

**ANALISIS BUTIR SOAL EVALUASI PESERTA
DIKLAT PRAJABATAN CPNS K1/K2 GOLONGAN III
PROVINSI DKI JAKARTA TAHUN 2018**

***THE ANALYSIS OF EVALUATION ITEMS OF CPNS K1/K2
PRE-SERVICE TRAINING PARTICIPANTS GROUP III
DKI JAKARTA PROVINCE 2018***

Yanto Suharto

Widyaiswara BPSDM Provinsi DKI Jakarta, Jakarta Pusat, 10610

yanhart.67@gmail.com

ABSTRAK

Analisis butir soal evaluasi dilakukan untuk mengetahui kualitas instrumen evaluasi akhir peserta diklat prajabatan CPNS K1/K2. Penelitian ini menggunakan aplikasi Anates versi 4.00 dengan pendekatan kuantitatif dan metode deskriptif. Subyek penelitian adalah peserta diklat prajabatan CPNS K1/K2 Golongan III Provinsi DKI Angkatan 131-132 tahun 2018 sebanyak 60 orang dengan butir soal evaluasi berbentuk pilihan ganda sebanyak 50 butir. Hasil analisis menunjukkan bahwa dari 50 butir soal terdapat 26 butir soal valid berdasarkan koefisien korelasi 0,354 pada taraf signifikansi 1%, dan koefisien korelasi 0,273 pada taraf signifikansi 5 %. Berdasarkan hasil rekap secara keseluruhan, maka diperoleh soal yang dapat digunakan sebanyak 26 butir soal (52%) terdiri dari soal yang bisa langsung dipakai 18 butir soal (36%) dan soal yang dapat dipakai tetapi harus diperbaiki sebanyak 8 butir soal (16%), sedangkan yang tidak dapat dipakai atau didrop sebanyak 24 butir soal (48%).

Kata kunci: Analisis, evaluasi, Anates

ABSTRACT

The analysis of evaluation items is performed to determine the quality of the final evaluation instrument for pre-service training participants. This research is conducted using the Anates version 4.00 application with quantitative approaches and descriptive methods. The subjects of the study are CPNS K1/K2 Group III pre-service training participants of DKI Province, class 131-132 in 2018 as many as 60 people with 50 items of multiple choice evaluation items. The results of the analysis show that out of 50 items, there are 26 valid questions based on the correlation coefficient of 0.354 at the 1% significance level, and the correlation coefficient of 0.273 at the 5% significance level. Based on the results of the overall recap, 26 questions (52%) can be used. Those consist of questions 18 items (36%) that can be directly used and 8 items (16%) questions that can be used but must be corrected, while those that cannot be used or dropped are 24 items (48%).

Keywords: Analysis, Evaluation, Anates

PENDAHULUAN

Pendidikan dan Pelatihan Prajabatan bagi Calon Pegawai Negeri Sipil yang diangkat dari Tenaga Honorer Kategori 1 (K1)/Kategori 2 (K2) diselenggarakan untuk membentuk PNS yang memiliki pengetahuan dan wawasan sebagai pelayan masyarakat yang baik (Deguci, 2013).

Menurut Raharjo (2016), kompetensi yang dibangun dalam Diklat Prajabatan bagi

CPNS yang diangkat dari Tenaga Honorer K1/K2 adalah kompetensi sebagai pelayan masyarakat yang baik, yang diindikasikan dengan kemampuan:

1. memahami wawasan kebangsaan sebagai dasar mengutamakan kepentingan nasional dalam pelaksanaan tugas jabatannya;

2. memahami sikap untuk tidak korupsi dan mendorong percepatan pemberantasan korupsi di lingkungan instansinya;
3. memahami ketentuan kepegawaian berkaitan dengan peran dan fungsi ASN, dan kedudukan, kewajiban dan hak PNS;
4. memahami pola pikir ASN sebagai Pelayan Masyarakat.

Untuk mengukur keberhasilan peserta Diklat Prajabatan CPNS Pengangkatan dari Tenaga Honorar K1/K2 dilakukan evaluasi pemahaman peserta diklat melalui ujian tulis atau tes yang merupakan salah satu teknik yang digunakan untuk melakukan evaluasi terhadap hasil belajar (Rahayu & Djazari, 2016). Hasil evaluasi pemahaman tersebut digunakan sebagai salah satu penentu kelulusan peserta diklat.

Evaluasi menurut Kumano (2001), merupakan penilaian terhadap data yang dikumpulkan melalui kegiatan asesmen. Evaluasi juga merupakan suatu proses yang sistematis untuk menentukan atau membuat keputusan sampai sejauhmana tujuan-tujuan pengajaran telah dicapai (Purwanto, 2011). Sedangkan Arikunto (2009) mengungkapkan bahwa evaluasi adalah serangkaian kegiatan yang ditujukan untuk mengukur keberhasilan program pendidikan.

Evaluasi pemahaman peserta diklat prajabatan dilaksanakan secara tertulis dengan menggunakan instrumen berupa soal-soal baik pilihan ganda maupun uraian.

Soal bentuk pilihan ganda menurut Surapranata (2009) adalah soal yang jawabannya harus dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Pilihan ganda adalah salah satu bentuk dari jenis tes obyektif yang pada waktu ini mendapat perhatian dan sering digunakan dalam evaluasi pendidikan. Dari sejumlah pilihan jawaban yang disediakan, hanya ada satu jawaban yang benar, yang disebut kunci jawaban, sedangkan kemungkinan jawaban yang lain disebut pengecoh (*distractor*).

Kelebihan bentuk tes pilihan ganda menurut Slameto (2001) adalah:

- a. Mengukur berbagai jenjang kognitif (dari ingatan sampai dengan evaluasi)
- b. Penskorannya mudah, cepat, obyektif, dan dapat mencakup ruang lingkup bahan/materi yang luas dalam suatu tes untuk suatu kelas atau jenjang pendidikan.

- c. Bentuk ini sangat tepat untuk ujian yang pesertanya sangat banyak atau yang sifatnya masal, sedangkan hasilnya harus segera diumumkan, seperti ujian semester, ujian sekolah dan ujian akhir semester.

Sedangkan kelemahan bentuk tes pilihan ganda antara lain:

- a. Memerlukan waktu yang relatif lama untuk menulis soalnya;
- b. Sulit membuat pengecoh yang homogen dan berfungsi;
- c. Terdapat peluang untuk menebak kunci jawaban.

