

PENGEMBANGAN TES LITERASI SAINS

Buku ini membekali pengetahuan dan keterampilan pembaca dalam memahami pengembangan instrumen penilaian, khususnya untuk penilaian literasi sains.

Secara umum buku ini dapat dipergunakan oleh dosen, mahasiswa S1, S2, dan S3, peneliti, dan semua pihak yang berkepentingan dengan pengembangan alat evaluasi/instrumen penilaian dalam bidang pendidikan.

Contoh instrumen literasi sains tersaji dalam buku ini, mulai dari tingkat sekolah dasar hingga perguruan tinggi khususnya mahasiswa calon guru.

PENGEMBANGAN TES LITERASI SAINS



Prof. Dr. Ani Rusilowati, M.Pd.
Dr. Budi Astuti, M.Sc.

PENGEMBANGAN TES LITERASI SAINS

Hak cipta dilindungi undang-undang.
Dilarang mengutip atau memperbanyak sebagian
atau seluruh isi buku ini tanpa izin tertulis dari penerbit.

PENGEMBANGAN TES LITERASI SAINS

**Prof. Dr. Ani Rusilowati, M.Pd.
Dr. Budi Astuti, M.Sc.**

Penerbit

LPPM UNNES

Gedung Prof. Retno Sriningsih Satmoko
Sekaran, Gunungpati, Semarang 50229

Hak Cipta © pada penulis dan dilindungi Undang-Undang
Penerbitan
Hak Penerbitan pada LPPM UNNES.
Gedung Prof. Retno Sriningsih Satmoko, Kampus Unnes Sekaran,
Gunungpati, Semarang 50229

Dilarang mengutip sebagian atau seluruh buku ini dalam bentuk
apapun tanpa izin dari penerbit.

PENGEMBANGAN TES LITERASI SAINS

Prof. Dr. Ani Rusilowati, M.Pd.
Dr. Budi Astuti, M.Sc.

Editor
Dr. Sanyoto Eko Nugroho, M.Si.

Penyunting Bahasa
Harits Agung Wicaksono, S.Pd.

Layout & Desain Cover
Harits Agung Wicaksono, S.Pd.

Pengembangan Tes Literasi Sains/ Prof. Dr. Ani Rusilowati,
M.Pd.; Dr. Budi Astuti, M.Sc.; -Cet. 1-illus-Semarang: Lppm
Unnes, 2019;
xii + 115 hlm; 14,8 x 21 cm

Keanggotaan IKAPI No. 175/ALB/JTE/2019

ISBN : 978-623-90096-4-9

PRAKATA

Puji syukur penulis panjatkan kepada Allah swt. atas segala rahmat-Nya sehingga buku ini dapat terselesaikan. Secara umum buku ini dapat dipergunakan oleh dosen, guru, mahasiswa S1, S2, dan S3, peneliti, dan semua pihak yang berkepentingan dengan pengembangan alat evaluasi, penelitian pengembangan, dan penelitian pendidikan, serta para praktisi pendidikan.

Hasil tulisan yang ada di buku ini berasal dari berbagai buku, literatur yang dicantumkan pada daftar pustaka, dan hasil-hasil penelitian pengembangan alat evaluasi, khususnya literasi sains. Namun tentunya telah diolah dan ditulis berdasarkan gaya bahasa penulis. Beberapa tabel dan gambar yang disertakan adalah hasil kreasi penulis dari merangkum bahan bacaan yang digunakan.

Penulis menyadari bahwa selain kehendak Allah swt., juga bantuan dan dorongan dari semua pihak, sehingga penyusunan buku ini dapat berjalan dengan lancar.

Pada kesempatan ini penulis menyampaikan terima kasih kepada Rektor, Dekan FMIPA, dan Ketua Jurusan Fisika Universitas Negeri Semarang yang telah memberi kesempatan kepada penulis untuk menyusun buku ini, serta semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan banyak bantuan.

Penulis sadar bahwa masih banyak kekurangan yang ada dalam penulisan buku ini. Oleh karena itu, penulis sangat mengharapkan kritik, saran, serta masukan guna perbaikan dan penyempurnaan, sehingga buku yang sederhana ini dapat memberikan manfaat bagi semua pihak yang membutuhkannya.

Semarang, Februari 2019

Penulis

DAFTAR ISI

PRAKATA	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN	1
BAB II PENILAIAN DAN EVALUASI	5
A. Pengertian.....	5
B. Perencanaan Penilaian Hasil Belajar.....	7
1. Perencanaan Umum.....	7
2. Perencanaan Khusus.....	7
C. Teknik Penilaian.....	12
1. Kuis.....	12
2. Tes Harian.....	12
3. Teknik Penilaian Jenis Tagihan Nontes.....	13
BAB III BENTUK ALAT PENILAIAN	15
A. Bentuk Alat Penilaian.....	15
1. Tes.....	15
2. Bentuk Alat Penilaian Nontes.....	19
B. Pelaksanaan Penilaian.....	19
1. Test Tertulis.....	19
2. Tes Lisan.....	20
3. Pelaksanaan Observasi untuk Asesmen Otentik.....	20
BAB IV KUALITAS INSTRUMEN	21
A. Validitas Tes.....	21
1. Validitas Isi.....	22
2. Validitas Konstruk.....	28
3. Validitas Berdasar Kriteria.....	34
B. Reliabilitas Tes.....	34
1. Reliabilitas Tes (bentuk pilihan ganda).....	35
2. Reliabilitas Tes Uraian.....	39
BAB V ANALISIS BUTIR SOAL	41
A. Analisis Butir Soal Secara Kualitatif.....	42
B. Analisis Butir Soal Secara Kuantitatif.....	42
1. Tingkat Kesukaran (TK).....	46

2. Daya Pembeda (DP)	48
3. Penyebaran (Distribusi) Jawaban.....	50
C. Analisis Butir Soal dengan Program Komputer	51
1. ITEMAN	51
2. EXCEL	57
D. Analisis Butir Soal Secara Modern (<i>Item Response Theory/IRT</i>)	58
1. Kelebihan Analisis IRT.....	58
BAB VI PENGEMBANGAN TES	63
A. Langkah Pengembangan Tes	63
1. Penentuan Tujuan Tes	64
2. Penyusunan Spesifikasi Tes	65
3. Penulisan Soal Tes	69
4. Penelaahan Soal Tes	72
5. Penganalisisan Soal Uji Coba.....	74
6. Perbaikan Tes.....	75
7. Perakitan Tes	75
8. Pelaksanaan Tes.....	75
9. Interpretasi Hasil Tes	76
10. Pelaporan Hasil Tes.....	76
B. Pengembangan Tes.....	76
1. Tes Uraian	77
2. Tes Objektif	79
BAB VII LITERASI SAINS.....	83
A. Pengertian Literasi Sains	84
B. Aspek Literasi Sains.....	84
BAB VIII PENGEMBANGAN TES LITERASI SAINS	95
A. Pengembangan Instrumen Literasi Sains SD	97
B. Pengembangan Instrumen Literasi Sains SMP	99
C. Pengembangan Instrumen Literasi Sains SMA	100
1. Pengembangan Instrumen Literasi Sains Mahasiswa Calon Guru.....	100
DAFTAR PUSTAKA.....	105
GLOSARIUM.....	111
INDEKS.....	114
BIODATA PENULIS.....	117

DAFTAR TABEL

Tabel 1.1 Posisi Literasi Sains Anak Indonesia dalam Penilaian PISA	1
Tabel 1.2 Perbandingan Skor Rata-rata Literasi Sains Indonesia dan Malaysia Berdasarkan Hasil Penilaian PISA.....	2
Tabel 4.1 Right-Tail Probability (p) for Selected Values of the Validity Coefficient (V).....	25
Tabel 4.2 Minimum Values of CVR and CVRt One Tailed Test, $p = .05$	27
Tabel 4.3 <i>Cut off value</i> Uji Kecocokan Model (GoF)	33
Tabel 5.1. Format Penelaahan Butir Soal Pilihan Ganda	43
Tabel 5.2. Format Penelaahan Butir Soal Uraian	45
Tabel 6.1 Matriks tentang Tujuan Tes dan Ranah Kognitif untuk Tes Prestasi Sains.....	71
Tabel 8.1. Indikator Literasi Sains Berdasarkan PISA	101
Tabel 8.2. Hasil Validasi Instrumen Literasi Sains	102
Tabel 8.3. Taraf Kesukaran Butir Soal Instrumen Literasi Sains .	103
Tabel 8.4. Daya Pembeda Butir Soal Instrumen Literasi Sains	103

DAFTAR GAMBAR

Gambar 4.1. Scree Plot.....	30
Gambar 5.1. Kurva Ciri Soal Model Satu Parameter	60
Gambar 5.2. Kurva Karakteristik Soal Model 2-PL	62
Gambar 8.1. Format Soal Literasi Sains.....	96
Gambar 8.2. Contoh Soal Literasi Sains untuk SD	98
Gambar 8.3. Contoh Soal Literasi Sains untuk SMP	99
Gambar 8.4. Bentuk Soal Literasi Sains untuk Calon Guru.....	104

PENDAHULUAN

Lemahnya kemampuan siswa Indonesia dan Malaysia dalam bidang sains, khususnya literasi sains ditunjukkan oleh penelitian tentang asesmen hasil belajar sains pada level internasional yang diselenggarakan oleh *Organization for Economic Co-operation and Development (OECD)* melalui *Programme for International Student Assessment (PISA)* untuk anak usia 15 tahun. Indonesia ikut berpartisipasi dalam penelitian PISA tersebut sejak tahun 2000. Hasil yang dilaporkan oleh OECD (2003, 2004, 2007, 2010, 2013, 2016) terkait hasil literasi sains anak Indonesia dapat dilihat pada Tabel 1.1 berikut:

Tabel 1.1 Posisi Literasi Sains Anak Indonesia dalam Penilaian PISA

No	Tahun	Rata-rata Skor Indonesia	Rata-rata Skor Internasional	Posisi	Jumlah Negara peserta
1	2000	393	500	38	41
2	2003	395	500	38	40
3	2006	393	500	53	60
4	2009	383	500	57	65
5	2012	382	500	64	65
6	2015	403	493	62	70

Berdasarkan data tersebut tampak bahwa dari tahun ke tahun, Indonesia selalu berada pada peringkat rendah.

Hasil penilaian PISA terhadap literasi sains siswa Malaysia juga belum begitu menggembirakan. Perbandingan perolehan skor literasi sains pada tahun 2012 dan 2015 untuk Malaysia dan Indonesia dapat dilihat pada Tabel 1.2 berikut:

Tabel 1.2 Perbandingan Skor Rata-rata Literasi Sains Indonesia dan Malaysia Berdasarkan Hasil Penilaian PISA

No	Tahun	Rata-rata Skor Indonesia	Rata-rata Skor Malaysia	Rata-rata Skor Internasional
1	2012	382	421	500
2	2015	403	443	493

Capaian siswa Malaysia dibandingkan dengan siswa Indonesia lebih unggul kurang lebih 40 point. Namun demikian, capaian Indonesia dan Malaysia masih berada di bawah skor rata-rata internasional. Hal ini berarti, kedua negara masih memiliki permasalahan dalam penguasaan literasi sains.

Pemerintah telah berupaya meningkatkan kemampuan literasi siswa, baik literasi membaca, sains, maupun matematika. Salah satu kebijakan pemerintah adalah menerapkan kurikulum yang menuntut pembelajaran terpadu dengan harapan siswa dapat memahami suatu materi pelajaran secara *holistik* dan *integratif*. Namun tidak diikuti dengan penyediaan tes kemampuan literasi sains sebagai alat evaluasinya. Buku ajar IPA yang digunakan di sekolah-sekolah pun belum memuat soal yang memiliki ciri literasi sains seperti yang diukur dalam soal PISA. Ketika siswa tidak dibiasakan mengerjakan soal yang menuntut kemampuan literasi dalam memahami suatu wacana/bacaan yang menyertai soal, maka siswa tidak akan berhasil menjawab soal literasi sains.

Hasil wawancara dengan beberapa guru SMP di Jawa Tengah, Sumatera Utara, dan Sulawesi Selatan dan beberapa dosen LPTK bidang Sains menunjukkan bahwa mereka masih kesulitan dalam mengembangkan perangkat pembelajaran, termasuk pengembangan soal literasi sains. Kondisi ini tentunya menjadi kendala bagi siswa Indonesia untuk dapat bersaing dalam penilaian PISA. Sementara itu, persaingan global menuntut siswa untuk mampu bersaing di kancah dunia. Kemampuan literasi, baik membaca matematika ataupun sains perlu diasah agar mampu bersaing dengan siswa dari berbagai negara.

Berdasarkan penelitian awal terhadap ketersediaan instrumen untuk mengukur kemampuan literasi sains di SMP wilayah Jawa Tengah, Makassar, dan Sumatera Barat menunjukkan bahwa instrumen tersebut belum tersedia. Kemampuan guru dalam membuat soal literasi sains juga masih rendah (Rusilowati & Basam, 2017a). Upaya membuat soal literasi sains telah dilakukan dalam rangka untuk mengetahui profil kemampuan literasi sains siswa SMP di kota Kudus Jawa Tengah. Soal yang dikembangkan bertema Interaksi (Rusilowati, dkk., 2016a), dan Siklus (Rusilowati, dkk., 2016b). Buku ajar sains terpadu berbasis literasi juga telah dikembangkan, dan hasil penelitian menyimpulkan bahwa buku ajar berbasis literasi sains efektif untuk meningkatkan literasi sains siswa (Rusilowati, dkk., 20015, 2016c, 2017b).

Pengembangan instrumen literasi sains dapat mengacu pada materi buku ajar berbasis literasi sains, memperhatikan tuntutan kurikulum IPA untuk SD, SMP, SMA dan sederajat, serta memperhatikan perkembangan isu pendidikan secara internasional. Instrumen dikembangkan berdasarkan tujuan pembelajaran dan indikator setiap materi, aspek literasi sains, serta indikator yang dituntut dalam TIMSS dan/atau PISA.

Pengembangan instrumen berbasis literasi sains sama dengan pengembangan instrumen pada umumnya. Pengetahuan

tentang istilah-istilah dalam evaluasi perlu dipahami lebih dahulu agar memiliki persamaan persepsi. Analisis terhadap hasil pengembangan instrumen juga perlu disampaikan agar dapat memahami dengan baik kualitas karakteristik butir dari hasil pengembangan instrumen.

Buku ini mengulas tentang konsep dasar evaluasi, bentuk instrumen, standar evaluasi, pengembangan tes, analisis butir, konsep literasi sains dan pengembangan tes literasi sains. Contoh soal berbasis literasi sains disajikan mulai dari tingkat SD, SMP, SMA, dan mahasiswa calon guru.

PENILAIAN DAN EVALUASI

A. Pengertian

Ada beberapa istilah yang berhubungan dengan evaluasi hasil belajar, yaitu tes, pengukuran (*measurement*), asesmen, penilaian, dan evaluasi. Tes merupakan salah satu instrumen yang digunakan untuk mengumpulkan informasi yang diperlukan dalam evaluasi terdiri atas sejumlah pertanyaan atau butir soal yang digunakan untuk memperoleh data atau informasi melalui respons peserta tes. Tes merupakan salah satu instrumen yang populer dan penting untuk mengevaluasi belajar siswa.

Pengukuran dapat diartikan sebagai proses mendapatkan informasi dan biasanya diwujudkan dengan angka untuk individu atau menentukan karakteristik individu menurut aturan tertentu. Salah satu alat ukurnya dapat berupa tes.

Asesmen adalah sekelompok pengertian yang mengacu pada pengumpulan data dan informasi untuk tujuan menjelaskan tingkat pengetahuan, penampilan/kinerja, dan prestasi dari individu. Asesmen mencakup pengertian yang lebih luas bukan hanya konsep pengukuran, tetapi interpretasi terhadap satu atau lebih pengukuran untuk tujuan pembuatan keputusan. Asesmen merupakan prosedur yang digunakan untuk memperoleh informasi mengenai penampilan siswa, termasuk tes dan penampilan tugas-tugas autentik.

Penilaian diartikan sebagai proses sistematis untuk mengumpulkan, menganalisis, menginterpretasikan informasi, dan menentukan tingkat keberhasilan siswa terhadap tujuan

pembelajaran. Kadang kala perbedaan antara asesmen, penilaian dan evaluasi tidak begitu jelas, sehingga para guru merasa kesulitan dalam mengaplikasikannya. Penilaian biasanya dikaitkan dengan proses dan hasil pembelajaran.

Evaluasi biasanya dikaitkan dengan keberhasilan suatu program, termasuk program pendidikan. Hasil evaluasi biasanya digunakan untuk memberikan rekomendasi kepada pemangku kebijakan dalam mengambil keputusan. Apakah suatu program layak dilanjutkan, dihentikan, ataukah dimodifikasi.

Tes berisi suatu pertanyaan/tugas/seperangkat tugas yang direncanakan untuk memperoleh informasi tentang *trait*/atribut pendidikan/psikologi. Setiap butir pertanyaan atau tugas tersebut memiliki jawaban atau ketentuan yang dianggap benar. Setiap tes menuntut keharusan adanya respons dari peserta yang dapat disimpulkan sebagai suatu *trait* yang dimiliki oleh peserta yang sedang dicari informasinya. Jadi, jika ada tugas atau pertanyaan yang harus dikerjakan oleh seseorang, tetapi tidak ada jawaban atau cara mengerjakan yang benar atau salah, atau suatu usaha pengukuran yang tidak mengharuskan peserta tidak menjawab atau mengerjakan suatu tugas, maka bukanlah tes.

Dalam setiap program penilaian dibutuhkan tahapan-tahapan:

- a. perencanaan: menetapkan tujuan dan memilih alat evaluasi yang layak (tes dan nontes)
- b. pengumpulan data : melaksanakan pengukuran
- c. verifikasi data: memeriksa hasil dan memberi skor
- d. analisis data: analisis dan interpretasi skor dengan Penilaian Acuan Norma (PAN) atau Penilaian Acuan Patokan (PAP)
- e. penafsiran data dan pengambilan keputusan: menggunakan hasil pengukuran untuk membuat keputusan

B. Perencanaan Penilaian Hasil Belajar

1. Perencanaan Umum

Perencanaan umum merupakan segenap rencana kegiatan evaluasi hasil belajar dalam suatu jenis pendidikan. Karena dalam sistem pendidikan untuk SMP/MTs, SMA/SMK/MA dan sekolah-sekolah yang setingkat keduanya menggunakan guru mata pelajaran, berarti pencapaian tujuan pendidikan untuk siswa bukan merupakan tanggung jawab seorang guru. Oleh karena itu, setiap guru di suatu sekolah harus memahami kompetensi lulusan suatu tingkat sekolah (SD/SMP/SMA) dan menyadari langkah yang dilakukan dalam mencapai kompetensi tersebut melalui mata pelajaran yang diajarkannya.

