

Buku Ajar Evaluasi Pembelajaran Bahasa Indonesia

by Eva Eri Dia

Submission date: 15-Mar-2024 01:19PM (UTC+0700)

Submission ID: 2320984676

File name: 3._Buku_Eva_Eri_Dia_Evaluasi_Pembelajaran_Bahasa_Indonesia.pdf (4.97M)

Word count: 25561

Character count: 158806

BUK U A J A R

Evaluasi Pembelajaran

BAHASA INDONESIA



Dr. Eva Eri Dia, S.Pd., M.Pd.

EVALUASI PEMBELAJARAN BAHASA INDONESIA | Dr. Eva Eri Dia, S.Pd., M.Pd.



Dr. **Eva Eri Dia** S.Pd., M.Pd.

DOSEN



PENDIDIKAN

- S1 STKIP PGRI Jombang
- Jurusan Pendidikan Bahasa dan Sastra Indonesia
- S2 Universitas Negeri Surabaya
- Jurusan Pendidikan Bahasa dan Sastra Indonesia
- S3 Universitas Negeri Malang
- Jurusan Pendidikan Bahasa Indonesia

Aktif sebagai dosen di STKIP PGRI Jombang sejak tahun 2007 sampai sekarang

Beberapa buku yang telah ditulis diantaranya: 1) Analisis Praanggapuan: Konsep Tindak Tutur (Presuposition) dalam Program Talk Show", dan 2) Pengantar Semantik Bahasa Indonesia. Telah banyak artikel yang dipublikasikan: 1) Model Sinekistik dalam Pembelajaran Kemampuan menggunakan Bahasa Indonesia untuk Memparkan Pendapat, 2) Pemahaman Konsep dan Ideologi Guru, 3) Penerapan Implementasi Perangkat Mengajar Model Kurikulum 2013, 3) Rethorika Menulis Kata Pengantar Skripsi, 4) The Gong Traveling, 5) Analisis Bentuk Praanggapuan 6) ICOLE, 7) Penamaan Anak sebagai Bentuk Kekuasaan, 8) Ideologi Nama Anak, 9) Kajian Semantik Makna Konotasi Pada Rubrik Opini "Jai Diri" Harian Jawa Pos, 10) Analisis Relaksi Makna yang Terdapat dalam Novel Hujan", karya Tere Liye, 11) Kajian Semantik dan Pragmatik dalam Buku Pembelajaran Tasawuf Melalui Program Al Karim Learning, 12) Penerjemahan Makna Semantik Versus Pragmatik dalam Buku Guru 4.0 Karya Lukman Hakim, 13) Katergori Makna Leksikal Verba Pada Iklan Produk Kecantikan MS Glow, 14., Peningkatan Perhalalan Belajar Melalui Model Kooper Learning pada Siswa Mts, 15) Peranan pada Nama Warcap di Kawasan Kesambi, 16) Tingkat Kesukaran Soal Ulangan Kelas XI di SMA As- Sulaimanyah Mojolugung Jombang, 17) Medan Makna Dalam Sajak Cinia Karya Mustola Bisipada Akun Instagram @Napas_Sastra, 18) Perituan Bunyi dalam Novel Magic Hour Karya Tisa Ts dan Stanley Meulen, 19) Personification Story Analysis In The Novel Dilan: Dia Adalah Dilanku Tahun 1990 By Pidi Baiq.

<https://lmaaksara.com/shop/>

peneliti_lmaaksara

5 aksara

031 4560279

peneliti aksara



LIMA AKSARA



ISBN 978-623-4149-35-3

9 | 786238 | 149353

BUKU AJAR
EVALUASI PEMBELAJARAN
BAHASA INDONESIA

Dr. Eva Eri Dia, S.Pd., M.Pd



PENERBIT
LIMA AKSARA

**Judul:**

Buku Ajar Evaluasi Pembelajaran
Bahasa Indonesia

Penulis:

Dr. Eva Eri Dia, S.Pd., M.Pd

ISBN 978-623-8149-35-3

Editor:

Prof. Dr. Erni Munastiwi, M.M
M. Syaifuddin S.

Penyunting:

Gendis Laras Kinasih

Desain sampul dan tata letak

Limax Media

Penerbit:

Lima Aksara

26

Redaksi:

Pratama Residence Blok C23/B19 Plosogeneng-Jombang |
0814-5606-0279 | <https://limaaksara.com>

Distributor tunggal:

CV. Lima Aksara | Pratama Residence Kav C23/B19 Plosogeneng-
Jombang | 0857-4666-6795 | IG@limaaksara | Fb: Lima Aksara
Indonesia

Cetakan Pertama Juni 2023

Hak cipta dilindungi undang-undang. Plagiasi
dipertanggungjawabkan secara utuh oleh penulis. Dilarang
memperbanyak isi buku ini, baik sebagian maupun seluruhnya
dalam bentuk apapun tanpa izin tertulis dari Penerbit.

Kata Pengantar

Buku Ajar Evaluasi Pembelajaran Bahasa Indonesia ini ditulis sebagai panduan para guru Bahasa Indonesia di berbagai tingkat dan para mahasiswa jurusan Pendidikan Bahasa Indonesia dalam melaksanakan Evaluasi pembelajaran di kelas. Beberapa hal yang dibahas diantaranya adalah; 1) Konsep Penilaian, Pengukuran Dan Evaluasi, 2) Tujuan Dan Prinsip Penilaian, 3) Alat Penilaian, 4) Norma Dan Skala Penilaian, 5) Validitas Dan Reliabilitas, 6) Pengolahan Hasil Tes Dan Non Tes, 7) Tingkat Kesukaran, Daya Beda Tes, Dan Pengecoh, 8) Penilaian Berbasis Kelas, 9) Pbk Dalam Ketrampilan Reseptif, Produktif, Dan Apresiasi, 9) Pengembangan Alat Penilaian Dan Penyusunan Soal, 10) Analisis Butir Soal Validitas, 11) Analisis Butir Soal Reliabilitas, 12) Analisis Butir Soal Pilihan Ganda, Dan 12) Analisis Butir Soal Uraian.

53 Kami menyadari buku ini masih banyak kekurangan karena keterbatasan pengetahuan dan pengalaman kami. Oleh karena itu, kami sangat mengharapkan kritik dan saran yang membangun dari pembaca demi kesempurnaan buku ini.

Jombang, 10 Agustus 2022

Penulis

Daftar Isi

KATA PENGANTAR | iii

DAFTAR ISI | iv

BAB 1 KONSEP PENILAIAN, PENGUKURAN DAN EVALUASI | 1

- A. Penilaian | 1
- B. Pengukuran | 3
- C. Evaluasi | 4

BAB 2 TUJUAN DAN PRINSIP PENILAIAN | 9

- A. Pengertian Penilaian | 9
- B. Tujuan Penilaian | 11
- C. Prinsip-prinsip Penilaian | 12

BAB 3 ALAT PENILAIAN | 16

- A. Pengertian Tes Dan Non-Tes | 15
- B. Macam-macam Tes | 18
- C. Pengertian Non-tes | 20
- D. Jenis-Jenis Tes | 26
- E. Tes Kebahasaan | 28
- F. Tes Kesastraan | 29

BAB 4 NORMA DAN SKALA PENILAIAN | 31

- A. Pengertian Penilaian Acuan Norma (PAN) | 31
- B. Tujuan dan Manfaat Penilaian Acuan Norma (PAN) | 32
- C. Kriteria Penyusunan Penilaian Acuan Norma (PAN) | 32
- D. Ciri-ciri Penilaian Acuan Norma (PAN) | 33
- E. Pengertian Penilaian Acuan Patokan (PAP) | 33
- F. Tujuan dan Manfaat Penilaian Acuan Patokan (PAP) | 34
- G. Penerapan Penilaian Acuan Patokan (PAP) | 35
- H. Kelebihan dan Kekurangan Penilaian Acuan Norma dan Penilaian Acuan Patokan | 35
- I. Persamaan dan Perbedaan Penilaian Acuan Norma dan Penilaian Acuan Patokan | 37

BAB 5 VALIDITAS DAN RELIABILITAS | 39

- A. Pengertian Validitas | 39
- B. Pengertian Reliabilitas | 45

BAB 6 PENGOLAHAN HASIL TES DAN NON TES | 49

- A. Pengertian Tes dan Non Tes | 49
- B. Pengolahan Hasil Tes | 50
- C. Pengolahan Hasil Non-Tes | 56

BAB 7 TINGKAT KESUKARAN, DAYA BEDA TES, DAN PENGECOH | 63

- A. Tingkatan Kesukaran | 63
- B. Cara Menentukan Tingkatan Kesukaran Tes | 63
- C. Daya Beda Butir Tes | 65

BAB 8 PENILAIAN BERBASIS KELAS | 69

- A. Hakikat dan Jenis Penilaian Berbasis Kelas | 69
- B. Objek dan Alat Penilaian Berbasis Kelas | 74
- C. Fungsi, tujuan, dan prinsip dari Penilaian Berbasis Kelas | 78
- D. Manfaat Hasil Penilaian Berbasis Kelas | 81
- E. Implikasi Penilaian Berbasis Kelas | 81

BAB 9 PBK DALAM KETRAMPILAN RESEPTIF BERBAHASA | PBK DALAM KETRAMPILAN PRODUKTIF BERBAHASA | PBK DALAM APRESIASI SASTRA | 84

- A. PBK dalam keterampilan reseptif berbahasa | 84
- B. PBK dalam keterampilan produktif berbahasa | 89
- C. PBK dalam apresiasi sastra | 95

BAB 10 PENGEMBANGAN ALAT PENILAIAN DAN PENYUSUNAN SOAL | 97

- A. Pengembangan Alat Penilaian | 97
- B. Penyusunan Soal | 102

BAB 11 ANALISIS BUTIR SOAL VALIDITAS | 108

- A. Jenis-jenis Validitas | 108
- B. Validitas Isi (Content Validity) | 112

- C. Validitas Konsep (Concept / Construct Validity) | 114
- D. Validitas Pengukuran Setara (Congruent Validity) | 115
- E. Validitas Pengukuran Serentak (Concurrent Validity) | 115
- F. Validitas Ramalan (Predictive Validity) | 116

BAB 12 ANALISIS BUTIR SOAL RELIABILITAS | 118

- A. Pengertian Analisis Reliabilitas | 118
- B. Jenis - Jenis Reliabilitas | 122
- C. Karakteristik Reliabilitas | 123
- D. Metode Pengujian Reliabilitas | 125

BAB 13 ANALISIS BUTIR SOAL PILIHAN GANDA | 127

- A. Tingkat Kesukaran Soal | 127
- B. Daya Pembeda Soal | 137
- C. Pengecoh Soal | 139

BAB 14 ANALISIS BUTIR SOAL URAIAN | 144

- A. Tingkat Kesukaran | 144
- B. Daya Pembeda (discriminating power) | 164

DAFTAR PUSTAKA | 172

I

KONSEP PENILAIAN, PENGUKURAN DAN EVALUASI

25

A. Penilaian

Istilah penilaian merupakan alih bahasa dari istilah *assessment*. Penilaian adalah suatu kegiatan untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai peserta didik. Kata “menyeluruh” mengandung arti bahwa penilaian tidak hanya ditunjukkan pada penguasaan salah satu bidang tertentu saja, tetapi mencakup aspek pengetahuan, keterampilan, sikap, dan nilai-nilai. Selanjutnya, Gronlund mengartikan “penilaian adalah suatu proses yang sistematis dari pengumpulan, analisis, dan interpretasi informasi/data untuk menentukan sejauh mana peserta didik telah mencapai tujuan pembelajaran. Dari pengertian di atas jelas menunjukkan bahwa penilaian lebih difokuskan pada peserta didik sebagai subjek belajar dan tidak sedikitpun menyinggung komponen-komponen pembelajaran lainnya.

Dengan demikian, dapat disimpulkan bahwa penilaian adalah suatu proses atau kegiatan yang sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil

belajar peserta didik dalam rangka membuat keputusan-keputusan berdasarkan kriteria dan pertimbangan tertentu. Keputusan yang dimaksud adalah keputusan tentang peserta didik, seperti nilai yang akan diberikan atau juga keputusan tentang kenaikan kelas atau kelulusan (Arifin, 2014:4).

Keputusan tentang peserta didik meliputi juga pengelolaan belajar, penempatan peserta didik sesuai dengan jenjang atau jenis program pendidikan, bimbingan dan konseling, dan menyeleksi peserta didik untuk pendidikan lebih lanjut. Keputusan penilaian terhadap suatu hasil belajar sangat bermanfaat untuk membantu peserta didik merefleksikan apa yang mereka ketahui, bagaimana mereka belajar, dan mendorong tanggung jawab dalam belajar. Keputusan penilaian dapat dibuat oleh guru, sesama peserta didik (peer) atau dirinya sendiri (self-assessment). Pengambilan keputusan perlu menggunakan pertimbangan yang berbeda-beda dan membandingkan hasil penilaian. Pengambilan keputusan harus membimbing peserta didik untuk mel²⁵kan perbaikan pencapaian hasil belajar.

Penilaian harus dipandang sebagai salah satu faktor penting yang menentukan keberhasilan proses dan hasil belajar, bukan hanya sebagai cara yang digunakan untuk menilai hasil belajar. Kegiatan penilaian harus dapat memberikan informasi kepada guru untuk meningkatkan kemampuan mengajarnya dan membantu peserta didik mencapai perkembangan belajarnya secara optimal. Implikasinya adalah kegiatan penilaian harus digunakan sebagai cara atau teknik untuk mendidik

sesuai dengan prinsip pedagogis. Guru harus menyadari bahwa kemajuan belajar peserta didik merupakan salah satu indikator keberhasilannya dalam pembelajaran. Jika sebagian besar peserta didik tidak berhasil dalam belajarnya berarti pula merupakan kegagalan bagi guru itu sendiri (Arifin, 2014:4).

B. Pengukuran

Pengukuran adalah suatu proses atau kegiatan untuk menentukan kuantitas sesuatu. Kata “sesuatu” bisa berarti peserta didik, guru, gedung sekolah, meja belajar, white board, dan sebagainya. Dalam proses pengukuran, tentu guru harus menggunakan alat ukur (tes atau non-tes). Alat ukur tersebut harus standar, yaitu memiliki derajat validitas dan reliabilitas yang tinggi. Dalam bidang pendidikan, psikologi, maupun variabel-variabel sosial lainnya, kegiatan pengukuran biasanya menggunakan tes.

Pengukuran merupakan proses yang mendeskripsikan performance peserta didik dengan menggunakan suatu skala kuantitatif (system angka) sedemikian rupa sehingga sifat kualitatif dari performance peserta didik tersebut dinyatakan dengan angka-angka (Alwasilah et al.1996). Dapat didefinisikan sebagai proses penetapan angka terhadap individu atau karakteristiknya menurut aturan tertentu (Ebel &Frisbie, 1986). Pengukuran adalah usaha untuk mengetahui keadaan sesuatu hal menurut apa adanya, yang biasanya dinyatakan dalam bilangan. Sedangkan menurut Endang Purwanti (2008:4) pengukuran dapat diartikan

sebagai kegiatan atau upaya yang dilakukan untuk memberikan angka-angka pada suatu gejala atau peristiwa, atau benda, sehingga hasil pengukuran akan selalu berupa angka. ⁶

Dari pendapat ahli beberapa ahli tersebut dapat disimpulkan bahwa pengukuran adalah suatu kegiatan yang dilakukan untuk menentukan fakta kuantitatif yang disesuaikan dengan kriteria-kriteria tertentu sesuai dengan objek yang akan diukur.

C. Evaluasi ¹⁹

Evaluasi lebih luas ruang lingkungannya daripada penilaian, sedangkan penilaian lebih terfokus pada aspek tertentu saja yang merupakan bagian dari ruang lingkup tersebut. Jika hal yang ingin dinilai adalah sistem pembelajaran, maka ruang lingkungannya adalah semua komponen pembelajaran, dan istilah yang tepat untuk menilai sistem pembelajaran adalah evaluasi, bukan penilaian. Jika hal yang ingin dinilai satu atau beberapa bagian/komponen pembelajaran, misalnya hasil belajar, maka istilah yang tepat digunakan adalah penilaian bukan evaluasi. Di samping itu, ada juga istilah pengukuran. Kalau evaluasi dan penilaian bersifat kualitatif, maka pengukuran bersifat kuantitatif (skor/angka) yang diperoleh dengan menggunakan suatu alat ukur atau instrumen. Dalam konteks hasil belajar, alat ukur atau instrumen tersebut dapat berbentuk tes atau non-tes ³⁶ (Zainal, 2014:2).

Dalam sistem pembelajaran (maksudnya pembelajaran sebagai suatu sistem), evaluasi merupakan salah satu komponen penting dan tahap

yang harus ditempuh oleh guru untuk mengetahui keaktifan pembelajaran. Hasil yang diperoleh dari evaluasi dapat dijadikan balikan (feed-back) bagi guru dalam memperbaiki dan menyempurnakan program dan kegiatan pembelajaran (Zainal, 2014:2). Evaluasi merupakan suatu proses menyediakan informasi yang dapat dijadikan sebagai pertimbangan untuk menentukan harga dan jasa (the world and merit) dari tujuan yang dicapai, desain, implementasi dan dampak untuk membantu membuat keputusan, membantu pertanggungjawaban dan meningkatkan pemahaman terhadap fenomena. Menurut rumusan tersebut, inti dari evaluasi adalah penyediaan informasi yang dapat dijadikan sebagai bahan pertimbangan dalam mengambil keputusan (Eko, 2009:3).

Evaluasi merupakan suatu proses atau kegiatan pemilihan, pengumpulan, analisis dan penyajian informasi yang dapat digunakan sebagai dasar pengambilan keputusan serta penyusunan program selanjutnya. Proses yang menentukan sejauh mana tujuan pendidikan dapat dicapai. Menurut Eko (2014:4), dalam melaksanakan evaluasi ada tujuh elemen yang harus dilakukan, yaitu:

- a) Penentuan fokus yang akan dievaluasi (*focusing the evaluation*)
- b) Penyusunan desain evaluasi (*designing the evaluation*)
- c) Pengumpulan informasi (*collecting information*)
- d) Analisis dan inteprestasi informasi (*analyzing and interpreting*)
- e) Pembuatan laporan (*reporting information*)

- f) Pengelolaan evaluasi (*managing evaluation*), dan
- g) Evaluasi untuk evaluasi (*evaluating evaluation*).

Dalam pengertian tersebut menunjukkan bahwa dalam melakukan evaluasi, evaluator pada tahap awal harus menentukan fokus yang akan dievaluasi dan desain yang akan digunakan. Hal ini berarti harus ada kejelasan apa yang akan dievaluasi yang secara implisit menenkankan adanya tujuan evaluasi, serta adanya perencanaan bagaimana melaksanakan evaluasi. Selanjutnya, dilakukan pengumpulan data, menganalisis dan membuat interpretasi terhadap data yang terkumpul serta membuat laporan. Selain itu, evaluator juga harus melakukan pengaturan terhadap evaluasi dan mengevaluasi apa yang telah dilakukan dalam melaksanakan evaluasi secara keseluruhan (Eko, 2014:5).

Ber²⁹sarkan pendapat diatas dapat disimpulkan bahwa evaluasi merupakan proses yang sistematis dan berkelanjutan untuk mengumpulkan, mendeskripsikan, menginterpretasikan dan menyajikan informasi tentang suatu program untuk dapat digunakan sebagai dasar membuat keputusan, maupun menyusun program selanjutnya. Adapun tujuan evaluasi adalah untuk memperoleh informasi yang akurat dan objektif tentang suatu program. Informasi tersebut dapat berupa proses pelaksanaan program, dampak/hasil yang dicapai, efisiensi serta pemanfaatan hasil evaluasi yang difokuskan untuk program itu sendiri, yaitu untuk mengambil keputusan apakah dilanjutkan, diperbaiki atau dihentikan. Selain itu, juga digunakan untuk kepentingan penyusunan program berikutnya

maupun penyusunan kebijakan yang terkait dengan program (Eko, 2014:6).

Wujud dari hasil evaluasi adalah adanya rekomendasi dari evaluator untuk pengambil keputusan (*decision maker*). Menurut Anas (2011:22), ada empat kemungkinan kebijakan yang dapat dilakukan berdasarkan hasil evaluasi pelaksanaan program, yaitu:

- a) Menghentikan program, karena dipandang bahwa program tersebut tidak ada manfaatnya, atau tidak dapat dilaksanakan sebagaimana diharapkan.
- b) Merevisi program, karena ada bagian-bagian yang kurang sesuai dengan harapan (terdapat kesalahan tetapi sedikit).
- c) Melanjutkan program, karena pelaksanaan program menunjukkan bahwa segala sesuatu sudah berjalan sesuai dengan harapan dan memberikan hasil yang bermanfaat.
- d) Menyebarkan program (melaksanakan program di tempat lain atau mengulangi lagi di lain waktu), karena program tersebut berhasil dengan baik, maka sangat baik jika dilaksanakan lagi ditempat dan waktu lain.

29 Dalam bidang pendidikan ditinjau dari sasarannya, evaluasi ada yang bersifat makro dan ada yang mikro. Evaluasi yang bersifat makro sasarannya adalah program pendidikan, yaitu program yang direncanakan untuk memperbaiki bidang pendidikan.

44 Evaluasi mikro sering digunakan ditingkat kelas. Jadi sasaran evaluasi mikro adalah program pembelajaran di kelas dan yang menjadi penanggung jawabnya

adalah guru untuk sekolah atau dosen untuk perguruan tinggi (Daryanto, 2008:2). Guru mempunyai tanggungjawab menyusun dan melaksanakan program pembelajaran di kelas, sedangkan pimpinan sekolah bertanggung untuk mengevaluasi program pembelajaran yang disusun dan dilaksanakan oleh guru.

2

TUJUAN DAN PRINSIP PENILAIAN

A. Pengertian Penilaian

Dalam kegiatan pembelajaran, seorang guru harus menguasai beberapa pengetahuan terkait dengan penilaian pendidikan, diantaranya: (1) Mampu memilih prosedur-prosedur penilaian yang tepat untuk membuat keputusan pembelajaran, (2) Mampu mengembangkan prosedur penilaian yang tepat untuk membuat keputusan pembelajaran, (3) Mampu dalam melaksanakan, melakukan penskoran, serta menafsirkan hasil penilaian yang telah dibuat, (4) Mampu menggunakan hasil-hasil penilaian untuk membuat keputusan-keputusan di bidang pendidikan, (5) Mampu mengembangkan prosedur penilaian yang valid dan menggunakan informasi penilaian, dan (6) Mampu dalam mengkomunikasikan hasil hasil penilaian.

“Evaluation is a systematic process determining the extent to which instructional objectives are achieved by pupils”. Kalimat tersebut menjelaskan bahwa penilaian adalah suatu proses dalam mengumpulkan informasi dan membuat keputusan berdasarkan

informasi tersebut. Dalam proses mengumpulkan informasi, tentunya tidak semua informasi bisa digunakan untuk membuat sebuah keputusan. Informasi-informasi yang relevan dengan apa yang dinilai akan mempermudah dalam melakukan sebuah **49** penilaian dalam kegiatan pembelajaran.

Penilaian merupakan suatu proses atau kegiatan yang sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dalam rangka membuat keputusan-keputusan berdasarkan kriteria dan pertimbangan tertentu. Definisi dari penilaian juga disampaikan oleh Ralph Tyler yang mengungkapkan bahwa penilaian merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagaimana tujuan pendidikan sudah tercapai.

Menurut Griffin dan Nix, penilaian adalah suatu pernyataan berdasarkan sejumlah fakta untuk menjelaskan tentang karakteristik seseorang atau sesuatu. Haryati berpendapat lain, ia mengungkapkan bahwa penilaian (assessment) merupakan istilah yang mencakup semua metode yang biasa dipakai untuk mengetahui keberhasilan belajar siswa dengan cara menilai unjuk kerja individu peserta didik atau kelompok.

Penilaian merupakan bagian integral dari pembelajaran matematika dan memberikan kontribusi yang signifikan pada pembelajaran matematika. Oleh sebab itu, guru juga harus merencanakan penilaian yang akan digunakan sebagai bagian dari pembelajaran. Dalam dunia pendidikan, Gronlund

dan Linn mendefinisikan tentang sebuah penilaian sebagai suatu proses yang sistematis dan mencakup kegiatan mengumpulkan, menganalisis, serta menginterpretasikan informasi untuk menentukan seberapa jauh seorang siswa atau sekelompok siswa mencapai tujuan pembelajaran yang telah ditetapkan, baik aspek pengetahuan, sikap maupun keterampilan. Berdasarkan beberapa pendapat di atas, maka dapat disimpulkan bahwa penilaian adalah suatu proses pengumpulan informasi secara menyeluruh yang dilakukan secara terus menerus untuk mengetahui kemampuan atau keberhasilan siswa dalam pembelajaran dengan menilai kinerja siswa baik kinerja secara individu maupun dalam kegiatan kelompok. Penilaian itu harus mendapatkan perhatian yang lebih dari seorang guru. Dengan demikian, penilaian tersebut harus dilaksanakan dengan baik, karena penilaian merupakan komponen vital (utama) dari pengembangan diri yang sehat, baik bagi individu (siswa) maupun bagi organisasi/kelompok (Suprananto, 2012).

B. Tujuan Penilaian

Chittenden (1994) secara simpel mengklasifikasikan tujuan penilaian (assessment purpose) adalah untuk (1). keeping track, (2). checking up, (3). finding-out, and (4). summing-up (Asrul ¹ al., 2014).

Keempat tujuan tersebut oleh Arifin (2013:15) diuraikan sebagai berikut:

1. Keeping track, yaitu untuk menelusuri dan melacak proses belajar peserta didik sesuai

dengan rencana pelaksanaan pembelajaran yang telah ditetapkan. Untuk itu, guru harus mengumpulkan data dan informasi dalam kurun waktu tertentu melalui berbagai jenis dan teknik penilaian untuk memperoleh gambaran tentang pencapaian kemajuan belajar peserta didik.

2. **Checking-up**, yaitu untuk mengecek ketercapaian kemampuan peserta didik dalam proses pembelajaran dan kekurangan-kekurangan peserta didik selama mengikuti proses pembelajaran. Dengan kata lain, guru perlu melakukan penilaian untuk mengetahui bagian mana dari materi yang sudah dikuasai peserta didik dan bagian mana dari materi yang belum dikuasai.
3. **Finding-out**, yaitu untuk mencari, menemukan dan mendeteksi kekurangan kesalahan atau kelemahan peserta didik dalam proses pembelajaran, sehingga guru dapat dengan cepat mencari alternatif solusinya.
4. **Summing-up**, yaitu untuk menyimpulkan tingkat penguasaan peserta didik terhadap kompetensi yang telah ditetapkan. Hasil penyimpulan ini dapat digunakan guru untuk menyusun laporan kemajuan belajar ke berbagai pihak yang berkepentingan.

C. Prinsip-prinsip Penilaian

(Irhamni, 2018) menyebutkan ada beberapa prinsip penilaian yang yang disadurkan oleh M.Ngalim Purwanto, dalam bukunya, "Prinsip-prinsip dan Teknik Evaluasi Pengajaran".

5

1. Penilaian hendaknya didasarkan pada hasil pengukuran yang komprehensif
2. Harus dibedakan antara penskoran (scoring) dan penilaian (grading)
3. Dalam proses pemberian nilai hendaknya diperhatikan adanya dua macam orientasi, yaitu penilaian norms-referenced dan criterion-referenced.

Norms-referenced evaluation adalah penilaian yang diorientasikan kepada suatu kelompok tertentu; jadi, hasil hasil evaluasi perseorangan siswa atau mahasiswa dibandingkan dengan prestasi kelompoknya. Prestasi kelompoknya itulah yang dijadikan patokan atau norm dalam menilai siswa atau mahasiswa secara perorangan.

Penilaian norms-referenced kompetitif intrakelompok. Criterion-referenced evaluation adalah penilaian yang diorientasikan kepada suatu standar absolute, tanpa dihubungkan dengan suatu kelompok tertentu. Penilaian Criterion-referenced sangat relevan bagi lembaga pendidikan yang telah menggunakan kurikulum yang berdasarkan kompetensi (competency ased education).

4. Kegiatan pemberian nilai hendaknya merupakan bagian integral dari proses belajar mengajar.

Ini berarti bahwa tujuan penilaian di samping untuk mengetahui status siswa dan menaksir kemampuan belajar serta penguasaannya terhadap bahan pelajaran, juga

digunakan sebagai feedback (umpan balik), baik kepada siswa sendiri maupun bagi guru atau pengajar. Dengan demikian pengajar dapat mengetahui kelebihan dan kelemahan siswa tertentu sehingga selanjutnya ia dapat melakukan koreksi terhadap kesalahan yang dilakukannya dan atau member reinforcement

12 bagi prestasinya yang baik.

5. Penilaian harus bersifat komparabel.

Artinya, setelah tahap pengukuran yang menghasilkan angka-angka itu dilaksanakan, prestasi-prestasi yang menduduki skor yang sama harus memperoleh nilai 23 ng sama pula. Dengan kata lain, penilaian harus dilakukan secara adil, jangan sampai terjadi penganakemasan atau penganaktirian.

6. Sistem penilaian yang dipergunakan hendaknya jelas bagi siswa dan pengajar sendiri.

(Irhamni, 2018) menyebutkan ada beberapa prinsip penilaian yang yang disadurkan oleh M.Ngalim Purwanto, dalam bukunya, "Prinsip-prinsip dan 5 Teknik Evaluasi Pengajaran".

1. Penilaian hendaknya didasarkan pada hasil pengukuran yang komprehensif

2. Harus dibedakan antara penskoran (scoring) dan penilaian (grading)

3. Dalam proses pemberian nilai hendaknya diperhatikan adanya dua macam orientasi, yaitu penilaian norms-referenced dan criterion-referenced.

4. Kegiatan pemberian nilai hendaknya merupakan bagian integral dari proses belajar mengajar.
5. Penilaian harus bersifat komparabel.
6. Sistem penilaian yang dipergunakan hendaknya jelas bagi siswa dan pengajar sendiri.

3

ALAT PENILAIAN

A. Pengertian Tes Dan Non-Tes

Istilah "tes" berasal dari bahasa latin "testum" yang berarti sebuah piring atau jambangan dari tanah liat. Istilah "tes" kemudian dipergunakan dalam lapangan psikologi dan selanjutnya hanya dibatasi sampai metode psikologi, yaitu suatu cara untuk menyelidiki seseorang. Penyelidikan tersebut dilakukan mulai dari pemberian suatu tugas kepada seseorang atau untuk menyelesaikan suatu masalah tertentu (Zainal Arifin:2010: hlm 2).

Gilbert Sax (1980) mengemukakan "tes dapat didefinisikan sebagai tugas atau serangkaian tugas yang digunakan untuk memperoleh pengamatan sistematis yang dianggap mewakili ciri pendidikan atau psikologis atau pada upeti". Dalam pengertian tersebut, Sax lebih menekankan tes sebagai suatu tugas atau rangkaian tugas. Istilah tugas dapat berbentuk soal atau perintah/suruhan lain yang

harus dikerjakan oleh peserta didik. Hasil kuantitatif ataupun kualitatif dari pelaksanaan tugas itu digunakan untuk menarik kesimpulan-kesimpulan tertentu terhadap seseorang.

S. Hamid Hasan (1988) menjelaskan *"tes adalah alat pengumpulan data yang dirancang secara khusus. Kekhususan tes dapat terlihat dari konstruksi 33 tir (soal) yang dipergunakan"*. Hamid Hasan lebih terfokus pada tes sebagai alat pengumpul data. Pengumpulan data bukan hanya ada dalam prosedur penelitian, tetapi juga ada dalam prosedur evaluasi. Untuk mengumpulkan data evaluasi tentu orang memerlukan suatu alat, antara lain tes. Tes dapat berupa pertanyaan. Oleh sebab itu, setiap jenis pertanyaan yang dipergunakan, rumusan pertanyaan yang diberikan, pola jawaban yang disediakan atau dirancang harus memenuhi suatu perangkat kriteria yang ketat.

