukkan Gambar 2. Sementara untuk menghitung koefisien reliabilitas soal pilihan ganda pada Gambar 3.

$$r_{11} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum_{i=1}^n s_i^2}{s_{total}^2} \right)$$

Keterangan:

n = banyaknya soal

s_i^2 = varians skor tiap soal

s_{total}^2 = varians skor total

Gambar 2. Rumus koefisien reliabilitas soal *essay* (Suherman, 2003)

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{s_{total}^2 - \sum_{i=1}^n p_i q_i}{s_{total}^2} \right)$$

Keterangan:

n = banyaknya soal

p_i = proporsi banyak subyek yang menjawab benar pada butir soal ke- i

q_i = proporsi banyak subyek yang menjawab salah pada butir soal ke- i

s_{total}^2 = varians skor total

Gambar 3. Rumus koefisien reliabilitas soal pilihan ganda (Suherman, 2003)

Rumus untuk menghitung varians skor total tertera pada Gambar 4 dan

Tabel 2 menunjukkan klasifikasi koefisien reliabilitas

$$s^2 = \frac{n \sum X_i^2 - (\sum X_i)^2}{n(n-1)}$$

Gambar 4. Rumus perhitungan varians skor total (Sudjana, 2005)

Tabel 2. Klasifikasi koefisien reliabilitas

Besarnya r_{II}	Interpretasi
$r_{II} \leq 0,20$	Koefisien reliabilitas sangat rendah
$0,20 < r_{II} \leq 0,40$	Koefisien reliabilitas rendah
$0,40 < r_{II} \leq 0,70$	Koefisien reliabilitas sedang
$0,70 < r_{II} \leq 0,90$	Koefisien reliabilitas tinggi
$0,90 < r_{II} \leq 1,00$	Koefisien reliabilitas sangat tinggi

Sumber: Guilford (1956, p. 145)

Seperti halnya koefisien validitas soal, koefisien reliabilitas juga memiliki interpretasi dan bisa digunakan sebagai pertimbangan apakah soal akan tetap digunakan atau dibuang. Kriteria soal buruk menurut koefisien reliabilitas adalah

koefisien reliabilitas memiliki interpretasi reliabilitas sangat rendah.

2. Indeks Kesukaran

Indeks Kesukaran (IK) menyatakan koefisien kesukaran suatu butir soal. Rumus untuk menentukan indeks kesukaran ditunjukkan Gambar 5. Sedangkan klasifikasinya ada pada Tabel 3.

$$IK = \frac{\bar{x}}{SMI}$$

Keterangan:
 \bar{x} = nilai rata-rata tiap butir soal
 SMI = skor maksimal ideal

Gambar 5. Rumus perhitungan Indeks Kesukaran (Suherman, 2003)

Tabel 3. Klasifikasi indeks kesukaran

Besarnya IK	Interpretasi
$IK = 0,00$	Soal terlalu sukar
$0,00 < IK \leq 0,30$	Soal sukar
$0,30 < IK \leq 0,70$	Soal sedang
$0,70 < IK < 1,00$	Soal mudah
$IK = 1,00$	Soal terlalu mudah

Sumber: Suherman (2003)

Berdasarkan kriteria indeks kesukaran soal, maka sepatutnya buang soal yang memiliki interpretasi soal yang terlalu mudah dan terlalu sukar. Idealnya suatu tes memiliki komposisi soal 25 persen soal yang mudah, 50 persen soal yang sedang, dan 25 persen soal yang sukar.

3. Contoh Kasus

Sebagai contoh, suatu lembaga diklat akan memberikan ujian

komprehensif sebanyak delapan soal berbentuk pilihan ganda kepada sepuluh orang peserta pada kegiatan Pelatihan Dasar Pekerjaan Sosial (PDPS). Karena pelatihan tersebut masih akan dilaksanakan beberapa angkatan lagi dan lembaga diklat menginginkan soal-soal yang bagus pada ujian komprehensif, maka soal-soal ujian komprehensif tersebut dianalisis. Hasil pemeriksaan tertera pada Tabel 4.