Seringkali hasil evaluasi pemahaman dengan menggunakan soal-soal bentuk pilihan ganda tidak sesuai dengan harapan, padahal soal-soal sudah disusun sedemikian rupa. Jika demikian halnya, maka perlu dilakukan telaah atau analisis terhadap instrumen yang digunakan atau analisis butir soal (*item*) untuk mengevaluasi kualitas perangkat tes tersebut (Nuswowati, Binadja, Soeprodjo, & Ifada, 2010).

Analisis butir soal adalah suatu kegiatan analisis untuk menentukan tingkat kebaikan butir-butir soal yang terdapat dalam suatu tes sehingga informasi yang dihasilkan dapat kita pergunakan untuk memperbaiki butir soal dan tes tersebut.

Analisis tes adalah salah satu kegiatan dalam rangka mengkonstruksi tes untuk mendapatkan gambaran tentang mutu tes, baik mutu keseluruhan tes maupun mutu tiap butir soal. Analisis dilakukan setelah tes disusun dan diujikan kepada seluruh subyek, dan hasilnya menjadi umpan balik untuk perbaikan mutu tes bersangkutan. Oleh karena itu, kegiatan analisis tes merupakan keharusan dalam keseluruhan proses mengkontruksi tes (Karnoto, 2003).

Menurut Purwanto (2011), dengan membuat analisis soal sedikitnya tenaga edukatif dapat mengetahui empat hal penting yang dapat diperoleh dari tiap soal. Pertama, sampai dimana pemahaman tingkat kesukaran soal; kedua, apakah soal tersebut mempunyai daya pembeda sehingga dapat membedakan peserta diklat yang kompeten dan kurang kompeten; ketiga, apakah alternatif jawaban menarik; dan keempat, apakah soal tersebut memiliki hasil korelasinya tinggi atau rendah.

Berikut ini akan dijelaskan mengenai analisis butir soal:

1. Validitas Butir Soal

Validitas butir soal digunakan untuk mengetahui dukungan suatu butir soal terhadap skor total. Untuk menguji validitas setiap butir soal, skor-skor yang ada pada butir soal yang dimaksud dikorelasikan dengan skor total. Sebuah soal akan memiliki validitas yang tinggi jika skor soal tersebut memiliki dukungan yang besar terhadap skor total. Dukungan setiap butir soal dinyatakan dalam bentuk korelasi sehingga untuk mendapatkan validitas suatu butir soal digunakan rumus korelasi. Untuk soal-soal bentuk obyektif, biasanya diberikan skor 1 (untuk jawaban benar), atau skor 0 (untuk jawaban salah), sedangkan skor totalnya diperoleh dari menjumlahkan skor tiap butir yang membangun perangkat tes tersebut sehingga diperoleh data interval (Sugiyono, 2018).

Perhitungan dilakukan dengan menggunakan rumus korelasi *product moment Pearson* (Arikunto, 2009) sebagai berikut:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

Keterangan :

r_{xy} = koefisien korelasi antara dua variabel yaitu X dan Y

X = skor butir soal

Y = skor total

N = jumlah peserta

Interpretasi besarnya koefisien korelasi dapat dilihat pada Tabel 1 (Arikunto, 2009) berikut.

Tabel 1. Kategori Validitas Butir Soal (r_{xy})

Koefisien	Kategori
$0.80 < r_{xy} \leq 1.00$	Sangat tinggi
$0.60 < r_{xy} \leq 0.80$	Tinggi
$0.40 < r_{xy} \leq 0.60$	Cukup
$0.20 < r_{xy} \leq 0.40$	Rendah
$0.00 \leq r_{xy} \leq 0.20$	Sangat rendah

2. Tingkat kesukaran Butir Soal

Tingkat kesukaran menurut Masriyah (1999) dinyatakan dalam Indeks Kesukaran (*difficulty index, P*), yaitu bilangan yang menunjukkan proporsi peserta yang menjawab benar soal

tersebut. Semakin besar indeks kesukaran, semakin mudah butir tersebut. Sebaliknya, semakin kecil indeks kesukaran, semakin sukar butir tersebut. Indeks kesukaran suatu butir tes dapat dihitung dengan rumus berikut:

$$P = \frac{B_a + B_b}{J_a + J_b}$$

Keterangan:

P = Indeks kesukaran butir tes

B_a = Jumlah peserta kelompok atas yang menjawab benar

B_b = Jumlah peserta kelompok bawah yang menjawab benar

J_a = Jumlah seluruh peserta kelompok atas

J_b = Jumlah seluruh peserta kelompok bawah

Tabel 2. Kriteria Indeks Kesukaran (P)

Indeks Kesukaran	Penafsiran butir soal
$P < 0,30$	Sukar
$0,30 \leq P \leq 0,70$	Sedang
$P > 0,70$	Mudah

3. Indeks daya pembeda

Daya pembeda butir tes menurut Ratumanan (2003) menyatakan seberapa jauh kemampuan butir tersebut mampu membedakan antara kelompok peserta pandai dengan kelompok lemah.

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara peserta yang berkemampuan tinggi dengan peserta yang berkemampuan rendah (Zein, Fadillah, & Novianti, 2013). Angka yang menunjukkan besarnya daya pembeda disebut Indeks Diskriminasi Tes atau Daya Pembeda (D). Rumus untuk menentukan Diskriminasi atau Daya Pembeda adalah sebagai berikut:

$$D = \frac{B_a - B_b}{N_a} \times 100\%$$

Keterangan:

D = Daya pembeda

B_a = Jumlah peserta tes kelompok atas yang menjawab benar

B_b = Jumlah peserta tes kelompok bawah yang menjawab benar

Na = Jumlah peserta tes pada salah satu kelompok A atau B
(Arikunto, 2009)
Kategori daya pembeda dapat dilihat pada tabel 3

Tabel 3. Kategori Daya Pembeda (D)

Batasan	Kategori
0% < D ≤ 20%	Buruk
20% < D ≤ 40%	Cukup
40% < D ≤ 70%	Baik
D > 70%	Baik Sekali

4. Efektivitas pengecoh (distractor)

Untuk mengetahui efektivitas tiap *option* jawaban dapat dilakukan dengan menghitung berapa banyak peserta yang memilih *option* tersebut. Selain itu dapat dilihat pengecoh mana yang berfungsi efektif, pengecoh mana yang tidak efektif, dan pengecoh mana yang menyesatkan. Jika ternyata lebih banyak peserta yang memilih suatu pengecoh tertentu dan hanya sedikit yang memilih kunci, maka ada kemungkinan kesalahan kunci jawaban, dan mungkin pengecoh tersebut sebenarnya adalah kunci jawaban. Namun mungkin pula kuncinya sudah benar, tetapi pengecoh terlalu menarik untuk dipilih. Menurut Masriyah (1999), jika ditinjau dari pemilih kelompok atas dan kelompok bawah, maka:

- 1) Pengecoh efektif jika Na < Nb
- 2) Pengecoh menyesatkan jika Na > Nb
- 3) Pengecoh tidak efektif jika Na = Nb

Keterangan :

Na = banyaknya peserta kelompok atas yang memilih pengecoh
Nb= banyaknya peserta kelompok bawah yang memilih pengecoh

5. Reliabilitas

Reliabilitas adalah tingkat keajegan atau kestabilan dari hasil pengukuran suatu tes atau dengan kata lain konsisten dan dapat diandalkan (Nuriyah, 2014). Alat ukur yang reliabel adalah alat ukur yang apabila digunakan untuk mengukur hal yang sama berulang-ulang, hasilnya relatif sama. Reliabilitas suatu tes pada umumnya dinyatakan dengan koefisien reliabilitas atau kesalahan pengukuran

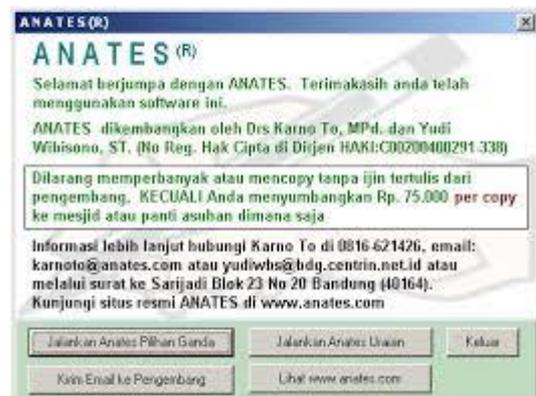
standar yang dihitung berdasarkan koefisien reliabilitas. Besaran koefisien reliabilitas dapat diketahui dengan menggunakan metode belah dua untuk mengetahui reliabilitas seluruh tes dapat digunakan rumus Spearman-Brown sebagai berikut:

$$r_{tt} = \frac{2r_{1/2/2}}{1 + |r_{1/2/2}|}$$

Keterangan:

- r_{tt} = koefisien reliabilitas tes
- r_{1/2/2}=koefisien reliabilitas separuh tes
- X = skor ganjil
- Y = skor genap
- N = jumlah peserta

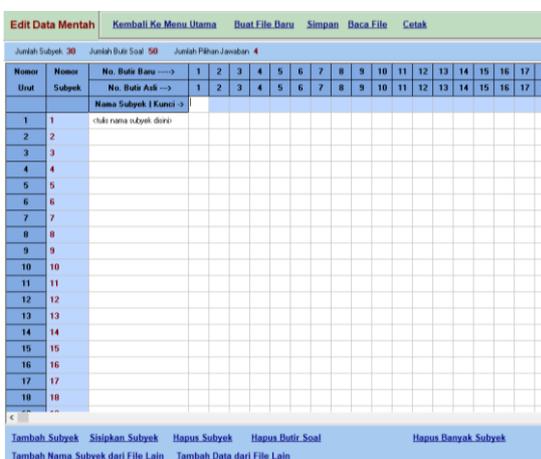
Pada penelitian ini hanya akan menganalisis butir soal dengan empat aspek, yaitu validitas, indeks daya pembeda, tingkat kesukaran, dan reliabilitas, sehingga diperoleh data kuantitatif yang dapat digunakan sebagai indeks untuk menentukan kategori butir soal. Untuk menganalisis keempat aspek tersebut digunakan aplikasi Anates, yaitu sebuah aplikasi yang dikembangkan oleh Drs. Karno, M.Pd dan Yusuf Wibisono, ST. Anates mampu menganalisis butir soal secara cepat, mudah dan akurat, Anates juga mampu menampilkan beberapa fitur dan perhitungan di antaranya: skor dengan bobot, reliabilitas Tes, kelompok atas dan kelompok bawah, daya pembeda, tingkat kesukaran, dan korelasi skor butir dengan skor total (validitas) (Sari, Candra, & Herawati, 2014).



Gambar 1. Halaman Muka



Gambar 2. Fitur Anates



Gambar 3. Format input data

METODOLOGI

Penelitian ini dilaksanakan pada Pendidikan dan Pelatihan Prajabatan bagi CPNS K1/K2 Golongan III Pemerintah Provinsi DKI Jakarta tahun 2018. Subjek penelitian adalah peserta Diklat Prajabatan CPNS K1/K2 Golongan III Angkatan 131 dan 132.

Analisis butir soal ini menggunakan parameter yang terdiri dari validitas butir soal, tingkat kesukaran butir soal, daya pembeda butir soal, dan reliabilitas butir soal.

Penelitian ini adalah penelitian kuantitatif dengan metode deskriptif. Pelaksanaan penelitian ini menganalisis butir soal yang disusun oleh tim widyaiswara BPSDM Provinsi DKI Jakarta dan digunakan sebagai instrumen evaluasi pemahaman peserta diklat prajabatan bagi Calon Pegawai Negeri Sipil (CPNS) di lingkungan

Pemerintah Provinsi DKI Jakarta. Penelitian ini menggunakan aplikasi Anates versi 4.0 for Windows (Alpusari, 2014).

HASIL DAN PEMBAHASAN

Analisis butir soal dilakukan pada soal ujian (evaluasi) pemahaman yang terdiri dari 50 butir soal dengan empat pilihan jawaban (*option*) yang dimaksudkan untuk melakukan standarisasi dan meningkatkan kualitas soal sehingga dapat digunakan sebagai alat evaluasi yang akurat.

Aplikasi Anates dapat menganalisis butir soal secara kuantitatif dan kualitatif. Hasil penelitian ini mengungkapkan butir soal yang memenuhi standar dan dapat digambarkan dari validitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan reliabilitas (Sumaningsih, 2015).