Program penilaian dapat disusun untuk jangka waktu yang lama atau pun jangka pendek. Program tersebut harus bersifat singkat dan cukup jelas bagi setiap orang yang menggunakannya. Program penilaian untuk suatu sekolah hendaknya memuat hal-hal sebagai berikut.

- a. Perincian kompetensi lulusan, kompetensi inti (KI), kompetensi dasar (KD)
- b. Perincian mengenai aspek peningkatan yang harus diperhatikan dalam setiap tindakan penilaian.
- c. Metode penilaian yang dapat digunakan.
- d. Masalah alat penilaian yang dapat digunakan.
- e. Kriteria dan skala yang digunakan.
- f. Jadwal kegiatan penilaian.

2. Perencanaan Khusus

Perencanaan khusus yang dimaksud adalah perencanaan yang dilakukan oleh setiap pengajar setiap kali akan mengadakan penilaian. Sistem penilaian menurut kurikulum dikembangkan dengan *competency based assessment* dan dilakukan secara berkelanjutan. Oleh karena itu, perlu ada langkah-langkah yang harus ditempuh, yaitu membuat rencana penilaian selama satu

semester (promes) dan satu tahun (prota), membuat kisi-kisi tes berdasarkan kompetensi yang akan dicapai, menyiapkan tes untuk kuis, ulangan harian, dan ujian semester.

Sesuai panduan penilaian Badan Standar Nasional Pendidikan (BSNP), perencanaan khusus yang sesuai dengan prosedur penilaian oleh pendidik menggunakan langkah-langkah sebagai berikut:

a. Penentuan Tujuan Penilaian

Penentuan tujuan penilaian merupakan langkah awal dalam rangkaian kegiatan penilaian secara keseluruhan seperti untuk penilaian harian, tengah semester, akhir semester, kenaikan kelas, atau penilain akhir dari satuan pendidikan.

b. Penyusunan Kisi-kisi

Kisi-kisi sangat menentukan cakupan aspek yang akan diukur dan menggambarkan validitas instrumen. Ada bermacam-macam kisi-kisi. Untuk tes prestasi ada kisi-kisi yang modelnya didasarkan pada tingkat kesulitan item dan model kognitif. Model kognitif berdasarkan penguasaan domain didasarkan pada domain-domain yang menjadi sasaran dalam kurikulum sehingga disebut juga tes berbasis kurikulum. Tes ini digunakan untuk menilai prestasi pengetahuan dan ketrampilan siswa secara menyeluruh dalam suatu program pembelajaran pada tingkatan kelas tertentu dan dapat digambarkan sebagai satu "termometer akademis" untuk menilai kemajuan siswa dalam satu set komponen pengetahuan dan ketrampilan yang luas. Jika distandardisasikan, maka perubahan di dalam skor tes dapat dipakai untuk menggambarkan kemajuan siswa. Materi di dalam tes juga disampel dari keseluruhan materi kurikulum, kemudian serangkaian tes dirancang untuk dilaksanakan pada waktu-waktu tertentu agar dapat mengukur cakupan materi dan ketrampilan yang sesungguhnya, termasuk semua komponen kurikulum. Dengan demikian, mustahil jika dapat diukur hanya dengan satu

bentuk tes. Berikut ini disajikan contoh format kisi-kisi penilaian yang menyatu dengan silabus.

Silabus Pembelajaran

Sekolah :
 Mata Pelajaran :
 Kelas/Semester :
 Kompetensi Inti :

Kompetensi Dasar	Materi Pokok/ Materi Pembelajaran	Kegiatan Pembelajaran	Indikator Pencapaian	Penilaian		Alokasi Waktu	Sumber Belajar
				Teknik Penilaian	Bentuk Instrumen		

Perencanaan penilaian yang sudah dilengkapi dengan contoh instrumen disajikan secara menyatu dengan Rencana Pelaksanaan Pembelajaran (RPP). Berikut ini adalah contoh kisi-kisi penilaian yang menyatu dengan RPP.

Rencana Pelaksanaan Pembelajaran (RPP)

Sekolah :
 Mata Pelajaran :
 Kelas/Semester :
 Alokasi Waktu : ... jam pelajaran (... x pertemuan)
 A. KI :
 B. KD :
 C. Indikator :
 D. Pendekatan dan Metode :
 E. Skenario/Langkah-langkah Kegiatan Pembelajaran
 Pertemuan 1 : Awal
 Inti.....
 Penutup.....
 Pertemuan 2 : dst.
 F. Evaluasi
 G. Sumber Belajar

Untuk menilai pencapaian kompetensi inti dalam satu semester, pendidik merancang penilaian untuk semester yang bersangkutan. Kisi-kisi ulangan akhir semester memuat KI, KD, dan indikator pencapaiannya yang dapat dijadikan dasar penyusunan tes pada akhir semester. Kisi-kisi ulangan akhir semester dapat dirancang dengan memuat tes tertulis dan tes praktik yang formatnya disajikan sebagai berikut.

Kisi-kisi Ulangan Akhir Semester

Sekolah :
 Mata Pelajaran :
 Kelas/Semester :
 Alokasi waktu :

Standar Kompetensi	Kompetensi Dasar	Indikator Pencapaian	Teknik Penilaian	
			Tes Tertulis	Tes Praktik
Dituliskan seluruh SK dalam semester bersangkutan	Dituliskan KD yang esensial dari KI yang bersangkutan	Dituliskan indikator pencapaian yang esensial dari KD yang bersangkutan.	Dicantumkan bentuk tes obyektif/essai	Dituliskan bentuk tes yang dipilih seperti: tes keterampilan, simulasi, unjuk kerja

c. Perumusan Indikator Pencapaian

Indikator pencapaian dikembangkan oleh pendidik berdasarkan KD mata pelajaran dengan memperhatikan hal-hal sebagai berikut:

- a. Rumusan indikator menggunakan kata kerja operasional.
- b. Tiap KD dikembangkan dua atau lebih indikator
- c. Tiap indikator dapat dibuat lebih dari satu butir instrumen.
- d. Indikator memiliki aspek manfaat atau terkait dengan kehidupan sehari-hari.

d. Penyusunan Instrumen

Instrumen yang digunakan dalam penilaian meliputi tes dan nontes. Langkah-langkah penyusunan instrumen disesuaikan dengan karakteristik teknik dan bentuk butir instrumennya.

e. Telaah Instrumen

Instrumen penilaian dalam bentuk tertulis, lisan, maupun perbuatan perlu analisis baik secara kualitatif maupun secara kuantitatif. Analisis kualitatif dilakukan secara rasional bersama teman sejawat, sedangkan analisis kuantitatif dilakukan secara statistik dengan menggunakan data hasil uji coba yang akan diuraikan tersendiri pada analisis butir soal.

f. Pelaksanaan Penilaian

Dalam buku panduan penilaian yang dikeluarkan BSNP tahun 2008, penilaian untuk kelompok mata pelajaran IPTEK dilakukan melalui ulangan harian, ulangan tengah semester, ulangan akhir semester, penugasan, dan pengamatan dengan menggunakan instrumen yang sesuai dengan SK dan KD. Penilaian melalui ulangan dapat dilakukan dalam bentuk tes tertulis dan/atau tes praktik bergantung pada karakteristik mata pelajaran. Penilaian harus dilaksanakan dalam situasi dan kondisi yang memungkinkan siswa menunjukkan kemampuan optimalnya. Untuk itu, penilaian harus dilakukan sesuai dengan prinsip-prinsip penilaian. Guru kelompok mata pelajaran IPTEK juga bertanggung jawab menilai aspek afektif siswa, baik yang berkaitan dengan akhlak maupun kepribadian. Hasil penilaian terhadap akhlak siswa akan dijadikan pertimbangan pada saat guru mata pelajaran pendidikan agama menentukan nilai akhlak siswa untuk dilaporkan pada buku laporan pendidikan atau rapor. Demikian pula hasil penilaian terhadap kepribadian siswa, juga akan dijadikan pertimbangan pada saat guru mata pelajaran pendidikan kewarganegaraan menentukan nilai kepribadian siswa untuk dilaporkan pada buku

laporan pendidikan atau rapor. Untuk menilai akhlak siswa, guru mata pelajaran iptek melakukan pengamatan terhadap perilaku siswa, baik di dalam maupun di luar kelas yang berkait antara lain dengan kedisiplinan, kejujuran, tanggung jawab, sopan santun, dan hubungan sosial. Untuk menilai kepribadian siswa, guru mata pelajaran iptek melakukan pengamatan terhadap perilaku siswa, baik di dalam maupun di luar kelas. Pengamatan ini dimaksudkan untuk menilai perilaku siswa yang mencerminkan kepribadian seperti percaya diri, harga diri, motivasi diri, kompetisi, saling menghargai, dan kerjasama.

C. Teknik Penilaian

Teknik penilaian adalah cara penilaian itu dilaksanakan. Teknik penilaian dapat dibedakan menjadi teknik penilaian tes dan nontes. Teknik penilaian tes meliputi jenis tagihan kuis dan tes harian, sedangkan teknik penilaian nontes antara lain meliputi: observasi, angket, wawancara, tugas, proyek, dan portofolio.

1. Kuis

Kuis merupakan sejumlah pertanyaan singkat (lisan atau tertulis) yang menanyakan hal-hal yang prinsip (pengetahuan dan pemahaman). Dilakukan sebelum pelajaran dimulai atau menjelang akhir pelajaran pada satu pertemuan selama 10 – 15 menit. Bentuk instrumen yang dapat digunakan misalnya pertanyaan lisan/tertulis yang menghendaki jawaban singkat, pertanyaan isian singkat, menjodohkan, pilihan ganda, bahkan berupa unjuk kerja yang dilakukan secara singkat, misalnya menyampaikan hasil percobaan gerak fototropisme dan lainnya.

2. Tes Harian

Tes yang dilakukan dalam proses dan akhir pembelajaran suatu kompetensi dan dilakukan untuk mengetes pencapaian pada satu indikator/KD. Bentuk tes yang dapat digunakan misalnya

pilihan ganda, isian singkat, menjodohkan, benar salah, unjuk kerja, dan uraian.

3. Teknik Penilaian Jenis Tagihan Nontes

Jenis tagihan non-tes meliputi observasi, angket, wawancara, tugas, proyek, dan portofolio.

a. Melakukan Observasi

Penilaian dengan teknik ini dilakukan dengan cara mengamati untuk menilai/mengetahui perkembangan (proses) dan atau pencapaian kompetensi, sikap, dan motivasi siswa baik secara individu maupun kelompok. Bentuk instrumen yang dapat digunakan adalah pedoman observasi yang berisi sejumlah pertanyaan, isian, atau *check list* yang diisi oleh guru.

b. Memberikan Angket/Kuesioner/Skala

Bentuk instrumen pada teknik penilaian ini berupa angket/kuesioner/skala yang berisi sejumlah pertanyaan/ pernyataan/isian singkat untuk menilai/mengetahui perkembangan (proses) dan atau pencapaian kompetensi, sikap, dan motivasi siswa yang diisi oleh siswa.

c. Melakukan Wawancara

Bentuk instrumen yang digunakan pada teknik penilaian dengan wawancara adalah pedoman wawancara yang berisi sejumlah pertanyaan yang disampaikan secara lisan kepada siswa dan dijawab secara lisan pula. Biasanya untuk mengungkap sikap/aspirasi/motivasi siswa terhadap PBM dan atau nilai-nilai pedagogis pada umumnya. Wawancara dalam hal ini bukan termasuk tes yang dilakukan secara lisan.

d. Memberikan Tugas

Pemberian tugas dimaksudkan agar pembelajaran, *reinforcement* (penguatan), dan pengayaan untuk KD tertentu

dapat terjadi. Tugas ini dapat diberikan secara individu atau kelompok. Secara periodik, sebaiknya berupa tugas aplikasi, bila mungkin sampai sintesis dan evaluasi yang dapat dikerjakan. Untuk menilai tugas, bentuk instrumen penilaiannya disesuaikan dengan jenis tugas yang diberikan.

e. Memberikan Proyek

Proyek merupakan tugas yang dikerjakan agar pembelajaran, *reinforcement* (penguatan), dan pengayaan terjadi dengan melibatkan sejumlah KD secara terintegrasi (tematik) yang dikerjakan secara individu atau kelompok (terdiri atas 2-4 orang) untuk periode tertentu. Proyek dapat melibatkan serangkaian kegiatan sesuai dengan sifat dari proyek, misalnya pengumpulan dan pengorganisasian data, analisis data, dan penyajian data dalam bentuk laporan. Proyek sering kali melibatkan pencarian data primer dan sekunder, mengevaluasi secara kritis hasil penyelidikan, dan kerjasama dengan orang lain. Penilaian dilakukan terhadap proses (perencanaan dan proses penyelesaiannya) dan produk proyek (hasil akhir). Dalam satu semester bisa diberikan satu atau dua proyek. Bentuk instrumen penilaian yang digunakan berupa rubrik.

f. Menerapkan Portofolio

Kumpulan tugas, karya, prestasi akademik/non-akademik, komentar/catatan guru tentang prestasi di dalam maupun luar sekolah baik yang sudah dinilai maupun yang belum. Kumpulan ini menggambarkan perkembangan kemampuan/kompetensi yang dicapai siswa. Setiap siswa memiliki *folder*. Kriteria dan pembobotan penilaian terhadap dokumen dalam portofolio yang data nilainya belum ada (misalnya prestasi dalam kejuaraan lomba pidato bahasa Inggris) hendaknya dikomunikasikan kepada siswa. Bentuk alat penilaian yang digunakan berupa lembar penilaian yang dilengkapi dengan rubrik penskoran.

BENTUK ALAT PENILAIAN

A. Bentuk Alat Penilaian

Alat yang digunakan untuk melakukan pengukuran/penilaian/evaluasi terhadap pencapaian kompetensi siswa dapat berupa tes dan nontes. Berikut adalah bentuk-bentuk instrumen yang dikelompokkan menurut jenis tagihan dan teknik penilaian.

1. Tes

Teknik-teknik penilaian yang termasuk dalam jenis tagihan tes meliputi kuis dan tes harian. Tingkat berpikir yang digunakan dalam mengerjakan tes harus mencakup mulai yang rendah sampai yang tinggi, dengan proporsi yang sebanding sesuai dengan jenjang pendidikan. Pada jenjang pendidikan menengah, tingkat berpikir yang terlibat sebaiknya diperbanyak pada tingkat pemahaman, aplikasi, dan analisis. Namun hal ini bergantung pada karakteristik bidang studi. Bentuk-bentuk tes yang dapat dikembangkan adalah tes objektif dan uraian.

a. Tes Objektif

Tes objektif merupakan tes yang dibuat sedemikian rupa sehingga hasil tes tersebut dapat dinilai secara objektif, dinilai oleh siapapun akan menghasilkan skor yang sama. Peserta tes tinggal memilih, mengisi, atau menjodohkan, sesuai dengan perintah yang tertera dalam soal. Bentuk-bentuk tes objektif meliputi:

- 1) soal bentuk melengkapi (*completion*)

- 2) salah-benar (*true-false*)
- 3) pilihan ganda (*multiple choice*)
- 4) soal bentuk klasifikasi atau penggolongan (*classification*)
- 5) pasangan/menjodohkan (*matching*)
- 6) sebab-akibat, dan
- 7) asosiasi pilihan ganda

Bentuk-bentuk tes melengkapi, salah-benar, menjodohkan, dan pilihan ganda biasanya menilai kemampuan berpikir tingkat rendah, yaitu pengetahuan. Bentuk-bentuk ini dapat mencakup banyak materi pembelajaran, penskorannya bersifat objektif dan dapat dikoreksi dengan cepat. Oleh karena itu, bentuk ini digunakan untuk penilaian yang melibatkan banyak siswa dan waktu koreksi yang sempit. Salah satu kelemahannya adalah peserta didik tidak diberi kesempatan untuk mengembangkan sendiri jawabannya. Bentuk-bentuk tes ini kurang dianjurkan pemakaiannya dalam penilaian kelas karena tidak dapat menggambarkan dengan baik kemampuan peserta didik yang sesungguhnya. Namun demikian pilihan ganda (PG) dapat digunakan untuk mengukur kemampuan tingkat tinggi, asalkan butir soal pilihan ganda berkualitas.

Butir soal bentuk salah-benar berupa suatu pernyataan yang ditawarkan kepada peserta ujian untuk diputuskan benar atau salah, ya atau tidak. Pokok uji bentuk benar-salah akan berbobot jika pernyataan yang diajukan mengandung unsur analisis atau sintesis. Untuk menyatakan benar salahnya pernyataan yang diajukan dapat dengan melingkari atau menyilang huruf B jika benar atau S jika salah.

Butir soal bentuk pilihan ganda berupa satu pernyataan yang belum lengkap (disebut *stem*). Untuk melengkapinya, ditambahkan jawaban yang dipilih di antara beberapa alternatif atau pilihan yang telah disediakan dalam soal. Alternatif jawaban yang benar disebut kunci jawaban dan yang salah atau kurang benar disebut pengecoh (*distraktor*). Selain berupa pernyataan

yang belum lengkap, *stem* dapat juga berupa suatu pernyataan. Ada yang hanya berupa pernyataan dan ada yang diawali dengan uraian kasus atau grafik.

Butir soal bentuk klasifikasi atau penggolongan menggiring peserta ujian untuk mengelompokkan obyek atau hal berdasarkan kesamaan ciri yang dimiliki. Jadi, dasarnya adalah prinsip kesamaan dan perbedaan.

Butir soal bentuk pasangan bukan merupakan butir soal tunggal, tetapi disusun berseri, yakni satu seri pertanyaan dan satu seri jawaban. Peserta ujian harus mencari atau menempatkan jawaban sehingga cocok dengan pertanyaannya. Cara menjawabnya dapat dengan menyilang atau melingkari nomor/tanda jawaban yang sesuai atau diminta menulis secara lengkap.

Butir soal bentuk sebab akibat memiliki tingkat kesukaran yang relatif tinggi dibandingkan bentuk yang lain. Pada dasarnya bentuk sebab akibat terdiri atas dua pernyataan (*stem*). Pernyataan pertama akibat dan pernyataan kedua berperan sebagai sebab atau alasan. Oleh karena itu, antara pernyataan pertama dan kedua selalu disisipi atau dihubungkan dengan kata sebab atau karena. Peserta ujian diminta untuk memutuskan apakah memang kedua pernyataan tersebut menunjukkan hubungan sebab akibat, kedua pernyataan benar tetapi tidak menunjukkan hubungan sebab akibat, kedua pernyataan salah satu diantaranya salah atau kedua pernyataan salah. Oleh karena itu, alternatif jawabannya dapat empat atau lima. Alternatif akan sebanyak lima jika alternatif salah satu pernyataan salah dipisah menjadi pernyataan pertama benar dan pernyataan kedua (alasan) salah serta pernyataan pertama salah dan pernyataan kedua (alasan) benar.