Conny Semiawan S. (1986) mengemukakan tes adalah *"... alat pengukur untuk menetapkan apakah berbagai faset dari kesan yang kita perkirakan dari seseorang adalah benar merupakan fakta, juga adalah cara untuk menggambarkan bermacam-macam faset ini seobjektif mungkin"*.

Dari beberapa pendapat para ahli tersebut, dapat disimpulkan, bahwa tes adalah *suatu alat yang berisi serangkaian tugas yang harus dikerjakan atau soal-soal yang harus dijawab oleh peserta didik untuk mengukur suatu aspek perilaku tertentu*. Dengan demikian, *fungsi tes adalah sebagai alat ukur*. Dalam tes prestasi belajar, aspek perilaku yang hendak diukur adalah tingkat kemampuan peserta didik

dalam menguasai materi pelajaran yang telah disampaikan.

B. Macam-macam Tes

a) Dilihat dari jumlah peserta didik

- Tes Kelompok adalah tes yang dapat melayani sekelompok testis dalam suatu waktu. Tes kelompok ini lebih ekonomis jika dibandingkan dengan tes individual sebab dalam waktu singkat dapat diperoleh banyak individu yang diteliti.
- Tes Perorangan adalah tes yang dilaksanakan dengan tujuan mengungkap ciri-ciri khas dari seseorang yang banyak sedikitnya bersifat lahiriyah, seperti gaya bicara, cara berpakaian, nada suara, hobi atau kesenangan dan lain-lain. Kepribadian adalah organisasi dinamik dalam individu atas sistem-sistem psikososial yang menentukan penyesuaian dirinya yang khas terhadap lingkungan.

b) Dilihat dari kajian psikologi

- Test kemampuan (aptitude test), yaitu tes bakat. Test ini bertujuan untuk mengungkap bakat khusus yang dimiliki oleh seseorang. Tes bakat adalah pengukuran kognitif yang dirancang untuk memprediksikan sejauh mana individu akan mencapai kesuksesan sebelum mereka dilatih, dipilih dan ditempatkan. Tes bakat yang digunakan untuk memprediksi kesuksesan dalam suatu program khusus disebut tes bakat khusus.

- 65
- Tes Intelligensia Umum, tes intelegensi pertama yang terkenal adalah yang dikem⁴ngkan oleh Alfred Binet (1911) tahun 1905. Secara operasional, intelegensi adalah kemampuan yang tidak dapat diukur secara langsung dengan tes. Ketika tes memuat butir-butir pertanyaan, seorang individu dapat diketahui kadar inteleg⁴nsinya.
 - Tes Prestasi Belajar, tes prestasi akademik mengukur kemampuan saat itu, tes pres⁴si akademik mengandung validitas isi. Tes prestasi akademik disajikanm memuat ukuran prestasi dalam persentil, usia, tingkat, dan skor standar. Tes prestasi akademik standar dapat digunakan untuk keperluan seleksi dan penempatan, diagnosis, feedback, evaluasi program (STANDAR, T).
- c) Dilihat dari cara penyusunannya¹⁸
- Tes Buatan Guru adalah tes yang disusun sendiri oleh guru yang akan mempergunakan tes tersebut. Tes tersebut biasanya digunakan untuk ulangan harian, formatif, dan ulangan umum (sumatif). Berdasarkan isi dan tujuan bersifat kh³¹s.
 - Tes Baku adalah tes yang dikaji berulang-ulang kepada sekelompok besar peserta didik, dan item-itemnya relevan serta mempunyai daya pembeda yang tinggi. Tes baku biasanya telah dianalisis secara statistik dan diuji secara empiris oleh para pakar, oleh karena itu dapat dinyatakan valid untuk digunakan

secara umum. Berdasarkan isi dan tujuan bersifat umum.

d) Dilihat dari bentuk jawaban peserta didik

- Tes tertulis (written tes) adalah tes yang mengajukan butir-butir pertanyaan dengan mengharapkan jawaban tertulis. Tes ini biasanya digunakan untuk mengukur aspek kognitif peserta didik.
- Test lisan (oral test) adalah tes yang mengajukan pertanyaan-pertanyaan dengan menghendaki jawaban secara lisan. Tes ini biasanya dilakukan untuk aspek kognitif peserta didik.
- Test perbuatan (performance test) adalah tes yang mengajukan pertanyaan-pertanyaan dengan menghendaki jawaban dalam bentuk perbuatan. Tes ini biasanya digunakan untuk menilai aspek psikomotor/ keterampilan peserta didik.

C. Pengertian Non-tes

a) Observasi

Observasi adalah suatu proses pengamatan dan pencatatan secara sistematis, logis, objektif, dan rasional mengenai berbagai fenomena, baik dalam situasi yang sebenarnya maupun dalam situasi buatan untuk mencapai tujuan tertentu. Alat yang digunakan dalam melakukan observasi disebut pedoman observasi. Observasi tidak hanya digunakan dalam kegiatan evaluasi, tetapi juga dalam bidang penelitian, terutama penelitian kualitatif (*qualitative research*).

Tujuan utama observasi adalah:

1. Untuk mengumpulkan data dan informasi mengenai suatu fenomena, baik yang berupa peristiwa maupun tindakan, baik dalam situasi yang sesungguhnya maupun dalam situasi buatan;
2. Untuk mengukur perilaku kelas (baik perilaku guru maupun perilaku peserta didik), interaksi antara peserta didik dan guru, dan faktor-faktor yang dapat diamati lainnya, terutama kecakapan sosial (*social skills*).

b) Wawancara

Wawancara merupakan salah satu bentuk alat evaluasi jenis non-tes yang dilakukan melalui percakapan dan tanya jawab, baik langsung maupun tidak langsung dengan peserta didik. Pengertian wawancara langsung adalah wawancara yang dilakukan secara langsung antara pewawancara (interviewer) atau guru dengan orang yang diwawancarai (interviewee) atau peserta didik tanpa melalui perantara, sedangkan wawancara tidak langsung artinya pewawancara atau guru menanyakan sesuatu kepada peserta didik melalui perantaraan orang lain atau media. Jadi, tidak menemui langsung kepada sumbernya.

Tujuan wawancara adalah:

1. Untuk memperoleh informasi secara langsung guna menjelaskan suatu hal atau situasi dan kondisi tertentu.
2. Untuk melengkapi suatu penyelidikan ilmiah.

3. Untuk memperoleh data agar dapat memengaruhi situasi atau orang tertentu.

Kelebihan wawancara antara lain:

1. Dapat berkomunikasi secara langsung kepada peserta didik sehingga informasi yang diperoleh dapat diketahui objektivitasnya.
2. Dapat memperbaiki proses dan hasil belajar.
3. Pelaksanaan wawancara lebih fleksibel, dinamis, dan personal.

Kelemahan wawancara antara lain:

1. Jika jumlah peserta didik cukup banyak, maka proses wawancara banyak menggunakan waktu, tenaga, dan biaya,
2. Adakalanya terjadi wawancara yang berlarut-larut tanpa arah, sehingga data kurang dapat memenuhi apa yang diharapkan,
3. Sering timbul sikap yang kurang baik dari peserta didik yang diwawancarai dan sikap overaction dari guru sebagai pewawancara, karena itu perlu adanya adaptasi diri antara pewawancara dengan orang yang diwawancarai.

c) Skala Sikap

Sikap merupakan suatu kecenderungan tingkah laku untuk berbuat sesuatu dengan cara, metode, teknik, dan pola tertentu terhadap dunia sekitarnya, baik berupa orang-orang maupun berupa objek-objek tertentu. Sikap mengacu

kepada perbuatan atau perilaku seseorang, tetapi tidak berarti semua perbuatan identik dengan sikap. Perbuatan seseorang mungkin saja bertentangan dengan sikapnya. Guru perlu mengetahui norma-norma yang ada pada peserta didik, bahkan sikap peserta didik terhadap dunia sekitarnya, terutama terhadap mata pelajaran dan lingkungan sekolah. Jika terdapat sikap peserta didik yang negatif, guru perlu mencari suatu cara atau teknik tertentu untuk menempatkan sikap negatif itu menjadi sikap yang positif.

Dalam mengukur sikap, guru hendaknya memperhatikan tiga komponen sikap, yaitu:

1. Kognisi, yaitu berkenaan dengan pengetahuan peserta didik tentang objek
2. Afeksi, yaitu berkenaan dengan perasaan peserta didik terhadap objek
3. Konasi, yaitu berkenaan dengan kecenderungan berperilaku peserta didik terhadap objek.

Di samping itu, guru juga harus memilih salah satu model skala sikap. Adapun model, model skala sikap yang biasa digunakan untuk menilai sikap peserta didik terhadap suatu objek, antara lain:

1. Menggunakan bilangan untuk menunjukkan tingkat-tingkat dari objek sikap yang dinilai, seperti 1, 2, 3, 4 dan seterusnya.
2. Menggunakan frekuensi terjadinya atau timbulnya sikap itu, seperti: selalu, sering kali, kadang-kadang, pernah, dan tidak pernah.
3. Menggunakan istilah-istilah yang bersifat kualitatif, seperti bagus sekali, baik, sedang, dan

kurang. Ada juga istilah-istilah lain, seperti: sangat setuju, setuju, ragu-ragu (tidak punya pendapat), tidak setuju, dan sangat tidak setuju.

4. Menggunakan istilah-istilah yang menunjukkan status/kedudukan, seperti sangat rendah, di bawah rata-rata, di atas rata-rata, dan sangat tinggi.
5. Menggunakan kode bilangan atau huruf, seperti selalu (diberi kode 5), kadang-kadang (4), jarang (3), jarang sekali (2), dan tidak pernah (1).

d) Daftar Cek

Daftar cek adalah suatu daftar yang berisi subjek dan aspek-aspek yang akan diamati. Daftar cek dapat memungkinkan guru sebagai penilai mencatat tiap-tiap kejadian yang betapapun kecilnya, tetapi dianggap penting. Ada bermacam-macam aspek perbuatan yang biasanya dicantumkan dalam daftar cek, kemudian tinggal memberikan tanda centang (V) pada tiap-tiap aspek tersebut sesuai dengan hasil penilaiannya. Daftar cek banyak manfaatnya, antara lain membantu guru untuk mengingat-ingat apa yang harus diamati; dan dapat memberikan informasi kepada stakeholder. Namun, penilai tetap harus waspada kemungkinan perilaku penting yang belum tercakup di dalam daftar cek, karena itu penilai jangan terlalu kaku dengan apa yang sudah tertulis pada daftar cek tersebut.

e) Skala Penilaian

Dalam daftar cek, penilai hanya dapat mencatat tingkah laku tertentu, sedangkan dalam skala

penilaian fenomena fenomena yang akan dinilai itu disusun dalam tingkatan-tingkatan yang telah ditentukan. Jadi, tidak hanya mengukur secara mutlak ada atau tidaknya variabel tertentu, tetapi lebih jauh mengukur bagaimana intensitas gejala yang ingin diukur. Pencatatan melalui daftar cek termasuk pencatatan yang kasar. Fenomena-fenomena hanya dicatat ada atau tidak ada. Hal ini agak kurang realistik. Perilaku manusia, baik yang berwujud sikap jiwa, aktivitas, maupun prestasi belajar timbul dalam tingkat-tingkat tertentu. Oleh karena itu, untuk mengukur hal-hal tersebut ada baiknya digunakan skala penilaian.

Namun demikian, skala penilaian juga mempunyai kelemahan, antara lain "ada kemungkinan halo effects, generosity effects, dan carry-over effects":

1. Ada kemungkinan terjadinya halo effects, yaitu kelemahan yang akan timbul jika dalam pencatatan observasi terpicat oleh kesan-kesan umum yang baik pada peserta didik sementara ia tidak menyelidiki kesan-kesan umum itu.
2. Generosity effects, yaitu kelemahan yang akan muncul bila ada keinginan untuk berbuat baik.
3. Carry-over effects, yaitu kelemahan akan muncul jika guru tidak dapat memisahkan satu fenomena dengan fenomena yang lain.

f) Angket

Angket termasuk alat untuk mengumpulkan dan mencatat data atau informasi, pendapat, dan paham dalam hubungan kausal. Angket mempunyai kesamaan dengan wawancara, kecuali dalam implementasinya. Angket dilaksanakan

secara tertulis, sedangkan wawancara dilaksanakan secara lisan.

Keuntungan angket antara lain

1. Responden dapat menjawab dengan bebas tanpa dipengaruhi oleh hubungan dengan peneliti atau penilai, dan waktu relatif lama, sehingga objektivitas dapat terjamin
2. Informasi atau data terkumpul lebih mudah karena itemnya homogeny.
3. Dapat digunakan untuk mengumpulkan data dari jumlah responden yang besar yang dijadikan sampel.

Kelemahan angket antara lain:

1. Ada kemungkinan angket diisi oleh orang lain
2. Hanya diperuntukkan bagi yang dapat melihat saja
3. Responden hanya menjawab berdasarkan jawaban yang ada.

Angket terdiri atas beberapa bentuk, yaitu:

1. Bentuk angket berstruktur, yaitu angket yang menyediakan beberapa kemungkinan jawaban.
2. Bentuk angket tak berstruktur, yaitu bentuk angket yang memberikan jawaban secara terbuka.

D. Jenis-Jenis Tes

- a) **Tes Seleksi**, Tes seleksi sering dikenal dengan istilah "*ujian saringan*" atau "*ujian masuk*". Tes ini dilaksanakan dalam rangka penerimaan calon mahasiswa baru, dimana hasil tes digunakan untuk memilih calon peserta didik yang tergolong paling baik dari sekian banyak calon yang

mengikuti tesateri tes pada tes seleksi ini merupakan materi prasyarat untuk mengikuti program pendidikan yang akan diikuti oleh calon. Sesuai dengan sifatnya, yaitu menyeleksi atau melakukan penyaringan, maka materi tes seleksi terdiri atas butir-butir soal yang cukup sulit, sehingga hanya calon-calon yang tergolong memiliki kemampuan tinggi sajalah yang dimungkinkan dapat menjawab butir-butir soal tes dengan betul. Tes seleksi dapat dilaksanakan secara lisan, secara tertulis, dengan tes perbuatan, dan dapat pula dilaksanakan dengan mengkombinasikan ketiga jenis tes tersebut secara

- 2)rempak.
- b) **Tes formatif (formative test)**, yaitu test yang dilaksanakan setelah selesainya satu pokok bahasan. Test ini berfungsi untuk menentukan 2)ntas tidaknya satu pokok bahasan.
 - c) **Tes sumatif (summative test)**, yaitu test yang diberikan setelah sekumpulan satuan program pembelajaran selesai diberikan. Disekolah test ini 2)kenal sebagai ulangan umum.
 - d) **Test diagnostik (Diagnostic test)**, yaitu test yang dilakukan untuk menentukan secara tepat, jenis kesulitan yang dihadapi oleh peserta didik dalam suatu mata 2)elajaran tertentu.
 - e) **Pra test (pre test)**, yaitu test yang diberikan sebelum proses pembelajaran. Test ini bertujuan untuk mengetahui sejauh manakah materi yang akan diajarkan telah dapat dikuasai oleh peserta didik.

- f) **Test akhir (Post test)** yaitu test yang diberikan setelah dilaksanakan proses pembelajaran. Tes tersebut bertujuan untuk mengetahui tingkat kemajuan intelektual (tingkat penguasaan materi) peserta didik

E. Tes Kebahasaan

Tes kebahasaan merupakan bagian dari kegiatan pengajaran bahasa secara keseluruhan. Kegiatan tes sangat diperlukan dalam pengajaran bahasa karena berdasarkan informasi tes itulah dapat dilakukan penilaian secara objektif, khususnya terhadap hasil belajar bahasa siswa. Informasi tentang hasil belajar siswa tersebut, pada giliran selanjutnya, juga dapat dimanfaatkan sebagai masukan untuk meningkatkan pengajaran bahasa selanjutnya.

Berdasarkan kriteria bagaimana bahasa dikaji dan ditelaah, maka tes dikembangkan berdasarkan pandangan yang berbeda dalam memahami hakikat bahasa. Dari latar belakang pendekatan bahasa, jenis tes bahasa dapat dikelompokkan menjadi:

a. Tes Diskrit

Tes diskrit adalah tes yang hanya menekankan atau menyangkut satu aspek kebahasaan pada satu waktu. Tiap satu butir soal hanya dimaksudkan untuk mengukur satu aspek kebahasaan, misalnya fonologi, morfologi, sintaksis, atau kosa kata.

b. Tes Integratif

Tes integrative adalah suatu tes kebahasaan yang berusaha mengukur beberapa aspek kebahasaan atau keterampilan berbahasa pada satu waktu.

c. Tes Pragmatik

Dalam tes pragmatik tidak lagi ditemui adanya tes struktur atau kosa kata atau unsur-unsur kebahasaan yang lain secara sendiri, melainkan semua unsure kebahasaan dan bahkan langsung dikaitkan dengan unsur ekstralinguistik sekaligus.

d. Tes Komunikatif

Tes bahasa komunikatif ialah tes yang biasanya tidak digunakan untuk mengukur kemampuan gramatikal, yang lebih menitikberatkan pada komunikasi. Tes yang dimaksud untuk memberi tugas kepada peserta tes melakukan kegiatan dengan kemampuan bahasa tertentu, dengan kaitan yang jelas dengan konteks nyata termasuk kemampuan komunikatif, tes komunikatif perlu dikembangkan.

F. Tes Keastraan

Pengajaran sastra (Indonesia) di sekolah tidak berdiri sendiri sebagai sebuah mata pelajaran yang mandiri, melainkan "hanya" menjadi bagian mata pelajaran bahasa Indonesia. Dengan demikian, seorang guru bahasa Indonesia juga berarti guru apresiasi sastra. Ia bertugas mengukur hasil belajar bahasa dan sastra siswa yang menjadi asuhannya. Hal ini juga berarti ia dituntut untuk mampu

menyusun tes kebahasaan dan kesastraan sebagai salah satu sarana mengungkap hasil belajar siswa.

Penggabungan sastra ke dalam pengajaran bahasa (Indonesia) memang wajar dan dapat dimengerti. Untuk memahami karya sastra yang merupakan salah satu atau langkah dalam usaha mengapresiasi karya sastra, penguasaan terhadap bahasa yang bersangkutan merupakan karya sastra, penguasaan terhadap bahasa yang bersangkutan merupakan suatu hal yang tak dapat ditawar.

Idealnya terjadi kaitan yang erat antara pengajaran bahasa dengan pengajaran sastra yang bersifat saling mengisi dan menunjang. Dengan demikian, terdapat korelasi antara kemampuan berbahasa dengan kemampuan berapresiasi sastra. Kemampuan berbahasa yang tinggi yang dimiliki seorang siswa akan menjadi petunjuk bahwa ia juga tinggi kemampuan apresiasinya.

4

NORMA DAN SKALA PENILAIAN

A. Pengertian Penilaian Acuan Norma (PAN)

Penilaian acuan norma adalah penilaian yang dilakukan dengan mengacu pada norma kelompok. Hal ini berarti nilai-nilai yang diperoleh siswa dibandingkan dengan nilai-nilai siswa yang lain yang termasuk didalam kelompok tersebut. Dalam hal ini yang dimaksud dengan norma adalah kapasitas atau prestasi kelompok, sedangkan yang dimaksud dengan kelompok adalah semua siswa yang mengikuti tes tersebut. Dapat disimpulkan bahwa kata kelompok yang dimaksud adalah sejumlah siswa dalam satu kelas, sekolah, rayon dan provinsi atau wilayah.

Menafsirkan skor tes dengan acuan norma berarti memberikan gambaran kepada kita bagaimana seorang siswa bila dibandingkan dengan siswa lain yang mengambil tes yang sama. Cara membandingkan yang paling sederhana dan paling banyak digunakan dalam penilaian adalah mengurutkan skor dari yang tertinggi ke terendah. Kemudian menentukan dimana posisi siswa berada. Hal penting yang perlu diperhatikan saat

menggunakan acuan norma adalah kelompok acuan harus menggambarkan atau mewakili keseluruhan siswa yang mengambil tes. Bila kita ingin membandingkan kemampuan siswa dalam tes, maka perlu melihat siswa lain yang memiliki umur, kelas dan latar belakang pendidikan yang sama.

B. Tujuan dan Manfaat Penilaian Acuan Norma (PAN)

- Tujuan dari penilaian acuan norma adalah untuk membedakan peserta didik atas kelompok-kelompok berdasarkan tingkat kemampuan dari yang terendah sampai ke tertinggi. Secara ideal, pendistribusian tingkat kemampuan dalam satu kelompok menggambarkan suatu kurva normal.
- Penilaian acuan norma pada umumnya digunakan untuk seleksi. Penilaian acuan norma biasanya digunakan pada akhir unit pembelajaran untuk menentukan tingkat hasil peserta didik, peringkat dan klasifikasi peserta didik.

C. Kriteria Penyusunan Penilaian Acuan Norma (PAN)

- a) Soal-soal penilaian acuan norma tidak didasarkan atas pengajaran yang diterima siswa atau keterampilan serta tingkah laku yang diidentifikasi sebagai sesuatu yang dianggap relevan bagi belajar siswa.
- b) Pada proses belajar, penilaian norma pada umumnya banyak dilakukan oleh guru.

- c) Penekanan pada penilaian untuk proses belajar, seseorang mengacu pada ketentuan atau norma yang berlaku di sekolah.
- d) Seorang guru dapat menggunakan acuan norma nasional.

D. Ciri-ciri Penilaian Acuan Norma (PAN)

- a) Penilaian acuan norma digunakan untuk mengetahui kemampuan peserta didik dalam komunitasnya, seperti di kelas, sekolah dan sebagainya.
- b) Penilaian acuan norma menggunakan kriteria yang bersifat relatif.
- c) Nilai hasil dari penilaian acuan norma digunakan untuk menunjukkan kedudukan peserta didik (peringkat) dalam komunitasnya (kelompok).
- d) Penilaian norma memiliki kecenderungan untuk menggunakan rentangan tingkat penguasaan seseorang terhadap kelompoknya, mulai dari yang sangat istimewa sampai dengan yang mengalami kesulitan serius.

E. Pengertian Penilaian Acuan Patokan (PAP)

Pendekatan penilaian acuan patokan lebih menitikberatkan pada apa yang dapat dilakukan peserta didik. Penilaian acuan patokan meneliti apa yang dapat dikerjakan oleh peserta didik, bukan membandingkan seorang peserta didik dengan teman sekelasnya, melainkan dengan suatu kriteria atau patokan spesifik. Contohnya kriteria yang digunakan 70% atau 80%. Pendekatan penilaian acuan patokan ini dalam melakukan penilaiannya mengacu pada

suatu kriteria pencapaian tujuan pembelajaran atau indikator pembelajaran yang telah dirumuskan sebelumnya. Nilai-nilai yang diperoleh siswa dihubungkan dengan tingkat pencapaian penguasaan siswa terhadap materi pembelajaran sesuai dengan indikator pembelajaran atau tujuan pembelajaran yang telah ditetapkan.

Penilaian acuan patokan ini disebut juga dengan penilaian acuan kriteria. Penilaian acuan ini berupaya mendeskripsikan apa yang telah diketahui atau apa yang dapat dilakukan oleh peserta didik. Hasil penilaian dengan menggunakan acuan kriteria dapat berupa tercapainya atau tidak tercapainya, tuntas atau tidak tuntas, dan lulus atau tidak lulus. Penilaian acuan patokan memiliki asumsi bahwa semua orang bisa belajar apa saja namun memerlukan jumlah waktu yang berbeda. Dengan demikian, variabel pada acuan ini adalah waktu. Sebagai konsekuensi, maka muncul adanya program remedial atau pengayaan dan juga program akselerasi atau percepatan.

F. Tujuan dan Manfaat Penilaian Acuan Patokan (PAP)

Tujuan dari pendekatan penilaian acuan patokan adalah untuk mengukur secara pasti tujuan atau kompetensi yang ditetapkan sebagai kriteria keberhasilannya. Manfaat dari penilaian acuan patokan ini adalah untuk meningkatkan kualitas hasil belajar, sebab peserta didik diusahakan untuk mencapai standar yang telah ditentukan, dan hasil belajar peserta didik dapat diketahui derajat

pencapaiannya. Untuk menentukan batas lulus dengan pendekatan ini, setiap skor peserta didik dibandingkan dengan skor ideal yang mungkin dicapai oleh peserta didik.

G. Penerapan Penilaian Acuan Patokan (PAP)

Pendekatan penilaian acuan patokan tidak berorientasi pada “apa adanya”. Pertama, pendekatan ini mempergunakan angka rata-rata dengan terlebih dahulu menetapkan kriteria keberhasilan, yaitu “batas lulus” penguasaan bahan pelajaran atau kriteria pencapaian tujuan (TKP). Kedua, dalam proses pengajaran, tenaga pengajar tidak begitu saja membiarkan siswa menjalani sendiri proses belajarnya, melainkan terus menerus secara langsung ataupun tidak langsung merangsang dan memeriksa kemajuan belajar siswa serta membantunya melewati tahap-tahap pengajaran secara berhasil.

10

H. Kelebihan dan Kekurangan Penilaian Acuan Norma dan Penilaian Acuan Patokan

Kelebihan penilaian acuan norma adalah sebagai berikut:

- a) Dapat digunakan untuk menetapkan nilai secara maksimal.
- b) Dapat membedakan kemampuan peserta didik yang pintar dan kurang pintar. Membedakan kelompok atas dan bawah.
- c) Fleksibel: dapat menyesuaikan dengan kondisi yang berbeda-beda.
- d) Mudah menilai karena tidak ada patokan.

- e) Dapat digunakan untuk menilai ranah kognitif, afektif dan psikomotor.

Adapun kelebihan penilaian acuan patokan adalah sebagai berikut:

- a) Dapat membantu guru merancang program remidi.
- b) Tidak membutuhkan perhitungan statistik yang rumit.
- c) Dapat mengukur ketercapaian tujuan pembelajaran.
- d) Nilainya bersifat tetap selama standar yang digunakan sama.
- e) Hasil penilaian dapat digunakan untuk umpan balik atau untuk mengetahui apakah tujuan pembelajaran sudah tercapai atau belum.
- f) Banyak digunakan untuk kelas dengan materi pembelajaran berupa konsep.
- g) Mudah menilai karena ada patokan.

Selain memiliki kelebihan, kedua pendekatan tersebut memiliki kekurangan. Diantara kekurangan penilaian acuan norma adalah sebagai berikut.

- a) Sedikit menyebutkan kompetensi siswa apa yang mereka ketahui atau dapat mereka lakukan.
- b) Tidak fair karena peringkat siswa tidak hanya bergantung pada tingkat prestasi, tetapi juga atas prestasi siswa lain.
- c) Tidak dapat diandalkan siswa yang gagal sekarang mungkin dapat lulus tahun berikutnya.

Selanjutnya ada pula kekurangan dari penilaian acuan patokan yaitu:

- a) Sulit mengembangkan suatu kriteria, kriteria sulit dilakukan karena perlu waktu untuk menyetujui sebuah kriteria dan standar.
- b) Guru sulit melakukan perbandingan antar siswa dan lebih menekankan hasil dari pada proses karena dengan adanya patokan siswa bisa melakukan apapun untuk mencapai patokan tersebut. Contohnya seperti mencontek teman yang pintar.

10

I. Persamaan dan Perbedaan Penilaian Acuan Norma dan Penilaian Acuan Patokan

Penilaian acuan norma dan penilaian acuan patokan mempunyai beberapa persamaan sebagai berikut:

- a) Penilaian acuan norma dan penilaian acuan patokan memerlukan adanya tujuan evaluasi spesifik sebagai penentuan fokus item yang diperlukan.
- b) Kedua pengukuran memerlukan sample yang relevan, digunakan sebagai subjek yang hendak dijadikan sasaran evaluasi. Sample yang diukur mempresentasikan populasi siswa yang hendak menjadi target akhir pengambilan keputusan.
- c) Untuk mendapatkan informasi yang diinginkan tentang siswa, kedua pengukuran sama-sama memerlukan item-item yang disusun dalam satu tes dengan menggunakan aturan dasar penulisan instrument.
- d) Keduanya mempersyaratkan perumusan secara spesifik perilaku yang akan diukur.

- e) Keduanya menggunakan macam tes dan instrument yang sama.
- f) Keduanya dinilai kualitasnya dari segi validitas dan reliabilitasnya.
- g) Keduanya digunakan dalam pendidikan walaupun untuk maksud berbeda.

Perbedaan kedua penilaian adalah sebagai berikut:

- a) Penilaian acuan norma biasanya mengukur sejumlah besar perilaku khusus dengan sedikit butir tes untuk setiap perilaku. Penilaian acuan patokan biasanya mengukur perilaku khusus dalam jumlah yang terbatas dengan banyak butir tes untuk setiap perilaku.
- b) Penilaian acuan norma menekankan perbedaan diantara peserta tes dari segi tingkat pencapaian belajar secara relatif. Penilaian acuan patokan menekankan penjelasan tentang apa perilaku yang dapat dan tidak dapat dilakukan oleh setiap peserta tes.
- c) Penilaian acuan norma lebih mementingkan butir-butir tes yang mempunyai tingkat kesulitan sedang dan biasanya membuang tes yang terlalu mudah dan terlalu sulit. Penilaian acuan patokan mementingkan butir-butir tes yang relevan dengan perilaku yang akan dikur tanpa peduli dengan tingkat kesulitannya.
- d) Penilaian acuan norma digunakan terutama untuk survey. Penilaian acuan patokan digunakan terutama untuk penguasaan.

5

VALIDITAS DAN RELIABILITAS

A. Pengertian Validitas

Azwar (1987: 173) menyatakan bahwa validitas berasal dari kata validity yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu instrumen pengukur (tes) dalam melakukan fungsi ukurnya (Zhang et al., 2014). Suatu tes dikatakan memiliki validitas yang tinggi apabila alat tersebut melaksanakan fungsi ukur secara benar atau memberikan hasil ukur yang sesuai dengan maksud dilakukannya pengukuran tersebut. Artinya hasil ukur dari pengukuran tersebut menjadikan kuantitas yang mencerminkan secara benar fakta atau keadaan sesungguhnya dari apa yang diukur.

Suryabrata (2000: 41) menyatakan bahwa validitas tes pada dasarnya menunjuk kepada derajat fungsi pengukurannya suatu tes, atau derajat kecermatan ukurnya sesuatu tes. Sudjana (2004: 12) menyatakan bahwa validitas berkenaan dengan ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai. Di dalam buku Evaluasi Pendidikan (200:26)

27 menyatakan bahwa suatu instrumen evaluasi dikatakan valid, apabila instrumen yang digunakan dapat mengukur apa yang hendak diukur. Jadi, jika tes tersebut ialah tes pencapaian hasil belajar maka hasil tes tersebut apabila diinterpretasi secara intensif, hasil yang dicapai memang benar menunjukkan ranag evaluasi pencapaian belajar.