Tabel 4. Contoh kasus jawaban peserta diklat

	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	Soal 8	Total
Peserta A	1	1	1	0	1	0	1	1	6
Peserta B	1	0	1	1	0	0	1	1	5
Peserta C	1	1	1	1	1	1	1	1	8
Peserta D	0	1	0	1	1	1	1	0	5
Peserta E	1	1	0	0	0	1	0	0	3
Peserta F	1	0	0	1	0	0	1	1	4
Peserta G	1	0	1	0	1	0	0	1	4
Peserta H	1	1	1	0	1	1	0	1	6
Peserta I	0	1	0	1	0	1	1	0	4
Peserta J	1	0	1	1	0	0	1	0	4

Keterangan: 1 artinya peserta menjawab benar dan 0 artinya peserta menjawab salah.

a. Validitas

Koefisien validitas butir soal nomor 1, menggunakan rumus

koefisien validitas seperti di atas (Gambar 6).

$$r_{xy} = \frac{n \sum x_i y_i - \left(\sum x_i \right) \left(\sum y_i \right)}{\sqrt{\left(n \sum x_i^2 - \left(\sum x_i \right)^2 \right) \left(n \sum y_i^2 - \left(\sum y_i \right)^2 \right)}}$$

Gambar 6. Rumus koefisien validitas

Isian tabel yang dibutuhkan untuk mencari koefisien validitas soal nomor 1 tampak pada Tabel 5.

Tabel 5. Isian tabel untuk menghitung koefisien validitas

	Soal 1 (X _i)	Total (Y _i)	X _i Y _i	(X _i) ²	(Y _i) ²
Peserta A	1	6	6	1	36
Peserta B	1	5	5	1	25
Peserta C	1	8	8	1	64
Peserta D	0	5	0	0	25
Peserta E	1	3	3	1	9
Peserta F	1	4	4	1	16
Peserta G	1	4	4	1	16
Peserta H	1	6	6	1	36
Peserta I	0	4	0	1	16
Peserta J	1	4	4	1	16
Jumlah	Σ X _i =8	Σ Y _i =49	Σ X _i Y _i =40	Σ X _i ² =8	Σ Y _i ² =259

Gambar 7 adalah perhitungan koefisien validitasnya. Itu artinya soal nomor 1 memiliki koefisien validitas sangat rendah sesuai dengan kriteria interpretasi koefisien

validitas sehingga soal nomor 1 perlu dibuang atau ditinjau ulang. Tabel 6 adalah data validitas soal untuk soal kasus di atas.

$$r_{xy} = \frac{10.40 - 8.49}{\sqrt{(10.8 - 8^2)(10.259 - 49^2)}} = 0,145$$

Gambar 7. Perhitungan koefisien validitas

Tabel 6. Contoh kasus interpretasi koefisien validitas

	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	Soal 8
Koefisien validitas	0,145	0,386	0,535	0,089	0,655	0,218	0,270	0,535
Interpretasi	Sangat Jelek	Jelek	Cukup	Sangat Jelek	Cukup	Jelek	Jelek	Cukup

b. Reliabilitas

Untuk menghitung koefisien reliabilitas, langkah pertama adalah menentukan varians skor total yang

akan digunakan untuk soal nomor 1 sampai dengan soal nomor 8. Perhatikan Tabel 7.

Tabel 7. Isian tabel untuk menghitung varians

	Total (X_i)	X_i^2
Peserta A	6	36
Peserta B	5	25
Peserta C	8	64
Peserta D	5	25
Peserta E	3	9
Peserta F	4	16
Peserta G	4	16
Peserta H	6	36
Peserta I	4	16
Peserta J	4	16
Jumlah	$\sum X_i=49$	$\sum X_i^2=259$

Sehingga, diperoleh varians skor totalnya tampak pada Gambar 8.

Tabel 8 diperlukan untuk menentukan koefisien reliabilitas.

$$s^2 = \frac{n \sum X_i^2 - (\sum X_i)^2}{n(n-1)} = \frac{10.259 - 49^2}{10.9} = 2,1$$

Gambar 8. Perhitungan varians skor

Tabel 8. Isian tabel untuk menghitung koefisien reliabilitas

	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	Soal 8
Menjawab benar	8	6	6	6	5	5	7	6
Menjawab salah	2	4	4	4	5	5	3	4
P	0,8	0,6	0,6	0,6	0,5	0,5	0,7	0,6
Q	0,2	0,4	0,4	0,4	0,5	0,5	3	0,4
p.q	0,16	0,24	0,24	0,24	0,25	0,25	0,21	0,24

Diperoleh $\sum pq = 0,16+0,24+0,24+0,24+0,25+0,25+0,21+0,24 = 1,83$. Jadi, koefisien

reliabilitasnya adalah seperti ditunjukkan Gambar 9.