Butir soal bentuk asosiasi pilihan ganda merupakan bentuk pilihan ganda kompleks. Bentuk ini menuntut peserta ujian mencari seluruh alternatif jawaban yang benar, mungkin semua

alternatif jawaban benar, mungkin hanya sebagian yang benar, dan mungkin tak ada yang benar. Alternatif jawaban yang biasa digunakan adalah sebagai berikut:

- a. pernyataan (1), (2), dan (3) benar
- b. pernyataan (1) dan (3) benar
- c. pernyataan (2) dan (4) benar
- d. pernyataan (4) benar

b. Tes Uraian

Bentuk tes uraian memberi kesempatan kepada peserta tes untuk mengorganisasikan gagasan dan atau hal-hal yang sudah dipelajari dengan menggunakan kata-katanya sendiri. Bentuk uraian dibedakan menjadi uraian terbatas dan uraian bebas. Uraian terbatas cocok untuk mata pelajaran yang jawabannya cenderung tidak memiliki variasi misalnya matematika dan IPA. Agar hasil penskorannya objektif diperlukan pedoman penskoran. Penskoran dilakukan pada setiap langkah pengerjaan. Misalnya menuliskan rumus, menghitung hasil, menafsirkan dan menyimpulkan hasilnya.

Penskoran bersifat hirarkis sesuai dengan langkah pengerjaan soal. Bobot skor untuk tiap butir soal ditentukan berdasarkan tingkat kesulitannya. Soal yang sulit diberi bobot lebih besar daripada soal yang mudah.

Bentuk uraian bebas cocok untuk bidang studi ilmu-ilmu sosial. Walaupun hasil penskoran cenderung subjektif, bila disediakan pedoman penskoran hasilnya diharapkan dapat lebih objektif. Tingkat berpikir yang diukur biasanya tingkat tinggi.

c. Unjuk kerja

Penilaian unjuk kerja didasarkan pada hasil pengamatan terhadap unjuk kerja, tingkah laku, atau interaksi siswa dalam tugas tertentu. Penilaian ini lebih tepat digunakan untuk menilai kompetensi siswa dalam berpidato, membaca puisi, memecahkan

masalah dalam kelompok, partisipasi siswa dalam diskusi kelompok, menari, memainkan alat musik, melakukan aktivitas berbagai cabang olah raga, dan menggunakan peralatan laboratorium. Untuk menilai unjuk kerja diperlukan instrumen dan rubrik penilaian.

2. Bentuk Alat Penilaian Nontes

Pada Bab 1 telah disebutkan teknik penilaian nontes dapat berupa kegiatan observasi, pemberian kuesioner, melalukan wawancara, pemberian proyek, penerapan portofolio. Dengan demikian, alat penilaian yang sesuai dengan teknik penilaian tersebut adalah:

- a. Panduan observasi
- b. Kuesioner
- c. Panduan Wawancara
- d. Lembar penilaian

Alat-alat penilaian tersebut tentunya harus dilengkapi dengan rubrik penskoran. Rubrik berisi ketentuan mengenai komponen/aspek, kriteria, dan bobot penilaian. Rubrik penskoran juga diperlukan ketika menilai tes uraian, unjuk kerja, portofolio, tugas, dan proyek.

B. Pelaksanaan Penilaian

1. Test Tertulis

Dalam pelaksanaan tes tertulis perlu diperhatikan hal-hal sebagai berikut: (a) ruang tes diusahakan tenang dan bangku-bangku dalam ruangan tes disusun cukup longgar, (b) siswa diperingatkan agar tidak bekerja sebelum tanda dimulai, untuk itu soal dibagikan terbalik, (c) pengawas tes berlaku adil dan tidak mengganggu suasana testing, (d) pengawas tes mengakhiri sampai batas waktu testing yang telah ditentukan dan membuat catatan kejadian selama testing/mengisi berita acara. Agar data yang diperoleh adalah hasil kemampuan siswa yang sebenarnya, maka pengawas harus melakukan pengawasan secara ketat dan dipastikan peserta tes tidak bekerjasama atau berlaku curang.

2. Tes Lisan

Hal-hal yang perlu diperhatikan pada pelaksanaan tes lisan adalah sebagai berikut: (a) tester/guru jangan mengubah suasana tes lisan menjadi suasana diskusi, (b) guru hendaknya tidak berkata keras dan mengeluarkan kata tidak senonoh seperti 'tolol', 'bebal', dsb, (c) guru jangan memberi 'kunci-kunci' tertentu karena simpati atau kasihan, (d) menyiapkan pertanyaan dan *scope* jawaban yang diminta, (e) berikan skoring secara teliti untuk setiap jawaban yang diberikan peserta tes. Perlu disiapkan kunci jawaban dan rubrik penskoran.

3. Pelaksanaan Observasi untuk Asesmen Otentik

Untuk pelaksanaan observasi perlu memperhatikan beberapa hal agar tidak menimbulkan *bias*. Hal-hal yang perlu diperhatikan adalah: (a) menetapkan situasi observasi yang dikehendaki (situasi bebas/*free situation*, dibuat/*manipulated situation*, ataukah campuran/*partially controlled*), (b) cara mencatatkan observasi (bergantung apakah observasi terstruktur atau bukan), (c) lamanya observasi, dan (d) posisi pengamat (partisipan atau bukan). Lembar observasi perlu dipersiapkan.

KUALITAS INSTRUMEN

Alat penilaian/instrumen yang baik harus memenuhi standar validitas dan reliabilitas yang memadai. Baik instrumen tes ataupun nontes perlu diuji validitas dan reliabilitasnya. Tes sebagai alat seleksi maupun evaluasi diharapkan dapat menghasilkan nilai atau skor yang objektif dan akurat. Sejak awal, tes hendaknya disusun sesuai dengan prosedur dan prinsip penyusunan tes. Baik buruknya tes dapat dilihat dari hasil analisis terhadap validitas tes, reliabilitas tes, dan karakteristik butir soal.

Pengujian validitas dan reliabilitas tidak hanya berlaku pada alat penilaian berupa tes, tetapi alat penilaian yang nontes pun harus diuji validitas dan reliabilitasnya. Cara pengujiannya relatif sama.

A. Validitas Tes

Validitas tes ditentukan oleh ketepatan dan kecermatan hasil pengukuran. Pengukuran sendiri dilakukan untuk mengetahui seberapa banyak suatu aspek psikologis terdapat dalam diri seseorang, yang dinyatakan oleh skornya pada instrumen pengukur yang bersangkutan. Walaupun kita terbiasa melekat predikat valid bagi suatu tes, akan tetapi perlu dipahami bahwa sebenarnya pengertian validitas menyangkut masalah hasil ukur bukan masalah alat ukurnya sendiri. Sebutan validitas tes hendaknya diartikan sebagai validitas hasil pengukuran yang diperoleh oleh tes tersebut.

Berdasarkan cara estimasinya, tipe validitas pada umumnya digolongkan menjadi tiga kategori, yaitu *content validity* (validitas isi), *construct validity* (validitas konstruk), dan *criterion-related validity* (validitas berdasar kriteria).

1. Validitas Isi

Validitas isi merupakan validitas yang diestimasi lewat pengujian terhadap isi tes dengan analisis rasional atau lewat *professional judgement*. Pertanyaan yang dicari jawabannya dalam validasi ini adalah "sejauh mana butir soal dalam tes mencakup keseluruhan kawasan isi objek yang hendak diukur" atau "sejauh mana isi tes mencerminkan ciri atribut yang hendak diukur".

Validitas isi ini dilakukan sejak awal penyusunan tes, bukan setelah tes selesai dirancang. Validitas isi terkait dengan sejauh mana kesesuaian suatu item dengan indikator berperilaku dari atribut yang diukur (Azwar, 2012). Layak tidaknya suatu item ditentukan oleh hasil penilaian (*judgement*) yang dilakukan oleh ahli/*rater* berdasarkan logika. Objektivitas *judgement* dapat ditingkatkan dengan menambahkan banyak *rater*.

Estimasi validitas isi tidak melibatkan perhitungan statistik, melainkan hanya analisis rasional, maka tidaklah diharapkan setiap orang akan sependapat mengenai sejauh mana validitas isi suatu tes telah tercapai. Selanjutnya, validitas isi terbagi menjadi dua tipe yaitu *face validity* (validitas muka) dan *logical validity* (validitas logik).

Validitas muka merupakan tipe validitas yang paling rendah signifikansinya karena hanya didasarkan pada penilaian terhadap format penampilan tes. Apabila penampilan tes sudah meyakinkan dan memberikan kesan mampu mengungkap apa yang hendak diukur maka dikatakan validitas muka telah terpenuhi.

Validitas logik menunjuk sejauh mana isi tes merupakan representasi dari ciri-ciri atribut yang hendak diukur. Untuk memperoleh validitas logik yang tinggi suatu tes harus dirancang sedemikian rupa sehingga benar-benar hanya berisi butir soal yang relevan dan perlu mejadi bagian tes secara keseluruhan. Pada tes kognitif, kawasan perilaku yang hendak diukur dapat dikembalikan pada tujuan instruksional (standar komptensi dan kompetensi dasar). Dengan bantuan silabus, cakupan tes yang akan dibuat menjadi lebih jelas. Dalam perencanaan tes prestasi, penyusun tes perlu membuat kisi-kisi (*blue print*) agar validitas logik dapat tercapai.

Analisis terhadap validitas isi suatu instrumen yang dilakukan oleh beberapa *rater* dapat menggunakan beberapa rumus, antara lain Aiken's V dan CVR.

a. Aiken's V

Perhitungan validitas instrumen menurut Aiken (1985) menggunakan rumus:

$$V = \frac{\sum s}{n(c-1)}$$

Keterangan :

- V : nilai validitas
- s : r - lo
- r : skor kategori pilihan ahli
- lo : skor penilaian terendah
- n : banyaknya ahli
- c : skor penilaian tertinggi

Rumus di atas digunakan jika suatu instrumen divalidasi sebanyak n ahli. Rumus tersebut juga berlaku apabila sejumlah n butir divalidasi oleh satu validator (ahli). Kriteria validitas mengacu pada nilai validitas hitung yang dibandingkan dengan nilai koefisien validasi tabel seperti yang tertera pada Tabel 4.1. Apabila nilai $V_{hitung} \geq V_{tabel}$ berarti instrumen valid.

Kolom pertama Tabel 4.1 menunjukkan jumlah *rater* atau jumlah *item* butir. Kolom-kolom berikutnya menunjukkan pilihan skala ketika memberikan penilaian. Apabila menyusun instrumen/skala dengan pilihan sangat tidak relevan, tidak relevan, relevan, dan sangat relevan, maka ada empat kategori pilihan.

Jumlah *rater* atau butir mempengaruhi besarnya nilai minimum indeks validitas V . Semakin banyak *rater*, semakin kecil nilai V yang disyaratkan. Dari tabel tersebut terlihat bahwa sebenarnya secara teoretis batas minimal untuk menetapkan jumlah *rater* adalah dua orang. Namun demikian, untuk dapat diterima suatu item harus memperoleh nilai V yang sempurna (1).

Tabel 4.1 juga memberikan dua pilihan nilai p yang ditetapkan. Jika peneliti menetapkan nilai $p < 0,01$ yang artinya mengizinkan peluang kesalahan sebesar 1%, maka dilihat baris pertama tiap jumlah *rater*. Jika peneliti menetapkan nilai $p < 0,05$, maka dilihat baris kedua tiap jumlah *rater*.

Tabel 4.1 Right-Tail Probability (p) for Selected Values of the Validity Coefficient (V)

Right-Tail Probabilities (p) for Selected Values of the Validity Coefficient (V)

No. of Items (m) or Raters (n)	Number of Rating Categories (c)													
	2		3		4		5		6		7			
	V	p	V	p	V	p	V	p	V	p	V	p		
2							1.00	.040	1.00	.028	1.00	.020		
3							1.00	.008	1.00	.005	1.00	.003		
3			1.00	.037	1.00	.016	.92	.032	.87	.046	.89	.029		
4					1.00	.004	.94	.008	.95	.004	.92	.006		
4			1.00	.012	.92	.020	.88	.024	.85	.027	.83	.029		
5			1.00	.004	.93	.006	.90	.007	.88	.007	.87	.007		
5	1.00	.031	.90	.025	.87	.021	.80	.040	.80	.032	.77	.047		
6			.92	.010	.89	.007	.88	.005	.83	.010	.83	.008		
6	1.00	.016	.83	.038	.78	.050	.79	.029	.77	.036	.75	.041		
7			.93	.004	.86	.007	.82	.010	.83	.006	.81	.008		
7	1.00	.008	.86	.016	.76	.045	.75	.041	.74	.038	.74	.036		
8	1.00	.004	.88	.007	.83	.007	.81	.008	.80	.007	.79	.007		
8	.88	.035	.81	.024	.75	.040	.75	.030	.72	.039	.71	.047		
9	1.00	.002	.89	.003	.81	.007	.81	.006	.78	.009	.78	.007		
9	.89	.020	.78	.032	.74	.036	.72	.038	.71	.039	.70	.040		
10	1.00	.001	.85	.005	.80	.007	.78	.008	.76	.009	.75	.010		
10	.90	.001	.75	.040	.73	.032	.70	.047	.70	.039	.68	.048		
11	.91	.006	.82	.007	.79	.007	.77	.006	.75	.010	.74	.009		
11	.82	.033	.73	.048	.73	.029	.70	.035	.69	.038	.68	.041		
12	.92	.003	.79	.010	.78	.006	.75	.009	.73	.010	.74	.008		
12	.83	.019	.75	.025	.69	.046	.69	.041	.68	.038	.67	.049		
13	.92	.002	.81	.005	.77	.006	.75	.006	.74	.007	.72	.010		
13	.77	.046	.73	.030	.69	.041	.67	.048	.68	.037	.67	.041		
14	.86	.006	.79	.006	.76	.005	.73	.008	.73	.007	.71	.009		
14	.79	.029	.71	.035	.69	.036	.68	.036	.66	.050	.66	.047		
15	.87	.004	.77	.008	.73	.010	.73	.006	.72	.007	.71	.008		
15	.80	.018	.70	.040	.69	.032	.67	.041	.65	.048	.66	.041		
16	.88	.002	.75	.010	.73	.009	.72	.008	.71	.007	.70	.010		
16	.75	.038	.69	.046	.67	.047	.66	.046	.65	.046	.65	.046		
17	.82	.006	.76	.005	.73	.008	.71	.010	.71	.007	.70	.009		
17	.76	.025	.71	.026	.67	.041	.66	.036	.65	.044	.65	.039		
18	.83	.004	.75	.006	.72	.007	.71	.007	.70	.007	.69	.010		
18	.72	.048	.69	.030	.67	.036	.65	.040	.64	.042	.64	.044		
19	.79	.010	.74	.008	.72	.006	.70	.009	.70	.007	.68	.009		
19	.74	.032	.68	.033	.65	.050	.64	.044	.64	.040	.63	.048		
20	.80	.006	.72	.009	.70	.010	.69	.010	.68	.010	.68	.008		
20	.75	.021	.68	.037	.65	.044	.64	.048	.64	.038	.63	.041		
21	.81	.004	.74	.005	.70	.010	.69	.008	.68	.010	.68	.009		
21	.71	.039	.67	.041	.65	.039	.64	.038	.63	.048	.63	.045		
22	.77	.008	.73	.006	.70	.008	.68	.009	.67	.010	.67	.008		
22	.73	.026	.66	.044	.65	.035	.64	.041	.63	.046	.62	.049		
23	.78	.005	.72	.007	.70	.007	.68	.007	.67	.010	.67	.009		
23	.70	.047	.65	.048	.64	.046	.63	.045	.63	.044	.62	.043		
24	.79	.003	.71	.008	.69	.006	.68	.008	.67	.010	.66	.010		
24	.71	.032	.67	.030	.64	.041	.64	.035	.62	.041	.62	.046		
25	.76	.007	.70	.009	.68	.010	.67	.009	.66	.009	.66	.009		
25	.72	.022	.66	.033	.64	.037	.63	.038	.62	.039	.61	.049		

(Sumber: Aiken, 1985)

Contoh menentukan validitas isi suatu instrumen/skala dengan 5 pilihan kategori divalidasi oleh 3 *rater*. Hasil penilaian sebagai berikut.

No Butir	<i>Rater 1</i>	<i>Rater 2</i>	<i>Rater 3</i>
1	4	4	5
2	4	3	3
3	5	4	4

Dari contoh tersebut dapat ditentukan besarnya $n = 3$, $c = 5$, $l_0 = 1$. Hasil perhitungan V adalah sebagai berikut.

No Butir	<i>Rater 1</i>		<i>Rater 2</i>		<i>Rater 3</i>		$\sum S$	V
	r	s	R	S	r	s		
1	4	3	4	3	5	4	10	0,83
2	5	4	5	4	4	3	11	0,92
3	5	4	4	3	4	3	10	0,83

Butir 1 dan 3 menghasilkan $V_{hitung} < V_{tabel}$, karena nilai V_{tabel} untuk 3 *rater* dan 5 pilihan dengan taraf kesalahan 5% adalah 0,92. Dengan demikian *item* tersebut belum valid. Butir nomor 2 sudah valid, karena $V_{hitung} = V_{tabel}$.

b. CVR

CVR (*Content Validity Ratio*) merupakan cara lain untuk menilai validitas isi suatu instrumen. Untuk menentukan CVR diperlukan sejumlah ahli materi (*Subject Matter Experts/SME*) yang diminta untuk menilai apakah item esensial dan merepresentasikan dengan baik tujuan pengukuran. Lawshe (1975) menentukan nilai CVR dengan rumus sebagai berikut.

$$CVR = (2n_e/n) - 1$$

- n_e : banyaknya SME yang menilai suatu item adalah esensial
- n : banyaknya SME yang melakukan penilaian

Hasil perhitungan dengan CVR selanjutnya dirujuk pada tabel nilai minimum yang ditetapkan oleh Lawshe seperti tertera pada Tabel 3.2.