1. Konsep Validitas

Menurut (Zhang et al., 2014) konsep Validalitas Tes dibedakan menjadi tiga macam, yaitu :

a. Validitas Isi (*content validity*)

Validitas isi suatu tes mempermasalahkan seberapa jauh suatu tes mengukur tingkat penguasaan terhadap isi atau konten atau materi tertentu yang seharusnya dikuasai sesuai dengan tujuan pengajaran. Dengan kata lain tes yang mempunyai validitas isi yang baik ialah tes yang benar-benar mengukur penguasaan materi yang seharusnya dikuasai sesuai dengan konten pengajaran yang tercantum dalam Garis-garis Besar Program Pengajaran (GBPP). Menurut Wahidmurni menyatakan bahwa dalam praktik validitas isi, ada beberapa tahap yang bisa dilakukan untuk mengevaluasi validitas isi sebuah tes:

- Mengidentifikasi dan membuat sebuah daftar wilayah atau domain yang menjadi tujuan ukur
- Mengumpulkan ahli dalam domain yang dimaksud

- Mengembangkan metodologi pencocokan yang konsisten
 - Melakukan analisa dari prosedur pencocokan
- b. Validitas Kriteria (*criterion-related validity*)

Menurut Wahidmurni (2010: 88) menyatakan bahwa bukti validnya hasil ukur berdasarkan suatu kriteria tertentu disebut sebagai validitas kriteria. Kriteria merupakan variabel yang menjadi perhatian utama. Intinya ialah pada pemilihan variabel kriteria dan prosedur pengukurannya. Secara logis, nilai dari studi validitas kriteria tergantung pada relevan tidaknya pemakaian kriteria tertentu. Oleh sebab itu validitas kriteria didekati dengan metode statistik, tepatnya metode korelasi antara skor hasil tes dengan skor suatu kriteria. Dua tipe validitas kriteria yang biasanya digunakan yakni:

- c. Validitas *prediktif*

Validitas prediktif merupakan hal yang mengacu pada sejauh mana efektifitas sebuah skor tes mampu memprediksi kinerja individu pada sebuah kriteria tertentu pada masa yang akan datang. Jadi pengukuran kriteria dilakukan belakangan. Validitas produktif sebuah skor tes hanya dapat diketahui setelah jangka waktu tertentu dari waktu tes tersebut diadminitrasikan.

- d. Validitas *Konkuren*

Kriteria validitas sebuah skor tes dapat dipandang sebagai sejauh mana sebuah skor tes

hasil ukur konstruk tertentu memiliki korelasi dengan skor pada skor tes hasil ukur konstruk lain yang secara teoritik seharusnya berkorelasi. Korelasi ini mengindikasikan tingkat kesesuaian hasil ukur sebuah konstruk dengan kriteria sebuah konstruk lain yang dimaksud.

e. *Validitas Konstruk (construct validity)*

Validitas konstruk (*construct validity*) adalah validitas yang mempermasalahkan seberapa jauh butir-butir tes mampu mengukur apa yang benar-benar hendak diukur sesuai dengan konsep khusus atau definisi konseptual yang telah ditetapkan. Validitas konstruk biasa digunakan untuk instrumen yang dimaksudkan mengukur variabel konsep, baik yang sifatnya performansi tipikal seperti instrumen untuk mengukur sikap, minat konsep diri, locus kontrol, gaya kepemimpinan, motivasi berprestasi, dan lain-lain, maupun yang sifatnya performansi maksimum seperti instrumen untuk mengukur bakat (tes bakat), inteligansi (kecerdasan intelektual), kecerdasan, emosional dan lain-lain. Validitas konstruk dapat dibedakan menjadi dua bentuk:

f. *Validitas Divergen dan Konvergen*

Validitas divergen berarti rendahnya korelasi antar pengukuran pada atribut-atribut yang berhubungan namun membentuk konstruk yang berbeda Sedangkan validitas konvergen ditunjukkan dengan tingginya korelasi antar pengukuran yang berbeda namun

mengukur konstruk yang sama. validasi konstruk umumnya merupakan sebuah pengukuran yang dilakukan melalui serangkaian teknik-teknik tertentu (Fogarty, 2006), namun hal itu dapat diringkas menjadi proses yang terdiri dari tahap-tahap berikut (Crocker & Algina, 1986).

- Menetapkan satu atau lebih hipotesis tentang bagaimanakah konstruk seharusnya berkorelasi dengan konstruk lain serta bagaimanakah arah perbedaan antar kelompok yang relevan.
- Memilih atau mengembangkan instrumen pengukuran yang terdiri dari butir-butir yang merepresentasikan perilaku-perilaku tertentu sebagai manifestasi dari konstruk yang dimaksud.
- Melakukan studi empirik guna menyelidiki hubungan atau perbedaan yang ditetapkan sebagai hipotesis-hipotesis.
- Menentukan apakah data pengukuran konsisten dengan hipotesis-hipotesis yang ditetapkan kemudian melakukan evaluasi sejauh mana temuan berkenaan dengan variabel-variabel tersebut dapat dijelaskan oleh hipotesis lain.

2. Confirmatory Factor Analysis

Terdapat satu teknik yang akhir-akhir ini lebih ditekankan dalam validasi konstruk, yakni analisis faktor. Analisis faktor hanya merupakan

salah upaya menegakkan bukti empiris validitas struktur yaitu mengungkap struktur hubungan antar dimensi yang diukur oleh tes. Jadi analisis faktor merupakan salah satu prosedur guna mendapatkan bukti empiris mengenai struktur data skor hasil pengukuran. Harapannya, struktur data skor ini akan mengikuti model yang disodorkan pengembang tes, tentu dengan dasar teoritik yang memadai. Ada dua jenis analisis faktor yaitu *exploratory factor analysis* (EFA) dan *confirmatory factor analysis* (CFA).

Faktor analisis didesain untuk menjelaskan hubungan sejumlah variabel amatan (*observed variables*) dengan sejumlah variabel laten tak teramati (*unobserved variables*). Dalam pengembangan tes baru variabel laten ini merupakan konstruk hipotetis yang disuguhkan pengembang tes. Secara teknis, analisis faktor dapat dipandang sebagai varian dari model regresi. Perbedaannya, pada analisis faktor tidak semua variabel diukur secara langsung. Persamaan regresi standardized yang dituangkan dalam bentuk matriks:

$$Y = BX + E$$

di mana:

Y = vektor matriks skor variabel dependen (1 x m)

B = matriks koefisien regresi standardized (m x b)

X = vektor matriks skor variabel independen (b x 1)

E = vektor matriks skor error (1 x m)

Sedangkan dalam analisis faktor:

$$Y = Pf + U$$

Di mana:

Y = vektor matriks skor variabel amatan ($1 \times m$)

P = matriks koefisien regresi standardized ($m \times b$)

F = vektor matriks skor variabel laten ($b \times 1$)

U = vektor matriks skor eror ($1 \times m$)

B. Pengertian Reliabilitas

Kata reliabilitas dalam bahasa Indonesia diambil dari kata reliability dalam bahasa Inggris, berasal dari kata asal reliable yang artinya dapat dipercaya. Menurut Nur (1987:47) menyatakan bahwa reliabilitas ukuran menyangkut seberapa jauh skor deviasi individu, atau skor-z, relatif konsisten apabila dilakukan pengulangan pengadministrasian dengan tes yang sama atau tes yang ekuivalen. Azwar (2003: 176) menyatakan bahwa reliabilitas merupakan salah-satu ciri atau karakter utama instrumen pengukuran yang baik. Arifin (1991: 122) menyatakan bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda (Zhang et al., 2014).

Reliabilitas dalam bahasa Indonesia diambil dari kata reliability dalam bahasa Inggris, berasal dari kata reliable yang artinya dapat dipercaya. Instrumen dikatakan dapat dipercaya jika digunakan berkali-kali pada responden yang sama memberikan hasil yang tetap atau ajek (konsisten).

Nur (1987: 47) menyatakan bahwa reliabilitas ukuran menyangkut seberapa jauh skor deviasi individu, atau skor-z, relatif konsisten apabila

dilakukan pengulangan pengadministrasian dengan tes yang ekivalen.

Menurut Azwar (2003: 176) reliabilitas adalah salah satu ciri atau karakter utama instrumen pengukuran yang baik. sedangkan menurut Arifin (1991: 122) menyatakan bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda.

Berdasarkan beragam makna tersebut, dalam bidang pengukuran ada aneka ragam istilah untuk menunjuk pada istilah konsistensi, keajekan, ketetapan, kestabilan dan keandalan. Instrumen yang reliabel belum tentu valid. Meteran yang putus dibagian ujungnya, bila digunakan berkali-kali akan menghasilkan data yang sama (reliabel) tetapi selalu tidak valid. hal ini disebabkan karena instrumen (meteran) tersebut rusak. Reliabilitas merupakan syarat untuk pengujian validitas instrumen. Oleh karena itu walaupun instrumen yang valid pada umumnya pasti reliabel, tetapi pengujian reliabilitas instrumen perlu dilakukan.

Berdasarkan cara-cara melakukan pengujian tingkat reliabilitas instrumen, dibagi menjadi dua, diantaranya:

1. Reliabilitas Eksternal (Eksternal Reliability)

Ada dua cara untuk menguji realibilitas eksternal:

a. Metode bentuk paralel (equivalent method)

Metode paralel dilakukan dengan cara menyusun dua instrumen yang hampir sama (equivalent), kemudian diujicobakan pada

sekelompok responden yang sama (responden mengerjakan dua kali) kemudian hasil uji coba tersebut dikorelasikan dengan teknik korelasi product moment. data dari hasil uji coba, yang satu dianggap sebagai nilai X, sedangkan yang lainnya dianggap nilai Y.

b. Metode tes berulang (test-retest method)

Metode ini dilakukan untuk menghindari penyusunan instrumen dua kali. dengan menggunakan metode ini kita hanya menyusun satu perangkat instrumen. instrumen tersebut diujicobakan pada sekelompok responden. hasilnya dicatat. kemudian kedua hasil tersebut dikorelasikan. perhitungan dan penafsiran hasil korelasi menggunakan aturan yang sama dengan metode paralel.

2. Reliabilitas Internal (Internal Reliability)

Reliabilitas internal diperoleh dengan cara menganalisis data dari satu kali pengumpulan data. Berdasarkan sistem pemberian skor instrumen, ada dua metode analisis reliabilitas internal.

a. Instrumen skor diskrit

Instrumen skor diskrit adalah instrumen yang skor jawaban/ responnya hanya dua, yaitu 1 (satu) dan 0 (nol). dengan kata lain hanya dua jawaban yaitu benar dan salah. Untuk instrumen yang skornya diskrit (1 dan 0) tingkat reliabilitasnya dapat dicari dengan menggunakan (1) metode belah dua, (2) rumus Flanagan, (3) Rumus Rulon, (4)

rumus K - R. 20, (5) rumus K - R. 21, (6) rumus Hoyt.

b. Instrumen skor non diskrit

Instrumen skor non diskrit adalah instrumen pengukuran yang dalam sistem skoringnya bukan 1 dan 0, tetapi bersifat gradual, yaitu ada penjenjangan skor, mulai dari skor tertinggi sampai skor terendah. Hal ini biasanya terdapat pada instrumen tes bentuk uraian dan angket dengan skala Likert maupun skala bertingkat. Untuk instrumen skor non diskrit ini analisis reliabilitasnya menggunakan rumus alpha.

Rumus alpha

$$r_{11} = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$
$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

Keterangan:

r_{11} = reliabilitas instrumen

K = banyaknya butir pertanyaan

$\sum \sigma_b^2$ = jumlah varians butir

σ_t^2 = varians total

X = skor total

6

PENGOLAHAN HASIL TES DAN NON-TES

4

A. Pengertian Tes dan Non-Tes

Secara bahasa tes berasal dari bahasa Perancis Kuno: *testum* yang memiliki arti “piring untuk menyisihkan logam – logam mulia” (yang berarti menggunakan alat berupa piring itu akan dapat diperoleh jenis – jenis logam mulia yang nilainya sangat tinggi). Sedangkan dalam bahasa Inggris ditulis dengan kata *test* yang memiliki arti “ujian” atau “percobaan”. Menurut istilah *test* merupakan alat yang digunakan dalam rangka suatu pengukuran dan penilaian; sedangkan *testing* berarti saat dilaksanakannya atau peristiwa berlangsungnya pengukuran dan penilaian; lalu *tester* memiliki arti orang yang melakukan tes tersebut.

Pengertian Non tes adalah suatu cara penilaian hasil belajar peserta didik yang dilakukan tanpa menguji peserta didik tetapi, dengan cara melakukan pengamatan secara sistematis. Tujuan non tes yaitu untuk menilai kepribadian anak secara menyeluruh yang meliputi, sikap, tingkah laku, sifat, dan sikap

sosial. Penilaian ini bisa dilakukan secara individu maupun kelompok.

Dengan demikian, dapat disimpulkan bahwa tes adalah suatu alat pengumpul informasi yang bersifat lebih resmi bila dibandingkan dengan alat informasi yang lainnya karena penuh batasan – batasan. Tes ini dilakukan dengan pemberian tugas, ataupun ujian yang diberikan guru. Sedangkan non tes adalah cara penilaian peserta didik tanpa melakukan tes. Cara ini dilakukan dengan melakukan pengamatan (observasi), wawancara (interview), dan melalui angket (kuisisioner).

B. Pengolahan Hasil Tes

Menurut Zainal Arifin (2009) dalam mengolah data hasil tes terdapat 4 pokok yang harus dilakukan yaitu:

1. Menskor yaitu memberi nilai atau skor pada hasil tes yang diperoleh oleh peserta didik. Untuk memperoleh skor mentah diperlukan tiga jenis alat bantu yaitu kunci jawaban, kunci skoring dan pedoman konversi.
2. Mengubah skor mentah menjadi skor standart sesuai dengan norma tertentu.
3. Mengkonversikan skor standart ke dalam nilai, baik berupa huruf maupun angka.
4. Melakukan analisis soal jika diperlukan untuk mengetahui drajat validitas dan reabilitas soal, tingkat kesukaran soal, dan daya pembeda.

Setelah melaksanakan kegiatan tes dan lembar pekerjaan peserta didik telah diperiksa kebenarannya, kesalahan, dan kelengkapannya.

Langkah selanjutnya yaitu menghitung skor mentah untuk setiap peserta didik berdasarkan rumus tertentu dan bobot setiap soal. Kegiatan ini harus dilakukan dengan ekstra hati - hati karena kegiatan ini merupakan dasar dalam melakukan pengolahan hasil tes sampai menjadi prestasi. Sebelum tes dilakukan, guru harus sudah menyusun pedoman pemberian skor terutama untuk bentuk soal esai (Zainal Arifin, 209:223).

Pada pengolahan hasil tes menurut (Skor et al., n.d.) ada beberapa cara antara lain: a. Pemberian Skor

a) Skor tes uraian

Ditentukan oleh guru atau tim penilai yang didasarkan pada tingkat kesulitan dan kemudahan suatu soal. Dalam satu jenis soal, bisa memiliki skor yang sama, karena bobot kesulitan atau kemudahannya relatif sama, dan bisa juga memiliki skor yang berbeda dari masing-masing soal.

Misalnya:

Nomor 1 sampai 5 rentang skor antara 0 sampai 5

Nomor 6 sampai 10 rentang skor 0 sampai 10

Nomor 11 sampai 15 rentang skor 0 sampai 15

Pada tes uraian, jawaban benar dapat dianggap skor bersih.

$Sk = \Sigma B$ (Skor sama dengan jumlah benar)

b) Skor tes pilihan ganda

Jumlah jawaban benar dari tes pilihan ganda belum dianggap sebagai skor bersih. Hal ini

disebabkan pada pertimbangan ada kemungkinan siswa menjawab dengan menebak saja, sehingga perlu dikurangi dengan perhitungan tertentu. Pada bentuk menjodohkan, bisa disamakan dengan bentuk jawaban uraian yakni:

$Sk = \Sigma B$ (Skor sama dengan jumlah benar)
atau bisa juga dikenakan seperti pada bentuk benar salah.

$Sk = \Sigma B - \Sigma S$ (Skor sama dengan jumlah benar dikurangi jumlah salah.) Sedangkan pada bentuk pilihan ganda: $Sk = \Sigma B - \Sigma S / o - 1$

$o - 1$ adalah bilangan pembagi yang diperoleh dari jumlah option (pilihan) - 1

a. Pengubahan Skor Menjadi Nilai

1) Skala Penilaian dengan dasar proporsi jawaban benar.

i. Skala nilai 10

$$\text{Nilai} = \frac{\text{Skor tercapai} \times 10}{\text{Skor ideal}}$$

ii. Skala nilai 100

$$\text{Nilai} = \frac{\text{Skor tercapai} \times 100}{\text{Skor ideal}}$$

Apabila dikonfersikan ke dalam nilai A-E

A = 91 sampai 100

B = 81 sampai 90

C = 71 sampai 80

D = 61 sampai 70 E = 0 sampai 60 atau

A = 86 sampai 100

B = 76 sampai 86

C = 65 sampai 75

D= 50 sampai 65

E = 0 sampai 49

Adapun cara lain dalam pengolahan hasil tes menurut (Ramadhani et al., 2021) adalah melalui PAP atau Penilaian Acuan Patokan dan PAN atau Penilaian Acuan Norma. Berikut ini penjelasannya:

b. Penilaian Acuan Patokan (PAP)

1) Pengolahan Skor Mentah Menjadi Nilai Huruf

Terdapat 60 item soal pilihan ganda pelajaran bahasa Arab, tiap item yang benar berbobot 1. Skor mentah yang diperoleh 20 siswa adalah 32, 36, 27, 50, 22, 34, 35, 37, 43, 17, 21, 42, 46, 32, 31, 28, 57, 57, 54, 51.

Prosedur yang dapat ditempuh adalah sebagai berikut:

a) skor ideal, yaitu skor yang mungkin dicapai jika semua item dapat dijawab dengan benar. Skor ideal diperoleh dengan jalan menghitung jumlah item yang diberikan serta bobot dari tiap-tiap item.

Dari contoh diatas diketahui skor idealnya adalah 60

b) Mencari rata-rata ideal (id) dengan rumus:

$$= \frac{1}{2} \times \text{skor ideal} = \frac{1}{2} \times 60 = 30$$

c) Mencari deviasi (SD) ideal dengan cara:

$$SD = \frac{1}{3} \times \text{SD} = \frac{1}{3} \times 30 = 10$$

d) Menyusun kebutuhan konversi sesuai dengan yang dibutuhkan.

Adapun pedoman konversi dengan adalah:

$$+ 1,5 (SD) = 30 + 1,5 \times 10 = 45 = A$$

- + 0,5 (SD) = $30 + 0,5 \times 10 = 35 = B$
- 0,5 (SD) = $30 - 0,5 \times 10 = 25 = C$
- 1,5 (SD) = $30 - 1,5 \times 10 = 15 = D$

Dari data tersebut dapat kita simpulkan bahwa siswa yang mendapat skor 45 – 60 mendapat nilai A, 35 – 44 = B, 25 – 34 = C, 15 – 24 = D, 0 – 14 = E.

Pemberian nilai dengan menggunakan huruf disesuaikan dengan huruf yang terdapat dalam urutan abjad. Huruf tidak hanya menunjukkan kuantitas, tetapi dapat juga digunakan sebagai simbol untuk menggambarkan kualitas.

Skor Angka	Nilai Huruf	Predikat
50	A	Sangat Baik
37	B	Baik
33	C	Cukup
22	D	Kurang
5	E	Sangat Kurang

- e) Skor mentah menjadi skor standar 1 – 10
 Untuk mengubah skor mentah menjadi skor terjabar dalam skala 1 – 10 dapat digunakan ketentuan-ketentuan berikut:
- + 2,25 (SD) = 10 = $30 + 2,25 \times 10 = 53 = 10$
 - + 1,75 (SD) = 9 = $30 + 1,75 \times 10 = 48 = 9$
 - + 1,25 (SD) = 8 = $30 + 1,25 \times 10 = 43 = 8$
 - + 0,75 (SD) = 7 = $30 + 0,75 \times 10 = 38 = 7$
 - + 0,25 (SD) = 6 = $30 + 0,25 \times 10 = 33 = 6$
 - 0,25 (SD) = 5 = $30 - 0,25 \times 10 = 28 = 5$

- 0,75 (SD) = 4 = 30 - 0,75 x 10 = 23 = 4
- 1,25 (SD) = 3 = 30 - 1,25 x 10 = 18 = 3
- 1,75 (SD) = 2 = 30 - 1,75 x 10 = 13 = 2
- 2,25 (SD) = 1 = 30 - 2,25 x 10 = 8 = 1

Dengan demikian dapat diambil kesimpulan bahwa siswa yang mendapat skor 53

- 60 mendapat nilai 10, 48 - 52 = 9, 43 - 47 = 8, 38 - 42 = 7, 33 - 37 = 6, 28 - 32 = 5, 23 - 27 = 4, 18 - 22 = 3, 13 - 17 = 2, 8 - 12 = 1, dan skor dibawahnya 0. Bila kita ingin agar skala tersebut lebih halus yakni ada nilai diantara nilai-nilai tersebut, seperti 9,5; 8,5; 7,5 dan seterusnya, kita bisa memperkecil jarak antar skala-skala itu. Diantara 2,25 (SD) dan 1,75 (SD) dapat ditempatkan 2,00 (SD) yang ekuivalen dengan nilai 9,5. Diantara 1,75 (SD) dan 1,25 (SD) dapat ditempatkan 1,50 (SD) yang ekuivalen dengan nilai 8,5 dan begitu seterusnya.

c. Penilaian Acuan Norma (PAN)

Penilaian acuan norma menskor peserta didik dengan membandingkan hasil belajar satu peserta dengan hasil peserta lainnya dalam satu kelompok kelas. Contoh diketahui 20 siswa mengikuti ujian akhir semester mata pelajaran bahasa Arab memperoleh skor mentah sebagai berikut:

32, 36, 27, 50, 22, 34, 35, 37, 43, 17, 21, 42, 46, 32, 31, 28, 57, 57, 54, 51.

Penyelesaian nilai peserta didik dengan pendekatan PAN:

- 1) Menyusun skor terkecil hingga terbesar

17, 21, 22, 27, 28, 31, 32, 32, 34, 35, 36, 37, 42, 43, 46, 50, 51, 54, 57, 57.

2) Mencari rentangan (range) yaitu skor terbesar dikurangi skor terkecil $57 - 17 = 40$

3) Mencari banyak kelas interval

4) Banyak kelas $= 1 + (3,3) \log n = 1 + (3,3) \log 20$

$$= 1 + (3,3) (1,3010)$$

$$= 1 + 4,2933 = 5,2933$$

$$= 6 \text{ (dibulatkan)}$$

(1) Mencari interval kelas

$$\text{Interval} = 6,666 = 7 \text{ (dibulatkan)}$$

(2) Menyusun daftar distribusi frekuensi

Kelas interval	Tabulasi	Frekuensi
52 - 58	III	3
45 - 51	III	3
38 - 44	II	2
31 - 37	IIIIII	7
24 - 30	II	2
17 - 23	III	3
Jumlah		20

C. Pengolahan Hasil Non-Tes

Menurut Benjamin (2019) berikut adalah pengolahan hasil non tes:

a. Pengolahan Hasil Data Wawancara dan Kuisisioner

Untuk pengolahan dan analisis data hasil wawancara dan kuisisioner pada umumnya dimulai dengan mencari frekuensi jawaban responden untuk setiap alternatif yang ada pada setiap pertanyaan. Frekuensi tertinggi ditafsirkan sebagai kecenderungan alat ukur tersebut. Sebaliknya,

frekuensi paling rendah dapat diartikan sebagai kecenderungan jawaban yang tidak menggambarkan pendapat kebanyakan responden (Sudjana, 1990). Sebagai contoh, melalui kuisisioner ataupun wawancara akan diungkap pandangan siswa/mahasiswa mengenai guru/dosen yang diharapkan dalam kemampuan mengajar serta hubungan guru/dosen dengan siswa/mahasiswa. Kuisisioner atau wawancara kemudian diajukan kepada 40 orang responden. Kuisisioner yang telah

**Frekuensi Jawaban Siswa
Mengenai Masalah Kemampuan Guru Mengajar**

Masalah yang diungkap	F	%	Peringkat Jawaban
1. Kemampuan Mengajar			
a. Menguasai bahan	4	10	3
b. Mampu menjelaskan bahan	12	30	2
c. Menguasai bahan dan mampu menjelaskannya	24	60	1
2. Prosedur mengajarkan bahan pelajaran			
a. Dimulai dari yang umum	10	25	2
b. Dimulai dari yang khusus	6	15	3
c. Harus sistematis	24	60	1

diisi oleh responden kemudian diperiksa dan diolah dengan menghitung frekuensi jawaban seluruh siswa terhadap tiap pertanyaan. Misalnya hasil pemeriksaan tersebut adalah sebagai berikut: Dari data diatas dapat ditafsirkan bahwa kemampuan guru mengajar yang diharapkan oleh siswa adalah gur yang menguasai bahan pelajaran dan dapat menjelaskannya agar dapat dipahami para siswa, sedangkan prosedur

mengajar yang diharapkan oleh siswa adalah harus sistematis. Cara lain dalam mengolah data diatas adalah dengan menggunakan (X^2). Rumus yang digunakan adalah:

$$X^2 = \sum \frac{(fo-fe)^2}{fe}$$

Dalam X^2 , yang dicari ialah adakah perbedaan yang berarti antara frekuensi hasil pengamatan atau jawaban nyata (fo) dengan frekuensi jawaban yang diharapkan. Jika ada perbedaan, artinya jawaban tersebut betul betul adanya, bukan faktor kebetulan. Contoh, kita ambil dari jawaban nomor 1 dari tabel 1

	Fo	Fe	$\frac{(fo-fe)^2}{fe}$
Kemampuan Mengajar			
Menguasai bahan	4	13,3	6,50
Mampu menjelaskan bahan	12	13,3	0,13
Menguasai bahan dan mampu menjelaskannya	24	13,3	8,61
$X^2 =$			15,24

Keterangan :

Fe = 13,3, diperoleh dari $40/3 = 13,3$

Harga $X^2 = 15,24$ kemudian dibandingkan dengan harga tabel untuk tingkat kepercayaan

0,05 (misalnya) dengan derajat bebas 3-1 (banyak alternatif jawaban 3)

Harga X^2 dalam tabel = 5,99

Dengan demikian, $X^2 = 15,24 > 5,99$ sehingga perbedaan itu cukup berarti. Ini berarti bahwa interpretasi yang menyatakan bahwa guru yang diharapkan adalah guru yang menguasai bahan dan dapat menjelaskannya kepada siswa adalah sah sebagai kesimpulan dari data tersebut.

b. Pengolahan Data Hasil Observasi

Pengolahan data hasil observasi sangat bergantung pada pedoman observasinya, terutama dalam mencatat hasil observasi. Hasil observasi dinyatakan dalam bentuk pernyataan pernyataan sebagaimana adanya yang tampak dari perilaku yang diobservasi, diolah dengan melakukan analisis dan interpretasi seluruh hasil amatan tersebut. Dengan kata lain, dengan menggunakan analisis kualitatif. Sudah barang tentu sifatnya subjektif, yakni dipengaruhi oleh pengamatannya (Sudjana, 1990). Namun, ada pula observasi yang hasil pengamatannya diberi nilai atau disediakan skala nilai, misalnya dengan huruf A, B, C, D atau dengan angka 4, 3, 2, 1 yang tersebut bermakna sebagai skala nilai. Contoh:

Observasi Kemampuan Guru Dalam Mengajar

Aspek yang diamati	Nilai Pengamatan			
	4	3	2	1
1. Penguasaan bahan		v		
2. Kemampuan menjelaskan bahan	v			
3. Hubungan dengan siswa		v		
4. Penguasaan kelas	v			
5. Keaktifan belajar siswa		v		

Dari contoh diatas, skor hasil observasi adalah:
 $3 + 4 + 3 + 4 + 3 = 17$

Nilai rata rata untuk kelima aspek tersebut adalah $17/5 = 3,4$. Skor ini cukup tinggi sebab maksimum rata rata atau skor maksimum untuk setiap aspek adalah 4 atau 20 untuk memuat aspek (5x4). Jika dibuat interpretasi untuk setiap aspek, maka dapat disimpulkan bahwa guru tersebut sangat istimewa dalam hal kemampuan menjelaskan dan penguasaan kelas, sedangkan dalam penguasaan bahan, komunikasi dengan siswa, dan dalam mengaktifkan siswa termasuk memuaskan (Sudjana, 1990).

c. Pengolahan Data Skala Penilaian atau Skala Sikap

Data hasil skala sikap, baik skala penilaian maupun skala sikap yang berbentuk skor atau data interval, pengolahannya hampir sama dengan pengolahan data hasil observasi yang menggunakan skor atau nilai dalam pengamatannya (Sudjana, 1990). Sudah tentu dalam pelaksanaannya diperhatikan ketentuan ketentuan yang berlaku disekolah, khususnya dalam sistem penilaian, baik dalam hal sistem pemberian angka maupun dalam menentukan batas kelulusannya. Sebagai contoh akan menilai skala sikap ilmiah. Lembar observasi

yang dapat digunakan untuk menilai sikap ilmiah siswa secara individu dapat dilihat pada tabel berikut:

No	Nama Siswa	Indikator Sikap							Jml skor
		keterbukaan	objektif	teliti	kedisiplinan	kerjasama	Kejujuran	Tanggung jawab	
1									
2									
3									

Skor untuk masing masing sikap diatas di rata rata dan dikonversi ke dalam bentuk kualitatif. Skala penilaian dibuat dengan rentangan dari 1 sampai dengan 5. Penafsiran angka angka tersebut adalah sebagai berikut:

- 1 = sangat kurang
- 2 = kurang
- 3 = cukup
- 4 = baik
- 5 = amat baik

Skor maksimum = skor maks setiap indikator
 X jumlah indikator
 = 5 x 7 = 35

Nilai sikap ilmiah ini dapat diberikan dalam bentuk huruf, oleh karena itu total skor yang telah diperoleh harus dikonversi.

$$\text{Konversi Nilai} = \frac{\text{Skor total jawaban siswa}}{\text{Skor maksimum}} \times 100$$

Jadi siswa yang memperoleh skor 28 setelah dikonversi nilainya menjadi :

$$\frac{28}{35} \times 100 = 80.$$

35

Banyak cara untuk mengkonversi skor menjadi nilai, salah satunya yang sederhana yaitu menggunakan kriteria sebagai berikut:

Skor Total	Nilai Konversi		Kategori
	Angka	Huruf	
29-35	81-100	A	Amat Baik
21-28	61-80	B	Baik
14-20	41-60	C	Cukup
7-13	20-40	D	Kurang

7

TINGKAT KESUKARAN, DAYA BEDA TES, DAN PENGECOH

A. Tingkatan Kesukaran

Analisis tingkat kesukaran dimaksudkan untuk mengetahui apakah soal tersebut tergolong mudah atau sukar. Tingkat kesukaran adalah bilangan yang menunjukkan sukar atau mudahnya sesuatu soal. (Arikunto, 1999). Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya. Adapun bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (difficulty index). Besarnya indeks kesukaran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah. (Solichin, 2017).

B. Cara Menentukan Tingkatan Kesukaran Tes

Untuk menghitung tingkat kesukaran tiap butir soal digunakan persamaan:

$$P = B/Jx$$

Keterangan :

P adalah indeks kesukaran,

B adalah banyaknya siswa yang menjawab soal dengan benar.

Jx adalah jumlah seluruh siswa

Indeks kesukaran diklasifikasikan seperti tabel berikut:

Tabel 1. Klasifikasi Tingkat Kesukaran

P-P	Klasifikasi
0,00 – 0,29	Soal sukar
0,30 – 0,69	Soal sedang
0,70 – 1,00	Soal mudah

Rumus lain yang digunakan untuk menentukan tingkat kesukaran soal uraian sama dengan soal pilihan ganda yaitu:

$$Tk = \frac{SA + SB}{IA + IB} \times 100\%$$

Keterangan:

Tk : Indeks tingkat kesukaran butir soal

SA : jumlah skor kelompok atas

SB : jumlah skor kelompok bawah

IA : jumlah skor ideal kelompok atas

IB : jumlah skor ideal kelompok bawah

Setelah indeks tingkat kesukaran diperoleh, maka harga indeks kesukaran tersebut diinterpretasikan pada kriteria sesuai tabel berikut:

Tabel 2. Interpretasi Tingkat Kesukaran

Indeks tingkat kesukaran	Kriteria
0 - 15%	Sangat sukar, sebaiknya dibuang
16 % - 30 %	Sukar
31 % - 70 %	Sedang
71 % - 85 %	Mudah
86 % - 100 %	Sangat mudah sebaiknya dibuang

C. Daya Beda Butir Tes

Secara operasional, daya beda butir didefinisikan sebagai efektivitas butir untuk membedakan peserta tes yang memperoleh skor tinggi dengan peserta tes yang memperoleh skor rendah. Ada beberapa indikator daya beda butir, baik secara tidak langsung maupun secara langsung. Secara tidak langsung, koefisien korelasi biserial (p_{bi}) antara skor butir dengan skor total juga menyatakan daya beda butir. Akan tetapi, dalam kesempatan ini daya beda butir dikaji menggunakan pengukuran langsung, yakni dengan menggunakan perbandingan kinerja peserta tes pada kelompok atas (upper region) dan kelompok bawah (lower region) pada skor total.