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{s_{total}^2 - \sum_{i=1}^n p_i q_i}{s_{total}^2} \right)$$

$$r_{11} = \left(\frac{10}{10-1} \right) \left(\frac{2,1 - 1,83}{2,1} \right) = 0,143$$

Gambar 9. Perhitungan koefisien reliabilitas

Ternyata, koefisien reliabilitas adalah 0,143 sehingga memiliki interpretasi reliabilitas sangat rendah. Untuk analisis soal ditinjau dari reliabilitas, yang dilihat adalah bukan per butir soal melainkan soal tes keseluruhan. Kalau berdasarkan contoh kasus, maka Koefisien reliabilitas ujian komprehensif sangat rendah dan soal-soalnya tidak cocok untuk digunakan ujian komprehensif berikutnya.

c. Indeks kesukaran

Untuk menghitung indeks kesukaran, maka yang kita perlukan adalah

$$IK = \frac{\bar{x}}{SMI} = \frac{8}{10} = 0,8$$

Gambar 10. Perhitungan indeks kesukaran

Berdasarkan tabel interpretasi indeks kesukaran, maka soal nomor 1 tergolong soal yang mudah. Tabel

nilai rata-rata butir soal dan Skor Maksimal Ideal. Seperti contoh kasus di atas, untuk soal nomor 1 diperoleh nilai rata-rata soal nomor 1 adalah $\bar{x} = 0,8$ dengan menghitung jumlah nilai dibagi banyaknya peserta diklat. Skor maksimal idealnya adalah 1 karena pilihan ganda pada kasus ini berbobot 1 untuk benar. Dengan demikian, indeks kesukaran untuk soal nomor 1 ditunjukkan Gambar 10.

9 menunjukkan indeks kesukaran tiap butir soal untuk contoh kasus di atas.

Tabel 9. Contoh kasus interpretasi indeks kesukaran

	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	Soal 8
\bar{x}	0,8	0,6	0,6	0,6	0,5	0,5	0,7	0,6
SMI	1	1	1	1	1	1	1	1
IK	0,8	0,6	0,6	0,6	0,5	0,5	0,7	0,6
Interpretasi	Mudah	Sedang	Sedang	Sedang	Sedang	Sedang	Sedang	Sedang

C. PENELITIAN YANG RELEVAN

1. Penelitian Pristiany Hadisuryani

Penelitian yang dilakukan oleh Pristiany Hadisuryani (2016) di Cigadung Kota Bandung pada siswa kelas IV dengan tujuan menganalisis peningkatan kemampuan pemecahan masalah matematis siswa yang memperoleh pembelajaran Pendidikan matematika realistik. Instrumen yang digunakan dalam penelitian ini adalah soal matematika untuk mengukur kemampuan pemecahan masalah matematis.

Untuk mendapatkan soal matematika yang benar-benar mengukur kemampuan pemecahan masalah matematis, maka sebelumnya dilakukan uji instrumen terlebih dahulu kepada siswa kelas V SDN Neglasari 4 Kota Bandung sebanyak 31 orang. Uji instrumen yang dilakukan adalah menghitung koefisien validitas, reliabilitas, daya pembeda, dan indeks kesukaran. Hasilnya adalah 8 soal yang dirancang oleh peneliti memiliki validitas yang memadai, derajat reliabilitas tes sangat tinggi, dan memiliki 7 indeks kesukaran sedang dan 1 sukar sehingga dari hasil analisis tersebut diperoleh

bahwa tes siap digunakan untuk mengukur kemampuan pemecahan masalah matematis.

2. Penelitian Elviana

Penelitian yang dilakukan oleh Elviana dengan tujuan untuk mengetahui karakteristik butir-butir soal Evaluasi Pembelajaran Pendidikan Agama Islam (PAI) di Universitas Islam Negeri (UIN) Ar-Raniry Banda Aceh. Soal diujikan kepada 26 mahasiswa prodi PAI fakultas Tarbiyah UIN Ar-Raniry Banda Aceh. Jumlah butir soal sebanyak 25 dengan bentuk soal pilihan ganda. Berdasarkan perhitungan koefisien validitas dan reliabilitas, terdapat 6 soal yang layak digunakan dan soal evaluasi tersebut memiliki koefisien reliabilitas 0,39 sehingga dikategorikan reliabilitasnya rendah. Soal tersebut memiliki indeks kesukarannya terlalu mudah berjumlah 2 butir, kriteria Mudah berjumlah 1 butir, kriteria Sedang berjumlah 15 butir, kriteria Sukar berjumlah 4 butir dan kriteria sangat sukar berjumlah 3 butir.