Tabel 4.2 Minimum Values of CVR and CVR_t One Tailed Test, p = .05

*Minimum Values of CVR and CVR_t
One Tailed Test, p = .05*

No. of Panelists	Min. Value*
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

Sumber: Lawshe, 1975

Kolom pertama Tabel 3.2 menunjukkan jumlah penilai dan kolom kedua menunjukkan nilai minimal CVR yang diterima. Dari tabel terlihat bahwa jumlah penilai minimal yang disarankan adalah 5 orang, dan seluruh penilai tersebut harus menyatakan bahwa item adalah esensial agar item tersebut diterima. Apabila

hasil $CVR_{hitung} \geq CVR_{tabel}$, maka dapat disimpulkan bahwa item esensial pada taraf signifikansi 5%.

Sebagai contoh instrumen divalidasi oleh 7 orang penilai dan ada 6 orang penilai menyatakan item adalah esensial, maka nilai $CVR_{hitung} = (2 \times 6 / 7) - 1 = 0,71$. CVR_{tabel} untuk $n = 7$ adalah 0,99. Hal ini berarti $CVR_{hitung} < CVR_{tabel}$, sehingga dapat disimpulkan bahwa item tidak valid. Pada Tabel 3.2 tampak bahwa bila jumlah penilai kurang dari 8, maka semua penilai harus menyatakan item yang dinilai esensial.

2. Validitas Konstruk

Validitas konstruk merupakan tipe validitas yang menunjukkan sejauh mana tes mengungkap suatu *trait* atau konstruk teoretik yang hendak diukurnya (Allen & Yen, 1979). Meskipun pengujian validitas konstruk menggunakan teknik statistik yang kompleks, tetapi hasil estimasinya tidak dinyatakan dalam bentuk koefisien validitas.

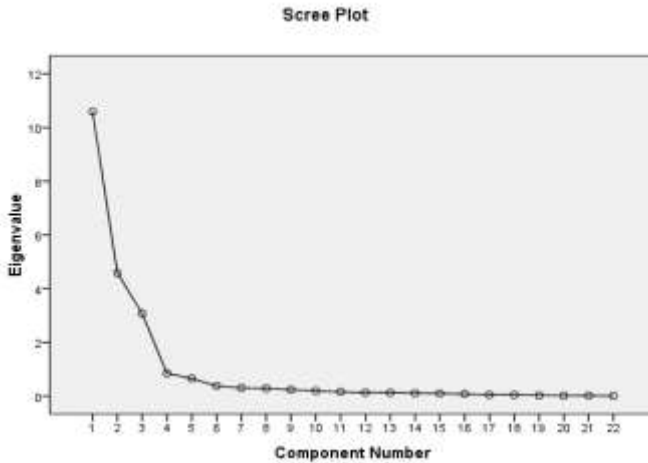
Pengujian validitas konstruk dapat menggunakan beberapa teknik, di antaranya adalah ANOVA (Meng, *et al.*, 1992; Rosenthal, *et al.*, 2000; Westen & Rosenthal, 2003); RSM (*Rule Space Model*) (Buck, *et al.*, 1997; Buck & Tatsuoka, 1998; Tatsuoka, Corter, & Tatsuoka, 2004); GDM (*General Diagnosis Model*) (von Davier, 2005; Xu & von Davier, 2006); *Confirmatory factor analysis*, *Structural Equation Modeling* (SEM) (Meng, *et al.*, 1992; Rosenthal, *et al.*, 2000; Westen & Rosenthal, 2003; M. Saefudin, 2005; Jones & Cauffman, 2006); CFA dengan program AMOS (Cooke & Michie, 2001; Jones & Cauffman, 2006).

Ada dua jenis analisis faktor, eksploratori dan konfirmatori. Ketika penyusunan alat evaluasi belum pasti tersusun oleh berapa aspek atau indikator, maka dipilih analisis faktor eksploratori. Analisis faktor eksploratori dapat menggunakan program SPSS. Ketika alat evaluasi sudah disusun oleh dimensi/aspek dan indikator yang sudah pasti, maka dipilih

analisis faktor konfirmatori. Program yang dapat digunakan untuk menguji analisis faktor konfirmatori adalah LISREL. Signifikansi hubungan antara setiap variabel laten (subkemampuan) dengan variabel kemampuan dapat dilihat dari besarnya koefisien jalur dari setiap dimensi (γ).

Setiap subkemampuan dikonstruksi dari beberapa aspek kemampuan. Signifikansi dapat dilihat dari hasil koefisien jalur dari variabel laten dan koefisien estimasi atau *factor loading* (λ) dari setiap aspek kemampuan. Pengujian signifikansi koefisien jalur dari dimensi (γ) dan koefisien estimasi (λ) tersebut dikonfirmasi dengan harga *t*. Bila harga *t* hitung > *t* tabel dengan $\alpha = 0,05$ yang sebesar 1,96, maka hasil uji adalah signifikan. Hal ini berarti indikator tersebut valid sebagai instrumen untuk mengukur variabel laten. Semakin besar λ semakin valid indikator tersebut mengukur variabel laten. Batasan yang dapat digunakan sebagai indikasi validitas instrumen pada LISREL 8.54 dapat digunakan hasil perhitungan, dan Prosedur analisis faktor konfirmatori dapat dilakukan apabila syarat data terpenuhi. Syarat tersebut antara lain: (1) *Keiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO MSA) >0,5. (2) Signifikansi *Chi-Square* (p) < 0,05, (3) *Anti Image Correlation* setiap butir > 0,5. Butir soal yang dianalisis yang nilai korelasi antar butirnya > 0,5.

Langkah selanjutnya yaitu melihat berapa banyak faktor yang mungkin terbentuk pada analisis faktor melalui *total variance explained*. Banyaknya faktor yang terbentuk dapat dilihat dari hasil *scree plot*. Berikut merupakan contoh hasil analisis faktor terhadap suatu instrumen yang terdiri atas 22 butir. Setelah dilakukan analisis faktor konfirmatori didapatkan 3 komponen yang terbentuk, 6 butir untuk faktor 1, 12 butir untuk faktor 2, dan 4 butir untuk faktor 3. Banyaknya faktor dapat dilihat pada *scree plot* Gambar 4.1.



Gambar 4.1. Scree Plot

Pada Gambar 4.1 tampak bahwa harga *eigenvalue* dari titik 1 ke 3 menurun namun masih di atas angka 1, sedangkan titik ke 4 dan seterusnya harga *eigenvalue* kurang dari 1. Hasil tersebut menunjukkan bahwa instrumen tersebut terkonstruksi dari 3 faktor.

Kecocokan model/*Goodness of Fit Index* suatu hasil pengembangan dapat diuji dengan aplikasi *Structural Equation Modelling* (SEM). Uji kecocokan model SEM meliputi uji kecocokan absolut, inkremental, dan parsimoni.

a. Uji Kecocokan Absolut

Kecocokan Absolut dapat dilihat dari komponen:

- (1) *Chi Square*, mengukur seberapa dekat matriks kovarian hasil prediksi model dengan matriks kovarian sampel data.
- (2) *Goodness of Fit Index* (GFI), menggambarkan tingkat kesesuaian model secara keseluruhan yang dihitung dari residual kuadrat dari model yang diprediksi dibandingkan dengan data yang sebenarnya.
- (3) *Root Mean Square Error of Approximation* (RMSEA), merupakan ukuran yang menggambarkan kecenderungan

chi square menolak model dengan ukuran sampel yang besar.

b. Uji Kecocokan Inkremental

Komponen yang diuji untuk kecocokan inkremental meliputi:

- (1) *Adjusted Goodness of Fit Index* (AGFI) merupakan pengembangan dari GFI yang telah disesuaikan dengan rasio dari *degree of freedom*.
- (2) *Tucker Lewis Index* (TLI) merupakan indeks kesesuaian inkremental yang membandingkan model yang diuji dengan *baseline model*.
- (3) *Normed Fit Index* (NFI) merupakan ukuran perbandingan antara *proposed model* dengan *null model*.
- (4) *Comparative Fit Index* (CFI) merupakan indeks kesesuaian inkremental yang relatif tidak sensitif terhadap besarnya sampel dan kurang dipengaruhi oleh kerumitan model.
- (5) *Incremental Fit Index* (IFI) digunakan untuk mengatasi masalah parsimoni dan ukuran sampel yang terkait dengan NFI.
- (6) *Relative Fit Index* (RFI), merupakan turunan dari NFI dan CFI.

c. Uji Kecocokan Parsimoni

Kecocokan parsimoni dapat dilihat dari hasil perhitungan terhadap komponen:

- (1) *Parsimonius Goodness of Fit Index* (PGFI) dan *Parsimonius Normed of Fit Index* (PNFI) digunakan untuk membandingkan kecocokan yang lebih baik pada model alternatif.
- (2) *Akaike Information Criterion* (AIC) dan *Consistent Akaike Information Criterion* (CAIC) merupakan indeks yang menggambarkan kecocokan perbandingan antarmodel.

Kriteria/*Cut off value* setiap hasil uji kecocokan disajikan pada Tabel 3.3.

Hasil analisis terhadap kecocokan model tidak mungkin memenuhi seluruh kriteria yang disyaratkan pada uji tersebut. Menurut Garson (2006), dukungan terhadap kecocokan model yang dikembangkan oleh data empirik paling tidak dilihat dari tiga ukuran kecocokan yang mewakili tiga kategori uji kecocokan model yang berbeda. Ketiga kategori uji kecocokan model tersebut masing-masing dihitung berdasarkan: (a) kovarians dugaan dan kovarians hasil observasi atau uji kecocokan absolut, (b) perbandingan model yang telah ditentukan dengan model alternatif atau uji kecocokan inkremental, dan (c) kovarians dugaan dan kovarians hasil observasi dengan mempertimbangkan kesederhanaan atau uji kecocokan parsimoni. Jika dua dari tiga kategori tersebut signifikan maka model yang dikembangkan cocok dengan data. Dari tiga kategori tersebut dapat dipilih masing-masing satu parameter untuk menentukan kecocokan model konstruk. Misalnya RMSEA untuk mewakili kategori (a), CFI untuk mewakili kategori (b), dan PGFI untuk mewakili kategori (c) (Rusilowati, 2012). Nilai indeks kesesuaian yang diperoleh menggambarkan benar/tidaknya suatu tes dikonstruksi oleh variabel-variabel laten, yang mempunyai indikator-indikator seperti yang telah ditentukan sebelumnya. Hair, *et al.* (1998: 656) menyatakan bahwa RMSEA merupakan ukuran paling baik digunakan dalam konfirmatori.

Popham (1995) mengemukakan tiga strategi yang dapat digunakan untuk menguji validitas konstruk, yaitu: melalui intervensi pembelajaran, perbedaan populasi uji coba, dan hubungan dalam pengukuran. Kajian intervensi pembelajaran dilakukan kepada dua kelompok siswa yang memiliki perbedaan pengetahuan (misalnya kelompok yang belum dan sudah menerima intervensi pembelajaran) diberi tes yang sama.

Tabel 4.3 Cut off value Uji Kecocokan Model (GoF)

Uji Kecocokan	Komponen	Cut off Value
Absolut	<i>Chi Square</i>	$\geq 0,05$
	<i>Goodness of Fit Index (GFI)</i>	$> 0,90$
	<i>Root Mean Square Error of Approximation (RMSEA),</i>	$< 0,08$
	<i>Root Mean Residual (RMR)</i>	$< 0,05$
Inkremental	<i>Adjusted Goodness of Fit Index (AGFI)</i>	$< 0,90$
	<i>Tucker Lewis Index (TLI)</i>	$> 0,90$
	<i>Normed Fit Index (NFI)</i>	$> 0,90$
	<i>Comparative Fit Index (CFI),</i>	$> 0,90$
	<i>Incremental Fit Index (IFI)</i>	$> 0,90$
	<i>Relative Fit Index (RFI)</i>	$\geq 0,95$
Parsimoni	<i>Parsimonius Goodness of Fit Index (PGFI)</i>	Semakin besar semakin baik
	<i>Parsimonius Normed of Fit Index (PNFI)</i>	Semakin besar semakin baik
	<i>Akaike Information Criterion (AIC)</i>	Positif dan lebih kecil
	<i>Consistent Akaike Information Criterion (CAIC)</i>	Positif dan lebih kecil

Hipotesis yang diajukan, kelompok siswa yang telah diintervensi akan merespons tes yang diberikan secara berbeda dengan kelompok siswa yang belum menerima intervensi.

3. Validitas Berdasar Kriteria

Prosedur pendekatan validitas berdasar kriteria menghendaki tersedianya kriteria eksternal yang dapat dijadikan dasar pengujian skor tes. Suatu kriteria adalah variabel perilaku yang akan diprediksi oleh skor tes atau berupa suatu ukuran lain yang relevan. Untuk melihat tingginya validitas dilakukan perhitungan terhadap korelasi antara skor tes dengan skor kriteria. Koefisien ini merupakan koefisien validitas bagi tes yang bersangkutan, yaitu r_{xy} , X melambangkan skor tes dan Y melambangkan skor kriteria.

Prosedur validitas berdasar kriteria menghasilkan dua macam validitas yaitu validitas prediktif (*predictive validity*) dan validitas konkuren (*concurrent validity*). Validitas prediktif sangat penting artinya jika tes dimaksudkan untuk berfungsi sebagai prediktor bagi performansi pada waktu yang akan datang. Misalnya tes digunakan untuk seleksi penerimaan mahasiswa baru. Untuk menguji validitas prediktifnya diperlukan kriteria indeks prestasi setelah calon mahasiswa diterima menjadi mahasiswa. Jadi, skor tes yang diperoleh sekarang baru dapat diuji validitasnya pada waktu yang akan datang, yaitu setelah skor kriterianya diperoleh.

Validitas konkuren dapat ditentukan apabila skor tes dan skor kriterianya dapat diperoleh pada waktu yang sama, sehingga korelasi antara kedua skor tersebut merupakan koefisien validitas konkuren. Misalnya kita akan menguji validitas skala *self-concept* yang baru. Untuk menguji validitas skala tersebut kita dapat menggunakan skala *self-concept* lain yang telah lebih dahulu teruji validitasnya, misalnya TSCS (*Tennessee Self Concept Scale*).

B. Reliabilitas Tes

Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan konsistensi (*consistency*) skor tes. Indeks reliabilitas berkisar antara 0 - 1.

Makin tinggi koefisien reliabilitas suatu tes (mendekati 1), makin tinggi pula konsistensi/ketepatannya. Ada 3 cara yang dapat dilakukan untuk menentukan reliabilitas tes, yaitu :

- a. *Test-retest*, yaitu tes yang sama dikerjakan oleh kelompok yang sama, sebanyak dua kali.
- b. Test paralel, satu kelompok mengerjakan 2 atau lebih tes berbeda tetapi berdasarkan kompetensi kisi-kisi yang sama.
- c. Tes belah dua: kesesuaian antara hasil pengukuran belahan pertama dan belahan kedua dari alat ukur yang sama.

1. Reliabilitas Tes (bentuk pilihan ganda)

Untuk mengetahui koefisien reliabilitas tes soal bentuk pilihan ganda digunakan rumus Kuder Richadson 20 (KR-20) seperti berikut ini.

$$KR - 20 = \frac{k}{k-1} \left[1 - \frac{\sum p(1-p)}{(SD)^2} \right]$$

Keterangan:

k : Jumlah butir soal

(SD)² : Varian

Contoh menghitung dengan KR-20:

Siswa	Soal				Σ Skor	\bar{X}	$X - \bar{X}$	$(X - \bar{X})^2$
	1	2	3	4				
A	1	0	0	0	1	2	-1	1
B	1	1	0	0	2	2	0	0
C	0	0	1	1	2	2	0	0
D	0	0	0	0	0	2	-2	4
E	1	1	0	1	3	2	-1	1
F	1	1	1	1	4	2	-2	4
p	0,67	0,50	0,33	0,50	12			10
1-p	0,33	0,50	0,67	0,50				
p(1-p)	0,22	0,25	0,22	0,25				

$$\Sigma p(1-p) = 0,22 + 0,25 + 0,22 + 0,25 = 0,944$$

$$\text{Jumlah siswa} = 6 \text{ orang}$$

$$\text{Jumlah skor} = 12$$

$$\text{Varians} = \Sigma (X - \bar{x})^2 / N = 10 / 6 = 1,67$$

$$\text{Standar Deviasi} = \sqrt{1,67} = 1,29$$

$$r = \frac{k}{k-1} \left[1 - \frac{\Sigma p(1-p)}{(SD)^2} \right]$$

$$r = \frac{4}{4-1} \left[1 - \frac{0,944}{1,67} \right]$$

$$= 0,58 \text{ (Artinya bahwa tingkat keajegan tes ini rendah)}$$

Rumus lain yang dapat digunakan adalah Spearman-Brown. Caranya adalah dengan mengelompokkan nomor butir yang ganjil dan genap. Perhatikan contoh berikut ini.

Nama Peserta Didik	Soal				Σ Skor
	1	2	3	4	
Ara	1	0	0	0	1
Benta	1	1	0	0	2
Clara	0	0	1	1	2
Danang	0	0	0	0	0
Ennis	1	1	0	1	3
Faiz	1	1	1	1	4
p	0,67	0,50	0,33	0,50	12

Nama Peserta Didik	Butir Ganjil (1+3)	$(x - \bar{x})^2$	Butir Genap (2+4)	$(x - \bar{x})^2$	Skor Z untuk		$Z_{ganjil} \times Z_{genap}$
					Ganjil	Genap	
Ara	1	0	0	1	0	-1,22	0
Benta	1	0	1	0	0	0	0
Clara	1	0	1	0	0	0	0
Danang	0	1	0	1	-1,72	-1,22	2,10
Ennis	1	0	2	1	0	+1,22	0
Faiz	2	1	2	1	+1,72	+1,22	2,10
Σ		2		4			4,2

N = 6
Mean = 1,0

Jumlah = 4,2
SD = 0,58 dan 0,82

$$r_{12} = \frac{\sum Z_{gan} x Z_{gen}}{n}$$

$$r_{12} = \frac{4,2}{6}$$

$$= 0,70$$

$$r = \frac{2r_{1.2}}{1 + r_{1.2}}$$

$$= \frac{2 \cdot (0,70)}{1 + 0,70}$$

$$= 0,82$$

(Artinya bahwa tingkat konsistensi tes ini adalah tinggi, sehingga skor tes ini dapat dipercaya penggunaannya)

Penggunaan rumus Sperman-Brown harus memenuhi syarat: (a) pembelahan random jika diambil dari tes homogen, (b) pembelahan tes gasal-genap jika isi dan tingkat kesukaran tes setara, (c) untuk soal dikotomi dan nondikotomi, (d) tes paralel. Rumus Spearman-Brown belah dua yang lain adalah sebagai berikut.

$$r_{xx'} = 2 (r_{xy}) / (1 + r_{xy})$$

r_{xy} = dapat dicari dengan rumus *product moment*

Rumus korelasi *product moment* adalah sebagai berikut.

$$r_{xy} = \frac{N \sum X_i Y_i - (\sum X_i) (\sum Y_i)}{\sqrt{\{N \sum X_i^2 - (\sum X_i)^2\} \{N \sum Y_i^2 - (\sum Y_i)^2\}}}$$

X_i = jumlah skor soal bernomor gasal

Y_i = jumlah skor soal bernomor genap

N = jumlah peserta tes

Untuk contoh di atas diperoleh besarnya $r_{xy} = 0,67$ dan $r_{xx'} = 0,78$. Berdasarkan contoh penggunaan rumus di atas, tampak bahwa setiap rumus memiliki tingkat keakuratan yang berbeda, bergantung terpenuhinya syarat batas setiap rumusnya.