Secara umum, kelompok atas dan kelompok bawah ditentukan berdasarkan skor total dengan rentangan 10 sampai 33 persen. Apabila skor total berdistribusi normal, maka 27 persen skor tertinggi dan 27 persen skor terendah masing-masing akan

menjadi kelompok atas dan kelompok bawah. Apabila distribusi sekor total lebih datar (landai) daripada kurve normal, maka persentase yang diambil lebih besar, yakni mendekati 33 persen. Allen & Yen (1979) memberi batasan aplikasi kelompok atas dan kelompok bawah antara 25 dan 33 persen. Indeks daya beda butir dinyatakan dengan d dan dihitung dengan

menggunakan rumus berikut ini.

$$D = \frac{U - L}{N}$$

keterangan:

U = banyak peserta tes pada kelompok atas yang menjawab butir dengan benar

L = banyak peserta tes pada kelompok bawah yang menjawab butir dengan benar

N = banyak peserta tes.

Pertanyaan yang muncul sekarang adalah seberapa indeks daya beda butir yang dianggap masuk akal. Ebel antara lain memberi batasan terhadap indeks daya beda butir seperti berikut.

Jika $d \geq 0,40$, maka butir tergolong sangat baik.

Jika $0,30 \leq d < 0,40$, maka butir tergolong cukup tetapi perlu perbaikan.

Jika $0,20 \leq d < 0,30$, maka butir tergolong kurang dan harus direvisi.

Jika $d < 0,20$, maka butir tergolong jelek, sehingga harus digugurkan.

Sebagai contoh, berikut ini disajikan sebuah ilustrasi indeks dayabeda butir. Pada ilustrasi tersebut instrumen hasil belajar yang terdiri dari 5 butir tes objektif diuji cobakan terhadap 100

peserta tes. Sekor total yang diperoleh diranking dan selanjutnya ditentukan 27% sekor tertinggi sebagai kelompok atas dan 27% sekor terendah sebagai kelompok bawah. Dengan demikian diperoleh 27 orang peserta tes masuk kelompok atas dan 27 orang peserta tes masuk kelompok bawah. Selanjutnya dikaji peserta tes pada kelompok atas yang menjawab butir dengan benar (U) dan peserta tes pada kelompok bawah yang menjawab butir dengan benar (L). Ternyata diperoleh hasil seperti tercantum pada tabel di bawah ini.

Tabel 3 ilustrasi daya beda butir

No. butir	U	L	$D = (U-N)/N$	keterangan
1	16	9	0,26	Baik, daya beda tinggi
2	15	4	0,41	Sangat baik, daya beda sangat tinggi
3	21	4	0,63	Hamper sempurna, tapi jarang diperoleh
4	27	0	1,00	Sempurna namun tidak pernah diperoleh
5	15	15	0,00	Jelek harus direvisi
6	4	20	0,59	Sangat jelek harus digugurkan
7	0	27	-1,00	Paling jelek harus digugurkan

a. Pengecoh (analisis Distraktor)

Menganalisis fungsi pengecoh (distractor) dikenal dengan istilah menganalisis pola penyebaran jawaban butir soal pada soal bentuk pilihan ganda. Pola tersebut diperoleh dengan menghitung banyaknya testee yang memilih pilihan jawaban butir soal atau yang tidak memilih pilihan manapun

(blangko). Dari pola penyebaran jawaban butir soal dapat ditentukan apakah pengecoh berfungsi dengan baik atau tidak. Suatu pengecoh dapat dikatakan berfungsi dengan baik jika paling sedikit dipilih oleh 5 % pengikut tes.

Cara melakukan analisis pengecoh

Pertimbangan terhadap analisis pengecoh: a. Diterima, karena sudah baik b. Ditolak, karena tidak baik c. Ditulis kembali, karena kurang baik Sebuah pengecoh dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% pengikut tes.

Contoh:

Pilihan jawaban	A	B	C	D	E	O	JUMLAH
Kelompok atas	5	7	15	3	3	0	33
Kelompok bawah	8	8	6	7	7	3	37
Jumlah	13	15	21	8	10	3	70

O = Omitted (tidak menjawab)

C = kunci jawaban Pengecoh

A : $13/70 \times 100\% > 5\%$, berfungsi

B : $15/70 \times 100\% > 5\%$, berfungsi

D : $8/70 \times 100\% > 5\%$, berfungsi

E : $10/70 \times 100\% > 5\%$. berfungsi

Untuk tes pilihan ganda dengan 5 alternatif jawaban dan $P = 0,8$, dilihat dari segi Omitted (O), sebuah butir soal dikatakan baik jika persentase O-nya $\leq 10\%$.

8

PENILAIAN BERBASIS KELAS

A. Hakikat dan Jenis Penilaian Berbasis Kelas

Pada pembahasan pertama, kita akan membahas pengertian atau hakikat dari penilaian berbasis kelas. Setelah mengetahui makna atau pengertian dari penilaian berbasis kelas, maka akan dilanjutkan mengenai jenis-jenis dari penilaian berbasis kelas tersebut.

1. Hakikat Penilaian Berbasis Kelas

Penilaian Berbasis Kelas ialah penilaian dengan memiliki makna yakni “assessment”, yaitu data yang diperoleh dari penilaian berbasis kelas adalah salah satu bukti yang bisa digunakan dalam mengukur keberhasilan sesuatu program atau metode pendidikan. Atau secara keseluruhan Penilaian Berbasis Kelas dapat didefinisikan menjadi suatu sistem pengumpulan, pelaporan, dan penggunaan data mengenai hasil belajar siswa dalam menetapkan kelas pencapaian dan penguasaan siswa mengenai tujuan pendidikan yang sudah ditetapkan. Penerapan Penilaian Berbasis Kelas ini mengharuskan guru atau

pengajar menerapkan prinsip-prinsip penilaian berkelanjutan dengan tujuan sebagai bukti yang autentik, akurat, dan konsisten sebagai akuntabilitas umum.

Dalam penerapan penilaian berbasis kelas ini memiliki bermacam unsur-unsur, yakni sebagai berikut:

- a.) Penilaian prestasi kelas adalah semacam metode penilaian yang memiliki fungsi dalam mengetahui kelas-kelas pencapaian prestasi yang dicapai oleh siswa dalam mata pelajaran tertentu yang sesuai dengan kompetensi dengan mengacu kurikulum.
- b.) Penilaian kinerja adalah semacam teknik penilaian dengan tujuan dalam mengetahui ukuran penguasaan ketrampilan siswa yang mana melewati semacam tes penampilan atau demonstrasi atau praktik kerja nyata. Contohnya: pengajar memberikan tugas kepada siswa untuk berpraktik pidato di depan kelas.
- c.) Penilaian autentik adalah semacam metode penilaian dengan memiliki tujuan dalam mengetahui ukuran pencapaian kompetensi siswa berbentuk kemampuan nyata, tidak hanya yang dibuat-buat atau yang hanya didapatkan dalam kelas saja.
- d.) Penilaian portofolio adalah semacam teknik dengan tujuan dalam mengetahui ukuran pencapaian kompetensi dan perkembangan siswa didasarkan pada sekumpulan hasil kerja dari waktu ke waktu.

Dalam Penilaian Berbasis Kelas memiliki empat aktivitas yang mana aktivitas tersebut harus dilaksanakan oleh pengajar, aktivitas-aktivitas tersebut terdiri dari; (1) Mengumpulkan informasi mengenai ukuran pencapaian hasil belajar siswa, (2) mempergunakan data-data sesuai dengan hasil belajar yang telah siswa dapatkan, (3) membuat keputusan yang tepat, dan (4) guru harus membuat laporan sebagai bentuk akuntabilitas umum. Pengumpulan-pengumpulan informasi tersebut dapat dilaksanakan dari luar kelas ataupun dari dalam kelas. Setelah data-data atau informasi tadi telah terkumpul, maka pengajar dapat membuat keputusan tentang hasil pencapaian belajar yang telah siswa-siswa lakukan.

Setelah pengajar membuat keputusan, maka tindakan selanjutnya ialah melaporkan keputusannya tadi kepada berbagai pihak, antara lain siswa, orang tua siswa, masyarakat, atasan, dan juga instansi yang terkait lainnya. Selain itu, penilaian berbasis kelas ini pun juga memiliki beberapa karakteristik. Karakteristik-karakteristik terbagi menjadi beberapa macam, yakni sebagai berikut:

- a. Menggeser tujuan penilaian dari keperluan dalam pengklasifikasian siswa menjadi pelayanan perorangan siswa pada pengembangan kemampuannya.
- b. Menggunakan PAP daripada PAN.
- c. Menjamin perolehan tujuan pendidikan yang termaktup dalam kurikulum, dikarenakan

sebuah kompetensi dasar yang telah dirumuskan pada kurikulum sebagai acuan utama.

- d. Menggunakan keseimbangan teknik dan alat penilaian baik tes tertulis, lisan, maupun tes tindakan atau perbuatan serta cara lain sebagai jaminan validitas penilaian.
- e. Memberikan informasi yang lebih lengkap dan mudah dipahami.
- f. Memanfaatkan berbagai macam cara dan metode penilaian dengan menerapkan berbagai pendekatan dan cara belajar siswa (Dr. Zainal Arifin, 2009).

Dalam pandangan Yahya Hairun (2020) berpendapat bahwa penilaian berbasis kelas memiliki beberapa alasan yang mendasarinya. Pertama, waktu pembelajaran lebih banyak dilakukan dalam ruangan/kelas. Kedua, kegiatan belajar-mengajar lebih kearah ke dalam kelas, sehingga pengajar mampu membuat penilaian sendiri dalam ruangan kelas tersebut. Ketiga, kegiatan penilaian berbasis kelas ini melibatkan pengajar dan peserta didik secara bersamaan. Keempat, penilaian berbasis kelas ini dianggap menjadi sangat penting bagi seorang pengajar dalam mengukur penguasaan proses belajar-mengajar. Kelima, karena penilaian merupakan bagian yang penting dalam kegiatan seorang pengajar, sehingga seorang pengajar wajib melakukan penilaiannya dalam kegiatan belajar-mengajar dalam ruanga kelas (Ina Magdalena1, PENERAPAN PENILAIAN BERBASIS KELAS DAN

2. Jenis-jenis Penilaian Berbasis Kelas

Dalam pandangan Sumarna Surapranata dan Muhammad Hatta (2004) berpendapat bahwa Penilaian Berbasis Kelas memiliki beberapa macam jenis, yakni “tes tertulis, tes perbuatan, pemberian tugas, penilaian kinerja, penilaian kinerja, penilaian proyek, penilaian hasil kerja siswa, penilaian sikap, dan penilaian protfolio”.

- a. Tes tertulis adalah alat penilaian berbasis kelas yang mana menyajikan penggunaannya berbentuk tertulis dan peserta didik akan memberikan jawaban atas pertanyaan yang telah diberikan.
- b. Tes perbuatan adalah tes yang dilaksanakan pada saat prosesi kegiatan belajar berlangsung yang memungkinkan terjadinya praktik.
- c. Pemberian tugas dilaksanakan pada semua mata pelajaran yang dimulai dari awal kelas hingga akhir kelas sesuai dengan materi yang telah diajarkan dan juga sesuai dengan perkembangan siswa.
- d. Penilaian proyek adalah penilaian yang dilakukan terhadap tugas yang harus diselesaikan dengan waktu yang telah ditentukan.
- e. Penilaian produk siswa adalah penilaian terhadap penguasaan ketrampilan siswa dalam membuat suatu produk dan penilaian kualitas

hasil kerja tertentu. Penilaian produk ini terbagi menjadi tiga tahapan:

1. Tahapan persiapan, yakni tahapan menilai ketrampilan merencanakan, menyusun, menggali atau mengembangkan ide.
 2. Tahap produksi, yakni menilai kemampuan memilih dan menggunakan bahan-bahan dan alat.
 3. Tahap penilaian.
- f. Penilaian sikap bisa dilaksanakan dan berkaitan dengan bermacam objek sikap, seperti sikap dengan mata pelajaran, sikap dengan guru, sikap kepada proses pembelajaran, sikap dengan materi pelajaran, dan lain-lain.
- g. Penilaian portofolio adalah penilaian berbasis kelas mengenai sekumpulan karya siswa yang telah tertata secara sistematis dan terorganisasi yang didapat selama proses pembelajaran (Dr. Zainal Arifin, 2009).

B. Objek dan Alat Penilaian Berbasis Kelas

Pada penilaian berbasis memiliki objek, sebagaimana yang telah dinyatakan oleh Departemen Pendidikan Nasional dalam petunjuk pengembangan Kurikulum Berbasis Kompetensi, maka objek penilaian berbasis kelas ada:

1. 34 nilai kompetensi dasar mata pelajaran, yaitu pengetahuan, ketrampilan, sikap, dan nilai-nilai yang direfleksikan dalam kebiasaan berpikirnya dan aktivitas setelah siswa menyelesaikan suatu aspek dari mata pelajaran tertentu.

2. ³⁴ nilai kompetensi Rumpun Pelajaran, adalah pengetahuan ketrampilan, sikap, dan nilai-nilai yang direfleksikan dalam kebiasaan berfikir dan beraktivitas yang mana mewajibkan dicapai oleh siswa setelah menyelesaikan rumpun pelajaran.
3. Penilaian kompetensi lintas kurikulum adalah pengetahuan, ketrampilan, sikap, dan nilai-nilai yang dicerminkan dari kebiasaan berpikir dan beraktivitas mengenai kecakapan belajar sepanjang hayat dan juga kecakapan hidup yang wajib dimiliki oleh siswa melalui pengalaman belajar yang saling berkaitan.
4. Penilaian kompetensi tamatan adalah pengetahuan, ketrampilan, sikap, dan nilai-nilai yang dicerminkan dalam kebiasaan berpikir dan beraktivitas setelah siswa menyelesaikan pendidikannya.
5. Penilaian Terhadap Pencapaian Ketrampilan Hidup. Kecakapan hidup yang dimiliki oleh siswa melalui bermacam pengalaman belajar perlu dinilai sejauh mana kesesuaiannya dengan kebutuhan mereka dalam bertahan hidup dan berkembang dalam kehidupan sehari-hari di masyarakat dan keluarga (Dr. Zainal Arifin, 2009).

Selain memiliki objek, penilaian berbasis kelas ini juga memiliki alat/domain dalam melaksanakan programnya. Penilaian yang autentik memerlukan pelaksanaan yang menyeluruh terhadap kompetensi yang sudah dipelajari oleh siswa. Domain-domain ini terbagi menjadi domain kognitif, domain afektif, dan domain psikomotor.

1. Domain Kognitif ⁵⁶

Mencakup beberapa hal, sebagai berikut:

- a. Tingkatan hafalan, meliputi kemampuan menghafal secara verbal atau paraphrase materi pembelajaran dalam bentuk fakta, konsep, prinsip dan prosedur atau langkah-langkah.
- b. Tingkatan pemahaman, mencakup kemampuan membandingkan karakteristik, menggeneralisasi, dan menyimpulkan.
- c. Tingkatan aplikasi, meliputi kemampuan menerapkan rumus, dalil mengenai masalah-masalah nyata yang terjadi di lapangan.
- d. Tingkatan analisis, mencakup kemampuan memadukan berbagai unsur atau komponen, menyusun, membentuk bangunan, mengarang, melukis, menggambar, dan lain sebagainya.
- e. Tingkatan evaluasi, mencakup kemampuan menilai mengenai objek studi dengan menggunakan kriteria tertentu.

Pada pengukuran penguasaan kognitif bisa menggunakan tes lisan, tes tertulis, dan juga portofolio. Portofolio sendiri adalah sekumpulan dari tugas-tugas yang telah dikerjakan oleh siswa. Tujuannya ialah mengukur kemampuan membaca dan menulis yang lebih luas.

2. Domain Afektif

Hubungan dengan ranah afektif, memiliki 2 hal yang wajib dinilai. Yang pertama kompetensi afektif yang ingin dicapai pada pembelajaran mencakup kelas pemberian respons, apresiasi,

penilaian, dan juga internalisasi. Yang kedua sikap dan minat peserta didik mengenai mata pelajaran dapat positif, dapat negative, dapat juga netral.

3. Domain Psikomotor

- 4alam domain ini mencakup beberapa hal, yakni:
- a. Tingkatan penguasaan gerakan awal yang mana mencakup kemampuan siswa dalam menggerakkan sebagian anggota tubuhnya.
 - b. Tingkatan gerakan semirutin mencakup kemampuan melakukan/menirukan gerakan yang melibatkan seluruh anggota tubuh.
 - c. Tingkatan gerakan rutin mencakup kemampuan melakukan gerakan secara menyeluruh dengan sempurna dan sampai padaa tingkatan otomatis (Dr. Zainal Arifin, 2009).

Alat yang digunakan dalam domain ini ialah tes penampilan atau kinerja yang sudah dikuasai oleh siswa, yakni:

- a. Tes paper and pencil, dalam tes ini sasarannya ialah kemampuan siswa yang menampilkan karyanya.
- b. Tes identifikasi, tes yang disasarkan dengan tujuan mengukur kemampuan siswa dalam mengenali sesuatu.
- c. Tes simulasi, tes ini dilaksanakan apabila tidak ada alat yang sesungguhnya yang bisa digunakan dalam memperagakan penampilan siswa.
- d. Tes petik kerja, tes ini menggunakan alat yang sesungguhnya, tujuannya ialah mengetahui

apakah siswa sudah menguasai dalam penggunaan alat tes tersebut (Dr. Zainal Arifin, 2009).

C. Fungsi, tujuan, dan prinsip dari Penilaian Berbasis Kelas.

Fungsi dari penilaian berbasis kelas bagi siswa dan pengajar ialah sebagai (1) membantu siswa dalam mewujudkan dirinya dengan cara mengubah sikapnya kearah yang lebih baik dan lebih maju, (2) membantu siswa dalam mendapatkan kepuasan atas apa yang sudah dikerjakannya, (3) membantu pengajar menetapkan apakah strategi yang digunakan telah memadai, dan (4) membantu pengajar dalam membuat pertimbangan dan keputusan mengenai administrasi. Adapun tujuan dari penilaian berbasis kelas ialah sebagai pemberian penghargaan mengenai pencapaian hasil belajar siswa dan memperbaiki metode dan aktivitas pembelajaran. Dalam dokumen Kurikulum Berbasis Kompetensi (2002) mengemukakan tujuan penilaian berbasis secara khusus ialah sebagai berikut:

- a. Data mengenai kemajuan hasil belajar siswa secara individual pada pencapaian tujuan belajar sesuai dengan kegiatan yang dilakukannya.
- b. Data yang digunakan dapat membina kegiatan belajar lebih lanjut.
- c. Data yang diperoleh dapat digunakan oleh siswa dan pengajar dalam mengetahui tingkatan kemampuan siswa.

- d. Motivasi belajar siswa dengan cara memberikan data tentang kemajuan pencapaian hasil belajarnya.
- e. Data meliputi semua aspek mengenai kemajuan siswa dan pada gilirannya pengajar bisa membantu pertumbuhan secara efektif menjadi anggota masyarakat dan pribadi yang utuh.
- f. Bimbingan yang tepat dapat membantu memilih sekola atau jabatan yang sesuai dnegan ketrampilan, minat, dan kemampuannya (Dr. Zainal Arifin, 34 09).

Selain itu penilaian berbasis kelas ini juga memiliki prinsip dan 34n menjalankan programnya. Prinsip-prinsip dari penilaian berbasis kelas, tealh dikemukakan Pusat Kurikulum Balibang Depdiknas (2002), sebagai berikut:

- 22 Valid (tepat).
- 2. Mendidik.
- 3. Berorientasi pada kompetensi.
- 4. Adil dan objektif.
- 5. Terbuka.
- 6. Berkesinambungan.
- 7. Menyeluruh.
- 8. Bermakna.

Adapun prinsip penilaian bagi siswa adalah untuk menunjukkan apa yang telah mereka ketahui dan pahami, serta mempraktikan kemampuannya. Sebagaimana penerapannya sebagai berikut:

- a. Pelaksanaan penilaian berbasis kelas seharusnya dilaksanakan pada suasana yang bersahabat.
- b. Semua siswa memiliki kesempatan dan perlakuan yang sama.

- c. Siswa harus mengetahui dan memahami secara jelas mengenai penilaian berbasis kelas.
- d. Kriteria membuat keputusan atas hasil penilaian berbasis kelas seharusnya disepakati dengan siswa dan wali.

Pengajar juga seharusnya mampu dalam melaksanakan langkah-langkah penilaian berbasis kelas. Sebagaimana penerapan sebagai berikut:

- a. Langkah-langkah penilaian berbasis kelas harus mudah dipahami.
- b. Langkah-langkah penilaian berbasis kelas dan catatan hasil belajar siswa hendaknya mudah dilaksanakan sebagai bagian dari aktivitas pembelajaran.
- c. Catatan harian harus mudah dibuat dan jelas.
- d. Data yang diperoleh untuk menilai semua pencapaian hasil belajar siswa dengan berbagai cara seharusnya digunakan sebagaimana mestinya.
- e. Penilaian pencapaian hasil belajar siswa yang bersifat positif untuk pembelajaran.
- f. Klasifikasi dan kesulitan belajar harus ditentukan sehingga siswa mendapatkan bimbingan dan bantuan belajar yang sewajarnya.
- g. Hasil penilaian hendaknya menunjukkan kemajuan dan berkelanjutan.
- h. Penilaian semua aspek yang berkaitan dengan pembelajaran.
- i. Peningkatan keahlian guru sebagai konsekuensi dari diskusi pengalaman dan membandingkan metode.

- j. Pelaporan penampilan siswa dari pengajar kepada wali murid dan atasannya yang dilakukan secara periodic (Dr. Zainal Arifin, 2009).

55

D. Manfaat Hasil Penilaian Berbasis Kelas

Penilaian berbasis kelas memiliki manfaat bagi pengajar yakni sebagai mengetahui kemajuan dan hasil belajar siswa, menganalisis kesulitan belajar, memberikan umpan balik dalam perbaikan proses belajar-mengajar, menentukan kenaikan kelas, dan memotivasi siswa. Bagi walimurid adalah mengetahui kelebihan dan kekurangan anaknya, peringkat anaknya dalam kelas, memberikan bimbingan, dan merangsang orang tua dalam menjalin komunikasi dengan pihak sekolah. Bagi siswa, penilaian berbasis kelas bermanfaat sebagai memantau hasil pencapaian kompetensi secara utuh, baik menyangkut aspek pengetahuan, ketrampilan, sikap dan nilai-nilai (Dr. Zainal Arifin, 2009).

E. Implikasi Penilaian Berbasis Kelas.

Pada sub-bab implikasi ini akan diambil dari beberapa jurnal yang melakukan penelitian tentang penerapan dan pemahaman dari seorang pengajar yang menggunakan program penilaian berbasis kelas. Yang pertama akan diambil dari jurnal yang berjudul **“Penerapan Penilaian Berbasis Kelas Dan Portofolio Pelajaran Bahasa Indonesia Kelas 6 SDN Selapajang Jaya 2”**. Pada jurnal ini peneliti melakukan penelitian tentang penerapan penilaian berbasis kelas dan juga portofolio yang mana subjeknya mengambil kelas 6 SD dengan mata

pelajaran Bahasa Indonesia ¹⁶sebut. Di penelitian ini memaparkan hasil tentang penilaian berbasis kelas dan portofolio pada pembelajaran mata pelajaran Bahasa Indonesia kelas 6 di SD Selapajang Jaya 2 Kota Tangerang. Hasil penelitian ini diperoleh dari wawancara dengan para pengajar SD Selapajang Jaya 2. Peneliti dalam hasil wawancara tersebut memperoleh hasil bahwasannya pengajar SD Selapajang dalam mengajar materi Bahasa Indonesia di kelas 6, menggunakan metode penilaian berbasis kelas yang mana terdiri dari non tes dan wawancara. Penilaian non tes diberikan untuk ¹⁶engetahui apakah efektif atau tidaknya dalam penggunaan penilaian berbasis kelas dan portofolio dalam pembelajaran Bahasa Indonesia. Dan rata-ratanya para pengajar mengatakan bahwa kelas 6 SD Negeri Selapajang Jaya 2 menjawab 95% efektif dalam menggunakan penilaian berbasis kelas. Sehingga para pengajar menerapkan metode penilaian ini dengan memberikan tes tertulis, pemberian tugas, penilaian ¹⁶ap dan juga penilaian portofolio pada peserta didik kelas 6 SDN Selapajang Jaya 2. (Magdalena et al., 2021).

Jurnal yang kedua adalah penerapan yang dapat dilihat dari Sekolah Menengah Pertama atau SMP. Dengan Judul "**PENERAPAN PENILAIAN BERBASIS KELAS DALAM PEMBELAJARAN IPS DI SMP NEGERI 1 BATULICIN**" dalam jurnal ini peneliti melakukan penelitian terhadap penerapan penilaian berbasis kelas didasarkan pada materi pembelajaran IPS. Dalam jurnal menyebutkan bahwasannya penilaian berbasis kelas telah dilaksanakan oleh para pengajar

SMP Negeri 1 Batulicin. Yang mana dilakukan dengan menggunakan cara menentukan kompetensi-kompetensi yang akan dinilai, menjelaskan kompetensi pada indikator, merumuskan tujuan kompetensi sesuai indikator, memilih teknik penilaian yang sesuai, menuliskan kisi-kisi yang sesuai dengan penilaian, mengembangkan instrument penilaian dan melaksanakan penilaian (Darwis, 2016).

Jurnal yang terakhir adalah penerapan penilaian berbasis kelas dengan materi pendidikan kelas 10 dari SMA Darul Mustofa Burneh Bangkalan. Jurnal tersebut memiliki judul **“PENERAPAN PENILAIAN BERBASIS KELAS DALAM BIDANG STUDI PENDIDIKAN AGAMA ISLAM DI KELAS X SMA DARUL MUSTOFA BURNEH BANGKALAN”**. Dalam jurnal ini, peneliti melakukan penelitian tentang penerapan penilaian berbasis kelas, apakah penilaian tersebut telah efektif atau tidak. Dalam penelitiannya, si peneliti dapat menyimpulkan bahwa penerapan penilaian berbasis kelas dari tingkatan SMA pada kelas 10 dengan mata pelajaran PAI ini bersifat holistik, yang mana artinya penilaiannya didasarkan pada keseluruhan proses kegiatan belajar-mengajar yang telah diikuti. Prinsip yang mendasari penilaian berbasis kelas ini ialah memotivasi, memantau, melaporkan, mengontrol dan mengevaluasi atas penugasan yang telah dicapai oleh para siswa. Jadi dari SMA Darul Mustofa Burneh ini, para pengajar telah menerapkan Penilaian Berbasis Kelas, yang mana penerapannya sesuai dengan prinsip-prinsip dasar dari PBK tersebut (Oktasari, 2012).

9

PBK DALAM KETRAMPILAN RESEPTIF BERBAHASA PBK DALAM KETRAMPILAN PRODUKTIF BERBAHASA PBK DALAM APRESIASI SASTRA

A. PBK dalam keterampilan reseptif berbahasa

Menurut Angelo (2001), dalam penelitian tindakan kelas, bahwa penilaian berbasis kelas bisa dipandang sebagai suatu metode penemuan, suatu pendekatan yang lebih meluas untuk meningkatkan kualitas belajar dan mengajar. Penilaian Berbasis Kelas dapat didesain untuk membantu guru menemukan bagaimana individu dan atau kelompok siswa sedang belajar dalam kelas. Guru dapat menerapkan hasil penilaiannya untuk memperbaiki mengajar sedangkan siswa dapat meningkatkan hasil pencapaian belajarnya.

Pendapat lain Penilaian Berbasis Kelas merupakan penilaian sebagai 'assessment', yaitu kegiatan yang dilakukan untuk memperoleh dan mengefektifkan informasi tentang hasil belajar siswa pada tingkat kelas selama dan setelah kegiatan pembelajaran. Data atau informasi dari penilaian di kelas ini merupakan salah satu bukti yang digunakan untuk mengukur keberhasilan suatu program pendidikan. Penilaian Berbasis Kelas merupakan

bagian dari evaluasi pendidikan karena lingkup evaluasi pendidikan secara umum jauh lebih luas dibandingkan Penilaian Berbasis Kelas

Penilaian Berbasis Kelas itu terdiri pada dasarnya suatu kegiatan penilaian yang dilaksanakan cara terpadu dalam kegiatan belajar mengajar yang dilakukan dengan mengumpulkan kerja siswa (portofolio), hasil karya (produk), penugasan (proyek), kinerja (performance), dan tes tertulis (paper and pen). Fokus penilaian diarahkan pada penguasaan kompetensi dan hasil belajar siswa sesuai dengan level pencapaian prestasi siswa

Menurut Supranata dan H (2004) dalam jurnal membaca Bahasa dan sastra Penilaian berbasis kelas bertujuan untuk mengetahui kemajuan dan hasil belajar peserta didik, mendiagnosa kesulitan belajar, memberikan umpan balik atau perbaikan proses pembelajaran, penentuan kenaikan kelas, dan memotivasi belajar siswa dengan cara mengenal dan memahami diri dan merangsang untuk melakukan usaha perbaikan Penilaian Berbasis Kelas berfungsi sebagai bahan pertimbangan dalam menentukan kenaikan kelas, umpan balik dalam perbaikan program pengajaran, alat pendorong dalam meningkatkan kemampuan peserta didik, dan sebagai alat untuk peserta didik melakukan evaluasi terhadap kinerjanya serta bercermin diri (instrospeksi) (Arip Sanjaya, 2018).

Empat jenis keterampilan berbahasa

	Lisan	Tulisan
<i>Reseptif</i>	Mendengarkan / menyimak	Membaca
<i>Produktif</i>	Berbicara	Menulis

Apa itu keterampilan reseptif berbahasa? Keterampilan reseptif dalam berbahasa artinya proses yang berlangsung pada diri pendengar yang menerima kode-kode bahasa yang bermakna dan berguna yang disampaikan oleh pembicara melalui alat-alat artikulasi dan diterima melalui alat-alat pendengar. Aspek keterampilan reseptif ini meliputi menyimak dan membaca. Jika dilihat dari polanya, kompetensi menyimak dan membaca merupakan kompetensi menerima buah pikir atau hasil kata orang lain. Dengan kata lain, seorang penyimak atau pembaca akan menikmati atau mengapresiasi produk-produk bahasa baik secara visual maupun auditif. Untuk itulah penguasaan kedua kompetensi ini dikategorikan sebagai keterampilan reseptif.

a. Menyimak

1) Pengertian Menyimak

Menyimak adalah salah satu keterampilan reseptif dalam berbahasa. Banyak sekali pengguna bahasa yang beranggapan bahwa menyimak sama dengan mendengar dan mendengarkan, padahal arti ketiga hal tersebut berbeda sekalipun maknanya bertalian erat. Mendengar diartikan sebagai menangkap bunyi (suara) dengan telinga. Mendengarkan berarti menangkap suatu suara

melalui telinga dengan sungguh sungguh. Sedangkan menyimak berarti mendengarkan dengan baik dan saksama apa yang diucapkan (atau dibacakan) bahkan yang diperlihatkan orang lain untuk dipahami maksud atau tujuannya.