1. Validitas Soal

Validitas butir soal yang dianalisis menggunakan aplikasi Anates versi 4.0 for Windows dapat digambarkan pada tabel berikut:

Tabel 4. Hasil Analisis Validitas

No. Soal	Korelasi	Signifikansi	Ket.
1	0.047	-	TV
2	NAN	NAN	TV
3	0.018	-	TV
4	0.395	Sangat Signifikan	V
5	0.111	-	TV
6	0.369	Sangat Signifikan	V
7	0.305	-	TV
8	0.380	Sangat Signifikan	V
9	0.354	Sangat Signifikan	V
10	0.483	Sangat Signifikan	V
11	0.337	Sangat Signifikan	V
12	-0.095	-	TV
13	0.487	Sangat Signifikan	V
14	0.516	Sangat Signifikan	V
15	0.456	Sangat Signifikan	V
16	0.446	Sangat Signifikan	V
17	-0.220	-	TV
18	0.347	Signifikan	V
19	0.535	Sangat Signifikan	V
20	0.254	-	TV
21	0.045	-	TV
22	0.626	Sangat Signifikan	V
23	0.285	Signifikan	V
24	0.506	Sangat Signifikan	V
25	0.421	Sangat Signifikan	V

No. Soal	Korelasi	Signifikansi	Ket.
26	0.264	-	TV
27	0,380	Sangat Signifikan	V
28	0.146	-	TV
29	0.033	-	TV
30	0.328	Signifikan	V
31	0.475	Sangat Signifikan	V
32	-0.007	-	TV
33	0.096	-	TV
34	0.253	-	TV
35	0.412	Sangat Signifikan	V
36	0.326	Signifikan	V
37	0.568	Sangat signifikan	V
38	0.105	-	TV
39	0.288	Signifikan	V
40	0.117	-	TV
41	0.084	-	TV
42	0.110	-	TV
43	-0.110	-	TV
44	0.186	-	TV
45	0.380	Sangat Signifikan	V
46	-0.076	-	TV
47	0.363	Sangat Signifikan	V
48	0.145	-	TV
49	0.478	Sangat Signifikan	V
50	0.206	-	TV

Hasil olahan Anates versi 4.0 for Windows Ket. V= Valid, TV = Tidak Valid, Jumlah subyek 60, Butir Soal 50. Valid atau tidak validnya suatu item soal, dapat digunakan dengan teknik korelasi sebagai teknik analisisnya. Butir soal yang valid, maka butir soal dapat digunakan untuk ujian evaluasi pemahaman, sedangkan butir soal yang tidak valid maka butir soal tidak dapat digunakan atau dibuang (*drop*).

Tabel 4 memperlihatkan hasil yang diperoleh dari uji validitas butir soal yang telah diujikan pada peserta Diklat Prajabatan CPNS K1/K2 Golongan III dengan melihat hasil uji korelasinya, maka terdapat butir soal yang valid berjumlah 26 butir sedangkan butir soal yang tidak valid berjumlah 24 butir berdasarkan hasil olahan Anates versi 4.0 for Windows. Butir soal yang valid merupakan soal yang memiliki kualitas karena menunjukkan hubungan antara pengukuran sesuai dengan keefektivan dari suatu tes.

Hasil dari analisis validitas tersebut diperoleh koefisien korelasi 0,354 pada taraf signifikansi 1%, sedangkan koefisien korelasi 0,273 pada taraf signifikansi 5 % = 0,273.

Korelasi pada taraf signifikansi 5% didapat butir soal yang valid berjumlah 26 soal, sedangkan soal yang tidak valid berjumlah 24 soal. Butir soal dengan validitas tertinggi terdapat 17 butir soal, sedangkan soal valid yang sedang terdapat 9 butir soal.

Butir-butir soal yang memiliki tingkat validitas tinggi, jika skor yang diperoleh pada tiap soal memiliki kesesuaian arah dengan skor total dapat dikatakan bahwa butir soal tersebut memiliki korelasi positif yang signifikan antar skor butir soal dengan skor keseluruhannya. Butir-butir soal yang tidak valid dan memiliki korelasi negatif yang signifikan maka butir soal tersebut tidak memiliki korelasi antara skor butir soal dengan skor keseluruhannya.

Untuk butir soal nomor 2, korelasi hasil analisis tidak bisa dihitung karena korelasinya 0,000 sehingga menampilkan kata NAN, sedangkan butir soal nomor 12, 17, 32, 43, dan 46 menunjukkan tidak berkorelasi signifikan dengan nilai koefisien $< r_{tabel}$, dengan demikian secara otomatis butir soal tersebut dinyatakan tidak valid dan tidak dapat digunakan untuk evaluasi pemahaman. Butir soal yang memiliki validitas tertinggi adalah nomor 22.

Butir tersebut merupakan salah satu soal yang valid, hal ini dapat dilihat dari korelasi yang diperoleh yaitu 0,626. Korelasi ini merupakan koefisien yang sangat tinggi. Dari derajat kebebasan sebesar 90 lalu di pada tabel nilai "r" product moment pada taraf signifikansi 5% dan taraf signifikansi 1% hasilnya sebagai berikut :

r_{tabel} atau r_1 pada taraf signifikansi 5 % = 0,273

r_{tabel} atau r_1 pada taraf signifikansi 1 % 0,354

Butir soal yang tidak valid dengan nilai koefisien korelasi terendah adalah nomor 32. Butir soal tersebut merupakan salah satu soal yang tidak valid dan memiliki korelasi negatif karena t_{hitung} lebih kecil dari r_{tabel} , dengan dengan

demikian maka soal tersebut tidak valid dan tidak bisa digunakan.

Soal evaluasi pemahaman yang disusun oleh tim widyaiswara ini merupakan gabungan dari sekian banyak butir soal, dengan harapan penyusun dapat mengetahui sejauh mana peserta diklat memahami apa yang telah dipelajari. Hal ini sesuai dengan pendapat Sudijono (2006) analisis butir soal yang dilakukan adalah untuk mengungkap hasil belajar peserta didik setelah mengikuti proses pembelajaran dalam jangka waktu tertentu. Setiap butir soal yang terdapat pada tes hasil belajar tersebut merupakan bagian yang tidak bisa terpisahkan dari tes hasil belajar sebagai suatu totalitas.

2. Daya Pembeda Soal

Daya Pembeda Soal adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta diklat yang sudah mampu menguasai materi diklat (kompeten) dengan peserta diklat yang belum menguasai materi diklat (belum kompeten) berdasarkan kriteria tertentu (Masriyah, 1999). Daya pembeda yang diperoleh dari hasil pengujian soal tersebut dapat dilihat pada tabel berikut.