Jika asumsi varians sama tidak terpenuhi, dapat menggunakan rumus Rulon sebagai berikut:

$$r_{xx'} = 1 - (SD^2 / SX^2) \dots\dots\dots SD^2 = \text{variens perbedaan skor}$$

$$SX^2 = \text{variens skor tes}$$

Jika belahan tidak sama panjang tetapi isi tes homogen, digunakan rumus Kristoff (belah tiga), sebagai berikut:

$$r_{xx'} = St^2 / SX^2 \dots\dots St^2 = \text{variens total tiap belahan}$$

$$SX^2 = \text{variens skor tes} = \{ \Sigma X^2 - (\Sigma X)^2 / n \} / (n-1)$$

$$St^2 = S_{12} S_{13} / S_{23} + S_{12} S_{23} / S_{13} + S_{13} S_{23} / S_{12} + 2 (S_{12} + S_{13} + S_{23})$$

S_{12} = kovarians belahan 1 dan 2

S_{13} = kovarians belahan 1 dan 3

S_{23} = kovarians belahan 2 dan 3

$$S_{12} = \{ \Sigma (X - M_x) (Y - M_y) \} / (n-1) \text{ atau } = \{ \Sigma XY - (\Sigma X)(\Sigma Y) / n \} / n-1$$

M_x = mean dari belahan X atau belahan 1

M_y = mean dari belahan Y atau belahan 2

2. Reliabilitas Tes Uraian

Rumus yang digunakan biasanya adalah rumus Alpha Cronbach, sebagai berikut:

$$\alpha = \{ k / (k-1) \} \{ 1 - \Sigma sj^2 / sx^2 \} \dots\dots\dots k = \text{banyak soal}$$

$$sj^2 = \text{variens butir ke } j$$

$$sx^2 = \text{variens skor tes}$$

$$sx^2 = \Sigma (X - M_x)^2 / (n-1) \text{ atau } sx^2 = \{ \Sigma X^2 - (\Sigma X)^2 / n \} / (n-1)$$

Kriteria Reliabilitas:

Koefisien Reliabilitas	Kategori
$r < 0,2$	Sangat rendah
$0,2 \leq r < 0,4$	Rendah
$0,4 \leq r < 0,6$	Sedang
$0,6 \leq r < 0,8$	Tinggi
$0,8 \leq r \leq 1,0$	Sangat tinggi

ANALISIS BUTIR SOAL

Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepatnya sesuai dengan tujuannya. Salah satu informasi yang dapat diberikan adalah menentukan peserta didik yang sudah atau belum menguasai materi yang diajarkan guru. Dalam melaksanakan analisis butir soal, para penulis soal dapat menganalisis secara kualitatif dan kuantitatif. Analisis kualitatif mencakup pertimbangan isi materi, konstruksi soal, dan kebahasaan, sedangkan analisis kuantitatif mencakup pengukuran kesulitan butir soal, daya beda soal, dan keberfungsian pengecoh untuk bentuk tes objektif (pilihan ganda).

Tujuan utama analisis butir soal dalam sebuah tes yang dibuat guru adalah untuk mengidentifikasi kekurangan-kekurangan dalam tes atau dalam pembelajaran. Berdasarkan tujuan ini, maka kegiatan analisis butir soal memiliki banyak manfaat, di antaranya adalah:

- (1) dapat membantu para pengguna tes dalam evaluasi atas tes yang digunakan,
- (2) sangat relevan bagi penyusunan tes informal dan lokal seperti tes yang disiapkan guru untuk siswa di kelas,
- (3) mendukung penulisan butir soal yang efektif,
- (4) secara materi dapat memperbaiki tes di kelas,
- (5) meningkatkan validitas soal dan reliabilitas,
- (6) menentukan fungsi butir soal sesuai dengan yang diharapkan,
- (7) memberi masukan kepada siswa tentang kemampuannya,

- (8) memberi masukan kepada guru tentang kesulitan siswa,
- (9) memberi masukan pada aspek tertentu untuk pengembangan kurikulum,
- (10) merevisi materi yang dinilai atau diukur,
- (11) meningkatkan keterampilan penulisan soal.

Analisis terhadap butir soal dapat dilakukan sebelum dan sesudah tes diujikan. Analisis secara kualitatif dilakukan sebelum tes diujikan, sedangkan analisis secara kuantitatif dilakukan setelah tes diujikan.

A. Analisis Butir Soal Secara Kualitatif

Analisis butir soal secara kualitatif dilakukan setelah pembuatan tes selesai dilakukan, dan sebelum tes diujicobakan/diujikan. Analisis dapat dilakukan oleh pembuat soal atau oleh orang lain misalnya teman sejawat yang sebidang.

Aspek yang diperhatikan dalam penelaahan secara kualitatif ini adalah setiap soal ditelaah dari segi materi, konstruksi, bahasa/budaya, dan kunci jawaban/pedoman penskorannya. Dalam melakukan penelaahan setiap butir soal, penelaah perlu mempersiapkan bahan-bahan penunjang seperti: (1) kisi-kisi tes, (2) kurikulum yang digunakan, (3) buku sumber, dan (4) kamus bahasa Indonesia.

Instrumen untuk memvalidasi dapat berupa lembar *checklist* (Ya/Tidak) atau *rating scale* (sangat baik, baik, kurang, sangat kurang). Contoh format penelaahan soal secara kualitatif dari Balitbang Sisjian Depdiknas dapat dilihat pada Tabel 4.1. untuk soal bentuk pilihan ganda dan Tabel 4.2. untuk bentuk uraian.

B. Analisis Butir Soal Secara Kuantitatif

Penelaahan soal secara kuantitatif maksudnya adalah penelaahan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empirik ini diperoleh dari soal yang

telah diujicobakan/diuji. Ada dua pendekatan dalam analisis secara kuantitatif, yaitu pendekatan secara klasik dan modern.

Analisis butir soal secara klasik adalah proses penelaahan butir soal melalui informasi dari jawaban peserta didik guna meningkatkan mutu butir soal yang bersangkutan dengan menggunakan teori tes klasik. Kelebihan analisis butir soal secara klasik adalah murah, dapat dilaksanakan sehari-hari dengan cepat, dapat menggunakan komputer, sederhana, sudah dikenal, dan data dari beberapa peserta didik atau sampel kecil.

Analisis dilakukan untuk menentukan karakteristik dari setiap butir soal yang menyusun suatu tes. Karakteristik soal meliputi tingkat kesukaran dan daya pembeda untuk soal uraian, sedangkan soal pilihan ganda ditambah dengan keberfungsian pengecoh.

Tabel 5.1. Format Penelaahan Butir Soal Pilihan Ganda

No	Kriteria Penilaian	Soal Nomor									
		1		2		3		4		...	
		Y	T	Y	T	Y	T	Y	T	Y	T
	A. MATERI										
1	Soal sesuai dengan indikator										
2	Pengecoh berfungsi										
3	Mempunyai satu jawaban yang benar										
4	Waktu yang tersedia mencukupi untuk menyelesaikan soal										
	B. KONSTRUKSI										
5	Pokok soal dirumuskan secara jelas										
6	Rumusan soal dan jawaban merupakan pernyataan yang diperlukan saja										
7	Pokok soal tidak memberi petunjuk ke arah jawaban benar										
8	Pokok soal tidak mengandung										

	pernyataan negatif ganda																		
9	Pilihan jawaban homogen dan logis ditinjau dari segi materi																		
10	Panjang rumusan jawaban relatif sama																		
11	Pilihan jawaban tidak mengandung pernyataan “semua jawaban benar” atau “semua jawaban salah”																		
12	Pilihan jawaban berbentuk angka disusun berdasarkan urutan besar kecilnya nilai atau kronologisnya																		
13	Gambar, grafik, tabel jelas dan berfungsi																		
14	Butir soal tidak bergantung pada jawaban soal sebelumnya																		
	C. BAHASA																		
15	Soal menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia																		
16	Bahasa yang digunakan komunikatif																		
17	Tidak menggunakan bahasa yang berlaku di daerah setempat																		
18	Pilihan jawaban tidak mengulang kata/frasa yang bukan merupakan satu kesatuan pengertian																		
Catatan:																			

Keterangan: Y = ya, T = tidak

Tabel 5.2. Format Penelaahan Butir Soal Uraian

No	Kriteria Penilaian	Soal Nomor									
		1		2		3		4		...	
		Y	T	Y	T	Y	T	Y	T	Y	T
	A. MATERI										
1	Soal sesuai dengan indikator										
2	Batasan pertanyaan dan jawaban yang diharapkan jelas										
3	Isi materi sesuai dengan tujuan pengukuran										
4	Isi materi yang ditanyakan sudah sesuai dengan jenjang sekolah dan tingkat kelas										
5	Waktu yang tersedia mencukupi untuk menyelesaikan soal										
	B. KONSTRUKSI										
6	Rumusan kalimat soal atau pernyataan menggunakan kata tanya atau perintah yang menuntut jawaban terurai										
7	Ada petunjuk yang jelas tentang cara mengerjakan soal										
8	Ada pedoman penskoran										
9	Gambar, grafik, tabel jelas dan berfungsi										
	C. BAHASA										
10	Rumusan kalimat soal komunikatif										

11	Soal menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia													
12	Rumusan soal tidak menggunakan kata-kata/kalimat yang menimbulkan penafsiran ganda													
13	Tidak menggunakan bahasa yang berlaku setempat													
Catatan:														

Keterangan: Y = ya, T = tidak

1. Tingkat Kesukaran (TK)

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 sampai dengan 1,00 (Aiken, 1994: 66). Makin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti makin mudah suatu soal. Rumus yang dipergunakan untuk menentukan TK soal objektif adalah sebagai berikut.

$$\text{Tingkat Kesukaran (TK)} = \frac{\text{Jumlah siswa yang menjawab benar butir soal}}{\text{Jumlah siswa yang mengikuti tes}}$$

Rumus yang dipergunakan untuk menentukan TK soal uraian adalah sebagai berikut.

$$\text{Tingkat Kesukaran(TK)} = \frac{\text{Mean}}{\text{Skormaksimal}}$$

$$\text{Mean} = \frac{\text{Jumah skorpadasoaltertentu}}{\text{Jumlah siswa yang mengikuti tes}}$$

Hasil perhitungan dengan menggunakan rumus di atas menggambarkan tingkat kesukaran suatu soal. Klasifikasi tingkat kesukaran soal adalah sebagai berikut.

- 0,00 ≤ TK ≤ 0,30 soal sukar
- 0,30 < TK ≤ 0,70 soal sedang
- 0,70 < TK ≤ 1,00 soal mudah

Tingkat kesukaran butir soal memiliki 2 kegunaan, yaitu bagi guru dan bagi pengujian dan pengajaran. Kegunaannya bagi guru adalah:

- (1) sebagai penekanan konsep terhadap pembelajaran ulang dan memberi masukan kepada siswa tentang hasil belajar mereka,
- (2) memperoleh informasi tentang penekanan kurikulum atau mencurigai terhadap butir soal yang *bias*.

Kegunaannya bagi pengujian dan pengajaran adalah:

- (1) penekanan konsep yang diperlukan untuk diajarkan ulang,
- (2) mengetahui kelebihan dan kelemahan kurikulum sekolah,
- (3) memberi masukan kepada siswa,
- (4) mengetahui kemungkinan adanya butir soal yang *bias*,
- (5) merakit tes yang memiliki ketepatan data soal.

Analisis secara klasik ini memang memiliki keterbatasan, yaitu tingkat kesukaran sangat sulit untuk mengestimasi secara

tepat karena estimasi tingkat kesukaran dibiarkan oleh sampel atau peserta tes. Jika peserta tes berkemampuan tinggi, maka soal akan sangat mudah ($TK \geq 0,90$). Jika peserta tes berkemampuan rendah, maka soal akan sangat sulit ($TK \leq 0,30$). Untuk mengatasi keterbatasan tersebut dikembangkan analisis secara modern yaitu dengan IRT (*Item Response Theory*). IRT dapat mengestimasi tingkat kesukaran soal tanpa menentukan siapa peserta tesnya (*invariance*). Dalam IRT, komposisi sampel dapat mengestimasi parameter dan tingkat kesukaran soal tanpa *bias*.

2. Daya Pembeda (DP)

Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Manfaat daya pembeda butir soal adalah:

- a. Untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak.
- b. Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru.

Apabila suatu butir soal tidak dapat membedakan kedua kemampuan tersebut, maka perlu dianalisis kemungkinan-kemungkinan sebagai berikut.

- a. Kunci jawaban butir soal itu tidak tepat.
- b. Butir soal memiliki 2 atau lebih kunci jawaban yang benar
- c. Kompetensi yang diukur tidak jelas
- d. Pengecoh tidak berfungsi
- e. Sebagian besar siswa menganggap ada yang salah informasi dalam butir soalnya.

Indeks daya pembeda setiap butir soal biasanya juga dinyatakan dalam bentuk proporsi. Makin tinggi indeks daya pembeda soal berarti makin mampu soal yang bersangkutan membedakan siswa yang telah memahami materi dengan yang belum memahami. Indeks daya pembeda berkisar antara -1,00 sampai dengan +1,00.

Untuk mengetahui daya pembeda soal bentuk pilihan ganda dapat menggunakan rumus berikut ini.

$$DP = \frac{BA - BB}{\frac{1}{2}N} \quad \text{atau} \quad DP = \frac{2(BA - BB)}{N}$$

DP = daya pembeda soal,

BA = jumlah jawaban benar pada kelompok atas,

BB = jumlah jawaban benar pada kelompok bawah,

N = jumlah siswa yang mengerjakan tes

Rumus di atas berlaku untuk N kurang dari 100 peserta tes. Apabila jumlah peserta tes lebih dari 100 orang, maka jumlah kelompok atas diambil sebanyak 27% dan kelompok bawah 27%. Rumus yang digunakan adalah:

$$DP = \frac{BA - BB}{27\%N}$$

Di samping rumus di atas, untuk mengetahui daya pembeda soal bentuk pilihan ganda dapat dipergunakan rumus korelasi *point biserial* (r_{pbis}) dan korelasi *biserial* (r_{bis}) (Millman and Greene, 1993: 359-360) dan (Glass and Stanley, 1970: 169-170) seperti berikut.

$$r_{pbis} = \frac{\bar{X}_b - \bar{X}_s}{SD} \sqrt{pq} \quad \text{dan} \quad r_{bis} = \frac{\bar{Y}_b - \bar{Y}_s}{SD} \cdot \frac{nb.ns}{u\sqrt{n^2 - n}}$$

\bar{X}_b , \bar{Y}_b adalah rata-rata skor warga belajar/siswa yang menjawab benar

\bar{X}_s, \bar{Y}_s adalah rata-rata skor warga belajar siswa yang menjawab salah

SD adalah simpangan baku skor total

nb dan ns adalah jumlah siswa yang menjawab benar dan jumlah siswa yang menjawab salah, serta $nb + ns = n$

p adalah proporsi jawaban benar terhadap semua jawaban siswa

q adalah $1 - p$

u adalah ordinat kurva normal.

Rumus untuk mengetahui daya pembeda soal bentuk uraian adalah sebagai berikut.

$$DP = \frac{\text{Mean kelompok atas} - \text{Mean kelompok bawah}}{\text{Skor maksimum soal}}$$

Hasil perhitungan dengan menggunakan rumus di atas dapat menggambarkan tingkat kemampuan soal dalam membedakan antar peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum/tidak memahami materi yang diujikan. Klasifikasinya daya pembeda menurut Crocker dan Algina, 1986: 315 adalah seperti berikut.

$0,40 \leq D \leq 1,00$	= soal diterima
$0,30 \leq D < 0,40$	= soal diterima, tetapi perlu diperbaiki
$0,20 \leq D < 0,30$	= soal diperbaiki
$0,00 \leq D < 0,20$	= soal tidak dipakai/dibuang

3. Penyebaran (Distribusi) Jawaban

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan soal. Hal ini dimaksudkan untuk mengetahui berfungsi tidaknya jawaban yang tersedia. Beberapa ahli menyatakan bahwa suatu pilihan jawaban (pengecoh) dapat dikatakan berfungsi apabila:

a. paling tidak dipilih oleh 2,5 % peserta tes,

b. lebih banyak dipilih oleh kelompok siswa yang belum paham materi.

C. Analisis Butir Soal dengan Program Komputer

Analisis butir soal secara kuantitatif dapat menggunakan bantuan program komputer. Analisis data dengan menggunakan program komputer lebih praktis dan tingkat keakuratan hitungan lebih tinggi jika dibandingkan dengan diolah secara manual atau menggunakan kalkulator. Program komputer yang digunakan untuk menganalisis data modelnya bermacam-macam bergantung tujuan dan maksud analisis yang diperlukan. Program yang sudah dikenal secara umum adalah EXCEL, SPSS (*Statistical Program for Social Science*), atau program khusus seperti ITEMAN, TAP (analisis secara klasik), RASCAL, ASCAL, BILOG (analisis secara item respons teori atau IRT), FACETS (analisis model Rasch untuk data kualitatif). Untuk memahami program-program komputer di atas perlu membaca manual programnya, kemudian praktikkan dengan menggunakan program komputer sebagai latihannya. Berikut ini akan disajikan contoh program analisis data dengan menggunakan komputer, seperti program ITEMAN, EXCEL.

1. ITEMAN

ITEMAN merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik. Program ini termasuk satu paket program dalam MicroCAT yang dikembangkan oleh *Assessment Systems Corporation* mulai tahun 1982 dan mengalami revisi pada tahun 1984, 1986, 1988, dan 1993; mulai dari versi 2.00 sampai dengan versi 3.50. Alamatnya adalah *Assessment Systems Corporation*, 2233 University Avenue, Suite 400, St Paul, Minesota 55114, United States of America.