Selain itu, untuk memudahkan pemahaman dapat pula dijelaskan bahwa peristiwa mendengar belum ada faktor kesengajaan, sedangkan pada proses mendengarkan sudah mulai ada faktor kesengajaan. Sehingga taraf mendengarkan bisa dikatakan lebih tinggi dibandingkan dengan mendengar. Sedangkan dalam peristiwa menyimak sudah berorientasi pada pemahaman akan hal-hal yang dibaca, dilihat, atau didengar oleh seseorang (T¹¹gan, 2000: 2.6).

2) Tes kompetensi menyimak

Kegiatan berbahasa yang berupa memahami bahasa yang dihasilkan orang lain melalui sarana lisan (atau pendengaran) merupakan k¹¹atan yang paling pertama yang dilakukan manusia. Berdasarkan materi dan tujuan menyimak sebenarnya ada banyak hal dari tujuan yang terkait dengan kegiatan menyimak contohnya me⁴³imak pembicaraan, menyimak siaran radio, tv, lagu-lagu, dan lain-lain tujuan menyimak pun ada macam macam contohnya untuk menangkap pesan yang disampaikan atau sekedar menikmati saja tanpa mengesampingkan tujuan tujuan yang lain pengujian menyimak lazimnya ditekankan ntuk mengukur kompetensi peserta didik memahami dan merespon pesan yang disampaikan secara lisan tersebut

b. Membaca

1) Pengertian Membaca

Membaca merupakan salah satu jenis keterampilan berbahasa ragam tulis yang bersifat reseptif. Diartikan demikian karena dengan membaca maka pembaca akan menerima berbagai ide, gagasan, maupun pesan dan informasi yang ingin disampaikan oleh penulis. Meskipun pola keterampilan ini berbeda dengan keterampilan menyimak dan berbicara, tetapi dalam literasi yang telah dikembangkan, seringkali keterampilan ini terintegrasi dengan keterampilan menyimak dan berbicara.

Kegiatan membaca merupakan yang kompleks. Selain membutuhkan kemampuan visual yang baik untuk membaca lambang-lambang huruf menjadi bermakna, kemampuan kognitif untuk memahami bacaan pun diperlukan. Dalam kegiatan membaca terjadi interaksi antara pembaca dan penulis secara tidak langsung. Meski demikian hubungan antara pembaca dan penulis tetaplah bisa dikatakan bersifat komunikatif.

Menurut Tarigan (2008: 07), "Membaca adalah suatu proses yang dilakukan sertadipergunakan oleh pembaca untuk memperoleh pesan yang hendak disampaikan oleh penulis melalui media kata-kata/ bahasa tulis."

2) Tes Kompetensi membaca

Dalam dunia pendidikan aktivitas dan tugas membaca merupakan suatu hal yang tidak dapat ditawar-tawar. Sebagian perolehan sebagian perolehan ilmu dilakukan peserta didik dan terlebih lagi mahasiswa dengan kegiatan membaca. Pada kenyataannya ada banyak tujuan membaca misalnya

untuk memperoleh informasi dan menanggapi informasi, memperluas pengetahuan, memperoleh hiburan dan kesenangan hati, dan lain-lain

Untuk meraih kompetensi membaca yang baik kemampuan dan kemauan membaca mesti baik pula. Hal itu mesti di prasarati oleh kemauan membaca berbagai bacaan. Intinya, peserta didik, juga guru dan dosen harus rajin membaca. Ini lebih banyak dipengaruhi oleh unsur sikap, ranah afektif. Maka, selain guru membelajarkan dan kemudian mengukur kompetensi membaca peserta didik, aspek sikap haruslah tidak dilupakan. Kita perlu mengetahui seberapa tinggi sikap kemauan membaca siswa. Pemerolehan data dapat melalui wawancara, pemberian angket, pengamatan atau gabungan dari kedua-ketiganya

Jika dalam angket itu ditemukan fakta bahwa sikap suatu kemauan membaca kurang baik, maka perlu memotivasi dan membantu mencari peseta didik solusi jika ada berbagai factor prnyrbab contoh membantu mengusahakan agar sekolah menyediakan berbagai bacaan

B. PBK dalam keterampilan produktif berbahasa

Komptensi berbahasa yang bersifat aktif produktif merupakan kemampuan yqng menuntut kegiatan encoding, kegiatan untuk menghasilkan bahasa kepada pihak lain, baik secara lisan maupun tertulis. Kegiatan berbahasa yang produktif adalah kegiatan menyampaikan gagasan, pikiran, perasaan, pesan, atau informasi oleh pihak penutur. Penutur

dapat bernama pembicara, dan dapat bernama penulis jika aktivitas menghaslkan bahasanya itu disampaikan melalui sarana tulisan (Nurgiyantoro, 2010)

a. Berbicara

1) Pengertian berbicara

Secara garis besar ada tiga jenis Berbicara, yaitu interaktif, semi interaktif, dan non interaktif, misalnya percakapan secara tatap muka dan berbicara lewat telepon yang memungkinkan adanya pergantian antara berbicara dan mendengarkan, dan memungkinkan kita meminta klarifikasi, pengulangan atau kita dapat meminta lawan bicara kita memperlambat tempo bicara dari lawan bicara. Kemudian, ada pula situasi semi bicara yang semi interaktif, misalnya berpidato dihadapan umum secara langsung. Dalam situasi ini khalayak memang tidak dapat melakukan interupsi terhadap pembicaraan, namun pembicara dapat melihat reaksi pendengar dari ekspresi wajah dan bahasa tubuh mereka. Beberapa situasi dapat dikatakan noninteraktif misalnya berpidato melalui radio atau televisi (Yeti Mulyati, 2007)

2) Tes kompetensi Berbicara

Berbicara adalah aktivitas berbahasa kedua yang dilakukan manusia dalam kehidupan bahasa setelah didengarkan. Berdasarkan bunyi bunyi (bahasa) yang didengarnya itulah kemudian manusia

belajar mengucapkan dan akhirnya mampu berbicara. Dalam kegiatan berbicara diperlukan penguasaan terhadap lambang bunyi baik untuk keperluan menyampaikan maupun menerima gagasan

Dibawah ini akan dicontohkan berbagai bentuk tes kompetensi berbicara, namun tugas tugas yang ditekankan adalah tugas yang bersifat pragmatik atau otentik yang menghendaki peserta didik telah menguasai tahap elementer dalam suatu bahasa, atau paling tidak sudah dapat mempergunakan bahasa itu untuk

13) aktivitas berbicara

1. Tugas berbicara otentik

Tugas berbicara otentik dimaksudkan sebagai tes berbicara yang memenuhi

13) kriteria asesmen otentik

2. Bentuk tugas kompetensi berbicara

Ada banyak tugas yang dapat diberikan kepada peserta didik untuk mengukur kompetensi berbicara pada bahasa target. Apapun bentuk tugasnya yang dipilih haruslah yang memungkinkan peserta didik untuk tidak saja mengekspresikan kemampuan berbahasanya melainkan juga gagasan, pikiran, perasaan, ataupun

13) menyampaikan informasi

a) Berbicara berdasarkan gambar

Untuk mengungkapkan kemampuan berbicara pembelajar dalam suatu bahasa,

gambar dapat dijadikan rangsang pembicaraan yang baik

- b) Berbicara berdasarkan rangsang suara
Tugas berdasarkan rangsang suara yang lazim digunakan adalah berupa radio atau rekaman yang sengaja dibuat untuk maksud itu
- c) Berbicara berdasarkan visual dan suara
Merupakan gabungan antara berbicara menggunakan gambar dan suara
- d) Bercerita
Tugas bercerita yang dimaksudkan disini ada kemiripan dengan tugas bercerita berdasarkan beberapa rangsang diatas, namun lebih luas cakupannya.
- e) Wawancara
Wawancara barangkali merupakan teknik yang paling banyak dipergunakan untuk menilai kompetensi berbicara seseorang dalam suatu bahasa, khususnya bahasa asing yang dipelajarinya (Nurgiyantoro, 2010).

b. Menulis

- 1) Pengertian menulis
Menulis adalah keterampilan produktif dengan menggunakan tulisan. Menulis dapat dikatakan suatu keterampilan berbahasa yang paling rumit di antara jenis jenis keterampilan berbahasa lainnya. Hal ini dikarenakan menulis bukan sekedar menyalin kata-kata dan kalimat-kalimat,

melainkan juga mengembangkan dan menuangkan pikiran pikiran dalam suatu struktur tulisan yang teratur (Yeti Mulyati, 2007)

2) Tes kompetensi menulis

Aktivitas menulis meruoakan suatu bentuk manifestasi kompetensi berbahasa paling akhir dikuasai pembelajar bahasa setelah setelah kompetensi mendengarkan, berbicara, dan membaca. Tes kemampuan menulis, sebagaimana halnya dengan tes kemampuan berbicara, cukup potensial untuk dijadikan tes yang bersifat pragmatik dan atau otentik. Pada umumnya, aktivitas orang menghasilkan bahasa tidak semata-mata hanya bertujuan demi produktivitas bahasa itu sendiri, melainkan karena ada sesuatu yang ingin dikomunikasikan lewat bahasa, dengan kata lain bahasa hanya merupakan sarana, dan gagasan yang ingin dikomunikasikan pada hakikatnya lebih penting daripada sarana bahasa itu sendiri.

Bentuk tugas kompetensi menulis dilihat dari segi kompetensi bahasa, menulis adalah aktivitas menghasilkan bahasa, dilihat dari pengertian secara umum, menulis adalah aktivitas mengemukakan pendapat atau gagasan melalui media bahasa.

Pada halnya dengan berbicara seperti dalam ujian yang hanya menyediakan

waktu waktu yang relatif singkat dan terbatas. Ujian tulis dapat diberikan dalam bentuk bentuk objektif. Ujian menulis bentuk objektif pilihan ganda telah menyediakan respon yang harus dipilih, maka kadar keotentikanya juga lebih rendah

Berikut beberapa tes kompetensi menulis

1. Tugas menulis dengan memilih jawaban

Walaupun tes kompetensi menulis yang lebih ideal dengan menyuruh peserta didik untuk menulis dalam arti yang sebenarnya, dalam praktiknya tes bentuk objektif masih dapat dilakukan, namun teks yang dimaksud harus memenuhi hal yang esensial dalam aktivitas menulis, yaitu adanya unsur memilah bahasa dan gagasan

2. Tugas menulis dengan membuat karya tulis

Tugas menulis untuk benar benar menghasilkan karya karya tulis, apapun bentuknya, haruslah mendapat prioritas dalam rangka mengukur kompetensi menulis peserta didik. Karya tulis yang dihasilkan dapat sekaligus menunjukkan kompetensi berbahasa tulis dalam arti yang sebenarnya

Contoh rubrik penilaiannya sebagai berikut

- 1) Menulis berdasarkan gambar
- 2) Menulis berdasarkan rangsang suara
- 3) Menulis berdasarkan rangsang visual dan suara
- 4) Menulis dengan rangsang buku
- 5) Menulis laporan
- 6) Menulis surat
- 7) Menulis berdasarkan tema tertentu

C. PBK dalam apresiasi sastra

Pembelajaran sastra (Indonesia) disekolah tidak berdiri sendiri sebagai sebuah mata pelajaran yang mandiri, melainkan hanya menjadi bagian mata pelajaran bahasa dan sastra indonesia. Penggabungan pembelajaran sastra kedalam bahasa indonesia dapat dimengerti karena bahasa merupakan sarana yang penting sebagai manifestasi teks kesastraan. Bahkan secara lahiriah, wujud formal yang tampak, wujud sastra adalah bahasa, sastra adalah karya seni yang bermediakan bahasa yang unsur unsur keindahannya menonjol. Akan tetapi, sebagai sebuah karya seni sastra tidak semata mata hanya berurusan dengan unsur bahasa saja, melainkan juga unsur-unsur sastra yang lain juga tak kalah penting

Jika pembelajaran sastra dimaknai dengan sebagai membelajarkan dan memberikan kesempatan pada peserta didik untuk memperoleh

berbagai pengalaman hidup, pengetahuan, dan hiburan yang menyenangkan lewat berbagai teks kesastraan, tes yang dimaksudkan untuk mengukur seberapa banyak mereka belajar, mestinya berupa tagihan atau tugas untuk mengungkap dan mengekspresikan seberapa banyak peserta didik memperoleh hal-hal tersebut.

Ada beberapa model penilaian yang dapat dipergunakan untuk mengukur seberapa banyak capaian peserta didik belajar kompetensi bersastra. Analog dengan tes kompetensi berbahasa yang dibedakan kedalam tugas merespon jawaban dan mengreasikan jawaban sendiri diatas, tes dan tugas tugas untuk mengukur hasil belajar kopetensi bersastra juga sengaja dibedakan kedalam golongan tersebut. Hal itu sengaja dimaksudkan untuk membedakan sekaligus menunjukan adanya tekanan capaian berkinerja sesuai dengan kompetensi mata pejaran.

Tes kompetensi bersastra dengan merespon jawaban Tes kompetensi bersastra yang dikukur dengan merspon jawaban yang telah disediakan mesti berupa tes objektif. Untuk mengerjakan tugas, peserta didik tinggal memilih jawaban itu, tugas ini mungkin dirasakan kurang ideal untuk tujuan pengukuran kompetensi bersastra, namun ia dapat dimanfaatkan sebagai pelengkap tugas yang lain seperti untuk ujian akhir yang waktu pengerjaanya terbatas. (Nurgiyantoro, 2010).

10

PENGEMBANGAN ALAT PENILAIAN DAN PENYUSUNAN SOAL

A. Pengembangan Alat Penilaian

Dalam pengertian umum, alat adalah sesuatu yang dapat digunakan untuk mempermudah seseorang untuk melaksanakan tugas atau mencapai tujuan agar secara efektif dan efisien. kata “Alat” biasa disebut juga dengan istilah “instrumen”. Dengan demikian, maka alat evaluasi juga dikenal dengan instrumen evaluasi. Untuk memperjelas pengertian pengertian “alat” atau “instrumen”. Misalnya, ada dua cara mengupas kelapa, yang satu menggunakan pisau parang, yang satu lagi tidak. tentu saja hasilnya akan lebih baik dan pekerjaannya berakhir lebih cepat yang menggunakan alat (parang tersebut). dalam kegiatan evaluasi, fungsi alat juga untuk memperoleh hasil yang lebih baik sesuai dengan kenyataan yang dievaluasi. Contohnya, jika yang dievaluasi seberapa siswa mampu mengingat nama kota atau sungai, hasil evaluasinya berupa berapa banyak siswa dapat menyebutkan nama kota dan sungai yang diingat.

Dengan pengertian tersebut, maka alat evaluasi dikatakan baik apabila mampu mengevaluasi sesuatu yang dievaluasi dengan hasil seperti keadaan yang dievaluasi. Pada umumnya alat evaluasi dibedakan

menjadi dua jenis, yakni **Tes dan Non-Tes**. Agar para guru mengetahui dan trampil dalam mengadakan penilaian, dibawah ini dibahas secara umum mengenai kedua jenis alat penilaian.

1. ¹⁸S

Tes adalah suatu alat atau prosedur yang sistematis dan objektif untuk memperoleh data-data atau keterangan-keterangan yang diinginkan tentang seseorang, dengan cara yang boleh dik¹⁵kan tepat dan cepat.

Tes pada umumnya digunakan untuk menilai dan mengukur hasil belajar siswa, terutama hasil belajar kognitif berkenaan dengan penguasaan bahan pelajaran sesuai dengan tujuan pendidikan dan pengajaran. Namun tes juga dapat digunakan untuk menilai hasil belajar bidang afektif dan psikomotoris.

Dilihat dari segi bentuknya, macam macam Tes yaitu:

1) Tes Bentuk Uraian

Tes uraian adalah tes (seperangkat soal yang berupa tugas, pertanyaan) yang menuntut peserta didik untuk mengorganisasikan dan menyatakan jawabannya menurut kata-kata (kalimat) sendiri.

2) Tes Bentuk Objektif

Tes Objektif sering juga disebut tes dikotomi (dichotomously scored item) karena jawabannya antara benar atau salah da skornya antara 1 atau 0. Disebut tes objektif karena penilaiannya. objektif,

Tes objek³¹ menuntut peserta didik untuk memilih jawaban yang telah disediakan, memberikan jawabansingkat, dan melengkapi pertanyaan pertanyaan yang belum sempurna. Tes Objektif terdiri atas beberapa bentuk yaitu, benar-salah, pilihan ganda, menjodohkan, dan melengkapi atau jawaban singkat.

3) Tes Lisan
Tes lisan adalah tes yang menuntut jawaban¹⁸ dari peserta didik dalam bentuk lisan. Peserta didik akan mengucapkan jawaban dengan kata-katanya sendiri sesuai dengan pertanyaan atau perintah yang diberikan.

4) ¹⁸s Perbuatan
Tes perbuatan atau tes praktik adalah tes yang menuntut jawaban peserta didik dalam bentuk perilaku, tindakan, atau perbuatan.

³²

2. NON-TES

Non tes adalah cara penilaian hasil belajar peserta didik yang dilakukan tanpa menguji peserta didik tetapi dengan melakukan pengamatan secara sistematis.

Macam-macam Non -Tes adalah:

a. Observasi

Observasi adalah suatu proses pengamatan dan pencatatan secara sistematis, logis, objektif, dan rasional mengenai berbagai fenomena, baik dalam situasi yang

sebernarnya maupun dalam situasi buatan untuk mencapai tujuan tertentu. Tujuan utama observasi adalah: (1) untuk mengumpulkan data dan informasi mengenai suatu fenomena, baik yang berupa peristiwa maupun tindakan, baik dalam situasi yang sesungguhnya maupun situasi buatan (2) Untuk mengukur perilaku kelas, baik perilaku guru maupun perilaku peserta didik. Terutama kecakapan ⁴¹cial atau social skill

b. Wawancara

Wawancara adalah percakapan dua orang atau lebih yang berlangsung antara narasumber dan pewawancara dengan tujuan mengumpulkan data-data berupa informasi. Oleh karena itu, teknik wawancara adalah salah satu cara pengumpulan data, misalnya untuk penelitian tertentu.

Tujuan wawancara adalah sebagai berikut:

- Untuk memperoleh informasi secara langsung guna menjelaskan suatu hal atau situasi dan kondisi tertentu.
- Untuk melengkapi suatu penyelidikan ilmiah
- Untuk memperoleh data agar dapat mempengaruhi situasi atau orang tertentu

c. Skala Sikap

Skala sikap merupakan salah satu bentuk dari tipe skala pengukuran yang digunakan untuk mengukur sikap.

d. Skala Penilaian

Skala penilaian adalah salah satu bentuk pedoman observasi yang dipergunakan untuk mengumpulkan data individu dengan menggolongkan, menilai tingkah laku individu atau situasi dalam tingkatan-tingkatan tertentu. Skala penilaian memiliki kesamaan dengan ceklis. Meskipun terdapat perbedaan-perbedaan dengan ceklis. Karena ceklis digunakan untuk menandai apakah sebuah perilaku hadir atau tidak, sedangkan skala penilaian menghendaki penilaian dilakukan menurut pertimbangan kualitatif menyangkut tingkat kehadiran sebuah perilaku.

e. Angket

Angket termasuk alat untuk mengumpulkan dan mencatat data atau informasi, pendapat, dan paham dalam hubungan kausal. Angket mempunyai kesamaan dengan wawancara, kecuali dalam implementasinya. Angket dilaksanakan secara tertulis, sedangkan wawancara dilaksanakan secara lisan.

f. Studi Kasus

Studi kasus secara sederhana diartikan sebagai proses penyelidikan atau pemeriksaan secara mendalam, terperinci, dan detail pada suatu peristiwa tertentu atau khusus yang terjadi. Studi kasus dapat diperoleh dari metode-metode penelitian formal.

g. Catatan Insidental

Catatan Insidental adalah catatan-catatan singkat tentang peristiwa-peristiwa sepiantas yang dialami peserta didik secara

perseorangan. Catatan ini merupakan pelengkap dalam rangka penilaian guru terhadap peserta didiknya, terutama yang berkenaan dengan tingkah laku peserta didik.

h. Sosiometri

Sosiometri adalah suatu prosedur untuk merangkum, menyusun, dan sampai batas tertentu dapat mengkuantifikasi pendapat-pendapat peserta didik tentang penerimaan teman sebayanya serta hubungan di antara mereka.

i. Inventori Kepribadian

Inventori kepribadian hampir serupa dengan tes kepribadian. Bedanya pada inventori, jawaban peserta didik tidak memakai kriteria benar-salah. Semua jawaban peserta didik benar selama dia menyatakan yang sesungguhnya.

j. Teknik Pemberian Penghargaan kepada Peserta Didik

Teknik pemberian penghargaan ini dianggap penting karena banyak respons dan tindakan positif dari peserta didik yang timbul sebagai sebab akibat.

B. Penyusunan Soal

Penyusunan soal yang baik yakni dalam pelaksanaan tes lisan maupun tes tertulis, soal atau pertanyaan hendaknya dibuat sedemikian rupa sehingga dapat mengukur dengan baik kemampuan peserta. Dalam tes lisan, soal atau pertanyaan yang diajukan oleh pengajar hendaknya mencakup

materi yang telah dibahas sebelumnya dan pertanyaan yang diajukan bersifat penguatan sehingga tidak terlalu mendalam. Berbeda dengan tes tertulis, dimana dalam pelaksanaan tes tersebut telah dipersiapkan sebelumnya, sehingga soal atau pertanyaan yang dibuat dapat mencakup materi secara keseluruhan dengan pertanyaan yang mendalam, tergantung dari apa yang mau diukur. Secara umum, ada beberapa hal yang harus diperhatikan dalam penyusunan soal :

- a. Mengikuti langkah-langkah dan prosedur yang benar
- b. Mengikuti berbagai kaidah yang ada agar soal-soal yang dihasilkan membentuk perangkat tes yang valid
- c. Mengikuti syarat-syarat dalam penyusunan soal

Langkah-langkah Penyusunan Soal

- a. Pilih materi atau topik yang tepat untuk diangkat menjadi bahan ujian
- b. Tentukan dari materi atau topik terpilih tersebut, bagian mana yang tepat untuk dijadikan soal dengan bentuk pilihan ganda atau benar-salah atau isian, atau uraian, atau ujian praktek
- c. Buat kisi-kisi soalnya yang disesuaikan dengan indikator capaian dalam setiap materi atau topik terpilih
- d. Tulis soal mengacu pada indikator sebagaimana dituangkan dalam kisi-kisi
- e. Tulis kunci jawaban (untuk soal selain uraian) atau pedoman penskoran (untuk soal uraian)

- f. Penelaahan dan perakitan soal beserta kunci jawaban atau pedoman penskorannya.

Kaidah Penyusunan Soal

Masing-masing bentuk soal mempunyai kaidah penulisan yang berbeda. Secara umum, kaidah penulisan soal mencakup materi, konstruksi, dan bahasa.

Bentuk Soal Pilihan Ganda

Bentuk soal pilihan ganda paling banyak digunakan dalam melakukan tes secara masal dengan cakupan materi yang banyak. Dalam pembuatan soal pilihan ganda, kaidah umum yang berlaku adalah:

Materi

- a. Soal yang dibuat harus sesuai dengan indikator yang telah ditetapkan;
- b. Pilihan jawaban yang diberikan harus homogen dan logis;
- c. Setiap soal hanya memiliki satu jawaban yang benar

Konstruksi

- a. Pokok soal harus dirumuskan secara jelas dan tegas;
- b. Rumusan pokok soal dan pilihan jawaban harus merupakan pernyataan yang diperlukan saja;
- c. Pokok soal hendaknya jangan memberikan petunjuk kearah jawaban yang benar;
- d. Pokok soal jangan mengandung pernyataan yang bersifat negatif ganda;
- e. Panjang rumusan pilihan jawaban harus relatif sama;

- f. Pilihan jawaban jangan mengandung “semua pilihan jawaban di atas salah” atau “semua pilihan jawaban di atas benar”;
- g. Pilihan jawaban berbentuk angka harus disusun berdasarkan urutan besar- kecilnya nilai angka tersebut;
- h. Gambar, grafik, tabel, diagram, dan sejenisnya yang terdapat pada soal harus jelas dan berfungsi;
- i. Butir materi soal jangan bergantung pada jawaban soal sebelumnya.

Bahasa

- a. Setiap butir soal harus menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia yang baik dan benar;
- b. Jangan menggunakan bahasa yang berlaku setempat;
- c. Pilihan jawaban jangan mengulang kata atau frase yang bukan merupakan satu kesatuan pengertian.

Bentuk Soal Benar-Salah

- a. Hindari penggunaan kata: terpenting, selalu, sebagian kecil, tidak pernah, hanya, sebagian besar dan kata-kata lain yang sejenis
- b. Rumusan soal harus jelas dan pasti benar atau pasti salah
- c. Hindari pernyataan negatif;
- d. Hindari penggunaan kata yang dapat menimbulkan penafsiran ganda;
- e. Jumlah rumusan butir soal yang kunci jawabannya benar dan salah hendaknya dibuat berimbang;

- f. Panjang rumusan butir soal hendaknya relatif sama;
- g. Susunan pernyataan benar dan pernyataan salah hendaknya disusun secara random dan tidak sistematis mengikuti pola tertentu.

Bentuk Soal Uraian

- a. Harus sesuai dengan indikator yang telah ditetapkan sebagaimana dalam kisi-kisi soal
- b. Batasan jawaban atau ruang lingkup yang akan diukur harus jelas;
- c. Rumusan soal atau pertanyaan hendaknya menggunakan kata tanya yang menuntut jawaban uraian, misalnya : mengapa, jelaskan, uraikan;
- d. Rumusan kalimat soal hendaknya komunikatif dan hindari kalimat atau istilah atau kata yang dapat menimbulkan tafsir ganda;
- e. Hal-hal yang menyertai soal, seperti tabel, diagram, gambar atau sejenisnya harus disajikan secara jelas dan berfungsi
- f. Buatlah petunjuk yang jelas tentang cara mengerjakan soal;
- g. Butir soal harus dilengkapi dengan kunci jawaban atau kriteria jawaban serta pedoman penskoran dan harus segera dibuat setelah soal ditulis.

Syarat Penyusunan Soal

- a. Soal-soal atau pertanyaan yang dibuat harus menanyakan tentang ilmu/materi/konsep/teori/pengetahuan yang telah dipelajari;
- b. Soal disusun mengikuti kaidah penyusunan soal;

- c. Penulis soal harus menguasai ilmu yang akan dituliskan soalnya;
- d. Penulis soal mengetahui adanya ragam/jenis/bentuk soal beserta keunggulan dan kelemahannya;
- e. Penulis soal paham akan kaidah penulisan soal untuk berbagai bentuk soal;
- f. Penulis soal paham bahwa soal yang ditanyakan berhubungan langsung dengan penskoran sehingga diperoleh penilaian yang objektif.

11

ANALISIS BUTIR SOAL VALIDITAS

A. Pengertian Analisis Validitas⁷²

Validitas adalah indeks yang menunjukkan sejauh mana suatu alat pengukur itu mengukur apa yang ingin diukur. Menurut Azwar (1986) validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya.

Suatu alat pengukur dapat dikatakan mempunyai validitas yang tinggi apabila alat tersebut menjalankan fungsi ukurnya atau memberikan hasil ukur yang sesuai dengan tujuan dilakukannya pengukuran tersebut. Sedangkan jika suatu tes memiliki validitas yang rendah maka akan menghasilkan data yang tidak relevan dengan tujuan pengukuran.

Validitas biasanya meliputi tes dan skala yang terdiri atas sejumlah tugas yang dipilih berfungsi sebagai indikator hasil belajar. Validitas berhubungan dengan ketepatan alat penilaian terhadap konsep yang dinilai. Contohnya ketika menilai kemampuan peserta didik dalam pelajaran bahasa Indonesia, peserta didik diberi soal dengan kalimat yang panjang dan berbelit-belit sehingga memiliki daya sukar yang sulit dimengerti maknanya. Sehingga pada

akhirnya peserta didik tidak dapat menjawab karena tidak dapat memahami pertanyaanya.

Rumus yang digunakan untuk analisis validitas butir soal menurut Arikunto (2010: 213), yaitu:

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

Keterangan :

r_{xy} = koefisien korelasi yang dicari

N = banyaknya peserta tes

X = nilai variabel X (skor item)

Y = nilai variabel Y (skor item)

Jika $r_{hitung} > r_{tabel}$ maka butir item valid.

Dalam menggunakan validitas suatu tes, ada beberapa hal yang perlu diperhatikan, yaitu :

- a. Mengacu pada materi yang hendak diujikan.
- b. Mengacu pada hasil dari suatu tes atau instrument evaluasi yang dikenakan pada sekelompok individu.
- c. Berkaitan dengan derajat dengan istilah validasi tinggi, sedang, rendah.
- d. Mengacu pada penggunaan hasil evaluasi.

Validitas suatu instrumen evaluasi mempunyai beberapa makna penting diantaranya adalah sebagai berikut:

- a. Validitas berhubungan dengan ketepatan interpretasi hasil tes atau instrumen evaluasi untuk grup individual dan bukan instrumen itu sendiri.

- b. Validitas diartikan sebagai derajat yang menunjukkan kategori yang bisa mencakup kategori yang bisa mencakup kategori rendah, menengah, dan tinggi.
- c. Prinsip suatu tes valid, tidak universal. Validitas suatu tes yang perlu diperhatikan oleh para peneliti adalah bahwa ia hanya valid untuk suatu tujuan saja.

Ada dua unsur penting dalam validitas tes. Unsur tersebut adalah sebagai berikut:

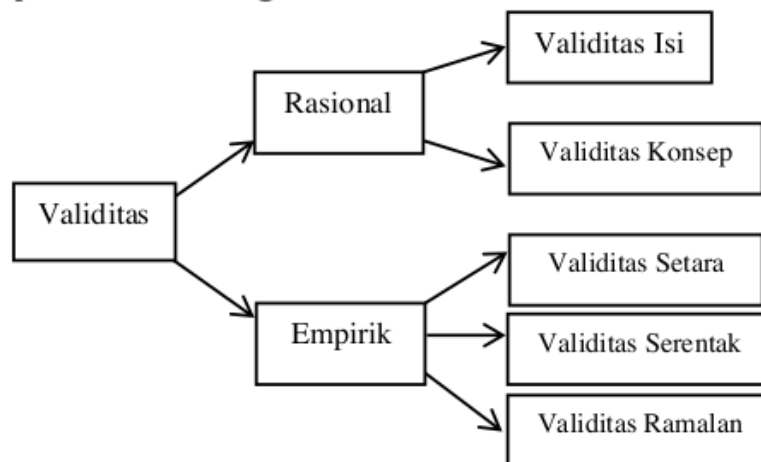
- a. Validitas suatu tes harus menunjukkan suatu derajat tertentu, ada yang sempurna, ada yang sedang, dan ada pula yang rendah.
- b. Validitas selalu dihubungkan dengan suatu putusan atau tujuan spesifik.

B. Jenis-jenis Validitas

Menurut Suryadi (2010) ada dua kenyataan pokok yang memperlihatkan taraf validitas suatu ujian, yaitu yang dipertimbangkan secara rasional dan yang dilihat melalui prosedur empirik. Analisis secara rasional dapat dilakukan terhadap topik dan bidang yang diujikan, yaitu isi ujian tersebut. Validitas yang diperoleh melalui analisis seperti ini disebut validitas isi (*contents validity*). Analisis rasional dapat juga dilakukan terhadap kegiatan-kegiatan dan proses-proses sesuai dengan konsep tertentu yang seharusnya menjadi isi dari ujian itu, dan hasilnya disebut validitas konsep atau konstruksi (*concept / construct validity*). Jenis kedua ialah kenyataan validitas yang bersifat empirik dan statistik. Jenis

ini diperoleh dengan memperhatikan hubungan yang ada antara alat (ujian) yang sedang dipelajari dengan pengukuran atau kenyataan-kenyataan yang lain. Bahanbahan pembanding ini mungkin diperoleh bersamaan waktunya dengan waktu penyelenggaraan ujian yang dimaksud mungkin juga tidak.