Tabel 5. Hasil Analisis Daya Pembeda

Soal No	Kelompok Atas	Kelompok bawah	Daya Pembeda (%)
1	10	8	12.50
2	16	16	0.00
3	0	0	0.00
4	16	9	43,75
5	14	14	0.00
6	16	14	12.50
7	16	12	25.00
8	16	13	18.75
9	16	12	25.00
10	16	8	50.00
11	10	3	43.75
12	8	11	-18.75
13	16	9	43.75
14	16	5	68.75
15	16	5	68.75
16	12	3	56.25
17	8	13	-31.25
18	14	7	43.75
19	16	6	62.50

Soal No	Kelompok Atas	Kelompok bawah	Daya Pembeda (%)
20	16	10	37.50
21	8	7	6.25
22	14	2	75.00
23	8	3	31.25
24	10	2	50.00
25	14	7	43,75
26	16	14	12.50
27	16	13	18.75
28	8	6	12.50
29	10	8	12.50
30	12	9	18.75
31	10	2	50.00
32	10	12	-12.50
33	12	10	12.50
34	12	6	37.50
35	12	4	50.00
36	16	12	25.00
37	16	8	50.00
38	8	5	18.75
39	16	11	31.25
40	6	3	18.75
41	8	6	12.50
42	12	11	6.25
43	4	6	-12.50
45	6	2	25.00
46	16	12	25.00
47	2	2	0.00
48	12	5	43.75
49	10	6	25.00
50	14	4	62.50

Berdasarkan Tabel 5, daya pembeda yang dianalisis dengan menggunakan program komputer Anates versi 4.0 for Windows didapatkan bahwa data yang dianalisis dibelah menjadi dua kelompok yaitu kelompok unggul (atas) dan kelompok asor (bawah) berdasarkan perolehan skor. Kelompok unggul adalah peserta dengan skor ≥ 35 , sedangkan kelompok asor adalah peserta dengan skor < 35 . Dengan adanya kelompok atas dan kelompok bawah, maka dapat diketahui daya pembeda peserta yang kompeten dan peserta yang kurang/belum kompeten.

Daya pembeda tersebut diperoleh dari hasil pengujian pada peserta diklat prajabatan dari instrumen evaluasi sebanyak 50 butir soal. Butir soal dengan daya pembeda nilai negatif pada butir soal nomor 12, 17, 32, dan 43 dianggap

sebagai butir soal dengan kategori sangat buruk, kemudian 15 butir soal-soal yang memperoleh kategori jelek sehingga soal tidak dipakai dan dianggap tidak memiliki daya pembeda yang baik, daya pembeda kategori cukup dengan jumlah 15 soal dan kategori baik 9 soal. Soal yang mendapatkan daya pembeda 0,000 pada analisis ini terdapat 4 butir soal, yaitu 2, 3, 5, dan 46 hal ini menunjukkan bahwa butir soal yang bersangkutan tidak memiliki daya pembeda sama sekali. Kelompok atas yang menjawab betul dan salah sama dengan jumlah kelompok bawah yang jawabannya betul. Dengan demikian butir-butir soal tersebut tidak memiliki daya pembeda ($DP = 0,00$). Oleh karena itu butir soal yang dapat digunakan untuk selanjutnya berjumlah 26 butir.

Peserta yang memperoleh kelompok unggul (atas) dalam daya pembeda berjumlah 16 orang dengan skor yang diperoleh dimulai dari 35 sampai skor 40, sedangkan peserta yang memperoleh kelompok asor (bawah) berjumlah 16 mahapeserta dengan skor yang diperoleh dimulai dengan skor 20 sampai 26. Indeks daya pembeda butir soal dengan kelompok unggul dan kelompok asor berjumlah 32 orang peserta, terdapat pembagian kelompok pada peserta yang dikelompokkan menjadi kelompok atas, kelompok tengah dan kelompok bawah.

3. Tingkat kesukaran soal

Perhitungan tingkat kesukaran soal dihitung dengan seberapa besar derajat kesukaran suatu soal ketika dikerjakan oleh peserta didik, dan soal yang dikatakan sukar apabila hasil yang diperoleh oleh peserta didik hanya sedikit yang dapat menjawabnya (Yuslita, Zulfan, & Arifin, 2016). Tingkat kesukaran yang diperoleh dari hasil pengujian soal dapat dilihat pada tabel berikut

Tabel 6. Hasil Analisis Tingkat Kesukaran

No	Jumlah Betul	Tingkat Kesukaran %	Tafsiran
1	40	66.67	Sedang
2	60	100.00	Sangat Mudah
3	2	3.33	Sangat Sukar
4	48	80.00	Mudah
5	52	86.67	Sangat Mudah

No	Jumlah Betul	Tingkat Kesukaran %	Tafsiran
6	58	96.67	Sangat Mudah
7	54	90.00	Sangat Mudah
8	56	93.33	Sangat Mudah
9	56	93.33	Sangat Mudah
10	52	86.67	Sangat Mudah
11	18	30.00	Sukar
12	30	50.00	Sedang
13	53	88.33	Sangat Mudah
14	40	66.67	Sedang
15	32	53.33	Sedang
16	34	56.67	Sedang
17	42	70.00	Sedang
18	38	63.33	Sedang
19	44	73.33	Mudah
20	40	66.67	Sedang
21	24	40.00	Sedang
22	32	53.33	Sedang
23	28	46.67	Sedang
24	28	46.67	Sedang
25	44	73.33	Mudah
26	58	96.67	Sangat Mudah
27	56	93.33	Sangat Mudah
28	26	43.33	Sedang
29	40	66.67	Sedang
30	40	66.67	Sedang
31	18	30.00	Sukar
32	38	63.33	Sedang
33	34	56.67	Sedang
34	28	46.67	Sedang
35	23	38.33	Sedang
36	54	90.00	Sangat Mudah
37	48	80.00	Mudah
38	22	36.67	Sedang
39	52	86.67	Sangat Mudah
40	18	30.00	Sukar
41	24	40.00	Sedang
42	42	70.00	Sedang
43	12	20.00	Sukar
44	8	13.33	Sangat Sukar
45	56	93.33	Sangat Mudah
46	10	16.67	Sukar
47	26	43.33	Sedang
48	30	50.00	Sedang
49	38	63.33	Sedang
50	24	40.00	Sedang

Berdasarkan data hasil analisis di atas bahwa setiap butir soal dikatakan sudah atau belum memadai tingkat kesukarannya dapat diketahui dari besar

kecilnya angka yang melambangkan tingkat kesukaran atau disebut indeks kesukaran dari butir soal tersebut. Angka indeks kesukaran butir soal berkisar antara 0.00% sampai dengan 100% atau 0 – 1, makin mendekati angka nol makin tinggi tingkat kesukarannya (Sangat Sukar) atau sebaliknya makin mendekati angka 1 maka makin tinggi tingkat kemudahannya (Sangat Mudah).