Program ini dapat digunakan untuk: (a) menganalisis data file (format ASCII) jawaban butir soal yang dihasilkan melalui *manual entry* data atau dari mesin *scanner*; (b) menskor dan

menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 siswa dan 250 butir soal; (c) menganalisis sebuah tes yang terdiri atas 10 skala (subtes), dan (d) memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option), reliabilitas (*Alpha*), *standar error of measurement*, *mean*, *varians*, standar deviasi, *skewness*, kurtosis untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor.

Sebelum menggunakan program ITEMAN, bacalah manual/buku petunjuk pengoperasinya secara saksama. Sebagai contoh, tahap awal adalah membuat "file data" yang berisi 5 komponen utama, yaitu:

- a. Baris pertama adalah baris pengontrol yang mendeskripsikan data.
- b. Baris kedua adalah daftar kunci jawaban setiap butir soal.
- c. Baris ketiga adalah daftar jumlah option untuk setiap butir soal.
- d. Baris keempat adalah daftar butir soal yang hendak dianalisis (jika butir yang akan dianalisis diberi tanda Y (yes), jika tidak diikuti dalam analisis diberi tanda N (no)).
- e. Baris kelima dan seterusnya adalah data siswa dan pilihan jawaban siswa.

Setiap pilihan jawaban siswa (untuk soal bentuk pilihan ganda) diketik dengan menggunakan huruf, misal ABCD atau angka 1234 untuk 4 pilihan jawaban atau ABCDE atau 12345 untuk 5 pilihan jawaban. Cara menggunakan program ini pertama data diketik di Windows. Cara termudah adalah menggunakan program Windows yaitu dengan mengetik data di tempat Notepad. Caranya adalah klik Start-Programs-Accessories-Notepad.

Contoh pengetikan data untuk soal bentuk pilihan ganda :

25 O N 10 [Jumlah soal, kode omit, kode tidak dijawab, jumlah karakter*]
43142442113424141324 ... [Kunci jawaban dapat ditulis dengan huruf]
44444444444444444444 ... [Jumlah pilihan]
YYYYYYYYYYYYYY ... [Soal yang dianalisis, bila tidak dianalisis ditulis N]
Aditya 1231232441134241433242 ... [dapat ditulis dengan huruf]
Pradipta 4231422421134241413242 ...
Grawira 1431422424334341413244 ...
Lulu 2431422424134341414112 ...
dst.

***Catatan:** jumlah karakter adalah ruang untuk menuliskan nama responden

Contoh pengetikan data untuk skala Likert:

30 x Y 20 [Jumlah soal, kode omit, kode tidak dijawab, jumlah karakter]
+++++-----+++++-----+++++-- [Positif/negatif pernyataan]
77777777777777777777777777777777 [Jumlah pilihan]
Nurul 211214123242343423111231243767
Imam 312214214242443423224562332565
Ali 2242123313324431243254624371YY
Kiki 22421112X432443323226556664122
Chanan 32421424234244344322653546X343

Langkah kedua data yang telah diketik disimpan, misal disimpan pada file: **Tes1.txt**. Selanjutnya untuk menggunakan program IteMan yaitu dengan mengklik icon **IteMan**. Kemudian isilah pertanyaan-pertanyaan yang muncul di layar komputer seperti berikut.

```

Enter the name of the input file: tes1.txt <enter>
Enter the name of the output file: haltes1.txt <enter>

Do you want the scores written to a file? (Y/N) Y
<enter>
Enter the name of the score file: scrtes1.txt <enter>
**ITEMAN ANALYSIS IS COMPLETE**

```

Langkah ketiga adalah membaca hasil, yaitu dengan mengklik *icon* hsltes1. Hasilnya adalah seperti pada contoh berikut.

MicroCAT (tm) Testing System
 Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corpora

Item and Test Analysis Program -- ITEMAN (tm) Version 3.00

Item analysis for data from file tes1.txt

Seq. No.	Scale -Item	Item Statistics			Alternative Statistic			
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Po Bi
1	0-1	0.850	-0.018	-0.012	A	0.850	-0.018	-0
					B	0.000	-9.000	-9
					C	0.100	0.047	0
					D	0.050	-0.040	-0
					E	0.000	-9.000	-9
					Other	0.000	-9.000	-9
					CHECK THE KEY A was specified, C works better			
2	0-2	0.450	0.534	0.425	A	0.050	-1.000	-0
					B	0.450	0.534	0
					C	0.300	-0.262	-0
					D	0.150	0.231	0
					E	0.050	0.121	0
					Other	0.000	-9.000	-9
3	0-3	0.600	0.515	0.406	A	0.050	-1.000	-0
					B	0.100	-0.142	-0
					C	0.250	0.039	0
					D	0.600	0.515	0
					E	0.000	-9.000	-9
					Other	0.000	-9.000	-9

4	0-4	0.400	0.172	0.135	A	0.050	-1.000	-0.856	
					B	0.200	-0.059	-0.041	*
					C	0.400	0.172	0.135	*
					D	0.200	0.474	0.332	?
					E	0.150	0.018	0.012	
					Other	0.000	-9.000	-9.000	
5	0-5	0.700	0.215	0.163	A	0.050	0.281	0.133	
					B	0.050	-1.000	-0.856	
					C	0.100	0.142	0.083	
					D	0.100	0.331	0.194	?
					E	0.700	0.215	0.163	*
					Other	0.000	-9.000	-9.000	
6	0-6	0.850	-0.089	-0.058	A	0.000	-9.000	-9.000	
					B	0.850	-0.089	-0.058	*
					C	0.050	-0.040	-0.019	
					D	0.100	0.142	0.083	?
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

Keterangan:

Prop. Correct = tingkat kesukaran butir;

Biser dan Point Biser = korelasi Biserial dan Korelasi Point Biserial,

Alt = alternative/pilihan jawaban,

Prop. Endorsing = proporsi jawaban pada setiap option

Selain menghitung tingkat kesukaran, daya pembeda butir soal, dan keberfungsian pengecoh, ITEMAN juga menghasilkan skala statistik yang lain. Contoh hasil perhitungan skala statistik dapat dilihat pada tampilan berikut.

There were 20 examinees in the data file.

Scale Statistics

```
-----  
Scale:          0  
-----  
N of Items      25  
N of Examinees  20  
Mean            16.250  
Variance        9.087  
Std. Dev.       3.015  
Skew            -2.463  
Kurtosis        6.976  
Minimum         5.000  
Maximum         20.000  
Median          17.000  
Alpha           0.437  
SEM             2.261  
Mean P          0.650  
Mean Item-Tot. 0.266  
Mean Biserial   0.352
```

Hasil skala statistik menunjukkan bahwa harga Alpha 0,437. Hal ini menunjukkan bahwa koefisien reliabilitas tes yang dibuat adalah 0,437 berada pada kategori sedang. Hasil korelasi point-biserial (r_{pbi}) dan korelasi biserial (r_{bis}) berasal dari perhitungan rumus yang telah dipaparkan sebelumnya. Hasil perhitungan menunjukkan rata-rata daya beda soal pada tes tersebut adalah 0,266 (r_{pbi}) atau 0,352 (r_{bis}).

Korelasi point-biserial (r_{pbi}) tidak sama dengan 0, korelasi biserial (r_{bis}) paling sedikit 25% lebih besar daripada (r_{pbi}) untuk perhitungan pada data yang sama. Korelasi point-biserial (r_{pbi}) merupakan korelasi *product moment* antara skor dikotomus dan pengukuran kriteria; sedangkan korelasi biserial (r_{bis})

merupakan korelasi *product moment* antara variabel latent distribusi normal berdasarkan dikotomi benar-salah dan pengukuran kriterion.

Menurut Millman dan Greene (1989) kedua korelasi ini memiliki kelebihan masing-masing. Kelebihan korelasi point biserial adalah: (1) memberikan refleksi kontribusi soal secara sesungguhnya terhadap fungsi tes. Maksudnya ini mengukur bagaimana baiknya soal berkorelasi dengan kriterion; (2) sederhana dan langsung berhubungan dengan statistik tes; (3) tidak pernah mempunyai value 1,00 karena hanya variabel-variabel dengan distribusi bentuk yang sama yang dapat berkorelasi secara sempurna, sedangkan variabel kontinyu (kriterion) dan skor dikotomus tidak mempunyai bentuk yang sama. Kelebihan korelasi biserial adalah: (1) cenderung lebih stabil dari sampel ke sampel, (2) penilaian lebih akurat tentang bagaimana soal dapat diharapkan untuk membedakan pada beberapa perbedaan point di skala abilitas/kemampuan, (3) value r_{bis} langsung berhubungan dengan indikator diskriminasi kurva karakteristik butir (*Item Characteristic Curve* atau ICC). Kebanyakan para ahli pendidikan, khususnya di Indonesia, banyak yang menggunakan korelasi point biserial daripada korelasi biserial.

2. EXCEL

Excel merupakan sebuah program pengolah data yang biasa dinamakan "*spreadsheet*". Karena program ini dapat digunakan untuk mengolah data yang berupa angka ataupun lainnya. Ada dua cara mengolah data dengan Excel, yaitu (1) melalui program bantu khusus perhitungan statistik dan (2) melalui fungsi statistik yang terdapat di dalam Excel. Oleh karena itu tidak semua program Statistik ada di program Excel, seperti halnya uji reliabilitas baik bentuk pilihan ganda, uraian maupun

reliabilitas nontes, dalam hal ini harus disalin secara manual. Karena di dalam program Excel tidak tersedia program tersebut.

D. Analisis Butir Soal Secara Modern (Item Response Theory/IRT)

Analisis butir soal secara modern yaitu penelaahan butir soal dengan menggunakan *Item Response Theory* (IRT). Teori ini merupakan suatu teori yang menggunakan fungsi matematika untuk menghubungkan antara peluang menjawab benar suatu soal dengan kemampuan siswa. Nama lain IRT adalah *latent trait theory* (LTT), atau *characteristics curve theory* (ICC).

Asal mula IRT adalah kombinasi suatu versi hukum *phi-gamma* dengan suatu analisis faktor butir soal (*item factor analysis*) kemudian bernama Teori *Trait Latent (Latent Trait Theory)*, sekarang secara umum dikenal menjadi *Item Response Theory* (McDonald, 1999: 8). Dalam subbab ini akan disajikan kelebihan analisis secara IRT dan kalibrasi butir soal dan pengukuran kemampuan orang.

1. Kelebihan Analisis IRT

Untuk mengetahui kelebihan analisis IRT, guru perlu mengetahui keterbatasan analisis butir soal secara klasik. Keterbatasan model pengukuran secara klasik bila dibandingkan dengan teori jawaban butir soal adalah seperti berikut Hambleton dan Swaminathan (1985: 1-3):

- a. tingkat kesukaran dan daya pembeda tergantung pada sampel;
- b. penggunaan metode dan teknik untuk desain dan analisis tes dengan memperbandingkan kemampuan siswa pada pernbagian kelompok atas, tengah, dan bawah. Meningkatnya validitas skor tes diperoleh dari tingkat kesukaran tes dihubungkan dengan tingkat kemampuan setiap siswa;

- c. konsep reliabilitas tes didefinisikan dari istilah tes paralel;
- d. tidak ada dasar teori untuk menentukan bagaimana siswa memperoleh tes yang sesuai dengan kemampuan siswa;
- e. Standar error of measurement (SEM) hanya berlaku untuk seluruh peserta didik.

Kelemahan teori tes klasik di atas diperkuat (Hambleton, Swaminathan, dan Rogers, 1991: 2-5) yaitu:

- a. Tingkat kemampuan dalam teori klasik adalah "*true score*". Jika tes sulit artinya tingkat kemampuan peserta didik rendah. Jika tes mudah artinya tingkat kemampuan peserta didik tinggi.
- b. Tingkat kesukaran soal didefinisikan sebagai proporsi peserta didik dalam grup yang menjawab benar soal. Mudah/sulitnya butir soal bergantung pada kemampuan peserta didik yang dites dan kemampuan tes yang diberikan.
- c. Daya pembeda, reliabilitas, dan validitas soal/tes didefinisikan berdasarkan grup peserta didik.

Selanjutnya Hambleton dan Swaminathan (1985: 13) menyatakan bahwa tujuan utama IRT adalah memberikan kesamaan antara statistik soal dan estimasi kemampuan. Ada tiga keuntungan IRT adalah:

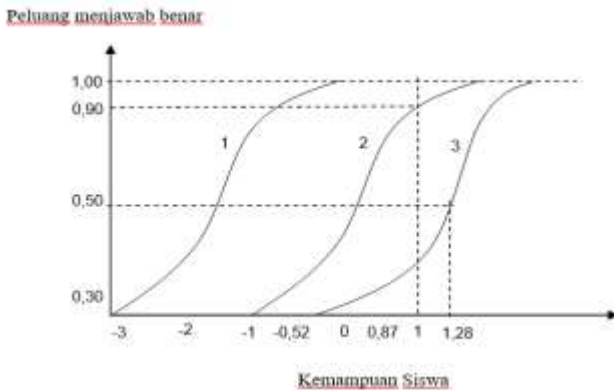
- a. asumsi banyak soal yang diukur pada trait yang sama, perkiraan tingkat kemampuan peserta didik adalah independen;
- b. asumsi pada populasi tingkat kesukaran, daya pembeda merupakan independen sampel yang menggambarkan untuk tujuan kalibrasi soal;
- c. statistik yang digunakan untuk menghitung tingkat kemampuan siswa diperkirakan dapat terlaksana.

Jadi IRT merupakan hubungan antara probabilitas jawaban suatu butir soal yang benar dan kemampuan siswa atau tingkatan/level prestasi siswa. Namun kelemahan bekerja dengan model IRT

adalah bekerja melalui suatu proses yang sulit karena kelebihan IRT adalah: (a) tanpa varian pada parameter butir soal, (b) tanpa varian pada parameter abilitas, (c) adanya ketepatan pada pengukuran lokal (Bejar, 1983: 3-4).

Ada empat macam model IRT (Hambleton, 1993: 154-157; Hambleton dan Swaminathan, 1985: 34-50), yaitu: (a) Model satu parameter (Model Rasch), yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran soal. (b) Model dua parameter, yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal. (c) Model tiga parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, dan menebak (*guessing*). (d) Model empat parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya beda soal, menebak, dan penyebab lain.

Hambleton dan Swaminathan (1985: 48) menjelaskan bahwa siswa yang memiliki kemampuan tinggi tidak selalu menjawab soal dengan betul. Kadang-kadang mereka sembrono (mengerjakan dengan serampangan), memiliki informasi yang berlebihan, sehingga mereka menjawab salah pada suatu soal. Untuk mengatasi masalah ini diperlukan model 4 parameter.



Gambar 5.1. Kurva Ciri Soal Model Satu Parameter

Dari keempat model itu tidak sama penekanannya dan sudah barang tentu tiap-tiap model itu memiliki kelebihan dan kekurangan. Kelebihan dan kekurangan itu dapat diklasifikasikan sesuai dengan jumlah parameter yang ditentukan pada masing-masing model dan tujuan menggunakan model yang bersangkutan.

Contoh kurva ciri soal model satu parameter atau *Rasch* terlihat seperti pada Gambar 5.1.

2. Kalibrasi Butir Soal dan Pengukuran Kemampuan Orang

Kalibrasi butir soal dan pengukuran kemampuan orang merupakan proses estimasi parameter pada model respons butir. Model persamaan dasar Rasch adalah model probabilistik yang mencakup hasil dari suatu interaksi butir soal-orang. Proses mengestimasi kemampuan orang dinamakan pengukuran, sedangkan proses mengestimasi parameter tingkat kesukaran butir soal dinamakan kalibrasi. Jadi kalibrasi soal merupakan proses penyamaan skala soal yang didasarkan pada tingkat kesukaran butir soal dan tingkat kemampuan siswa. Adapun ciri suatu skala adalah mempunyai titik awal, biasanya 0, dan mempunyai satuan ukuran atau unit pengukuran.

Untuk menghitung peluang menjawab benar setiap butir pada model Rasch atau model satu parameter (1-PL) digunakan rumus berikut ini.

$$P_i(\theta) = \frac{1}{1 + e^{-D(\theta - b_i)}}$$

Persamaan yang menggunakan model dua parameter (2-PL) adalah sebagai berikut:

$$P_i(\theta) = \frac{1}{1 + e^{-D a_i(\theta - b_i)}}$$

(Hambleton, *et al.*, 1991:16; Alen & Yen, 1979: 254)

$P_i(\theta)$ = probabilitas menjawab benar soal ke i oleh peseta dengan kemampuan θ .

D = 1,7

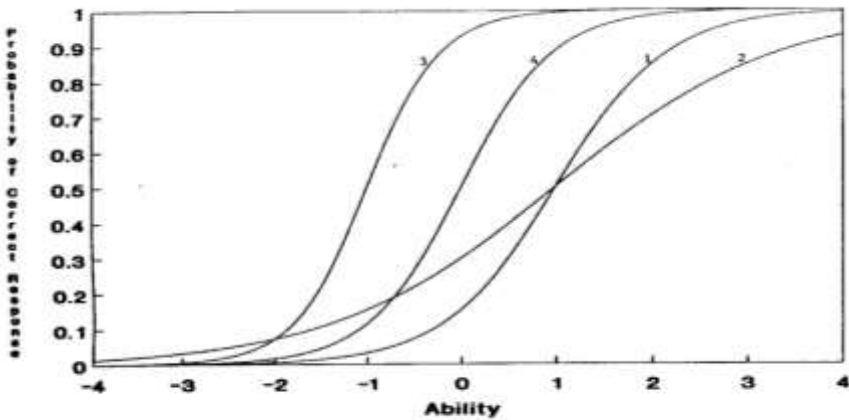
e = 2,718

a_i = parameter daya beda soal ke i (*slope*)

b_i = parameter tingkat kesulitan soal ke i (*threshold*)

I = 1, 2, 3, ..., n ; n = jumlah soal dalam tes

Bentuk kurva untuk model 2-PL dapat dilihat pada Gambar 5.2.



Gambar 5.2. Kurva Karakteristik Soal Model 2-PL
(Hambleton, *at al.* 1991: 16)

Dalam pelaksanaannya, analisis secara IRT tidak serumit seperti penjelasan di atas. Pelaksanaannya sangat mudah dipahami oleh para guru karena dalam analisis digunakan program komputer, seperti program RASCAL, PASCAL, BIGSTEPS, BILOG, atau QUEST.

PENGEMBANGAN TES

A. Langkah Pengembangan Tes

Pengembangan alat ukur psikologis, termasuk tes dan non-tes, menyangkut serangkaian kegiatan berurutan, langkah demi langkah, untuk menghasilkan alat ukur yang bermutu dan memadai. Sumadi Suryabrata (1998: 67) menyebutkan bahwa langkah-langkah pembuatan tes pada hakikatnya adalah rancangan dasar (*grand design*) yang akan menentukan tinggi rendahnya kualitas alat ukur. Langkah-langkah tersebut hendaknya menyeluruh dan secara spesifik menunjukkan keseluruhan kualitas dan ciri-ciri yang harus dimiliki oleh alat ukur. Langkah penyusunan diawali dari menentukan dasar konseptual teoretis hingga jadwal penerbitan alat ukur.