Sedangkan, validitas empiris instrumen diuji dengan cara membandingkan antara kriteria yang ada pada instrumen dengan fakta-fakta empiris yang terjadi di lapangan. Bila terdapat kesamaan antara criteria dalam instrumen dengan fakta di lapangan, maka dapat dinyatakan instrument tersebut mempunyai validitas yang tinggi. Untuk validitas jenis empirik ini akan dijelaskan tiga kenyataan validitas yaitu : validitas pengukuran setara (*congruent validity*), validitas pengukuran serentak (*concurrent validity*) dan validitas ramalan (*predictive validity*). Secara ilustrasi dapat dilihat sebagai berikut:



C. Validitas Isi (*Contens Validity*)

Validitas isi artinya ketepatan daripada suatu tes dilihat dari segi isi tersebut. Suatu tes hasil belajar dikatakan valid, apabila materi tes tersebut betul-betul merupakan bahan-bahan yang representatif terhadap bahan-bahan pelajaran yang diberikan. Dengan kata lain sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan. Oleh karena materi yang diajarkan tertera dalam kurikulum maka validitas isi ini sering disebut validitas kurikuler. Validitas isi dapat diusahakan tercapainya sejak saat penyusunan dengan cara memerinci materi kurikulum atau materi buku pelajaran.

Contoh Kasus Seorang peneliti ingin menguji validitas isi dari 1 item pernyataan pada kuesioner. sebuah skala yang terdiri dari 5 item. Sebanyak 12 orang panel ahli dijadikan penilai dengan memilih 3 pilihan jawaban yaitu "penting", "sesuai, tidak penting" dan "tidak berguna".

Data Contoh Validitas Isi Lawshe's CVR

Penilai	Penilaian
A	Penting
B	Penting
C	Tidak Berguna
D	Penting
E	Penting
F	Penting
G	Tidak Berguna
H	Penting
I	Penting
J	Penting
K	Sesuai, Tidak Penting
L	Penting

8

Dari 12 orang panel ahli, 9 orang menyatakan aitem tersebut “penting”, 1 orang menyatakan “sesuai, tidak penting”, dan 2 orang menyatakan “tidak berguna. Dari data ini kemudian dapat dihitung CVR sebagai berikut :

$$CVR = (2(9)/12) - 1 = 0.500$$

Formula ini menghasilkan nilai-nilai yang berkisar dari +1 sampai -1, nilai positif menunjukkan bahwa setidaknya setengah panelis (SME) menilai item sebagai penting/esensial. Semakin lebih besar CVR dari 0, maka semakin “penting” dan semakin tinggi validitas isinya. Dari contoh di atas diperoleh nilai CVR sebesar 0.500 yang menunjukkan bahwa aitem yang digunakan sudah memenuhi validitas isi yang baik (Lawshe, 1975).

D. Validitas Konsep (Concept / Construct Validity)

Sebuah tes dikatakan memiliki validitas konstruksi apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam Tujuan pembelajaran Khusus. Dengan kata lain jika butir-butir soal mengukur aspek berpikir tersebut sudah sesuai dengan aspek berpikir yang menjadi tujuan pembelajaran.

Untuk mengetahui apakah suatu tes memenuhi syarat-syarat validitas konstruksi atau tidak maka kita harus membandingkan susunan tes tersebut dengan syarat-syarat penyusunan tes yang baik. Apabila susunan tes tersebut telah memenuhi syarat-syarat penyusunan tes maka berarti tes tersebut memenuhi syarat validitas konstruksi, apabila tidak memenuhi syarat-syarat penyusunan tes berarti tidak memenuhi validitas konstruksi.

E. Validitas Pengukuran Setara (*Congruent Validity*)

Jenis kevalidan ini menunjukkan kenyataan yang diperoleh dengan menghubungkan hasil suatu ujian dengan pengukuran yang setara (mengukur fungsi yang sama). Begitu pula mengkorelasikan hasil sebuah tes intelegensi yang baru dengan hasil tes intelegensi yang sudah ada akan memberikan kenyataan validitas jenis ini.

Contoh validitas Konstruk

II. DUKUNGAN KELUARGA		
	A. Dukungan Informasi	SKOR
1.		
2.		
3.		
4.		
5.		
6.		
7..		
B. Dukungan Penilaian		
8.		
9.		
10.		
10.		
12.		
13.		
C. Dukungan Instrumental		
14.		
15.		
16.		
17.		
18.		
19		
D. Dukungan Emosional		
20.		
20.		
22.		
23.		

F. Validitas Pengukuran Serentak (*Concurrent Validity*)

Validitas ini lebih umum dikenal dengan validitas empiris. Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman. Jika istilah "sesuai" tentu ada dua

hal yang dipasangkan. Dalam hal ini hasil tes dipasangkan dengan hasil pengalaman. Pengalaman selalu mengenai hal yang telah lampau sehingga data pengalaman tersebut sekarang sudah ada. Dalam membandingkan hasil sebuah tes maka diperlukan suatu kriterium atau alat banding, maka hasil tes merupakan sesuatu yang dibandingkan.

Jenis validitas pengukuran serentak ini menunjukkan kenyataan yang diperhitungkan dengan mengkorelasikan hasil ujian yang dimaksud dengan suatu ukuran lain yang pengukurannya dilaksanakan bersamaan waktu dengan pelaksanaan ujian itu. Jika suatu tes yang memeriksa sosialibilitas misalnya, dikorelasikan dengan kedudukan sosialibilitas mahasiswa yang di tes itu menurut teman-temannya yang terdekat, maka hasilnya akan memperlihatkan kenyataan validitas jenis ini. Misalnya seorang guru ingin mengetahui apakah tes sumatif yang disusun sudah valid atau belum. Untuk ini diperlukan sebuah kriterium mas lalu yang sekarang datanya dia memiliki misalnya nilai ulangan harian atau nilai ulangan sumatif yang lalu.

G. Validitas Ramalan (*Predictive Validity*)

Memprediksi artinya meramal, dan meramal selalu mengenai hal yang akan datang jadi sekarang belum terjadi. Sebuah tes dikatakan memiliki validitas prediksi atau validitas ramalan apabila mempunyai kemampuan untuk meramalkan apa yang akan terjadi pada masa

yang akan datang. Jenis validitas ini menunjukkan kenyataan jika ujian yang dimaksud dihubungkan dengan kriteria-kriteria tentang hasil karya atau kesuksesan di masa depan. Demikianlah jika suatu tes bakat skolastik diberikan pada siswa-siswa SMA dikorelasikan dengan prestasi mereka di perguruan tinggi, maka kenyataan yang diperoleh itu akan menunjukkan validitas ramalan.

Misalnya tes masuk Perguruan Tinggi adalah sebuah tes yang diperkirakan mampu meramalkan keberhasilan peserta tes dalam mengikuti kuliah di masa yang akan datang. Calon yang tersaring berdasarkan hasil tes diharapkan mencerminkan tinggi-rendahnya kemampuan mengetahui kuliah. Jika nilai tesnya tinggi tentu menajmin keberhasilannya kelak. Sebaliknya seorang calon dikatakan tidak lulus tes karena memiliki nilai tes yang rendah diperkirakan akan tidak mampu mengikuti perkuliahan yang akan datang.

Sebagai alat pembanding validitas prediksi adalah nilai-nilai yang diperoleh setelah peserta tes mengikuti pelajaran di Perguruan Tinggi. Jika ternyata siapa yang memiliki nilai tes lebih tinggi gagal dalam ujian semester I dibandingkan dengan yang dahulu nilai tesnya lebih rendah maka tes masuk yang dimaksud tidak memiliki validitas prediksi.

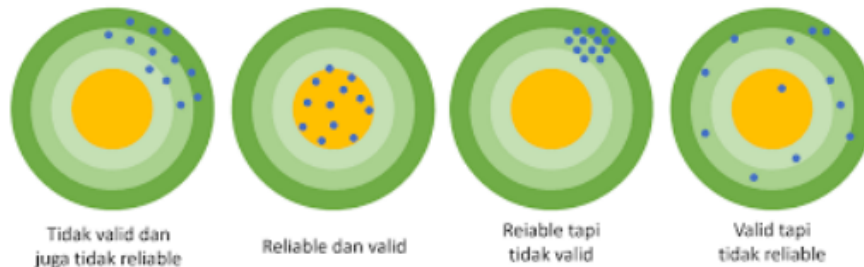
12

ANALISIS BUTIR SOAL RELIABILITAS

A. Pengertian Analisis Reliabilitas

Reliabilitas diterjemahkan dari kata reliability yang berarti hal yang dapat dipercaya (tahan uji). Sebuah tes dikatakan mempunyai reliabilitas yang tinggi jika tes tersebut memberikan data hasil yang ajeg (tetap) walaupun diberikan pada waktu yang berbeda kepada responden yang sama. Hasil tes yang tetap atau seandainya berubah maka perubahan itu tidak signifikan maka tes tersebut dikatakan reliabel. Oleh karena itu reliabilitas sering disebut dengan keterpercayaan, keterandalan, keajegan, konsistensi, kestabilan, dan sebagainya. Seperti diuraikan di atas sebuah alat ukur yang baik harus valid dan reliabel. Namun demikian validitas lebih penting dibandingkan dengan reliabilitas. Reliabilitas merupakan penyokong validitas. Sebuah alat ukur yang valid selalu reliabel. Akan tetapi alat ukur yang reliabel belum tentu valid, seperti digambarkan pada ilustrasi di atas. Seperti halnya validitas, reliabilitas juga merupakan tingkatan. Tingkat atau kadar reliabilitas dinyatakan dengan koefisien reliabilitas. Berikut ini akan dibahas macam-macam metode mencari besarnya koefisien reliabilitas.

Hubungan antara validitas dengan reliabilitas dapat digambarkan sebagaimana



tembakkan yang selalu tepat mengenai sasaran yang diinginkan, seperti ilustrasi pada gambar di bawah ini.

<https://berbagienergi.com/wp-content/uploads/2020/04/Validity.png>

Menurut Sugiono (2010) Reliabilitas adalah serangkaian pengukuran atau serangkaian alat ukur yang memiliki konsistensi bila pengukuran yang dilakukan dengan alat ukur itu dilakukan secara berulang. Reabilitas tes adalah tingkat keajegan (konsistensi) suatu tes, yakni sejauh mana suatu tes dapat dipercaya untuk menghasilkan skor yang ajeg, relatif tidak berubah walaupun ditekankan pada situasi yang berbeda-beda. Sedangkan Sukadji (2000) mengatakan bahwa reliabilitas suatu tes adalah seberapa besar derajat tes mengukur secara konsisten sasaran yang diukur. Reliabilitas dinyatakan dalam bentuk angka, biasanya sebagai koefisien. Koefisien tinggi berarti reliabilitas tinggi.

Menurut Nursalam (2003) Reliabilitas adalah kesamaan hasil pengukuran atau pengamatan bila fakta atau kenyataan hidup tadi diukur atau diamati

berkali-kali dalam waktu yang berlainan. Alat dan cara mengukur atau mengamati sama-sama memegang peranan penting dalam waktu yang bersamaan.

Menurut Arifin (1991), suatu tes dapat dikatakan andal (*reliable*) jika tes tersebut mempunyai hasil yang taat asas (konsisten). Sedangkan Sudjana (2004) mengatakan bahwa reliabilitas suatu tes adalah ketepatan atau kejegan tes tersebut dalam menilai apa adanya, artinya kapan pun tes tersebut digunakan akan memberikan hasil yang sama atau relatif sama. Berdasarkan beberapa pendapat tentang reliabilitas di atas, maka dapat diambil kesimpulan bahwa reliabilitas adalah tingkat konsistensi skor yang dicapai oleh orang yang sama dan tes yang sama pula ketika diuji pada waktu yang berbeda. Atau, konsistensi skor juga dapat diperoleh dengan soal yang berbeda tetapi memiliki kesamaan dari berbagai aspek.

B. Jenis – Jenis Reliabilitas

Salah satu syarat agar hasil suatu tes dapat dipercaya adalah tes tersebut harus mempunyai reliabilitas yang memadai. Oleh karena itu Jaali dan Pudji (2008) membedakan reliabilitas menjadi 2 macam, yaitu Reliabilitas Konsistensi tanggapan, dan Reliabilitas konsistensi gabungan item

➤ Reliabilitas Konsistensi Tanggapan

Reliabilitas ini selalu mempersoalkan mengenai tanggapan responden atau objek terhadap tes tersebut apakah sudah baik atau konsisten. Dalam artian apabila tes yang telah di cobakan

tersebut dilakukan pengukuran kembali terhadap obyek yang sama, apakah hasilnya masih tetap sama dengan pengukuran sebelumnya. Jika hasil pengukuran kedua menunjukkan ketidakonsistenan, maka hasil pengukuran tersebut tidak menggambarkan keadaan obyek yang sesungguhnya. Untuk mengetahui apakah suatu tes atau instrument tersebut sudah mantap atau konsisten, maka tes/instrument tersebut harus diuji kepada obyek ukur yang sama secara berulang-ulang. Ada tiga mekanisme untuk memeriksa reliabilitas tanggapan responden terhadap tes (Jaali ; 2008) yaitu:

1. **Teknik test-retest** ialah pengtesan dua kali dengan menggunakan suatu tes yang sama pada waktu yang berbeda.
 2. **Teknik belah dua** ialah pengtesan (pengukuran) yang dilakukan dengan dua kelompok item yang setara pada saat yang sama.
 3. **Bentuk ekivalen** ialah pengtesan (pengukuran) yang dilakukan dengan menggunakan dua tes yang dibuat setara kemudian diberikan kepada responden atau obyek tes dalam waktu yang bersamaan.
- Reliabilitas Konsistensi Gabungan Item
- Reliabilitas ini terkait dengan konsistensi antara item-item suatu tes atau instrument. Apabila terhadap bagian obyek ukur yang sama, hasil pengukuran melalui item yang satu kontradiksi atau tidak konsisten dengan hasil ukur melalui

item yang lain maka pengukuran dengan tes (alat ukur) sebagai suatu kesatuan itu tidak dapat dipercaya. Untuk itu jika terjadi hal demikian maka kita tidak bisa menyalahkan obyek ukur, melainkan alat ukur (tes) yang dipersalahkan, dengan mengatakan bahwa tes tersebut tidak reliable atau memiliki reliabilitas yang rendah.

Dalam menentukan reliabilitas sebuah alat evaluasi dalam hal ini instrumen tes, dapat dikelompokkan berdasarkan jenis instrumen tersebut, yaitu: (1) Tes Objektif, (2) Tes Uraian, dan (3) Tes Afektif.

C. Karakteristik Reliabilitas

1. Sebuah tes dianggap memiliki reliabilitas yang baik apabila memiliki karakteristik sebagaiberikut:
Reliabilitas merupakan milik dari satu set nilai tes bukan milik tes itu sendiri, artinya suatu tes dikatakan baik apabila dapat menghasilkan skor yang cukup akurat, apabila tes tersebut diberikan pada kelas tertentu, maka bisa juga menghasilkan skor yang cukup konsisten bila diberikan pada kelas yang berbeda atau ketika diberikan pada kelas yang sama pada waktu yang berbeda.
2. Suatu tes dikatakan reliable jika dua buah tes dilakukan pada jarak waktu yang berbeda dan menunjukkan skor yang tidak jauh berbeda.
3. Reliabilitas dapat dinyatakan untuk dua atau lebih pengukuran independen yang diperoleh dari tes yang sama untuk setiap anggota kelompok.

D. Metode Pengujian Reliabilitas

1. Metode tes ulang (tes re-tes estimate reliability)

Uji reliabilitas dengan metode tes ulang digunakan untuk mengetahui sejauh mana suatu pengukuran dapat diandalkan. Uji ini dilakukan sebanyak dua kali, pengukuran pertama dan ulangnya. Kedua pengukuran dapat dilakukan oleh orang yang sama atau berbeda. Dalam hal ini perlu diatur bahwa proses pengukuran kedua, keadaan yang diukur itu harus benar-benar sama. Selanjutnya hasil pengukuran yang pertama dan yang kedua dikorelasikan dan hasilnya menunjukkan reliabilitas dari tes ini. Hal penting yang perlu diperhatikan dalam pengukuran reliabilitas tes ulang adalah; jangka waktu antara kedua pengambilan penilaian, stabilitas yang diharapkan dari kinerja yang diukur. Secara umum, semakin lama antara interval pelaksanaan tes yang berulang, semakin rendah tingkat reliabilitasnya. Pendekatan tes ulang merupakan pemberian perangkat tes yang sama terhadap sekelompok subjek sebanyak dua kali dengan selang waktu yang berbeda. Asumsinya adalah bahwa skor yang dihasilkan oleh tes yang sama akan menghasilkan skor tampak yang relatif sama. Estimasi reliabilitas dengan pendekatan tes ulang akan menghasilkan koefisien stabilitas (stability). Untuk memperoleh koefisien reliabilitas melalui pendekatan tes ulang dapat dilakukan dengan menghitung koefisien korelasi linier antara distribusi skor subyek pada pemberian tes pertama dengan skor subjek pada pemberian tes kedua.

2. Metode Bentuk Paralel (Equivalent)

Tes paralel atau tes equivalent adalah dua buah tes yang mempunyai kesamaan tujuan, tingkat kesukaran dan susunan tetapi butir-butir soalnya berbeda, dalam istilah bahasa Inggris disebut alternate-forms method (parallel forms). Pengujian reliabilitas instrument dengan cara ini cukup dilakukan sekali, tetapi instrumennya dua, pada responden yang sama, waktu sama, instrument berbeda. Reliabilitas instrument dihitung dengan cara mengkorelasikan antara data instrument yang satu dengan data instrument yang dijadikan ekuivalen. Bila korelasi positif dan signifikan, maka instrument dapat dinyatakan reliable. Kelemahan dari metode ini adalah bahwa pengetes pekerjaannya berat karena harus menyusun dua seri tes. Lagipula harus tersedia waktu yang lama untuk mencobakan dua kali tes.

3. Metode Gabungan (parallel form and alternative form reliability estimate)

Pengujian reliabilitas ini dilakukan dengan cara mencobakan dua instrument yang ekuivalen itu beberapa kali, ke responden yang sama. Reliabilitas instrument dilakukan dengan mengkorelasikan dua instrument, setelah itu dikorelasikan pada pengujian kedua dan selanjutnya dikorelasikan silang. Jika dengan dua kali pengujian dalam waktu yang berbeda maka akan dapat dianalisis keenam koefisien reliabilitas. Bila keenam koefisien korelasi itu semuanya positif dan signifikan maka dapat dinyatakan bahwa instrument tersebut reliable.

D. Rumus Dalam Reliabilitas

Terdapat beberapa rumus dalam pengujian reliabilitas instrumen, antara lain; Spearman Brown, Flanagan, Rulon, Kuder Richardson (KR) dan Cronbanch Alpha.

a. Rumus Spearman-Brown

Keterangan:

r_i = reliabilitas instrument
 r_b = indeks korelasi antara dua belahan instrument
 N = banyaknya responden
 X = belahan pertama
 Y = belahan kedua

$$r_i = 2 \left(1 - \frac{V_1 - V_2}{V_t} \right) \quad V = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

b. Rumus Flanagan

Keterangan:

r_i = reliabilitas instrument

v_1 = varians belahan pertama (varian skor butir-butir ganjil)

v_2 = varians belahan kedua (varian skor butir-butir genap)

v_t = varians skor total

c. Rumus Rulon

$$r_i = 1 - \frac{V_d}{V_t}$$

$$r_i = \frac{k}{(k-1)} \left\{ \frac{s_t^2 - \sum p_i q_i}{s_t^2} \right\}$$

$$p_i = \frac{\text{banyaknya subjek yang skornya } 1}{N}$$

$$q_i = 1 - p_i \quad s_t^2 = \text{variens total}$$

$$r_i = \frac{2r_b}{1 + r_b} \quad r_b = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

Keterangan:

ri = reliabilitas instrument

Vt = varians total atau varians skor total

Vd = varians (variens difference)

d = skor pada belahan awal dikurangi skor pada belahan akhir

d. Rumus KR 20

Keterangan:

ri = reliabilitas instrument

k = banyaknya butir pertanyaan atau banyaknya soal

pi = proporsi subjek yang menjawab betul pada suatu butir (proporsi subjek yang mendapat skor 1)

e. Rumus KR 21

Keterangan:

ri = reliabilitas instrument

k = banyaknya butir pertanyaan atau banyaknya soal

p = skor rata-rata

f. Rumus Cronbanch Alpha

$$r_i = \frac{k}{k-1} \left(1 - \frac{\bar{p}(k-\bar{p})}{k\sigma_t^2} \right)$$

σ_t^2 = varians total

\bar{p} = skor rata-rata

Keterangan:

ri = reliabilitas instrumen

k = banyaknya butir pertanyaan atau banyaknya soal

$$r_i = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$

$\sum \sigma_b^2$ = jumlah varians butir

σ_t^2 = varians total

13

ANALISIS BUTIR SOAL PILIHAN GANDA

A. Tingkat Kesukaran Soal

Perhitungan tingkat kesukaran soal adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika suatu soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik. Suatu soal tes tidak terlalu sukar dan tidak pula terlalu mudah (Zainal Arifin:2013).

Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya.⁸

Seorang siswa akan menjadi hafal akan kebiasaan guru-gurunya dalam hal pembuatan soal ini. Misalnya saja guru A dalam memberikan ulangan soalnya mudah-mudah, sebaliknya guru B kalau memberikan ulangan soalnya sukar sukar. Dengan pengetahuannya tentang kebiasaan ini, maka siswa akan belajar giat jika menghadapi

ulangan dari guru B dan sebaliknya jika akan mendapat ulangan dari guru A, tidak mau belajar giat atau bahkan mungkin tidak mau belajar sama sekali.

Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (difficulty index) Besarnya indeks kesukaran antara 0,00 sampai dengan 1.0 Indeks kesukaran ini menunjukkan taraf kesukaran soal Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1.0 menunjukkan bahwa soalnya terlalu mudah.

Di dalam istilah evaluasi, indeks kesukaran ini diberi simbol P (p besar), singkatan dari kata "proporsi". Dengan demikian maka soal dengan P = 0,70 lebih mudah jika dibandingkan dengan P = 0,20. Sebaliknya soal dengan P = 0,30 lebih sukar daripada soal dengan P = 0,80.

Melihat besarnya bilangan indeks ini maka lebih cocok jika bukan disebut sebagai indeks kesukaran tetapi indeks kemudahan atau indeks fasilitas, karena semakin mudah soal itu, semakin besar pula bilangan indeksnya. Akan tetapi telah disepakati bahwa walaupun semakin tinggi indeksnya menunjukkan soal yang semakin mudah, tetapi tetap disebut indeks kesukaran.

$$P = \frac{a}{n}$$

21

Rumus mencari P adalah : Keterangan:

P = indeks kesukaran

B = banyaknya siswa yang menjawab soal itu dengan betul

JS = jumlah seluruh siswa peserta tes

Latihan:

Ada 20 orang dengan nama kode A s.d. T yang mengajarkan tes yang terdiri dari 20 soal. Jawaban tesnya dianalisis dan jawaban tertera seperti berikut ini

(1= jawaban betul; 0 = jawaban salah)

Siswa	Nomor Soal																				Skor siswa
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
A	1	1	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1	1	0	1	13
B	0	1	0	0	1	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	11
C	1	1	0	0	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	14
D	0	1	0	0	1	1	0	1	0	0	0	1	1	1	0	0	0	1	1	0	9
E	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	0	14
F	0	0	0	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0	0	8
G	1	0	0	1	0	0	1	1	0	1	0	1	1	1	1	1	1	0	1	1	13
H	0	0	0	1	0	0	1	1	0	0	0	1	1	0	1	0	1	0	1	1	9
I	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	17
J	0	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	0	1	1	0	13
K	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0	10
L	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	4
M	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	13
N	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	16
O	1	1	0	0	1	0	1	0	1	1	0	1	1	0	1	0	1	1	1	0	12
P	0	1	0	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0	1	1	10
Q	1	0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0	1	1	1	9
R	0	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1	0	1	11
S	1	1	0	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	1	1	14
T	0	1	0	1	1	0	1	1	0	0	0	1	1	1	0	0	0	1	0	1	10
Jumlah	10	14	4	9	15	6	18	17	3	11	10	18	20	10	9	7	10	14	13	13	

Contoh penggunaan

Misalnya jumlah siswa peserta tes dalam suatu kelas ada 40 orang. Dari 40 orang siswa tersebut 12 orang yang dapat mengerjakan soal nomor I dengan betul. Maka indeks kesukarannya adalah:

$$P = \frac{B}{JS}$$
$$= \frac{12}{40} = 0,30$$

Dari tabel yang disajikan tersebut, dapat ditafsirkan bahwa:

- Soal nomor 1 mempunyai taraf kesukaran $\frac{10}{20} = 0,5$
- Soal nomor 1 mempunyai taraf kesukaran $P = \frac{2}{20} = 0,1$
- Soal nomor 9 adalah soal yang tersukar karena hanya dapat dijawab betul

Menurut ketentuan yang sering diikuti, indeks kesukaran sering diklasifikasikan sebagai berikut:

- Soal dengan P 1,00 sampai 0,30 adalah soal sukar
- Soal dengan P 0,30 sampai 0,70 adalah soal sedang
- Soal dengan P 0,70 sampai 1,00 adalah soal mudah

Walaupun demikian ada yang berpendapat bahwa soal-soal yang dianggap baik, yaitu soal-soal sedang, adalah soal-soal yang mempunyai indeks kesukaran 0,30 sampai dengan 0,70.

Perlu diketahui bahwa soal-soal yang terlalu mudah atau terlalu sukar, lalu tidak berarti tidak

boleh digunakan. Hal ini tergantung dari penggunaannya. Jika dari pengikut yang banyak, kita menghendaki yang lulus hanya sedikit, kita ambil siswa yang paling top. Untuk ini maka lebih baik diambilkan butir-butir tes yang sukar.

Sebaliknya jika kekurangan pengikut ujian, kita pilihkan soal-soal yang mudah. Selain itu, soal yang sukar akan menambah gairah belajar bagi siswa yang pandai, sedangkan soal-soal yang terlalu mudah, akan membangkitkan semangat kepada siswa yang lemah.

Menghitung tingkat kesukaran soal bentuk objektif (pilihan ganda)

Rumus:

$$TK = ((WL + WH))/((nL + nH)) \times 100\%$$

Keterangan:

WL : jumlah peserta didik yang menjawab salah dari kelompok bawah

WH : jumlah peserta didik yang menjawab salah dari kelompok atas

nL : jumlah kelompok bawah

nH : jumlah kelompok atas

Sebelum menggunakan rumus di atas, harus ditempuh terlebih dahulu langkah-langkah sebagai berikut.

1. Menyusun lembar jawaban peserta didik dari skor tertinggi sampai dengan skor terendah.
2. Mengambil 27% lembar jawaban dari atas yang selanjutnya disebut kelompok atas (higher group), dan 27% lembar jawaban dari bawah

yang selanjutnya disebut kelompok bawah (lower group). Sisa sebanyak 46% disisihkan.

3. Membuat tabel untuk mengetahui jawaban (benar atau salah) dari setiap peserta didik, baik untuk kelompok atas maupun kelompok bawah. Jika jawaban peserta didik benar, diberi tanda + (plus), sebaliknya jika jawaban peserta didik salah, diberi tanda - (minus).

Contoh:

36 orang peserta didik SMP mengikuti Ujian Akhir Semester dalam mata pelajaran Teknologi Informasi dan Komunikasi. Berdasarkan hasil ujian tersebut kemudian disusun lembar jawaban peserta didik dari yang mendapat skor tertinggi sampai dengan skor terendah. Selanjutnya, diambil 27% dari skor tertinggi, yaitu $27\% \times 36 \text{ orang} = 9,72 = 10 \text{ orang}$ (dibulatkan) dan 27% dari skor terendah, yaitu $27\% \times 36 \text{ orang} = 9,72 = 10 \text{ orang}$ (dibulatkan). Setelah diketahui jumlah sampel kelompok atas dan kelompok bawah, kemudian membuat tabel untuk mengetahui jawaban (benar atau salah) dari setiap peserta didik dalam kelompok tersebut.

Tabel 1.1

Jawaban benar-salah dari kelompok atas

Peserta didik / no. soa	1	2	3	4	5	6	7	8	9	10
1	+	+	+	-	+	+	-	-	-	+
2	+	+	+	+	+	+	-	-	+	+
3	+	+	+	+	-	+	-	-	+	+
4	+	+	+	+	+	+	+	+	-	+
5	+	-	+	+	+	+	-	+	-	+
6	+	+	+	+	+	+	+	-	-	+

7	+	+	+	-	+	+	+	-	-	+
8	+	+	+	+	+	+	-	-	-	-
9	-	+	+	-	-	+	-	+	+	+
10	+	+	-	+	+	+	-	-	+	+

Tabel 1.2

Jawaban benar-salah dari kelompok bawah

Peserta didik / no. soal	1	2	3	4	5	6	7	8	9	10
1	+	-	+	-	-	+	+	-	-	-
2	+	-	-	-	+	+	+	-	-	-
3	+	+	-	+	-	+	-	-	-	-
4	+	-	-	-	-	+	-	+	-	+
5	+	-	+	-	+	+	-	-	-	-
6	+	-	+	-	+	+	+	+	-	+
7	+	+	-	+	-	+	-	-	+	-
8	-	+	+	+	+	-	-	+	-	+
9	+	+	-	+	+	-	-	+	-	-
10	-	-	+	-	+	+	-	-	+	+

1. Untuk soal nomor 1 pada kelompok bawah yang salah 6 orang, dan pada kelompok atas yang salah 4 orang.
2. Untuk soal nomor 2 pada kelompok bawah yang salah 6 orang, dan pada kelompok atas yang salah 2 orang.
3. Untuk soal nomor 3 pada kelompok bawah yang salah 6 orang, dan pada kelompok atas yang salah 3 orang.
4. Untuk soal nomor 4 pada kelompok bawah yang salah 6 orang, dan pada kelompok atas yang salah 1 orang.

5. Untuk soal nomor 5 pada kelompok bawah yang salah 6 orang dan pada kelompok atas yang salah 3 orang.
6. Untuk soal nomor 6 pada kelompok bawah yang salah 3 orang, dan pada kelompok atas yang salah 2 orang.
7. Untuk soal nomor 7 pada kelompok bawah yang salah 5 orang, dan pada kelompok atas yang salah 3 orang.
8. Untuk soal nomor 8 pada kelompok bawah yang salah 4 orang, dan pada kelompok atas yang salah 4 orang.
9. Untuk soal nomor 9 pada kelompok bawah yang salah 5 orang, dan pada kelompok atas yang salah 4 orang.
10. Untuk soal nomor 10 pada kelompok bawah yang salah 5 orang, dan pada kelompok atas yang salah 3 orang.

Berdasarkan data di atas dapat dibuat tabel seperti berikut:

Tabel 10.9
Perhitungan WL + WH dan WL - WH

No. Soal	WL	WH	WL + WH	WL - WH
1	6	4	10	2
2	6	2	8	4
3	6	3	9	3
4	6	1	7	5
5	6	3	9	3
6	3	2	5	1
7	5	3	8	2
8	4	4	8	0
9	5	4	9	1
10	5	3	8	2

Jadi, tingkat kesukaran setiap soal adalah sebagai berikut:

- 1) Untuk soal nomor 1: $TK = 10 / 20 \times 100\% = 50\%$
- 2) Untuk soal nomor 2: $TK = 8 / 20 \times 100\% = 40\%$
- 3) Untuk soal nomor 3: $TK = 9 / 20 \times 100\% = 45\%$
- 4) Untuk soal nomor 4: $TK = 7 / 20 \times 100\% = 35\%$
- 5) Untuk soal nomor 5: $TK = 9 / 20 \times 100\% = 45\%$
- 6) Untuk soal nomor 6: $TK = 5 / 20 \times 100\% = 25\%$
- 7) Untuk soal nomor 7: $TK = 8 / 20 \times 100\% = 40\%$
- 8) Untuk soal nomor 8: $TK = 8 / 20 \times 100\% = 40\%$
- 9) Untuk soal nomor 9 : $TK = 9 / 20 \times 100\% = 45\%$
- 10) Untuk soal nomor 10 : $TK = 8 / 20 \times 100\% = 40\%$

Adapun kriteria penafsiran tingkat kesukaran soal adalah:

- a. Jika jumlah persentase sampai dengan 27% termasuk mudah.
- b. jumlah persentase 28% - 72% termasuk sedang.
- c. Jika jumlah persentase 73% ke atas termasuk sukar.