Angka indeks kesukaran mendekati atau sama dengan 0.00% merupakan petunjuk bahwa butir soal tersebut termasuk dalam kategori terlalu sukar, sebab sebagian besar atau seluruh peserta diklat tidak dapat menjawab dengan benar butir soal tersebut. Jika angka indeks kesukaran mendekati atau sama dengan 100%, menunjukkan bahwa butir soal tersebut adalah butir soal dengan kategori terlalu mudah, sebab peserta diklat sebagian besar atau seluruhnya dapat menjawab dengan benar. Butir soal yang termasuk kategori sangat sukar terdapat pada nomor 3 dan 44, sedangkan kategori sangat mudah terdapat pada nomor 2, 5, 6, 7, 8, 9, 10, 13, 26, 27, 36, 39, dan 45.

Butir soal nomor 44 tersebut merupakan soal yang sangat sukar, hal ini diketahui dari hasil analisis dimana hampir seluruh peserta diklat tidak dapat menjawab dengan benar, hanya 2 orang dari 60 peserta diklat yang menjawab dengan benar, sehingga total skor yang didapat soal ini adalah 3.33 dan 13.33. Kesulitan yang terjadi pada soal ini dapat diakibatkan dari tata bahasa yang digunakan, istilah yang digunakan jarang didengar oleh peserta, pernyataan soal maupun pilihan jawaban multitafsir, dan kunci jawaban yang membingungkan peserta.

Tabel 6 memperlihatkan bahwa tingkat kesukaran soal yang didapat dari hasil pengujian dan dianalisis dengan kategori sangat mudah berjumlah 13 butir, kategori mudah 4 butir, kategori sedang sebanyak 26 butir, kategori sukar sebanyak 5 butir, dan kategori sangat sukar berjumlah 2 butir soal, dengan demikian persentase masing-masing kategori soal adalah 26% butir soal sangat mudah, butir soal dengan kategori mudah 8%, butir soal dengan kategori sedang sebanyak 52%, dan butir soal dengan kategori sukar sebanyak

10%, dan butir soal dengan kategori sangat sukar sebanyak 4%.

Dari keseluruhan butir soal yang diajukan dalam evaluasi pemahaman tersebut berdasarkan tingkat kesukaran, maka butir soal yang dapat dikategorikan baik dan dapat digunakan sebanyak 35 butir atau 70 % dan butir soal yang dibuang atau harus diperbaiki sebanyak 15 butir atau 30%.

Dengan demikian soal dengan kategori sangat mudah dan sangat sukar harus diperbaiki jika butir soal tersebut akan digunakan kembali, perbaikan soal secara total baik pernyataan soal, pernyataan pilihan jawaban, kunci jawaban, maupun tata bahasanya. Butir soal yang digunakan adalah butir soal dengan kategori mudah, sedang, dan sukar, hal ini sesuai dengan pendapat Suharsimi Arikunto (2005) yang menyatakan bahwa butir soal yang baik adalah butir soal yang tidak terlalu mudah dan terlalu sukar. Soal yang sangat mudah tidak merangsang peserta diklat untuk mempertinggi usaha memecahkan butir soal. Soal yang sangat sukar akan menyebabkan peserta diklat menjadi putus asa dan tidak mempunyai semangat untuk mencoba butir soal karena di luar jangkauannya.

Menurut Erman Suherman dan Yaya Sukjaya (1990), butir soal yang mendapatkan nilai indeks tingkat kesukaran 1–10%, maka soal harus dibuang, butir soal yang mendapatkan nilai indeks tingkat kesukaran 11%–20%, soal bisa dipakai dan diperbaiki. Butir soal yang mendapatkan indeks 21%–80% soal bisa dipakai langsung dan butir soal yang mendapatkan indeks 81%–90% bisa dipakai tetapi harus diperbaiki dan butir soal yang mendapatkan indeks kesukaran 91%–100% dapat langsung dibuang atau tidak dipakai. Tingkat kesukaran butir soal juga dapat digunakan untuk memprediksi alat ukur itu sendiri (soal) dan kemampuan peserta diklat dalam memahami materi yang dibelajarkan.

Misalnya satu butir soal termasuk kategori mudah, maka prediksi terhadap informasi ini adalah seperti berikut. 1) Pengecoh butir soal itu tidak berfungsi. 2) Sebagian besar peserta menjawab benar butir soal itu; artinya bahwa sebagian besar

peserta telah memahami materi yang ditanyakan.

Bila suatu butir soal termasuk kategori sukar, maka prediksi terhadap informasi ini adalah seperti berikut. 1) Butir soal itu "mungkin" salah kunci jawaban. 2) Butir soal itu mempunyai 2 atau lebih jawaban yang benar. 3) Materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, sehingga kompetensi minimum yang harus dikuasai peserta belum tercapai. 4) Materi yang diukur tidak cocok ditanyakan dengan menggunakan bentuk soal yang diberikan (misalnya menjelaskan kronologi atau menuliskan pendapat ditanyakan dalam bentuk pilihan ganda). 5) Pernyataan atau kalimat soal terlalu kompleks dan panjang.

4. Reliabilitas

Reliabilitas butir soal yang dianalisis menggunakan aplikasi Anates versi 4.0 for Windows dapat digambarkan pada tabel berikut:

Tabel 7. Hasil Analisis Reliabilitas

No Subyek	Skor Ganjil	Skor Genap	Skor Total
1	19	21	40
2	19	21	40
3	19	19	38
4	18	20	38
5	19	19	38
6	18	20	38
7	19	18	37
8	20	17	37
9	19	18	37
10	20	17	37
11	17	19	36
12	17	19	36
13	19	16	35
14	18	17	35
15	19	16	35
16	18	17	35
17	18	16	34
18	17	17	34
19	18	16	34
20	17	17	34
21	17	16	33
22	15	18	33
23	17	16	33
24	15	18	33
25	14	18	32
26	14	18	32
27	17	14	31
28	15	16	31

No Subyek	Skor Ganjil	Skor Genap	Skor Total
29	16	15	31
30	17	14	31
31	15	16	31
32	15	16	31
33	16	15	31
34	14	16	30
35	17	13	30
36	17	13	30
37	14	16	30
38	13	16	29
39	16	12	28
40	16	12	28
41	12	15	27
42	11	15	27
43	16	10	26
44	11	15	26
45	11	15	26
46	16	10	26
47	15	10	25
48	12	13	25
49	15	10	25
50	11	13	24
51	12	12	24
52	10	14	24
53	12	12	24
54	10	14	24
55	11	12	23
56	10	13	23
57	11	12	23
58	10	13	23
59	11	9	20
60	11	9	20