D. PENUTUP

Dalam Peraturan Pemerintah Republik Indonesia Nomor 39 Tahun 2012 tentang *Penyelenggaraan Kesejahteraan Sosial* bahwa tujuan pelaksanaan pendidikan dan pelatihan kesejahteraan sosial adalah agar sumber daya manusia memiliki dan meningkatkan pengetahuan, sikap, perilaku, dan keterampilan yang profesional dalam melaksanakan Penyelenggaraan Kesejahteraan Sosial. Terutama untuk mengukur unsur pengetahuan yang diperoleh oleh peserta diklat, perlu diberikan evaluasi hasil belajar berupa tes yang di dalamnya berisi soal-soal sesuai dengan indikator kompetensi jenis diklat. Supaya evaluasi hasil belajar tersebut benar-benar berfungsi, maka diperlukan alat ukur yang baik. Dengan kata lain, diperlukan soal-soal tes yang memenuhi kriteria valid dan reliabel. Oleh karena itu, analisis soal evaluasi perlu dilakukan dalam setiap kediklatan. Analisis soal tersebut setidaknya dilihat dari unsur validitas, reliabilitas, dan indeks kesukaran.

Adapun saran yang dapat diangkat di sini adalah sebagai berikut:

1. Dalam melakukan analisis soal evaluasi terdapat perhitungan matematis yang cukup rumit, untuk menghindari hal tersebut, analisis soal evaluasi bisa menggunakan aplikasi yang sudah beredar seperti anates atau SPSS (Hidayat, 2013)

bahkan bisa dibuatkan program analisis evaluasi baik berbasis web atau excel.

2. Untuk mendapatkan data skor hasil tes agar bisa dilakukan analisis soal, maka bisa dilakukan uji coba soal evaluasi atau data skor bisa diperoleh dari kegiatan *Training of Trainers* (TOT) sehingga pada saat pelaksanaan kegiatan Diklat dapat memiliki kualitas soal yang memenuhi kriteria.
3. Interpretasi hasil analisis soal evaluasi ini berbasis skor sehingga apabila ada soal yang divonis tidak memenuhi kriteria tidak berarti mutlak soal tersebut jelek tetapi sebagai pertimbangan untuk direvisi atau dibuang. Salah satu kasus yang menggambarkan hal ini adalah pada pelaksanaan Diklat Penanganan Korban Tindak Kekerasan terhadap Anak dan Perempuan di BBPPKS Banjarmasin tahun 2018, salah satu soal evaluasi hasil belajar yang berbentuk pilihan ganda telah divonis memiliki kriteria jelek. Setelah ditinjau kembali soal tersebut, ternyata terdapat salah kunci jawaban. Setelah diperbaiki kunci jawabannya, kemudian dilakukan analisis ulang lalu hasilnya soal tersebut berubah menjadi memiliki kriteria baik.

Referensi

- Ary, D. (2010). *Introduction to Research in Education*. Wadsworth Cengage Learning: Belmont USA.
- Elviana (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal MUDARRISUNA*, Volume 10 No. 2, 58-74.
- Guilford, J.P. (1956). *Fundamental Statistics in Psychology and Education*. New York: Mc Graw-Hill Book Co. Inc.
- Hadisuryani, P. (2016). *Peningkatan Kemampuan Pemecahan Masalah Matematis Siswa Melalui Pembelajaran Pendidikan Matematika Realistik: Studi Kuasi Eksperimen pada Siswa Kelas IV di SDN Cigadung Kota Bandung*. Skripsi di Universitas Pendidikan Indonesia Bandung
- Hidayat, A. (2013). *Tutorial Cara Uji Validitas Dengan SPSS – Validitas Instrumen*. Diakses dari <https://www.statistikian.com/2013/02/tutorial-uji-validitas-dengan-spss.html> (4 Juni 2021)
- Idrus, L. (2019). Evaluasi dalam Proses Pembelajaran. *Jurnal Manajemen Pendidikan Islam Adaara*, 9(2), 920-935.
- Peraturan Pemerintah Republik Indonesia Nomor 39 Tahun 2012 tentang *Penyelenggaraan Kesejahteraan Sosial*.
- Priatna, B.A. (2008). *Uji Coba Instrumen Penelitian dengan Menggunakan MS Excel dan SPSS*. Diakses dari http://file.upi.edu/Direktori/FPMIP/A/JUR._PEND._MATEMATIKA/196412051990031-BAMBANG_AVIP_PRIATNA/M/Makalah_November_2008.pdf (4 Juni 2021)
- Sudjana (2005). *Metode Statistika*. Bandung: Tarsito.
- Suherman, E. (2003). *Evaluasi Pembelajaran Matematika untuk Guru dan Calon Guru Matematika*. Bandung: JICA