Proses pengembangan tes menurut Depdiknas (2002: 3) meliputi tahapan sebagai berikut: (1) penentuan tujuan tes, (2) penyusunan kisi-kisi, (3) penulisan soal, (4) penelaahan soal (*review* dan revisi soal), (5) ujicoba soal, (6) analisis, (7) perakitan soal menjadi perangkat tes, dan (8) implementasi tes. Djemari Mardapi (2004: 88–95) merumuskan langkah-langkah pengembangan tes meliputi: (1) menyusun spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan uji coba tes, (5) menganalisis soal, (6) memperbaiki tes, (7) merakit tes, (8) melaksanakan tes, dan (9) menafsirkan hasil tes. Pada prinsipnya langkah yang dirumuskan oleh Depdiknas dan Djemari Mardapi adalah sama. Perbedaannya terletak pada langkah penentuan tujuan dan pembuatan kisi-kisi. Djemari Mardapi membagi

langkah menyusun spesifikasi tes menjadi beberapa sublangkah yaitu: (a) tujuan membuat tes, (b) menyusun kisi-kisi tes, (c) menentukan bentuk tes, dan (d) menentukan panjang tes. Perbedaan yang lain terletak pada perevisian tes. Depdiknas menetapkan revisi setelah dilakukan penelaahan soal, sedangkan Djemari Mardapi merencanakan revisi dilakukan setelah dilakukan uji coba.

Millman & Greene (1989: 335-366) menentukan langkah pengembangan tes meliputi 6 langkah, yaitu: (1) menentukan tujuan tes, (2) menentukan spesifikasi tes, (3) menulis soal tes, (4) menelaah soal tes, (5) memilih soal tes, dan (6) perakitan tes. Langkah-langkah tersebut senada dengan langkah pengembangan tes yang dikemukakan oleh Roid & Haladyna (1982) yaitu: (1) penetapan tujuan, (2) spesifikasi domain, (3) pengembangan item, (4) revisi item, dan (5) pengembangan tes. Berdasarkan teori-teori di atas, disimpulkan bahwa pengembangan tes meliputi (1) penetapan tujuan tes, (2) penyusunan spesifikasi dan kisi-kisi tes, (3) penulisan soal, (4) penelaahan dan uji coba tes, (5) penganalisisan hasil uji coba, (6) perbaikan tes, (7) perakitan tes, (8) interpretasi hasil tes. Berikut penjelasan setiap langkah pembuatan tes.

1. Penentuan Tujuan Tes

Pembuatan tes yang baik perlu memperhatikan tujuan dari pembuatan tes tersebut. Tujuan tes dapat dideskripsikan dengan mempertimbangkan kurikulum, domain kognitif, dan tindak lanjut setelah diberikan tes. Tujuan tes berdasarkan domain kurikulum meliputi tujuan: (a) sebelum pengajaran, (b) selama pengajaran, dan (c) setelah pengajaran. Tujuan tes berdasarkan domain kognitif adalah untuk menentukan kriteria kesuksesan performan siswa, biasanya dikaitkan dengan pemberian lisensi atau sertifikasi dan pelaporan kepada orang tua. Kaitan tujuan tes dengan tindak

lanjut adalah perlunya pemberian bimbingan dan konseling atau untuk kepentingan penelitian.

2. Penyusunan Spesifikasi Tes

Spesifikasi tes dari 3 hal, yaitu: (a) faktor kontekstual luar, (b) atribut internal tes, dan (c) spesifikasi tes pada masa yang akan datang.

a. Faktor Kontekstual Luar

Terkait dengan tujuan tes, perlu ditetapkan *setting* tes atau kondisi yang diwajibkan dalam pengadministrasian tes. Secara umum kondisi ini menggambarkan batas-batas kontekstual atau hal-hal yang membatasi kelayakan suatu tes. Hal penting dari faktor kontekstual luar tes yang dapat menunjukkan spesifikasi dari suatu tes antara lain adalah: (1) karakteristik populasi peserta tes atau kepada siapa tes ditujukan, (2) berapa lama waktu yang diperlukan untuk menyelesaikan tes, (3) pengadministrasian tes (diberikan secara individual atau kelompok), dan (4) standardisasi tes.

1) Karakteristik Peserta Tes

Karakteristik terpenting dari populasi peserta tes adalah kemampuan kognitif dan tingkat pengetahuan. Konseptualisasi dan operasionalisasi karakteristik peserta tes sangat berpengaruh pada proses spesifikasi atribut tes. Hal ini berkaitan juga dengan tujuan tes. Dalam hal *setting* tes, spesifikasi dititikberatkan pada perhitungan karakteristik kognitif dan nonkognitif seperti perkembangan kognitif, perhatian, hambatan fisik, ketidakcakapan menulis, kesulitan berbicara, hambatan budaya, dan kecakapan membaca.

Pertimbangan hati-hati terhadap karakteristik eksternal peserta tes secara tidak langsung memodifikasi spesifikasi atribut tes dalam pengembangan butir soal. Sebagai contoh, suatu tes untuk peserta tes yang memiliki hambatan dalam hal perhatian

(misalnya hiperaktif) sebaiknya panjang tesnya direduksi atau memberikannya secara bertahap, bagian demi bagian.

2) Batasan Waktu Tes

Batas waktu yang diperlukan untuk menyelesaikan suatu tes dapat ditentukan berdasarkan karakteristik peserta tes. Ketika kecepatan merespons bukan merupakan bagian dari domain yang diuji, pembatasan waktu sering berkait langsung dengan reliabilitas dan validitas tes.

Permasalahan penentuan jumlah soal dan waktu yang diperlukan untuk mengerjakan tes merupakan hal cukup krusial. Thorndike (1982: 15) menyatakan bahwa penekanan keringkasan tes adalah penting, tetapi penekanan untuk ketepatan dalam penilaian tes juga penting. Pembuat tes berada dalam pertentangan antara memproduksi tes yang dapat diterima oleh pengguna dengan menyediakan informasi yang cukup akurat untuk membuat keputusan.

Ketepatan jumlah soal dan waktu yang digunakan untuk menyelesaikan tes berkaitan langsung dengan tujuan tes. Relativitas jumlah soal merupakan spesifikasi yang diharapkan tes untuk memperoleh hasil yang reliabel dari setiap dimensi atau domain. Perbedaan tujuan tes memberi kesan perbedaan respons pada batas waktu eksternal seperti menggugurkan satu dimensi atau mengubah format soal. Jadi, batasan waktu merupakan masalah umum dalam pengembangan tes.

3) Pengadministrasian Tes

Pengadministrasian tes untuk kelompok berbeda dengan tes individual. Pengadministrasian tes untuk kelompok lebih praktis dan efektif daripada tes individual. Pengadministrasian kelompok mengizinkan tidak ada interaksi antara peserta tes dengan penguji atau menggunakan perluasan/elaborasi terhadap pertanyaan. Jadi, interaksi terjadi antara peserta tes dengan stimulus/pertanyaan dalam tes. Administrasi untuk kelompok memberi petunjuk secara jelas dan mudah untuk diikuti oleh

peserta tes. Penggunaan jumlah butir dengan beberapa tipe soal diharapkan tidak menimbulkan sumber kesalahan bagi hasil skor tes.

Pengadministrasian untuk tes individual secara praktis melatih dan memberi pertanyaan kepada peserta tes. Hal ini merupakan keuntungan khusus bagi administrator tes individual ketika melakukan interaksi personal, memanipulasi pertanyaan, menentukan skor, dan memilih soal berikutnya. Jadi, untuk mengadministrasikan beberapa tes individual, konstruksi tes harus bersifat responsif untuk mengembangkan prosedur pengadministrasian dan materi tes, serta kriteria penilaian.

4) Standardisasi Tes

Tes standar merupakan satu kondisi dari pengadministrasian dan prosedur penilaian yang didesain sama untuk semua pengguna tes. Kondisi dari administrasi meliputi: *setting* tes, petunjuk bagi peserta tes, materi tes, dan waktu yang digunakan. Prosedur penilaian meliputi derivasi dan transformasi dari skor mentah. Standardisasi nilai psikometri mereduksi variasi kesalahan, kondisi administrasi, dan praktek penskoran. Praktik standardisasi pengukuran merupakan hal yang perlu dipertimbangkan dan dipercayai di antara pengguna tes. Standardisasi merepresentasikan praktik pengukuran yang baik untuk semua tes dan berbagai tujuan tes termasuk spesifikasi dan pengembangan tes.

b. Atribut Internal Tes

Hal-hal yang berkaitan dengan atribut tes antara lain adalah: (1) kesesuaian tes dengan *subject matter*, (2) tipe/bentuk tes, (3) karakteristik soal (tingkat kesukaran dan daya beda tes), (4) kriteria dan prosedur penilaian, dan (5) jumlah soal. Hal yang perlu diperhatikan pada tahap awal penyusunan tes adalah penentuan kesesuaian tes dengan *subject matter* yang dapat dirangkum dalam suatu kisi-kisi. Bentuk tes dan jumlah soal

disesuaikan dengan tujuan pembelajaran yang ingin dicapai. Karakteristik soal ditentukan pada rentang yang sesuai dengan kemampuan siswa. Kriteria dan prosedur penilaian ditentukan sesuai dengan bentuk soal, misalnya skor benar-salah untuk objektif tes dan *partial credit* untuk tes esai.

1) Kesesuaian Tes dengan Materi

Sebelum membuat soal perlu dibuat kisi-kisi untuk menentukan kesesuaian tes dengan materi atau kurikulum. Kisi-kisi merupakan acuan bagi penulis soal, agar siapapun yang menulis soal akan menghasilkan soal yang isi dan tingkat kesulitannya relatif sama. Kisi-kisi biasanya berupa matriks (tabel) yang berisi spesifikasi soal. Langkah pembuatan kisi-kisi diawali dengan menentukan: tujuan pelajaran, pokok/subpokok bahasan, uraian materi, dan indikator (Djemari Mardapi, 2004: 89). Di samping itu, perlu ditentukan pula jumlah soal untuk setiap materi yang akan diteskan.

2) Bentuk Tes

Ada beberapa bentuk tes yang dapat dipilih dalam menyusun suatu perangkat tes, yaitu tes objektif dan nonobjektif. Bentuk tes objektif yang biasa digunakan adalah pilihan ganda, benar salah, menjodohkan, dan isian singkat. Tes uraian dapat dikategorikan menjadi dua, yaitu tes uraian objektif dan nonobjektif. Tes uraian objektif biasa digunakan pada bidang ilmu eksakta, sedangkan tes uraian nonobjektif biasa digunakan pada bidang ilmu sosial. Menurut Lane (2004) kesempatan mengembangkan tes pilihan ganda dengan desain soal berbasis asesmen kognitif akan makin terbuka pada masa yang akan datang.

3) Karakteristik Soal

Karakteristik soal meliputi daya beda dan tingkat kesukaran soal. Soal yang dikemas dalam suatu tes hendaknya memiliki daya beda yang baik. Artinya, indeks daya beda soal tidak negatif atau nol. Tes sebaiknya terdiri atas soal-soal yang memiliki

tingkat kesukaran yang proporsional. Jumlah soal dengan tingkat kesukaran sukar, sedang, dan mudah harus mengikuti kurva normal. Perbandingan antara soal sukar, sedang, dan mudah adalah 1:2:1, atau 25% untuk soal sukar, 50% untuk soal sedang, dan 25% untuk soal mudah. Soal yang terlalu sulit akan membuat siswa frustrasi, dan soal yang terlalu mudah tidak memberikan banyak informasi. Tentunya hal tersebut tidak berlaku jika tes digunakan untuk seleksi.

4) Kriteria dan Prosedur Penilaian

Kriteria dan prosedur penilaian disesuaikan dengan bentuk tes. Tes berbentuk objektif pilihan ganda kriteria penilaian ada dua kategori yaitu benar dan salah. Soal dijawab dengan benar diberi skor 1 dan soal dijawab salah diberi skor 0. Tes uraian memiliki kriteria penilaian yang berbeda. Oleh sebab itu, perlu dibuat rubrik penskoran agar memenuhi aspek keadilan.

5) Panjang Tes

Penentuan panjang tes didasarkan pada cakupan materi dan waktu yang disesuaikan dengan faktor kelelahan siswa. Pada umumnya waktu yang dibutuhkan untuk mengerjakan tes bentuk pilihan ganda adalah 1 sampai 2 menit untuk setiap soal.

c. Spesifikasi Tes pada Masa yang Akan Datang

Spesifikasi tes untuk masa mendatang dikaitkan dengan tiga hal, yaitu : (a) spesifikasi tes yang direncanakan, (b) ahli validasi dan pengguna tes, dan (c) *iterasi* perencanaan tes. Hal ini berkaitan dengan pengembangan tes yang lebih luas baik tes paralel maupun perluasan nomotetik.

3. Penulisan Soal Tes

Penulisan soal dapat dilakukan setelah menyusun spesifikasi tes yang meliputi penentuan tujuan, kisi-kisi, bentuk, dan panjang tes. Penulisan soal merupakan langkah menjabarkan

indikator menjadi pernyataan-pernyataan yang karakteristiknya sesuai dengan perincian pada kisi-kisi yang telah dibuat.

Dua hal penting yang harus diperhatikan ketika menulis soal, yaitu tipe soal dan teknik penulisan soal. Penentuan tipe soal penting karena untuk membedakan apa yang diukur, bagaimana cara mengukur, dan bagaimana substansi, serta format tes yang akan dibuat. Teknik penulisan soal terkait dengan aturan dalam menulis soal. Penulisan soal merupakan inti dari kegiatan pengembangan tes. Ketika menulis soal langkah-langkah dan aturan-aturan penulisan perlu diikuti agar diperoleh soal yang baik.

Penulisan soal diawali dengan mencermati langkah-langkah yang telah dilakukan sebelumnya, yaitu tujuan pembuatan tes dan spesifikasi tes. Dua hal tersebut merupakan panduan dalam menulis soal. Langkah selanjutnya meliputi: (a) pendataan kemampuan yang hendak diukur, (b) pembuatan matriks tentang isi tes dan domain kognitif, (c) penentuan spesifikasi soal.

a. Penentuan Kemampuan yang Hendak Diukur

Kemampuan yang hendak diukur disesuaikan dengan tujuan spesifik pembelajaran. Kemampuan ini dapat dikonstruksi dari relevansi kurikulum dan kriteria yang hendak dicapai. Jumlah soal harus proporsional untuk setiap aspek yang diukur. Tingkat kedetailan pendataan kemampuan sangat membantu penulis soal dalam mengembangkan soal.

b. Membuat Matriks tentang Tujuan Tes dan Ranah Kognitif

Matriks yang dibuat berbeda dengan kisi-kisi soal. Matriks ini hanya berisi rambu-rambu tentang tujuan tes, ranah kognitif yang hendak diukur dan jumlah soal untuk setiap ranah. Millman & Greene (1989: 351) mencontohkan matriks ini seperti yang disajikan dalam Tabel 6.1.

Tabel 6.1.
Matriks tentang Tujuan Tes dan Ranah Kognitif
untuk Tes Prestasi Sains

Tujuan	Ranah Kognitif				Jumlah total
	<i>Recall</i>	<i>Explicit Information Skill</i>	<i>Inferensial Reasoning</i>	<i>Evaluation</i>	
Siswa mampu mendemonstrasikan pemahamannya terhadap konsep atau metode untuk menjelaskan tentang					
Botani	2	6			8
Zoologi	3	2	3		8
Ekologi	3		1	2	6
Fisika	4		1	2	7
Kimia	4	1	1	1	7
Air, Tanah	1	1	2		4
Proses Total	17	10	8	5	40

(Millman & Greene, 1989: 351)

c. Menentukan Spesifikasi Soal

Spesifikasi soal secara umum merepresentasikan secara detail tentang petunjuk mengidentifikasi soal berdasarkan domain tes. Spesifikasi soal dapat berupa: (1) aturan untuk memilih jawaban benar termasuk format dan penempatan, (2) aturan menentukan distraktor termasuk jumlah distraktor dan formatnya, (3) karakteristik soal, (4) keterangan yang komplis dan detail tentang kondisi tes, termasuk perlengkapan wajib dan petunjuk spesifik untuk peserta tes, (5) ketentuan batas waktu, dan (6) prosedur dan kriteria penilaian yang digunakan.

4. Penelaahan Soal Tes

Penelaahan perlu dilakukan setelah tes dibuat. Hal ini perlu dilakukan untuk memperbaiki soal jika ternyata ada kesalahan dan kekurangan. Penelaahan biasanya dilakukan oleh orang lain, yang terdiri atas para ahli bidang studi. Dengan telaah soal ini diharapkan makin dapat memperbaiki kualitas soal yang terbentuk.

Ada dua metode yang biasa digunakan untuk mengevaluasi tes, yaitu ditinjau dari: (a) isi dan format tes, dan (b) respons siswa terhadap tes.

a. Evaluasi terhadap Isi dan Format Tes

Evaluasi isi dan format tes dilakukan setelah penulisan soal selesai, tanpa menunggu respons dari peserta tes. Evaluasi ini dapat dilakukan oleh beberapa evaluator yaitu ahli materi, ahli pengukuran pendidikan, ahli psikologi, dan guru sebagai praktisi pendidikan. Kriteria asesmen yang harus dipenuhi suatu tes adalah *accuracy*, *communicability*, *suitability*, dan *conformity*.

1) *Accuracy*

Accuracy (keakuratan) soal di antaranya dapat dilihat dari kunci jawaban. Setiap soal hendaknya hanya memiliki satu jawaban benar. Meskipun kadang-kadang jawaban benar dapat ditinjau dari berbagai perspektif, namun secara relatif diperlukan kunci jawaban yang akurat dari setiap soal.

2) *Communicability*

Communicability (kekomunikatifan) tes dapat dilihat dari kebenaran dan keterbacaan petunjuk pelaksanaan tes, kelugasan pertanyaan, konsistensi penulisan, dan beberapa ketentuan lain seperti yang disebutkan dalam cara penulisan soal.

3) *Suitability*

Suitability (kesesuaian) soal dapat dilihat dari kesesuaiannya dengan tingkat kesulitan yang direncanakan,

tingkat kepentingan dilakukannya pengukuran, dan ada atau tidak adanya bias.