Dari hasil analisis reliabilitas butir soal yang dinyatakan valid sebanyak 26 butir diperoleh data rata-rata (mean) sebesar 17,90, simpang baku sebesar 4,63, dan koefisien reliabilitas sebesar 0,86. Untuk menentukan kategori dapat menggunakan kriteria Guilford (1956) dalam (Martadiputra, 2015) yaitu:

- 1) $0,80 < r_{tt} \leq 1,00$ reliabilitas sangat tinggi
- 2) $0,60 < r_{tt} \leq 0,80$ reliabilitas tinggi
- 3) $0,40 < r_{tt} \leq 0,60$ reliabilitas sedang
- 4) $0,20 < r_{tt} \leq 0,40$ reliabilitas rendah
- 5) $-1,00 < r_{tt} \leq 0,20$ reliabilitas sangat rendah (tidak reliabel)

Berdasarkan hasil analisis dan kriteria tersebut maka instrumen tes evaluasi pemahaman peserta diklat prajabatan CPNS K1/K2 Golongan III Provinsi DKI Jakarta termasuk kategori reliabilitas sangat tinggi, artinya butir-butir soal

evaluasi yang dinyatakan valid sebanyak 26 butir reliabel.

5. Rekap analisis Butir Soal

Kegiatan analisis yang telah dilaksanakan di atas merupakan analisis secara parsial yaitu analisis validitas, daya pembeda, tingkat kesukaran, dan reliabilitas. Analisis yang terakhir adalah analisis rekap secara keseluruhan, sehingga butir soal yang akan digunakan untuk selanjutnya dapat terlihat. Rekapitulasi hasil analisis butir soal sebagai berikut:

Tabel 8. Rekapitulasi Hasil Analisis

No	DP	TK	Kor.	Sign.	V / TV	Ket
1	12.50	Sd	0.047	-	TV	D
2	0.00	SM	NAN	NAN	TV	D
3	0.00	SSk	0.018	-	TV	D
4	43,75	Md	0.395	SS	V	P
5	0.00	SM	0.111	-	TV	D
6	12.50	SM	0.369	SS	V	PB
7	25.00	SM	0.305	SS	V	PB
8	18.75	SM	0.380	SS	V	PB
9	25.00	SM	0.354	SS	V	PB
10	50.00	SM	0.483	-	TV	D
11	43.75	Sk	0.337	SS	V	P
12	-18.75	Sd	-0.095	-	TV	D
13	43.75	SM	0.487	SS	V	PB
14	68.75	Sd	0.516	SS	V	P
15	68.75	Sd	0.456	SS	V	P
16	56.25	Sd	0.446	SS	V	P
17	-31.25	Sd	-0.220	-	TV	D
18	43.75	Sd	0.347	S	V	P
19	62.50	Md	0.535	SS	V	P
20	37.50	Sd	0.254	-	TV	D
21	6.25	Sd	0.045	-	TV	D
22	75.00	Sd	0.626	SS	V	P
23	31.25	Sd	0.285	S	V	P
24	50.00	Sd	0.506	SS	V	P
25	43.75	Md	0.421	SS	V	P
26	12.50	SM	0.264	-	TV	D
27	18.75	SM	0.380	SS	V	P
28	12.50	Sd	0.146	-	TV	D
29	12.50	Sd	0.033	-	TV	D
30	18.75	Sd	0.328	S	V	P
31	50.00	Sk	0.475	SS	V	P
32	-12.50	Sd	-0.007	-	TV	D
33	12.50	Sd	0.096	-	TV	D
34	37.50	Sd	0.253	-	TV	D
35	50.00	Sd	0.412	SS	V	P

No	DP	TK	Kor.	Sign.	V / TV	Ket
36	25.00	SM	0.326	S	V	PB
37	50.00	Md	0.568	SS	V	P
38	18.75	Sd	0.105	-	TV	D
39	31.25	SM	0.288	S	V	PB
40	18.75	Sk	0.117	-	TV	D
41	12.50	Sd	0.084	-	TV	D
42	6.25	Sd	0.110	-	TV	D
43	-12.50	Sk	-0.110	-	TV	D
44	25.00	SS	0.186	-	TV	D
45	25.00	SM	0.380	SS	V	PB
46	0.00	Sk	-0.076	-	TV	D
47	43.75	Sd	0.363	SS	V	P
48	25.00	Sd	0.145	-	TV	D
49	62.50	Sd	0.478	SS	V	P
50	18.75	Sd	0.206	-	TV	D

Keterangan:

- DP = Daya Pembeda
 TK = Tingkat Kesukaran
 V/TV = Valid / Tidak Valid
 SM = Sangat mudah
 M = Mudah
 Sd = Sedang
 Sk = Sukar
 SSk = Sangat Sukar
 S = Signifikan
 SS = Sangat Signifikan
 P = Dipakai
 PB = Dipakai dengan Perbaikan
 D = Drop (Tidak Dipakai)

Berdasarkan hasil rekap di atas maka secara keseluruhan soal yang bisa dipakai 26 soal (52%) terdiri dari soal yang bisa langsung dipakai sebanyak 18 butir soal (36%), dapat dipakai tetapi harus diperbaiki sebanyak 8 butir soal (16%), sedangkan yang tidak dapat dipakai atau di-drop sebanyak 24 butir soal (48%).

Dalam kaitannya dengan hasil analisis butir soal secara keseluruhan, maka tindak lanjut yang harus dilaksanakan sebagai berikut:

- 1) Setiap butir soal berdasarkan hasil analisis termasuk dalam kategori valid dan siap dipakai (P), maka butir soal tersebut bisa langsung dipakai dan dimasukkan dalam bank soal pada bidang terkait. Butir-butir soal tersebut dapat digunakan lagi pada evaluasi akhir

- pemahaman peserta diklat di masa yang akan datang.
- 2) Butir soal yang termasuk dalam kategori soal dipakai diperbaiki (PB) dapat ditindaklanjuti dengan diteliti ulang, diperbaiki tata bahasanya, pernyataan soal dan pernyataan pilihan jawabannya sehingga pernyataan soal menjadi lebih jelas, soal mudah dipahami, istilah yang digunakan jelas. Setelah diperbaiki butir soal tersebut bisa digunakan kembali pada tes dimasa yang akan datang, alternatif lain butir soal tersebut dapat langsung dibuang dan tidak dipakai lagi dalam tes selanjutnya dimasa yang akan datang dan diganti dengan soal yang baru.
 - 3) Butir soal yang tidak dipakai dapat langsung dibuang dan tidak digunakan dalam tes selanjutnya.