Berdasarkan kriteria di atas, maka hasil perhitungan tingkat kesukaran soal dapat ditafsirkan seperti berikut:

Tabel 10.10
Penafsiran Hasil Perhitungan Tingkat Kesukaran Soal

Nomor Soal	Persentase Tingkat Kesukaran	Penafsiran
1	50 %	Sedang
2	40 %	Sedang
3	45 %	Sedang
4	35 %	Sedang
5	45 %	Sedang
6	25 %	Mudah
7	40 %	Sedang
8	40 %	Sedang
9	45 %	Sedang
10	40 %	Sedang

Tabel 10.11
Klasifikasi Soal Berdasarkan Proporsi Tingkat Kesukarannya

Tingkat Kesukaran Soal	Nomor Soal	Jumlah
Mudah P 27 %	6	1 (10 %)
Sedang P 28 % - 72%	1,2,3,4,5,7,8,9,10	9 (90%)
Sukar P 73 %	0	10 (100 %)

Untuk memperoleh prestasi belajar yang baik, sebaiknya proporsi antara tingkat kesukaran soal tersebar secara normal. Perhitungan proporsi tersebut dapat diatur sebagai berikut:

- a. Soal sukar 25%, soal sedang 50%, soal mudah 25%, atau
- b. Soal sukar 20%, soal sedang 60%, soal mudah 20%, atau
- c. Soal sukar 15%, soal sedang 70%, soal mudah 15%.

Seharusnya, penyelesaian suatu soal dilakukan dengan mempertimbangkan tingkat kesukaran soal, sehingga hasil yang dicapai peserta didik dapat menggambarkan prestasi yang sesungguhnya.

B. Daya Pembeda Soal

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah (Arikunto 1999). 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. Untuk menghitung daya pembeda setiap butir soal dapat digunakan rumus sebagai berikut:

$$DP = ((WL - WH))/n$$

Keterangan:

DP : daya pembeda

WL : jumlah peserta didik

WH : jumlah peserta yang gagal dari kelompok bawah

n : $27\% \times N$

Daya pembeda dihitung dengan mengurangi proporsi peserta kelompok atas yang menjawab benar dengan proporsi peserta kelompok bawah yang menjawab benar. Hasil perhitungan tersebut diinterpretasikan dalam empat kriteria yaitu:

0,00-0,19 = jelek (poor)

0,20-0,39 = cukup (satisfactory)

0,40-0,69 = baik (good)

0,70-1,00 = baik sekali (excellent)

Arifin (Arifin, 2012) menyatakan bahwa "perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai materi dengan peserta didik yang belum atau kurang menguasai materi berdasarkan kriteria tertentu." Jadi dapat disimpulkan bahwa soal post test pada pelatihan KKMN Angkatan 1 memiliki kualitas yang kurang baik dilihat dari segi daya pembeda yaitu lebih dari 50% keseluruhan soal, sehingga dapat membedakan peserta didik kelompok atas dan kelompok bawah. Sebagai tindak lanjut dari analisis daya pembeda, butir soal yang telah memiliki daya pembeda baik (kriteria cukup dan baik) sebaiknya disimpan ke dalam bank soal, dan butir soal yang memiliki daya pembeda jelek agar diperbaiki sehingga bisa disimpan ke dalam bank soal untuk digunakan pada tes hasil belajar yang akan datang. Sedangkan untuk butir soal dengan daya pembeda negatif sebaiknya

dibuang dan tidak digunakan pada tes hasil belajar yang akan datang, karena butir soal tersebut memiliki kualitas yang sangat jelek.

C. Pengecoh Soal

Pada soal bentuk pilihan ganda ada alternatif jawaban (opsi) yang merupakan pengecoh. Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya, butir soal yang kurang baik, pengecohnya akan dipilih secara tidak merata.

Pengecoh dianggap baik bila jumlah peserta didik yang memilih pengecoh itu sama atau mendekati jumlah ideal. Indeks pengecoh dihitung dengan rumus:

$$IP = P / ((N - B) / (n - 1)) \times 100\%$$

Keterangan:

IP : indeks pengecoh

P : jumlah peserta didik yang memilih pengecoh

N : jumlah peserta didik yang ikut tes

B : jumlah peserta didik yang menjawab benar pada setiap soal

n : jumlah alternatif jawaban (opsi)

1 : bilangan tetap

Catatan:

Jika, semua peserta didik menjawab benar pada butir soal tertentu (sesuai kunci jawaban), maka $IP = 0$ yang berarti soal tersebut jelek. Dengan demikian, pengecoh tidak berfungsi.

Contoh :

50 orang peserta didik di tes dengan 10 soal bentuk pilihan-ganda. Tiap soal memiliki 5 alternatif jawaban (a, b, c, d dan e). Kunci jawaban (jawaban yang benar) soal nomor 8 adalah c. Setelah soal nomor 8 diperiksa untuk semua peserta didik, ternyata dari 50 orang peserta didik, 20 peserta didik menjawab benar dan 30 peserta didik menjawab salah. Idealnya, pengecoh dipilih secara merata, artinya semua pengecoh secara merata ikut menyesatkan peserta didik.

Perhatikan contoh soal nomor 8 berikut ini:

Alternatif jawaban	a	b	c	d	e
Distribusi jawaban peserta didik	7	8	20	7	8
IP	93%	107%	**	93%	107%
Kualitas pengecoh	++	++	**	++	++
Keterangan :					

** = kunci jawaban

++ = sangat baik

+ = baik

- = kurang baik

- = jelek

- = sangat jelek

Pada contoh di atas, IP butir a, b, d, dan e adalah 93%, 107%, 93% dan 107%, Semuanya dekat dengan angka 100%, sehingga digolongkan sangat baik sebab semua pengecoh itu berfungsi. Jika pilihan jawaban

peserta didik menumpuk pada satu alternatif jawaban, misalnya seperti berikut:

Alternatif jawaban	a	b	c	d	e
Distribusi jawaban peserta didik	20	2	20	8	0
IP	267%	27%	**	107%	0%
Kualitas pengecoh	-	-	**	++	-

Dengan demikian, dapat ditafsirkan pengecoh (d) yang terbaik, pengecoh (e) dan (b) tidak berfungsi, pengecoh (a) menyesatkan, maka pengecoh (a) dan (e) perlu diganti karena termasuk jelek, dan pengecoh (b) perlu direvisi karena kurang baik. Adapun kualitas pengecoh berdasar indeks pengecoh adalah:

Sangat baik	IP = 76% -125%
Baik	IP = 51% -75% atau 126% 150%
Kurang baik	IP = 26% - 50% atau 151% -175%
Jelek	IP = 0%-25% atau 176% - 200%
Sangat jelek	IP = lebih dari 200%

Untuk analisis pengecoh perlu dibuat tabel khusus agar setiap butir soal diketahui berapa banyak peserta didik yang menjawab a, b, c dan seterusnya. Hal ini tentu saja sangat memakan waktu dan tenaga. Jika diolah dengan komputer dan data sudah dimasukkan dalam disket, pengolahan ini hanya memerlukan waktu beberapa detik saja.

Contoh :

Pengecoh dikatakan baik apabila dipilih $\geq 5\%$ dari jumlah peserta tes. Dalam menginterpretasikan

Tingkat pengecoh untuk soal nomor 1 :

Item Soal	Alternative jawaban					Keterangan
	A	B	C	D	E	
	0	7	3	0	1	B (Kunci Jawaban)

$A. P = \frac{P}{n-B} \times 100\%$ $= \frac{0}{11-7} \times 100\%$ $= \frac{0}{4} \times 100\% = \frac{0}{1} \times 100\% = 0\%$	$C. P = \frac{P}{n-B} \times 100\%$ $= \frac{3}{11-7} \times 100\%$ $= \frac{3}{4} \times 100\% = \frac{3}{1} \times 100\% = 3\%$
$D. P = \frac{P}{n-B} \times 100\%$ $= \frac{0}{11-7} \times 100\%$ $= \frac{0}{4} \times 100\% = \frac{0}{1} \times 100\% = 0\%$	$E. P = \frac{P}{n-B} \times 100\%$ $= \frac{1}{11-7} \times 100\%$ $= \frac{1}{4} \times 100\% = \frac{1}{1} \times 100\% = 100\%$

DISTRIBUSI JAWABAN PESERTA DIDIK

Kualitas Pengecoh	Presentasi	Opsi
Jelek	0%	A
-	-	B
Jelek	3%	C
Jelek	0%	D
Sangat baik	100%	E

Efektifitas pengecoh setiap butir soal, digunakan kriteria yang diadaptasi dari skala likert sebagai berikut:

- Efektivitas pengecoh dikatakan sangat baik apabila keempat pengecoh berfungsi.
- Efektivitas pengecoh dikatakan baik apabila terdapat tiga pengecoh yang berfungsi.

- c. Efektivitas pengecoh dikatakan cukup baik apabila terdapat dua pengecoh yang berfungsi.
- d. Efektivitas pengecoh dikatakan kurang baik apabila terdapat satu pengecoh yang berfungsi.
- c. Efektivitas pengecoh dikatakan tidak baik apabila semua pengecoh tidak berfungsi.

14

ANALISIS BUTIR SOAL URAIAN

50

A. Tingkat Kesukaran

Tingkat kesukaran (*difficulty*) merupakan peluang untuk menjawab benar suatu butir pertanyaan pada suatu tingkat kemampuan atau bisa dikatakan untuk mengetahui sebuah butir tersebut tergolong mudah atau sukar. Butir yang memiliki tingkat kesukaran yang rendah tergolong dalam tingkat kesukaran rendah sehingga kemampuan yang diperlukan untuk menyelesaikan permasalahan pada butir tersebut juga rendah. Sebaliknya, butir dengan tingkat kesukaran yang tinggi akan sulit diselesaikan oleh peserta yang memiliki kemampuan rendah. Perlu digaris bawahi bahwa tingkat kesukaran berbeda dengan level kognitif (C1-C6). Bisa saja butir yang menanyakan pertanyaan C1 sulit diselesaikan oleh peserta didik, sedangkan butir yang menanyakan pertanyaan C5 atau C6 justru bisa diselesaikan (D'Sa, 201) dalam (Bhakti, Y. B, dkk. 2022:137).¹

75

¹ Bhakti, Y. B, dkk. (2022). *Evaluasi Pembelajaran dalam Bidang Pendidikan*. Yogyakarta: CV. Bintang Semesta Media, 137.

Instrumen yang baik terdiri dari butir-butir instrumen yang tidak terlalu mudah dan tidak terlalu sukar. Butir soal yang terlalu mudah tidak mampu merangsang audiens untuk mempertinggi usaha memecahkannya. Sebaliknya, jika terlalu sukar membuat audiens putus asa dan tidak memiliki semangat untuk mencoba lagi karena di luar jangkauannya. Di dalam istilah evaluasi, indeks kesukaran ini diberi simbol "P" yang dinyatakan dengan "proporsi" (Febriana, R. 2019:128).² Untuk menghitung tingkat kesukaran soal bentuk uraian adalah dengan menghitung persentase peserta didik yang gagal menjawab benar atau di bawah batas lulus (Rukajat, A. 2018:139).³

Berikut beberapa cara untuk menganalisis butir soal uraian tingkat kesukaran:

1. Menurut Arifin. (2011:266), tingkat kesukaran dapat dihitung berapa persen peserta didik yang menjawab salah atau kurang dari batas lulus. Rumus menghitung tingkat soal uraian.⁴

$$TK = \frac{\sum s}{N} \times 100\%$$

Keterangan

$\sum s$: Jumlah peserta didik yang gagal/menjawab salah

² Febriana, R. (2019). *Evaluasi Pembelajaran*. Jakarta : PT Bumi Aksara, 128.

³ Rukajat, A. (2018). *Teknik Evaluasi Pembelajaran*. Yogyakarta: D35ublish, 139.

⁴ Arifin, Z. (2011). *Evaluasi Pembelajaran*. Bandung: P T Remaja Rosdakarya, 266.

N : Jumlah peserta didik yang mengikuti tes

Kriteria tingkat kesukaran soal uraian :

Jika jumlah peserta didik yang gagal mencapai 27% termasuk mudah.

Jika jumlah peserta didik yang gagal antara 28%-72% termasuk sedang.

Jika jumlah peserta didik yang gagal mencapai 75% ke atas termasuk sukar.

Contoh :

10 siswa mengikuti Ulangan Tengah Semester mata pelajaran Bahasa Indonesia sebanyak 10 soal uraian. Skor maksimal ialah 10 sedangkan skor minimal 0.

Tabel 2.1 Rekapitulasi Jawaban Soal Uraian

No.	Nama	Nomor Soal										Skor Total	Kuadrat Skor Total
		1	2	3	4	5	6	7	8	9	10		
1.	Ani	10	10	10	10	10	10	10	10	10	10	100	10000
2.	Bayu	10	8	10	0	10	10	10	10	10	10	88	7744
3.	Cantik	8	10	10	0	10	10	10	10	10	10	88	7744
4.	Dani	10	6	8	10	10	10	10	6	6	10	86	7396
5.	Eka	10	0	10	0	10	8	10	10	8	10	76	5776
6.	Fina	6	10	10	0	10	10	10	8	10	0	74	5476
7.	Gani	10	0	10	0	10	10	8	0	10	10	68	4624
8.	Hito	10	10	10	0	6	10	10	0	10	0	66	4356
9.	Ira	10	0	10	0	10	10	0	10	0	0	50	2500
10.	Jeny	8	10	10	10	0	10	0	0	0	0	48	2304
Jumlah Benar		92	64	98	30	86	98	78	64	74	60	744	57920
Jumlah Salah		0	3	0	7	1	0	2	3	1	4		

Tabel 2.2 Perhitungan Tingkat Kesukaran Soal Uraian

1	$TK = \frac{\sum s}{N} \times 100\% =$ $\frac{0}{10} \times 100\% = 0\%$ (Soal nomor 1 mempunyai tingkat kesukaran mudah)	6	$TK = \frac{\sum s}{N} \times 100\% = \frac{0}{10} \times 100 = 0\%$ (Soal nomor 6 mempunyai tingkat kesukaran mudah)
2	$TK = \frac{\sum s}{N} \times 100\% =$ $\frac{3}{10} \times 100\% = 30\%$ (Soal nomor 2 mempunyai tingkat kesukaran sedang)	7	$TK = \frac{\sum s}{N} \times 100\% = \frac{2}{10} \times 100\% =$ 20% (Soal nomor 7 mempunyai tingkat kesukaran mudah)
3	$TK = \frac{\sum s}{N} \times 100\% =$ $\frac{0}{10} \times 100\% = 0\%$ (Soal nomor 3 mempunyai	8	$TK = \frac{\sum s}{N} \times 100\% = \frac{3}{10} \times 100\% =$ 30% (Soal nomor 8 mempunyai tingkat

	tingkat kesukaran mudah)		kesukaran sedang)
4	$TK = \frac{\sum s}{N} \times 100\% =$ $\frac{7}{10} \times 100\% = 70\%$ (Soal nomor 4 mempunyai tingkat kesukaran sedang)	9	$TK = \frac{\sum s}{N} \times 100\% = \frac{1}{10} \times 100\% =$ 10% (Soal nomor 9 mempunyai tingkat kesukaran mudah)
5	$TK = \frac{\sum s}{N} \times 100\% =$ $\frac{1}{10} \times 100\% = 10\%$ (Soal nomor 5 mempunyai tingkat kesukaran mudah)	10	$TK = \frac{\sum s}{N} \times 100\% = \frac{4}{10} \times 100\% =$ 40% (Soal nomor 10 mempunyai tingkat kesukaran sedang)

Tabel 2.3 Penafsiran Hasil Perhitungan Tingkat Kesukaran Soal Uraian

Nomor Soal	Hasil Perhitungan	Penafsiran
1	0%	Mudah
2	30%	Sedang
3	0%	Mudah
4	70%	Sedang
5	10%	Mudah
6	0%	Mudah
7	20%	Mudah
8	30%	Sedang
9	10%	Mudah
10	40%	Sedang

2. Menurut Setiawan, D. F. (2018:63-65), cara menghitung tingkat kesukaran soal non objektif (soal uraian) sebenarnya tidak jauh berbeda dengan tingkat kesukaran pada soal objektif. Cara yang dapat digunakan adalah dengan menghitung persentase siswa yang gagal menjawab benar atau berada di bawah kriteria ketuntasan minimal untuk tiap soal. Kriteria ketuntasan minimal (KKM) adalah 75% dari jumlah skor maksimal ideal yang diperoleh. Sebagai contoh, siswa akan mendapatkan skor 10 jika mampu menjawab soal uraian nomor 1 sesuai kriteria. Berdasarkan ketentuan tersebut, KKM adalah $10 \times 75\% = 7,5$. Jika siswa mendapatkan skor di bawah 7,5 maka

dapat disimpulkan bahwa siswa tersebut berada dibawah kriteria ketuntasan minimal untuk soal nomor 1. Metode perhitungan dapat diformulasikan dengan persamaan 3 sebagai berikut.⁵

Rumus

$$TK = \frac{\sum G}{N}$$

Keterangan

42

TK : Tingkat Kesukaran

$\sum G$: Jumlah siswa yang gagal

N : Jumlah siswa

Kriteria yang dapat digunakan untuk memutuskan apakah soal berada pada tingkat sulit, sedang atau mudah sama seperti kriteria pada metode komparasi tunggal di mana pada tingkat kegagalan siswa dalam menjawab soal baik pada kelompok atas maupun kelompok bawah. Semakin tinggi tingkat kegagalan siswa maka soal dapat dikatakan semakin mudah. Kriteria tersebut dapat dirinci sebagai berikut.

74

⁵ Setiawan, D. F. (2018). *Prosedur Evaluasi dalam Pembelajaran*. Yogyakarta: Deepublish, 63-65.

1. Jika jumlah presentase kegagalan berada pada 0-27,9% maka soal masuk pada kategori mudah.
2. Jika jumlah presentase kegagalan berada pada 28%-72,9% maka soal masuk pada kategori sedang.
3. Jika jumlah presentase kegagalan berada diatas 73% maka soal masuk pada kategori sulit.

Contoh

10 siswa mengikuti tes formatif mata pelajaran ekonomi pada standar kompetensi menerapkan pemecahan masalah ekonomi sejumlah 2 soal uraian. Skor maksimal ideal pada soal nomor 1 adalah 4 sedangkan skor maksimal ideal soal nomor 2 adalah 10. Perolehan skor pada dua soal tersebut adalah sebagai berikut.

Tabel 2.4 Tingkat Kesukaran Soal Uraian

Siswa	Skor pada Soal Nomor	
	1	2
1	3	8
2	4	7
3	4	6
4	2	9
5	4	10
6	1	10
7	1	9
8	3	8
9	2	10
10	2	5
KKM	3	7,5
Jumlah siswa yang gagal pada soal nomor	5	3
TK	50%	30%

78

Berdasarkan tabel 2.5 dapat diketahui bahwa jumlah siswa yang gagal mendapatkan skor minimal 3 adalah 3 orang sedangkan pada soal nomor 2 jumlah siswa yang gagal mendapatkan skor di bawah 7,5 adalah 3 orang. Tingkat kesukaran pada soal nomor 1 dapat dihitung menggunakan persamaan 3 sebagai berikut.

$$TK = \frac{5}{10}$$

Hasil perhitungan tersebut menunjukkan bahwa soal nomor 1 memiliki tingkat kesulitan sedang karena berada pada tingkat keberhasilan 30%-70,9%. Perhitungan untuk soal nomor 2 dan seterusnya dapat menggunakan metode dan formula yang sama.

3. Menurut Suryadi, A. (2020:28-29), 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 tiap-tiap soal. Untuk menafsirkan tingkat kesukaran soalnya dapat digunakan kriteria sebagai berikut:⁶
 1. Jika jumlah peserta didik yang gagal mencapai 27% termasuk mudah.
 2. Jika jumlah peserta didik yang gagal antara 28% sampai dengan 72% termasuk sedang.
 3. Jika jumlah peserta didik yang gagal antara 72% keatas, termasuk sukar.

Contoh

33 orang peserta didik dites dengan lima soal bentuk uraian. Skor maksimum ditentukan 10 dan skor minimum 0. Jumlah peserta didik yang memperoleh nilai 0-5 = 10 orang (berarti gagal), nilai 6 = 12 orang, dan nilai 7-10 = 11

⁶ Suryadi, A. (2020). *Evaluasi Pembelajaran Jilid II*. Jawa Barat: CV Jejak, anggota IKAPI, 28-29.

orang. Jadi, tingkat kesukaran (TK) = $\frac{10}{33} \times 100\% = 30,3\%$ Tingkat kesukaran 30,3 berada diantara 28 dan 72, berarti soal tersebut termasuk sedang.

Catatan: Batas lulusan ideal = 6 (skala 0-10)

4. Menurut Nurgiyantoro, B. (2010:201), rumus yang digunakan untuk menganalisis tingkat kesukaran soal uraian adalah⁷

$$\text{Indeks Tingkat Kesukaran} = \frac{St + Sr - (2N \times \text{Skor min})}{2N \times (\text{Skor maks} - \text{skor min})}$$

Keterangan

St : Jumlah skor benar pada kelompok atas

Sr : Jumlah skor benar pada kelompok bawah

Skor maks : Skor maksimal suatu butir soal

Skor min : Skor minimal suatu butir soal

N : Jumlah siswa

Kriteria presentase tingkat soal uraian :

1. 0,00-0,30 : soal tergolong sukar
2. 0,31-0,70 : soal tergolong sedang
3. 0,71-1,00 : soal tergolong mudah

59

⁷ Nurgiyantoro, B. (2010). *Penilaian Pembelajaran Bahasa Indonesia (Berbasis Kompetensi)*. BPFE-YOGYAKARTA, 201.

Contoh

10 siswa mengikuti Ulangan Tengah Semester mata pelajaran Bahasa Indonesia sebanyak 10 soal uraian. Skor maksimal ialah 10 sedangkan skor minimal 0.

Tabel 2.5 Rekapitulasi Jawaban Soal Uraian

No.	Nama	Nomor Soal										Skor Total	Kuadrat Skor Total
		1	2	3	4	5	6	7	8	9	10		
1.	Ani	10	10	10	10	10	10	10	10	10	10	100	10000
2.	Bayu	10	8	10	0	10	10	10	10	10	10	88	7744
3.	Cantik	8	10	10	0	10	10	10	10	10	10	88	7744
4.	Dani	10	6	8	10	10	10	10	6	6	10	86	7396
5.	Eka	10	0	10	0	10	8	10	10	8	10	76	5776
6.	Fina	6	10	10	0	10	10	10	8	10	0	74	5476
7.	Gani	10	0	10	0	10	10	8	0	10	10	68	4624
8.	Hito	10	10	10	0	6	10	10	0	10	0	66	4356
9.	Ira	10	0	10	0	10	10	0	10	0	0	50	2500
10.	Jeny	8	10	10	10	0	10	0	0	0	0	48	2304
Jumlah Benar		92	64	98	30	86	98	78	64	74	60	744	57920
Jumlah Salah		0	3	0	7	1	0	2	3	1	4		

$$\text{Kelompok atas} = 27/100 \times 10\% = 2,7\% = 3$$

$$\text{Kelompok bawah} = 27/100 \times 10\% = 2,7\% = 3$$

Tabel 2.6 Kelas Atas Soal Uraian

No.	Nama	Nomor Soal										Skor Total	Kuadrat Skor Total
		1	2	3	4	5	6	7	8	9	10		
1.	Ani	10	10	10	10	10	10	10	10	10	10	100	10000
2.	Bayu	10	8	10	0	10	10	10	10	10	10	88	7744
3.	Cantik	8	10	10	0	10	10	10	10	10	10	88	7744

Tabel 2.7 Kelas Bawah Soal Uraian

No.	Nama	Nomor Soal										Skor Total	Kuadrat Skor Total
		1	2	3	4	5	6	7	8	9	10		
1.	Hito	10	10	10	0	6	10	10	0	10	0	66	4356
2.	Ira	10	0	10	0	10	10	0	10	0	0	50	2500
3.	Jeny	8	10	10	10	0	10	0	0	0	0	48	2304

Tabel 2.8. Perhitungan Tingkat Kesukaran Soal Uraian

1	<p>Indeks Tingkat Kesukaran =</p> $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks}-\text{skor min})}$ $\frac{28+28-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ <p>0,28</p> <p>(Soal nomor 1 mempunyai tingkat kesukaran sukar)</p>
2	<p>Indeks Tingkat Kesukaran =</p> $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks}-\text{skor min})}$ $\frac{28+20-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$

	0,24 (Soal nomor 2 mempunyai tingkat kesukaran sukar)
3	Indeks Tingkat Kesukaran = $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+30-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ 0,3 (Soal nomor 3 mempunyai tingkat kesukaran sukar)
4	Indeks Tingkat Kesukaran = $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{10+10-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ 0,1 (Soal nomor 4 mempunyai tingkat kesukaran sukar)
5	Indeks Tingkat Kesukaran = $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+16-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ 0,23 (Soal nomor 5 mempunyai tingkat kesukaran sukar)
6	Indeks Tingkat Kesukaran = $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+30-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ 0,3 (Soal nomor 6 mempunyai tingkat kesukaran sukar)
7	Indeks Tingkat Kesukaran = $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$

	$2N \times (\text{Skor maks-skor min})$ $\frac{30+10-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ $0,2$ <p>(Soal nomor 7 mempunyai tingkat kesukaran sukar)</p>
8	<p>Indeks Tingkat Kesukaran =</p> $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+10-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ $0,2$ <p>(Soal nomor 8 mempunyai tingkat kesukaran sukar)</p>
9	<p>Indeks Tingkat Kesukaran =</p> $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+10-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ $0,2$ <p>(Soal nomor 9 mempunyai tingkat kesukaran sukar)</p>
10	<p>Indeks Tingkat Kesukaran =</p> $\frac{St+Sr-(2N \times \text{Skor min})}{2N \times (\text{Skor maks-skor min})}$ $\frac{30+0-(2 \times 10 \times 0)}{2 \times 10 (10-0)}$ $0,15$ <p>(Soal nomor 10 mempunyai tingkat kesukaran sukar)</p>

Tabel 2.9. Penafsiran Hasil Perhitungan Tingkat Kesukaran Soal Uraian

Nomor Soal	Hasil Perhitungan	Penafsiran
1	0,28	Sukar
2	0,24	Sukar
3	0,3	Sukar
4	0,1	Sukar
5	0,23	Sukar
6	0,3	Sukar
7	0,2	Sukar
8	0,2	Sukar
9	0,2	Sukar
10	0,15	Sukar

- Dali (1992: 158) dalam (Hairun, Y. 2020:121-122), mengemukakan taraf kesukaran butir adalah ciri butir sedangkan keberhasilan peserta adalah ciri peserta. Hubungan proporsi jawaban benar pada butir dengan keberhasilan peserta pada uji tes dapat dikatakan sebagai hubungan di antara ciri butir dengan ciri peserta. Ciri butir selalu bergantung pada kelompok peserta. Pada kelompok peserta yang berbeda-beda ukuran taraf kesukaran butir yang sama akan berbeda-beda pula. Mengukur taraf kesukaran butir diukur melalui skor butir yang dihasilkan dari jawaban peserta tes.

Semakin banyak peserta tes yang menjawab butir dengan benar maka butir tersebut semakin mudah, dan sebaliknya semakin sedikit peserta tes yang menjawab butir itu dengan benar maka semakin sukar butir tersebut. Sehingga pada butir yang mudah probabilitas jawaban benar dari peserta tes lebih besar dibandingkan dengan butir yang sukar yang probabilitas jawaban benarnya sangat sedikit. Peserta didik yang dapat menjawab butir soal dengan benar menunjukkan kemampuan peserta didik tersebut. Artinya butir sesukar apapun jika dapat dijawab dengan benar maka menunjukkan bahwa kemampuan sama dengan tingkat kesukaran butir. Dalam responsi butir kesukaran butir disimbolkan dengan "b" dan kemampuan disimbolkan dengan "θ" sehingga secara matematis $b = \theta$, artinya tingkat kesukaran butir sama dengan kemampuan peserta didik dalam menjawab butir dengan benar. Rumus yang digunakan mengukur kesukaran butir sebagai berikut.⁸

$$\text{Tingkat kesukaran butir (TKB)} = \frac{\text{Rata-rata skor tiap butir seluruh peserta tes}}{\text{Skor maksimum tiap butir}}$$

Rumus tersebut dipersingkat menjadi:

⁸ Hairun, Y. (2020). *Evaluasi dan Penilaian dalam Pembelajaran*. Yogyakarta: Deepublish, 121-122.

$$\text{Tingkat kesukaran butir (TKB)} = \frac{\bar{X}}{X \text{ maks}}$$

Keterangan

\bar{X} : rata-rata skor butir ke-i

X maks : skor maksimum tiap butir

Skor maksimum merupakan skor yang ditetapkan oleh pembuat soal untuk setiap butir. Biasanya skor tersebut dicantumkan dalam rubrik penilaian. Dali (1992) dalam (Hairun, Y. 2020:122), mengemukakan bahwa butir dengan tingkat kesukaran di sekitar 0,5, sudah dianggap sebagai butir yang cukup baik. Selanjutnya Tucman dalam Dali (1992) dalam (Hairun, Y. 2020:122-124) mengemukakan bahwa sebaiknya taraf kesukaran butir terletak di antara:⁹

$$0,33 \leq TK \leq 0,67$$

Maka kriteria kesukaran butir, sebagai berikut:

⁹ Hairun, Y. (2020). Evaluasi dan Penilaian dalam Pembelajaran. Yogyakarta: Deepublish, 122-124.

Tabel 2.10 Kriteria Tingkat Kesukaran Butir

Interval Tingkat Kesukaran	Kategori Tingkat Kesukaran
0,00-0,30	Sukar
0,31-0,70	Sedang
0,71-1,00	Mudah

Contoh perhitungan tingkat kesukaran butir

Tabel 2.11 Skor Tes Butir Subjektif

Nama	Nomor dan Skor Butir						Skor	Kelompok
	1	2	3	4	5	6	Total	
Rais	9	8	10	12	2	2	63	Batas atas
Vivi	9	9	11	10	12	13	64	Batas atas
Rani	8	9	10	10	13	12	62	Batas atas
Wahyuni	6	6	6	6	2	2	28	Batas bawah
Yulia	6	6	7	6	3	4	32	Batas bawah
Munira	5	6	6	7	2	2	28	Batas bawah
Jumlah skor	43	44	50	51	44	45		
Skor maks. butir	10	10	15	15	25	25		
Rata-rata tiap butir	$43/6=7,16$	$44/6=7,3$	$50/6=8,3$	$51/6=8,5$	$44/6=7,3$	$45/6=7,5$		

Menghitung kesukaran butir:

$$\text{Tingkat kesukaran butir nomor 1, } \frac{7,16}{10} = 0,71$$

$$\text{Tingkat kesukaran butir nomor 2, } \frac{7,3}{10} = 0,73$$

$$\text{Tingkat kesukaran butir nomor 3, } \frac{8,3}{15} = 0,55$$

$$\text{Tingkat kesukaran butir nomor 4, } \frac{8,5}{15} = 0,56$$

$$\text{Tingkat kesukaran butir nomor 5, } \frac{7,3}{25} = 0,29$$

$$\text{Tingkat kesukaran butir nomor 6, } \frac{7,5}{25} = 0,3$$

Dari hasil perhitungan di atas dimasukkan dalam kriteria tingkat kesukaran, pada tabel berikut.