Tujuan utama analisis butir soal dalam sebuah tes yang dibuat tenaga edukatif adalah untuk mengidentifikasi kekurangan-kekurangan dalam tes atau dalam proses pembelajaran. Berdasarkan tujuan ini, maka kegiatan analisis butir soal memiliki banyak manfaat, diantaranya adalah: (1) mendukung penulisan butir soal yang efektif; (2) secara materi dapat memperbaiki tes yang digunakan; (3) meningkatkan validitas soal dan reliabilitas.

Di samping itu, manfaat lainnya adalah: (1) menentukan apakah suatu fungsi butir soal sesuai dengan yang diharapkan, sehingga dapat digunakan untuk menentukan tingkat keberhasilan peserta; (2) memberi masukan kepada pembuat soal tentang kesulitan belajar peserta; (3) memberi masukan pada aspek tertentu untuk pengembangan kurikulum; (4) merevisi materi yang dinilai atau diukur; (5) meningkatkan keterampilan penulisan soal.

Kegunaan analisis butir soal bukan hanya terbatas untuk meningkatkan butir soal, tetapi ada beberapa hal, yaitu bahwa data analisis butir soal bermanfaat sebagai dasar: (1) diskusi kelas efisien tentang hasil tes, (2) untuk kerja remedial, (3) untuk peningkatan secara umum pembelajaran di kelas, dan (4) untuk peningkatan keterampilan pada konstruksi tes.

SIMPULAN DAN SARAN

Analisis butir soal bertujuan untuk mengetahui kelayakan setiap butir soal yang akan diujikan kepada peserta ujian dengan mengukur validitas, indeks daya pembeda, indeks tingkat kesukaran, dan reliabilitas.

Berdasarkan hasil rekap di atas maka secara keseluruhan soal yang bisa dipakai 26 soal (52%) terdiri dari soal yang bisa langsung dipakai sebanyak 18 butir soal (36%), dapat dipakai tetapi harus diperbaiki sebanyak 8 butir soal (16%), sedangkan yang tidak dapat dipakai atau *didrop* sebanyak 24 butir soal (48%).

Saran bagi peneliti selanjutnya untuk melakukan analisis yang sama terhadap instrumen evaluasi pada berbagai kegiatan pendidikan dan pelatihan untuk mendapatkan kualitas instrumen yang lebih baik. Bagi penyusun butir soal agar memperhatikan konstruksi bahasa dan substansi materi dalam penyusunan instrumen agar butir soal yang dihasilkan memenuhi kaidah dan berkualitas.

UCAPAN TERIMA KASIH

Terima kasih saya ucapkan kepada Kepala BPSDM dan penyelenggara Diklat Prajabatan bagi CPNS K1/K2 Golongan III yang telah memberi kesempatan untuk melaksanakan kajian dan memfasilitasi publikasi karya tulis ilmiah ini.

DAFTAR PUSTAKA

- Alpusari, M. (2014). Analisis Butir Soal Konsep Dasar IPA 1 Melalui Penggunaan Program Komputer Anates Versi 4.0 For Windows. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Riau*, 106.
- Arif, M. (2014). Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda. *Jurnal Ilmiah Edutic Vol. 1 No. 1*.
- Arikunto, S. (2009). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Candra Sari, A. I., & Herawati, M. (2014). Aplikasi Anates Versi 4 Dalam Meanalisis Butir Soal. *Faktor Jurnal Ilmiah Kependidikan*, 203.
- Candra Sari, A. I., & Herawati, M. (2014). Aplikasi Anates Versi 4 Dalam Menganalisis Butir Soal. *Faktor*

- Jurnal Ilmiah Kependidikan*, 203-214.
- Deguci, E. E. (2013). Pengaruh Pendidikan dan Pelatihan Prajabatan Terhadap Motivasi Kerja Peserta Pendidikan dan Pelatihan Prajabatan. *Jurnal KBP*, 360-378.
- Karnoto. (2003). *Mengenal Analisis Tes*. Bandung: UPI.
- Martadiputra, B. A. (2015). *Analisis Data Sederhana dan Interpretasinya*. Jakarta: <http://eduhupakara.blogspot.com/2015/04/skp.html>.
- Masriyah. (1999). *Analisis Butir Soal*. Surabaya: Unesa University Press.
- Nuriyah, N. (2014). Evaluasi Pembelajaran. *Jurnal Edueksos*, 73-87.
- Nuswowati, M., Binadja, A., Soeprodjo, & Ifada, E. K. (2010). Pengaruh Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Terhadap Pencapaian Kompetensi. *Jurnal Inovasi Pendidikan Kimia*, 566-573.
- Purwanto. (2011). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Raharjo, T. (2016). Efektivitas Diklat Prajabatan Pola Baru Bagi Calon Pegawai Negeri Sipil Kementerian Keuangan. *Info Artha STAN*, 21-35.
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi. *Jurnal Pendidikan Akuntansi Indonesia*, 85-94.
- Ratumanan, G. T. (2003). *Evaluasi hasil belajar yang relevan dengan kurikulum berbasis kompetensi*. Surabaya: YP3IT dan Unipress.
- Sari, Candra, A. I., & Herawati, M. (2014). *Aplikasi Anates Versi 4 Dalam Menganalisis Butir Soal*.
- Slameto. (2001). *Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Sudijono. (2006). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Sugiyono. (2018). *Metode Penelitian*. Bandung: Alfabeta.
- Sumaningsih. (2015). Kualitas Butir Soal UAS Bahasa Inggris untuk Siswa MTs. *Lingua*, 224.
- Surapranata. (2009). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. Bandung: Remaja Rosdakarya.
- Yuslita, H., Zulfan, & Arifin, M. (2016). Analisis Tingkat Kesukaran Soal dan Daya Pembeda Soal Mata Pelajaran Sejarah Kelas XI Semester Ganjil di SMA 5 Banda Aceh Tahun 2015-2016. *Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 131-138.
- Zein, A., Fadillah, M., & Novianti, R. (2013). Hubungan Antara Validitas Butir, Reliabilitas, Tingkat Kesukaran Dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas XI Sma/Ma Negeri Di Kota Padang. *Semirata FMIPA Universitas Lampung*, 39-47.