Tabel 2.12 Perhitungan Tingkat Kesukaran Dibandingkan dengan Kriteria Tingkat Kesukaran

Nomor Butir	Skor Tingkat Kesukaran	Kategori Tingkat Kesukaran
1	0,71	Mudah
2	0,73	Mudah
3	0,55	Sedang
4	0,56	Sedang
5	0,29	Sukar
6	0,30	Sukar

B. Daya Pembeda (discriminating power)

Daya beda butir soal ialah indeks yang menunjukkan tingkat kemampuan butir soal membedakan kelompok yang berprestasi tinggi (kelompok atas) dari kelompok yang berprestasi rendah (kelompok bawah) diantara para peserta tes (Asmawi Zainul, dkk : 1997). Suryabrata (1999) menyatakan tujuan pokok mencari daya beda adalah untuk menentukan apakah butir soal tersebut memiliki kemampuan membedakan kelompok dalam aspek yang diukur, sesuai dengan perbedaan yang ada pada kelompok itu.

Menurut Arifin. (2011:273) Perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum/kurang menguasai kompetensi berdasarkan

kriteria tertentu.¹⁰ 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.

Daya beda butir soal yang sering digunakan dalam tes hasil belajar adalah dengan menggunakan indeks korelasi antara skor butir dengan skor totalnya. Daya beda dengan cara ini sering disebut validitas internal, karena nilai korelasi diperoleh dari dalam tes itu sendiri. Daya beda dapat dilihat dari besarnya koefisien korelasi biserial maupun koefisien korelasi point biserial.

Dalam analisis ini digunakan nilai koefisien korelasi biserial untuk menentukan daya beda butir soal. Koefisien korelasi biserial menunjukkan hubungan antara dua skor, yaitu skor butir soal dan skor keseluruhan dari peserta tes yang sama.

Koefisien daya beda berkisar antara $-1,00$ sampai dengan $+1,00$. Daya beda $+1,00$ berarti bahwa semua anggota kelompok atas menjawab benar terhadap butir soal itu, sedangkan kelompok bawah seluruhnya menjawab salah terhadap butir soal itu. Sebaliknya daya beda $-1,00$ berarti bahwa semua anggota kelompok atas menjawab salah butir soal itu, sedangkan kelompok bawah seluruhnya menjawab benar terhadap soal itu.

Daya beda yang dianggap masih memadai untuk sebutir soal ialah apabila sama atau lebih besar dari +0,30. Bila lebih kecil dari itu, maka butir soal tersebut dianggap kurang mampu membedakan peserta tes yang mempersiapkan diri dalam menghadapi tes dari peserta yang tidak mempersiapkan diri. Bahkan bila daya beda itu menjadi negatif, maka butir soal itu sama sekali tidak dapat dipakai sebagai alat ukur prestasi belajar. Oleh karena itu butir soal tersebut harus dikeluarkan dari perangkat soal. Makin tinggi daya beda suatu butir soal, maka makin baik butir soal tersebut, dan sebaliknya makin rendah daya bedanya, maka butir soal itu dianggap tidak baik (Asmawi Zainul, dkk : 1997).

Menurut Dali S Naga (1992) kriteria besarnya koefisien daya beda diklasifikasikan menjadi empat kategori. Secara lebih rinci dijelaskan dalam tabel di bawah ini.¹¹

Tabel 1. Klasifikasi Daya Beda Butir Soal

Kategori gaya beda	Koefisien Korelasi
Baik	0,40 - 1,00
Sedang (tidak perlu revisi)	0,30 – 0,39
Perlu direvisi	0,20 – 0,29
Tidak baik	-1,00 – 0,19

¹¹ ⁴⁵ Goleman et al., (2019). Analisis Butir Soal. *Journal of Chemical Information and Modeling*, 53(9), 1–26.

1. Menurut Arifin. (2011:277) Teknik yang digunakan untuk menghitung daya beda 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 soal.¹²

$$\text{Rumus : } t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{n(n-1)}\right)}}$$

Keterangan:

\bar{x}_1 = rata-rata dari kelompok atas

\bar{x}_2 = rata-rata dari kelompok bawah

$\sum x_1^2$ = jumlah kuadrat deviasi individual dari kelompok atas

$\sum x_2^2$ = jumlah kuadrat deviasi individual dari kelompok bawah

n = 27% x N (baik untuk kelompok atas maupun kelompok bawah)

Contoh:

Jumlah peserta didik (N) = 36 orang. Jumlah sampel (n) = 27% x 36 = 10 orang.

Skor soal nomor 1 dari kelompok atas : 8, 6, 8, 7, 7, 6, 9, 7, 8, 6.

Skor soal nomor 1 dari kelompok bawah : 4, 3, 3, 3, 4, 4, 5, 2, 4, 3.

35

¹² Arifin, Z. (2011). *Evaluasi Pembelajaran*. Bandung: P T Remaja Rosdakarya, 277.

Table 10.13
Perhitungan Perbedaan Dua Rata-rata

Skor HG (X ₁)	Skor LG (X ₂)	X ₁	X ₂	x ₁ ²	x ₂ ²
8	4	0,8	0,5	0,64	0,25
6	3	-1,2	-0,5	1,44	0,25
8	3	0,8	-0,5	0,64	0,25
7	3	-0,2	-0,5	0,04	0,25
7	4	-0,2	0,5	0,04	0,25
6	4	-1,2	0,5	1,44	0,25
9	5	1,8	1,5	3,24	2,25
7	3	-0,2	-1,5	0,04	2,25
8	4	0,8	0,5	0,64	0,25
6	3	-1,2	-0,5	1,44	0,25
72	35			9,66	6,5
$\bar{x} = 7,2$	$\bar{x}_2 = 3,5$				

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{n(n-1)}\right)}} = \frac{(7,2 - 3,5)}{\sqrt{\left(\frac{9,66 + 6,5}{10(10-1)}\right)}} = \frac{3,7}{\sqrt{\left(\frac{16,16}{10(9)}\right)}} = \frac{3,7}{\sqrt{0,1795555}}$$

$$= \frac{3,7}{0,423} = 8,747$$

dengree of freendom (df) = (n₁ - 1) + (n₂ - 1) = (10 - 1) = 9 + 9 = 18.

Jika ditentukan tingkat kepercayaan 0,01, maka nilai t dalam table menunjukkan 2,88. Ternyata nilai t hitung > nilai t tabel, yaitu 8,747 > 2,88. Hal ini berarti daya pembeda soal nomor 1 (satu) bentuk uraian itu signifikan.

2. Menurut Burhan Nurgiyantoro (2010:201) Perhitungan Indeks Tingkat Kesulitan (ITK) dan Indeks Daya Beda (IDB) dapat juga dilakukan untuk bentuk soal uraian, namun rumus yang dipergunakan berbeda dengan teks objektif. Walau demikian, prosedur perhitungannya tidak banyak berbeda. Rumus yang dipakai untuk menghitung ITK dan IDB yang dimaksud adalah sebagai berikut (Noll dkk. 1979:214-215).¹³

$$\text{Indeks Daya Beda} = \frac{St - Sr}{N(\text{Skormaks} - \text{Skormin})}$$

Keterangan:

St : Jumlah skor benar kelompok tinggi

Sr : Jumlah skor benar kelompok rendah

Skormaks : Skor maksimal suatu butir

Skormin : Skor minimal suatu butir

Langkah yang perlu dilakukan sebelum menghitung ITK dan IDB kurang sama dengan langkah-langkah soal objektif, yaitu dimulai dengan mengelompokkan jawaban peserta didik ke dalam kelompok tinggi dan rendah dan kemudian menganalisis jawaban. Pada soal uraian, jawaban biasanya dibuat berskala, missal 1 - 4, 1 - 5, atau 1-6 tergantung kompleksitas masing-masing soal. Di bawah dicontohkan analisis jawaban per soal per peserta didik

¹³ Nurgiyantoro, B. (2010). *Penilaian Pembelajaran Bahasa Indonesia (Berbasis Kompetensi)*. BPFE-YOGYAKARTA, 201.

dengan skala skor semua soal minimal 1 dan maksimal 5.

Tabel 5.3
Analisis Jawaban Soal Uraian Kelompok Tinggi dan Kelompok Rendah sebagai Persiapan Perhitungan Indeks Daya Beda

Kelompok Tinggi					Kelompok Rendah				
No.	Nomor Butir Soal				No.	Nomor Butir Soal			
Urut	1	2	3	4	Urut	1	2	3	4
1.	4	4	5	5	1.	3	2	2	4
2.	4	4	3	5	2.	3	1	2	4
3.	3	3	5	4	3.	2	2	1	4
4.	5	3	3	4	4.	1	3	2	3
5.	3	4	4	3	5.	2	1	2	3
6.	3	2	3	4	6.	2	1	2	3
Jumlah	22	20	23	25	Jumlah	13	10	11	21

Berdasarkan data-data tabel di atas kemudian dapat dihitung ITK dan IDB untuk keempat butir soal uraian yang ada.

$$(1) \text{ Butir nomor 1 : } IDB = \frac{22-13}{6 \times (5-1)}$$

$$= 0,38$$

$$(2) \text{ Butir nomor 2 : } IDB = \frac{20-10}{6 \times (5-1)}$$

$$= 0,42$$

$$(3) \text{ Butir nomor 3 : } IDB = \frac{23-11}{6 \times (5-1)}$$
$$= 0,50$$

$$(4) \text{ Butir nomor 4 : } IDB = \frac{25-21}{6 \times (5-1)}$$
$$= 0,178$$

Dengan mempergunakan kriteria kelayakan butir soal seperti di atas, yaitu IDB 0,20 ke atas, dari keempat butir soal uraian di atas terlihat hanya soal nomor 4 yang kurang layak karena IDB yang terlalu kecil. Hasil analisis butir soal uraian juga sebaiknya disajikan dalam tabel agar terlihat bagus.

DAFTAR PUSTAKA

- Arifin, Zainal (ED). 2013. *Evaluasi Pembelajaran*. Bandung. PT Remaja Rosdakarya.
- Arikunto, 1990. 1999. "Analisis Tingkat Kesukaran, Daya Beda, Dan Pengecoh."
- Junaidi, Junaidi, Juli Yani, and Rismayeti Rismayeti. 2016. "analisis butir soal untuk mengetahui seberapa mudah atau sulit soal yang dikerjakan siswa." *Jurnal Pustaka Budaya*
- Arikunto, Suharsimi. (2010) *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara : Yogyakarta.
- Arikunto, S. 1995. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Asrul, Ananda, R., & Rosinta. (2014). Evaluasi Pembelajaran. In *Ciptapustaka Media*.
- Azwar, S. 2002. *Tes Prestasi Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Edisi II. Yogyakarta: Pustaka Pelajar.
- Anne, Anastasi, Susaba Urbina (2007). *Tes Psikologi*. Edisi Ketujuh. Jakarta: PT indeks
- M. 1996. *Guru Dalam Proses Belajar Mengajar*. Bandung: Sinar Baru Algesindo.
- Arifin, Zainal. 2009. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya
- Abdul Kadir. (2015). MENYUSUN DAN MENGANALISIS TES HASIL BELAJAR Abdul Kadir. *Al-Ta'dib*, 8(2), 70-81.
- Arikunto, 1990. (1999). *dengan: P adalah indeks kesukaran, B adalah banyaknya siswa yang menjawab soal dengan benar, dan J. 1-6.*

- Arip Sanjaya, S. M. (2018). *JURNAL MEMBACA BAHASA DAN SASTRA INDONESIA*. BANTEN: MLI cabang Untirta dan HISKI Banten.
- Arikunto, Suharsimi. 2010. *Prosedur Penelitian Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, Saifudin. 1986. *Validitas dan Reliabilitas*. Jakarta: Rineka Cipta.
- Benjamin, W. (2019). No Title. *ペインクリニック学会治療指針 2*, 3, 1-9.
<https://www.scribd.com/document/409041396/Pengolahan-Data-Hasil-Penilaian-Notes>
- Darwis, D. (2016). Penerapan Penilaian Berbasis Kelas Dalam Pembelajaran Ips Di Smp Negeri 1 Batulicin. *Jurnal Socius*, 1(2).
<https://doi.org/10.20527/jurnalsocius.v1i2.2190>
- Daryanto, H. 2008. *Evaluasi Pendidikan*. Jakarta. Rineka Cipta.
- Depdiknas. 2008. *Peraturan Pemerintah RI No.19 Tahun 2005 tentang Standar Nasional Pendidikan*. Jakarta: Depdiknas.
- Dr. Elis Ratna Wulan, S. M. (2014). *Evaluasi Pembelajaran Dengan Pendekatan Kurikulum 2013*. Bandung : Pustaka Setia.
- Dr. Rina Febriana, M. (2019). *Evaluasi Pembelajaran*. Jakarta: Bumi Aksara.
- Dr. Zainal Arifin, M. (2009). *Evaluasi Pembelajaran* . Bandung: PT REMAJA ROSDAKARYA.
- Drs. Zainal Arifin, M. (2009). *Evaluasi Pembelajaran : prinsip teknik prosedur*. Bandung: PT. Remaja Rosdakarya Offset.

48

Fawzi, A. (2016). Pengembangan Alat Penilaian Pembelajaran Menyimak Eksposisi Kelas X Sma. *Bahasa Dan Seni: Jurnal Bahasa, Sastra, Seni Dan Pengajarannya*, 44(2), 125-134. <https://doi.org/10.17977/um015v44i22016p12>

14 5

Gregory, Robert J. 2013. *Tes Psikologi*. Edidi Keenam. Jakarta: PT Erlangga

Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS*, 6(1), 41-55.

Ina Magdalena¹, N. I. (2021). Penerapan Penilaian Berbasis Kelas Dan Portofolio Pelajaran Bahasa Indonesia Kelas 6 SDN Selapajang Jaya 2. *Jurnal Pendidikan dan Dakwah*, 240-242.

Ina Magdalena¹, N. I. (2021). PENERAPAN PENILAIAN BERBASIS KELAS DAN PORTOFOLIO PELAJARAN BAHASA INDONESIA KELAS 6 SDN SELAPANJANG JAYA 2. *Jurnal Pendidikan dan Dakwah*, 236.

Irhamni. (2018). Prinsip-Prinsip Dan Pendekatan Dalam Penilaian Hasil Belajar. *Jurnal Pendidikan*, 1(5), 108-119.

Irawati, H., Saifuddin, M. F., & Ma'rifah, D. R. (2017). Pengembangan Instrumen Tes Dan Non Tes Dalam Rangka Menyiapkan Penilaian Autentik Pada Kurikulum 2013 Di SMP/MTs Muhammadiyah Se-Kabupaten Bantul. *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian Kepada Masyarakat*, 1(2), 503-506.

46

Khairuddin Alfath, F. F. (2019). Teknik Pengolahan Hasil Asesmen: Teknik Pengolahan dengan Menggunakan Pendekatan Acuan Norma (PAN) dan Pendekatan Acuan Patokan (PAP). *journal.staimsyk.ac.id*, 13-22.

Kusuma, Mochtar. 2010. *Evaluasi Pendidikan*. Jakarta : PT. MULTI KREASI SATUDELAPAN.

25

Wahidmurni, Mustikawan Arifin, Ridho Ali. 2010. *EVALUASI PEMBELAJARAN KOMPETENSI DAN AKTIK*. Yogyakarta : Nuha Litera.

61

Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, (28), 563-575.

Linn, R.L, Grondlund, N.E. 2000. *Measurement and Assessment In Teaching . Eighth edition. New Jersey: Merrill an imprint of Prentice Hall*.

7

Magdalena, I., Kusumawati, N. I., Putri, R. H., & Fazriandina, A. (2021). Penerapan Penilaian Berbasis Kelas Dan Portofolio Pelajaran Bahasa Indonesia Kelas 6 Sdn Selapajang Jaya 2. *Jurnal Pendidikan Dan Dakwah*, 3(2), 234-243. <http://ejournal.stitpn.ac.id/index.php/pandawa>

60

35

Nurgiyantoro, Burhan. 1987. *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE
Nurgiyantoro, B. (2010). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: BPFE-YOGYAKARTA.

7

Oktasari, A. R. (2012). *Penerapan penilaian berbasis kelas dalam bidang studi pendidikan agama islam di kelas x sma darul mustofa burneh bangkalan*.

Pascasarjana Undiksha. (n.d.). *Taraf kesukaran, daya beda, dan efektivitas pengecoh*. 81-87.

71

Ramadhani, W., Murti, M., & Arsyam, M. (2021). Teknik pengolahan skor hasil evaluasi.

Osfpreprints, Ddi. <https://osf.io/gqx4s/download>

Ratna Pangastuti, K. M. (2018). Penilaian Acuan Norma, Penilaian Acuan Patokan, Kriteria Ketuntasan Minimal di Madrasah Ibtidaiyah An-Nur Plus Junwangi Krian Sidoarjo Jawa Timur. *ejournal. uinib.ac.id* , 210-214.

Sudjiono, A. 2005. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada

Supratiknya, A. (2012). Penilaian hasil belajar dengan teknik nontes. *Yogyakarta: Universitas Sanata Dharma*.

73

Sudjana, N. 1990. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.

Sukardi. 2011. *Metodologi Penelitian Pendidikan Kompetensi dan Praktiknya*. Yogyakarta: Bumi Aksara.

Suryadi. 2010. *Suplemen MK Evaluasi*. [Online] <http://file.upi.edu> Direktori FIPJUR_ADMINISTRASI_PENDIDIKAN196807291998021-SURYADI_VALIDITAS_tes.pdf. Diakses pada 23 Mei 2022.

Suprananto. (2012). Penilaian (Assessment) dalam Pembelajaran. *Penilaian Pendidikan*, 13-14.

27

Sudaryono. (2012). *Dasar-dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.

Sriyanto, A. (2019). Teknik Pengolahan Hasil Asesmen, Penentuan Standar Asesmen, Teknik Pengolahan dengan Menggunakan Pendekatan Acuan Patokan (PAP) dan Acuan Norma (PAN). *ejournal.kopertais4.or.id* , 255-256.

STANDAR, T. KONSEP DASAR TES STANDAR.

Skor, P., Skor, P., & Nilai, M.

60 (n.d.). *Pengolahan hasil tes. 100.*

http://file.upi.edu/Direktori/FPBS/JUR._PEND._BAHASA_JEPANG/196301041988031SUGIHARTONO/PERKULIAHAN/Evaluasi_Pembelajaran_Bahasa_Jepang/Dokumen/52_NGOLAHAN_HASIL_TES.pdf

Solichin, M. (2017). Analisis Daya Beda Soal Taraf Kesukaran, Butir Tes, Validitas Butir Tes, Interpretasi Hasil Tes Validitas Ramalan dalam Evaluasi Pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213.

Widoyoko, Putro, Eko. 2017. *EVALUASI PROGRAM PELATIHAN*. Yogyakarta : Pustaka Pelajar.

Yeti Mulyati, D. (2007). *Keterampilan Berbahasa Indonesia SD*. Tangerang Selatan : Universitas Terbuka.

37 Zhang, H. M., Peh, L. S., & Wang, Y. H. (2014). Servo motor control system and method of auto-detection of types of servo motors. *Applied Mechanics and Materials*, 496–500(1), 1510–1515.

37 <https://doi.org/10.4028/www.scientific.net/AMM.496-500.1510>

Buku Ajar Evaluasi Pembelajaran Bahasa Indonesia

ORIGINALITY REPORT

17%

SIMILARITY INDEX

11%

INTERNET SOURCES

7%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	Pranowo Narjosoeripto, Toni Harsan, Para Mitta Purbosari. "PENGARUH CLASSMARKER TERHADAP HASIL BELAJAR IPS SISWA SEKOLAH DASAR NEGERI 1 BOYOLALI", SCAFFOLDING: Jurnal Pendidikan Islam dan Multikulturalisme, 2020 Publication	1%
2	Submitted to IAIN Kudus Student Paper	1%
3	anyflip.com Internet Source	1%
4	docplayer.info Internet Source	1%
5	Submitted to Universitas Nasional Student Paper	1%
6	chabiboktafianjati.blogspot.com Internet Source	1%
7	ejournal.stit-ru.ac.id Internet Source	1%

8	es.scribd.com Internet Source	1 %
9	digilibadmin.unismuh.ac.id Internet Source	<1 %
10	pdfcoffee.com Internet Source	<1 %
11	mastiahumiaisyabilal.wordpress.com Internet Source	<1 %
12	Joko Widodo, Muhammad Nanang Qosim. "Penilaian Kinerja Pada Pembelajaran Maharah Al-Kalam Level 1 Di Kursus Bahasa Arab Al-Arobiya Surakarta", Uktub: Journal of Arabic Studies, 2021 Publication	<1 %
13	azmi648.blogspot.com Internet Source	<1 %
14	henzelitamaranatha.blogspot.com Internet Source	<1 %
15	Submitted to Universitas Muria Kudus Student Paper	<1 %
16	ejournal.stitpn.ac.id Internet Source	<1 %
17	Diana Kurniawati, Ridwan Nurazi, Lisa Martiah. "PENGARUH KUALITAS SISTEM INFORMASI, KUALITAS INFORMASI,	<1 %

PERCEIVED USEFULNESS, TERHADAP KEPUASAN PENGGUNA AKHIR SOFTWARE AKUNTANSI", JURNAL FAIRNESS, 2021

Publication

18

muhammadsamsularifin469.blogspot.com

Internet Source

<1 %

19

Siti Julaeha, Muhammad Maky, Uus
Ruswandi. "Desain, Pelaksanaan dan Evaluasi
Pembelajaran pada Sekolah Menengah",
Jurnal Dirosah Islamiyah, 2022

Publication

<1 %

20

core.ac.uk

Internet Source

<1 %

21

Okta Pita Dian Sari, Sunismi, Anies Fuady.
"ANALISIS INSTRUMEN SOAL PILIHAN GANDA
UJIAN AKHIR SEMESTER (UAS) MATA
PELAJARAN MATEMATIKA PEMINATAN",
JUMLAHKU: Jurnal Matematika Ilmiah STKIP
Muhammadiyah Kuningan, 2023

Publication

<1 %

22

Submitted to Universitas Muhammadiyah
Surakarta

Student Paper

<1 %

23

repository.uinjambi.ac.id

Internet Source

<1 %

24

Submitted to Academic Library Consortium

Student Paper

<1 %

25

jurnalpknfkipunlam.wordpress.com

Internet Source

<1 %

26

repository.stiedewantara.ac.id

Internet Source

<1 %

27

Laili Etika Rahmawati, Widya Puteri Kusumawati. "UJI VALIDITAS ASPEK KOMPETENSI BERSASTRA PADA SOAL ULANGAN AKHIR SEMESTER GANJIL MATA PELAJARAN BAHASA INDONESIA KELAS VII TAHUN PELAJARAN 2012/ 2013 SMP NEGERI WILAYAH NGAWI BARAT", Jurnal VARIDIKA, 2015

Publication

<1 %

28

Submitted to Universitas Islam Negeri Antasari Banjarmasin

Student Paper

<1 %

29

Purbadi Purbadi. "Manajemen Pembelajaran Berbasis Sekolah di SMK Muhammadiyah 1 Playen Gunungkidul", Media Manajemen Pendidikan, 2020

Publication

<1 %

30

KASMAH Kasmah. "KONSEP DASAR EVALUASI DALAM KURIKULUM BERBASIS KKNI", Adaara: Jurnal Manajemen Pendidikan Islam, 2018

Publication

<1 %

31

Submitted to Sogang University

Student Paper

<1 %

32

Submitted to Universitas Mataram

Student Paper

<1 %

33

repository.unpas.ac.id

Internet Source

<1 %

34

indonsc.blogspot.com

Internet Source

<1 %

35

Simun Ardianto. "ANALISIS KEMAMPUAN MENULIS KARANGAN NARASI DENGAN TEMA PERISTIWA DALAM KEHIDUPAN PADA MUATAN BAHASA INDONESIA", *Dinamika Jurnal Ilmiah Pendidikan Dasar*, 2019

Publication

<1 %

36

Aris Munandar. "EVALUASI PEMBELAJARAN TERHADAP PESERTA DIDIK", *eL-Muhbib: Jurnal Pemikiran dan Penelitian Pendidikan Dasar*, 2019

Publication

<1 %

37

journal.unj.ac.id

Internet Source

<1 %

38

Didi Yulistio. "Kemampuan Mahasiswa Menganalisis Tes (Soal) Bahasa Indonesia Sesuai Aspek Ranah Kognitif Model Anderson dan Krathwohl", *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, dan Asing*, 2022

Publication

<1 %

39	nhuynhuy1994.blogspot.com Internet Source	<1 %
40	Submitted to Universitas Negeri Makassar Student Paper	<1 %
41	Submitted to Universitas Pamulang Student Paper	<1 %
42	journal.maranatha.edu Internet Source	<1 %
43	repository.radenintan.ac.id Internet Source	<1 %
44	Yeni Widiyawati, Indri Nurwahidah, Akhmad Nayazik, Malikhatus Khasanah, Vimala Tantri. "Workshop Penyusunan Penilaian Pembelajaran Berbasis Online Bagi Guru SD di Kecamatan Gunung Pati Kota Semarang", Manggali, 2021 Publication	<1 %
45	jim.unindra.ac.id Internet Source	<1 %
46	Submitted to Universitas Negeri Semarang Student Paper	<1 %
47	Elda Fitri. "IMPLEMENTASI PENILAIAN BERBASIS KELAS DALAM PEMBELAJARAN IPA BERBASIS INQUIRI DI SMP NEGERI 4 MUARA BUNGO", Jurnal Tunas Pendidikan, 2022 Publication	<1 %

48

Lira Hayu Afdetis Mana. "Analisis Kompetensi Dasar Pembelajaran Keterampilan Menyimak pada Kurikulum 2013 Revisi di Tingkat SMP dan SMA", JIRA: Jurnal Inovasi dan Riset Akademik, 2021

Publication

<1 %

49

Marsela Renasari Presty, Dheska Arthyka Palifiana, Margono Margono. "hubungan kesiapan praktik mengajar dengan nilai ujian praktik real teaching pada mahasiswa semester VIII program studi DIV bidan pendidik universitas respati yogyakarta", Jurnal Formil (Forum Ilmiah) Kesmas Respati, 2019

Publication

<1 %

50

Andri Valen, Tio Gusti Satria. "Analisis Tingkat Kesulitan Soal PAS (Penilaian Akhir Semester) Mata Pelajaran IPS di Sekolah Dasar", Jurnal Basicedu, 2021

Publication

<1 %

51

Fauziatul Muniroh, Nurul Astuty Yensy, Effie Efrida Muchlis, Teddy Alfra Siagian. "DIAGNOSIS KESALAHAN PESERTA DIDIK DALAM MENYELESAIKAN SOAL MATEMATIKA POKOK BAHASAN LINGKARAN KELAS VIII SMP NEGERI 7 KOTA BENGKULU", Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS), 2022

<1 %

52

Submitted to UPH College - Jakarta

Student Paper

<1 %

53

Submitted to Universitas Andalas

Student Paper

<1 %

54

Ida Nurjelita Sani, Amrul Bahar, Elvinawati
Elvinawati. "PERBANDINGAN MODEL
PEMBELAJARAN PROBLEM SOLVING DAN
PROBLEM BASED LEARNING TERHADAP
KEMAMPUAN BERPIKIR KRITIS SISWA KELAS
XI MIA MAN 2 KOTA BENGKULU", Alotrop,
2020

Publication

<1 %

55

perangkat-mengajar.blogspot.com

Internet Source

<1 %

56

Submitted to IAIN Pontianak

Student Paper

<1 %

57

Mukhlas Habibi, Fina Aulika Lestari, Yusmicha
Ulya Afif. "Implementasi Penilaian Autentik
Kurikulum 2013 Pada Mata Pelajaran PAI dan
Budi Pekerti di SDN 1 Bangunrejo Ponorogo",
QALAMUNA: Jurnal Pendidikan, Sosial, dan
Agama, 2021

Publication

<1 %

58

repository.upi.edu

Internet Source

<1 %

59

Laili Etika Rahmawati, Nuraini Fatimah.
"Pengembangan Model Penilaian Autentik
Kompetensi Berbicara", Jurnal VARIDIKA, 2015
Publication

<1 %

60

Dinda Suzianti, Febrina Dafit. "Video Animasi
Berbasis Kearifan Lokal Riau untuk Membaca
Pemahaman Siswa Sekolah Dasar", Aulad:
Journal on Early Childhood, 2023
Publication

<1 %

61

P. Michael McCullough, Sam Faught. "Rational
Moralists and Moral Rationalists Value-Based
Management: Model, Criterion and
Validation", Journal of Business Ethics, 2005
Publication

<1 %

62

Submitted to Universitas Negeri Padang
Student Paper

<1 %

63

id.123dok.com
Internet Source

<1 %

64

Harri Gusnirwanda, Efrina Mora.
"Pelaksanaan Penilaian Pembelajaran Tematik
di MIs Nurul Hidayah Rawa Cangkuk Kec.
Medan Denai", Tarbiatuna: Journal of Islamic
Education Studies, 2023
Publication

<1 %

65

Submitted to Konsorsium Turnitin Relawan
Jurnal Indonesia
Student Paper

<1 %

- | | | |
|----|---|------|
| 66 | Submitted to Sriwijaya University
Student Paper | <1 % |
| 67 | www.jbasic.org
Internet Source | <1 % |
| 68 | Muh Syahrul Sarea, Rosnia Ruslan.
"KARAKTERISTIK BUTIR SOAL: CLASSICAL
TEST THEORY VS ITEM RESPONSE THEORY?",
DIDAKTIKA : Jurnal Kependidikan, 2019
Publication | <1 % |
| 69 | Submitted to UIN Syarif Hidayatullah Jakarta
Student Paper | <1 % |
| 70 | Submitted to Universitas Islam Negeri
Sumatera Utara
Student Paper | <1 % |
| 71 | jipp.unram.ac.id
Internet Source | <1 % |
| 72 | Eva Sartika Dasopang, Intan Purnama Sari.
"PERSEPSI KONSUMEN TERHADAP
PELAYANAN APOTEK DI KECAMATAN BINJAI
KOTA", Journal of Pharmaceutical And
Sciences, 2019
Publication | <1 % |
| 73 | Murni Yanto. "Manajemen dan Mutu
Pembelajaran Bahasa Indonesia Di SMP
Negeri 4 Rejang Lebong", Tadbir : Jurnal Studi
Manajemen Pendidikan, 2018
Publication | <1 % |

74

David Firna Setiawan. "MOTIF EKSTERNAL KEHADIRAN MAHASISWA PADA MATA KULIAH EVALUASI PEMBELAJARAN", PROMOSI (Jurnal Pendidikan Ekonomi), 2019

Publication

<1 %

75

digilib.uin-suka.ac.id

Internet Source

<1 %

76

jonedu.org

Internet Source

<1 %

77

Ika Puspitasari, Aliyatul Ilmi Suryani, Shokhibul Arifin. "Pengembangan Alat Penilaian Mata Pelajaran Pendidikan Agama Islam dengan Menggunakan Quizizz di Sekolah Menengah Pertama", Jurnal Basicedu, 2023

Publication

<1 %

78

123dok.com

Internet Source

<1 %

79

Zakiyah Zakiyah. "Karakteristik Tes Prestasi Bahasa Inggris sebelum Covid-19", JURNAL PENDIDIKAN, 2022

Publication

<1 %

80

e-journal.unipma.ac.id

Internet Source

<1 %

81

id.scribd.com

Internet Source

<1 %

82

repositories.uin-alauddin.ac.id

Internet Source

<1 %

83

Ahmad Fadilah, Abdul Muktadir, Puspa Djuwita. "Pengembangan Bahan Ajar Berbasis Pendekatan Content and Language Integrated Learning (CLIL) pada Pembelajaran Menulis Puisi Siswa Kelas IV SDN 86 Kaur", Jurnal Pembelajaran dan Pengajaran Pendidikan Dasar, 2023

Publication

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On