4) Conformity

Conformity (kecocokan) soal lebih menekankan pada kecocokan terhadap spesifikasi yang telah direncanakan. Pencocokan soal dengan spesifikasi merupakan teknik validasi terhadap isi atau konten. Di samping itu, kecocokan soal dengan kurikulum, materi pengajaran, dan kriteria kemampuan yang ditetapkan juga perlu mendapat perhatian.

b. Evaluasi Berdasarkan Respons Testee: Pelaksanaan Uji Coba Tes

Sebelum soal digunakan dalam tes yang sesungguhnya, uji coba perlu dilakukan agar makin memperbaiki kualitas soal. Uji coba dapat digunakan sebagai sarana memperoleh data empirik tentang kualitas soal yang telah disusun. Melalui uji coba dapat diperoleh data tentang kualitas soal. Jika ternyata soal yang dibuat belum memenuhi kriteria yang diharapkan, maka perlu dilakukan perbaikan.

Ketika melakukan uji coba, perlu didahului dengan: (1) uji coba awal, (2) pemilihan soal untuk dirangkai menjadi tes, (3) penentuan sampel uji coba, dan (4) penentuan jumlah soal.

1) Uji Coba Awal

Draf soal diujicobakan kepada kelompok kecil, selanjutnya dilakukan perevisian dan penulisan ulang. Uji coba awal tidak hanya bermanfaat untuk merevisi kalimat, tetapi juga untuk menentukan waktu yang diperlukan oleh peserta tes dalam menyelesaikan setiap soal, serta tingkat kesukaran soal. Karena jumlah peserta uji coba hanya sedikit dan tidak representatif, maka penginterpretasian hasil statistik harus hati-hati.

2) Pemilihan Soal Uji Coba

Berkaitan dengan uji coba awal, ada 3 strategi yang dapat digunakan untuk mengumpulkan data respons terhadap soal,

yaitu: (1) menganggap soal eksperimen sebagai tes operasional, (2) melekatkan soal eksperimen pada tes operasional, dan (3) soal eksperimen merupakan bagian dari uji coba. Setiap strategi memiliki kekurangan dan kelebihan. Apabila menggunakan soal eksperimen sebagai tes operasional, maka tidak diperoleh tambahan soal tetapi dapat menghemat waktu dan biaya. Apabila melekatkan soal eksperimen pada tes operasional, maka diperoleh soal lebih banyak dan bervariasi. Apabila digunakan soal eksperimen murni, biaya yang diperlukan lebih besar.

3) Penentuan Sampel Uji Coba

Sampel uji coba seharusnya seperti populasi target. Penentuan sampel uji coba bergantung pada 3 faktor, yaitu: (1) bagaimana sampel diambil, (2) informasi apa yang dibutuhkan, dan (3) berapa ukuran sampel yang diperlukan.

4) Jumlah Soal Uji Coba

Pembuatan soal hendaknya lebih banyak dari jumlah yang akan dieksperimenkan. Hal ini untuk mengantisipasi soal yang tidak memenuhi kriteria statistik soal. Thorndike (1982: 53) menyebutkan bahwa jumlah soal ekstra untuk mengantisipasi soal jelek sebanyak 50% dari jumlah soal yang akan dieksperimenkan.

5. Penganalisisan Soal Uji Coba

Data hasil uji coba dianalisis untuk diperoleh data empirik tentang kualitas soal. Hasil analisis soal diperoleh tingkat kesukaran dan daya beda soal. Teknik analisis data dapat menggunakan rumus probabilitas menjawab benar dan deskriminasi. Selain itu, penentuan parameter soal juga dapat dilakukan dengan menggunakan kurva karakteristik soal yang mengestimasi probabilitas jawaban benar sebagai fungsi dari tingkat kemampuan siswa.

6. Perbaikan Tes

Setelah dilakukan uji coba dan analisis, langkah selanjutnya adalah memperbaiki soal-soal yang belum sesuai dengan kriteria yang ditetapkan. Soal yang tidak sesuai dengan standar kualitas biasanya dibuang. Akan tetapi, dalam membuang soal tetap memperhatikan indikator yang akan diukur. Jika ternyata soal tersebut merupakan satu-satunya soal yang mengukur suatu indikator, maka soal tetap dipertahankan tetapi harus diperbaiki. Oleh sebab itu, ketika memilih soal harus memperhatikan kembali kurikulum dan domain kognitif.

7. Perakitan Tes

Setelah semua soal dianalisis dan diperbaiki, langkah berikutnya adalah merakit soal-soal menjadi satu kesatuan tes. Ada 4 langkah yang dapat dilakukan ketika merakit tes, yaitu: (1) mengurutkan soal, biasanya dari yang mudah ke yang lebih sulit, (2) membuat *layout*, (3) menggandakan tes, dan (4) melakukan pengamanan.

Prosedur pengembangan tes berakhir di perakitan tes. Namun, untuk mengetahui ketercapaian tujuan tes perlu melaksanakan dan menilai hasil tes. Agar testing dapat berjalan dengan baik diperlukan suatu petunjuk pelaksanaan yang dapat memandu pelaksana tes, siapapun orangnya. Petunjuk pelaksanaan tes selanjutnya disebut manual untuk pelaksanaan tes berisi karakteristik tes, pedoman penilaian, interpretasi hasil tes, dan cara pelaporan hasil tes.

8. Pelaksanaan Tes

Tes yang telah disusun diberikan kepada peserta tes. Pelaksanaan tes dilakukan sesuai dengan waktu yang telah ditetapkan. Dalam pelaksanaan tes ini perlu dilakukan pemantauan dan pengawasan agar tes tersebut dikerjakan dengan jujur dan sesuai dengan ketentuan yang ditetapkan.

9. Interpretasi Hasil Tes

Tes digunakan untuk mengindikasikan kemampuan seseorang. Oleh sebab itu, diperlukan penjelasan hasil tes secara psikologis. Tes menghasilkan data kuantitatif berupa skor. Hasil tes atau skor merupakan gambaran dari penguasaan siswa terhadap *latent traits* (Snow & Lohman, 1989: 316-317). Skor ini kemudian ditafsirkan sehingga menjadi nilai. Tinggi rendahnya nilai biasanya dikaitkan dengan acuan penilaian. Ada dua acuan penilaian, yaitu acuan norma dan acuan kriteria. Penilaian acuan kriteria (PAK), yaitu membandingkan nilai yang diperoleh dengan kriteria yang harus dicapai. PAK mampu mendeteksi siswa yang dapat mengerjakan atau mencapai kompetensi yang ditetapkan dan siswa yang belum (Kumaidi, 2002: 209).

10. Pelaporan Hasil Tes

Hasil tes dapat dilaporkan dalam bentuk nilai total dan penginterpretasian lebih detail. Frechtling (1989: 475) menyatakan bahwa fungsi pelaporan hasil tes bagi publik adalah menyajikan skor tes, jumlah indikator yang dapat dicapai siswa, dan menunjukkan posisi kemampuan siswa. Di samping itu, dengan adanya pelaporan hasil tes kepada publik, keefektifan dari program pembelajaran dan kualitas guru dapat diketahui.

B. Pengembangan Tes

Tes hasil belajar dapat dilakukan secara lisan dan tertulis. Tes tertulis merupakan suatu teknik penilaian yang menuntut jawaban secara tertulis, baik berupa isian atau pilihan. Tes tertulis dapat dibagi menjadi dua bentuk, yaitu tes uraian (*essay test*) dan tes objektif (*obyektive test*). Setiap bentuk tes mempunyai kelebihan dan kelemahan. Dengan demikian, bentuk tes tersebut sebaiknya tidak digunakan secara tunggal. Pada bab ini akan diuraikan tentang pengembangan tes uraian dan tes objektif.

1. Tes Uraian

Tes uraian adalah tes yang berbentuk pertanyaan lisan atau tulisan yang jawabannya merupakan karangan (essay) atau kalimat panjang-panjang. Panjang pendeknya kalimat atau jawaban tes itu relatif, sesuai dengan kecakapan dan pengetahuan si penjawab.

Telah disampaikan di atas bahwa setiap bentuk tes memiliki kelebihan dan kekurangan. Beberapa kelebihan tes bentuk uraian adalah:

- a. Sangat baik untuk mengukur proses mental tingkat tinggi, sehingga sampai sekarang masih tetap dipertahankan penggunaannya.
- b. Menyusunnya lebih mudah, karena jumlah butir soal (*test item*) terbatas.
- c. Peserta ujian didorong agar menjadi lebih siap, karena harus menguasai secara mendalam untuk dapat melakukan analisis, sintesis, dan evaluasi.

Kelemahan tes bentuk uraian antara lain:

- a. Kurang baik untuk mengukur jenjang kemampuan yang rendah, seperti ingatan dan pemahaman.
- b. Tidak mampu mewakili seluruh bahan yang diujikan, karena jumlah soalnya terbatas.
- c. Sukar untuk menyusun pertanyaan dan rambu-rambu jawaban yang baik.
- d. Pemberian skornya sangat subjektif. Subjektivitas dapat dipengaruhi oleh banyak faktor, misal jika pemeriksa cukup kenal dengan peserta ujian, cenderung akan memberikan skor lebih tinggi. Anak yang pandai merangkai kalimat dan atau tulisannya bagus akan mendapat skor yang tinggi pula.
- e. Ketidakkonsistenan dalam memeriksa jawaban. Misalnya, lembar jawaban yang diperiksa lebih awal akan cenderung dinilai lebih mahal atau lebih murah, tetapi umumnya

cenderung lebih mahal. Apa lagi jika dilakukan lebih dari seorang pemeriksa, walaupun rentang skor yang digunakan sudah menggunakan kesepakatan yang sama. Pemeriksa juga cenderung untuk mengoreksi gaya bahasa, tata tulisannya, jadi bukan semata-mata pemahaman terhadap materinya. Panjang uraian juga akan mempengaruhi skor. Uraian yang panjang cenderung diberi skor lebih tinggi walaupun tidak tepat jawabannya.

Agar dapat menyusun tes dengan benar berikut diuraikan cara menyusun butir soal bentuk uraian. Hal yang perlu dilakukan agar dapat menghasilkan tes bentuk uraian adalah:

- a. Terlebih dahulu harus ditentukan tingkatan kemampuan apa yang harus digunakan peserta ujian untuk menjawab pertanyaan yang akan diujikan. Misalnya, apakah untuk mengukur kemampuan peserta dalam menilai kelengkapan suatu rencana kegiatan/prosedur kerja, ataukah ingin mengukur kemampuan peserta dalam menyusun suatu rencana percobaan?
- b. Usahakan agar jawaban peserta ujian tidak keluar dari ruang lingkup (*scope*) pertanyaan yang dimaksud penguji. Untuk itu jika memang diperlukan, dapat dilengkapi dengan pedoman-pedoman tentang apa yang harus dikemukakan dalam menjawab pertanyaannya.

Syarat butir soal bentuk uraian yang baik harus mempertimbangkan segi materi, konstruksi, dan bahasa. Segi materi memperhatikan:

- a. kesesuaian soal dengan indikator.
- b. batasan pertanyaan dan jawaban yang diharapkan sesuai dengan materi dan tujuan pengukuran.
- c. isi materi yang ditanyakan sesuai dengan jenjang, jenis sekolah, dan tingkat kelas.

Segi konstruksi memperhatikan:

- a. Rumusan kalimat dalam bentuk kalimat tanya atau perintah yang menuntut jawaban terurai.
- b. Ada petunjuk yang jelas cara mengerjakan/menyelesaikan soal.
- c. Ada pedoman pensekoran.
- d. Tabel, grafik, diagram, kasus, atau yang sejenisnya bermakna (jenis-jenis keterangannya atau ada hubungannya dengan masalah yang ditanyakan).
- e. Butir soal tidak bergantung pada jawaban butir soal sebelumnya.

Segi bahasa memperhatikan:

- a. Rumusan kalimat soal komunikatif.
- b. Kalimat menggunakan bahasa yang baik dan benar sesuai dengan jenis bahasanya.
- c. Rumusan kalimat tidak menimbulkan penafsiran ganda atau salah pengertian.
- d. Menggunakan bahasa/kata yang umum (bukan kata lokal).
- e. Rumusan soal tidak mengandung kata-kata yang menyinggung perasaan siswa.
- f. Sebaiknya pertanyaan diawali oleh kata-kata seperti: bandingkan, berikan alasan, jelaskan mengapa, beri contoh, dan sebagainya.

2. Tes Objektif

Tes objektif adalah tes yang dibuat sedemikian rupa sehingga hasil tes itu dapat dinilai secara objektif. Siapapun yang melakukan penilaian akan menghasilkan skor yang sama. Si penjawab tes tinggal memilih mengisi, menjodohkan, dan sebagainya dengan menggunakan tanda-tanda seperti tertera dalam soal atau petunjuk pengerjaan soal.

Beberapa kelebihan atau keuntungan menggunakan tes bentuk objektif, adalah:

- a. Cara memeriksanya (*scoring*) cukup mudah, cepat dan benar-benar apa adanya (objektif).
- b. Dapat merangkum keseluruhan bahan, sehingga dari segi kesahan atas dasar kesahihan ini (*content validity*) lebih dapat dipertanggungjawabkan.
- c. Kerahasiaan butir soal relatif lebih terjamin dan beberapa model butir soal objektif dapat digunakan berulang hanya dengan mengubah sebagian alternatif jawaban.

Kelemahan penggunaan bentuk objektif antara lain:

- a. Sukar dalam menyusunnya, kecuali untuk bentuk melengkapi yang tidak memberikan alternatif/pilihan jawaban.
- b. Kurang dapat digunakan untuk mengukur jenjang kemampuan yang tinggi.
- c. Dapat menimbulkan kesempatan untuk menerka alternatif jawaban yang benar.

Beberapa hal yang perlu diperhatikan dalam membuat butir soal agar hasilnya memuaskan antara lain:

- a. Penyusunan butir soal untuk ujian formatif sebaiknya disusun secepatnya sesuai dengan kemajuan materi yang didiskusikan di kelas.
- b. Pembuatan butir soal sebaiknya lebih banyak daripada jumlah yang sudah direncanakan dalam tabel spesifikasi (tabel yang menunjukkan distribusi butir soal sesuai macam pokok bahasan dan kompetensi dasar). Artinya suatu KD dapat dibuat lebih dari satu butir soal. Jika setelah diujicobakan hasilnya memuaskan dapat dikoleksi ke dalam bank soal.

- c. Setiap butir soal ditulis dalam kartu pokok uji, sehingga mudah memperbaiki/menggantinya.
- d. Butir soal yang telah dibuat kemudian dikerjakan sendiri atau dikonsultasikan kepada teman sejawat terlebih dahulu untuk mencari jawaban kunci dan menaksir waktu serta tingkat kesukarannya. Bila perlu dikonsultasikan juga ke ahli bahasa.

Setiap butir soal yang sudah dibuat perlu dievaluasi. Hal yang perlu dipertanyakan antara lain:

- a. Apakah sudah mengukur *learning outcome* yang diinginkan dan sudah sesuai dengan kekhasan *learning outcome* yang diharapkan?
- b. Apakah butir soal tidak bermakna ganda?
- c. Apakah bahasanya sudah sesuai dengan bahasa subjek yang akan diuji?
- d. Apakah petunjuk mengerjakan sudah jelas?
- e. Apakah tingkat kesukarannya sudah memadai?
- f. Apakah soal yang sudah ada bebas dari unsur tumpang-tindih? (tumpang tindih terjadi jika seseorang mengerjakan suatu pokok uji akan dapat mengerjakan butir soal yang lainnya)
- g. Apakah urutan pokok uji sudah diatur dengan pola tertentu, diurutkan menurut pokok bahasan, menurut mudah sukarnya, menurut bentuknya, atau menurut dari hal-hal tersebut?
- h. Apakah sudah dicantumkan berapa lama mengerjakannya?
- i. Apakah sudah dicantumkan hal-hal yang penting yang lain yang harus diperhatikan para peserta ujian selain apa yang sudah disebutkan di atas, termasuk apa tujuannya, bagaimana sistem koreksinya.
- j. Apakah butir soal sudah sesuai dengan tujuannya?

Pertanyaan-pertanyaan tersebut dapat memandu penulis soal dalam upaya membuat soal yang baik dan benar dari langkah awal hingga perakitan soal menjadi tes.

Langkah pengembangan tes tersebut tidak hanya berlaku pada tes prestasi, tetapi juga untuk tes kemampuan yang lain.

LITERASI SAINS

Pendidikan sains sangat penting dalam kehidupan sehari-hari. Pendidikan sains membangun siswa untuk berpikir dalam memahami fenomena atau kejadian alam dengan metode ilmiah seperti yang dilakukan oleh ilmuwan (NRC, 1996). Pendidikan sains juga menyiapkan siswa menjadi warga negara yang bertanggung jawab terhadap kejadian di sekitar, seperti fenomena *global warming* atau pun kejadian lain (Sahlan & Rusilowati, 2012). Setidaknya ada empat elemen utama dalam pendidikan sains, yaitu (1) membangun pengetahuan siswa tentang konsep-konsep sains (kognitif), (2) mengembangkan kemampuan siswa berargumentasi secara kritis dalam kegiatan sains (psikomotorik), (3) memahamkan siswa tentang proses bagaimana peristiwa itu terjadi, bukan sekedar mengetahui (*how they know not what they know*), dan (4) melatih siswa untuk belajar bekerja sama dan membangun sikap ilmiah (aspek sosial dan afektif) (Osborne, 2007 : 177).

Holbrook & Rannikmae (2007) mengajukan tiga domain dalam pendidikan sains, yaitu domain sosial (*society domain*), domain personal (*personal domain*) dan domain hakikat sains (*nature of science domain*). Domain-domain ini sejalan dengan aspek literasi sains yang dituntut dalam PISA. *Science for all American* (AAAS, 1993) mendefinisikan literasi sains sebagai kemampuan (*ability*) memahami konsep dan prinsip sains (*concept and principle of science*) serta mempunyai kapasitas

berpikir ilmiah untuk memecahkan masalah sehari-hari kaitannya dengan sains.

A. Pengertian Literasi Sains

Literasi sains merupakan kemampuan menggunakan konsep sains untuk mengaplikasikannya dalam kehidupan sehari-hari, menjelaskan fenomena ilmiah serta menggambarkan fenomena tersebut berdasarkan bukti-bukti ilmiah (NRC, 1996; OECD, 2007; Bybee *et al.*, 2009; Rusilowati, 2013). Berkaitan dengan literasi sains, beberapa pakar sains membuat tingkat literasi sains dengan alasan literasi sains yang didefinisikan terlalu banyak dan kompleks untuk seseorang (Boujaude, 2002). Shwartz *et al* (2006) menambahkan bahwa literasi sains ini berbeda untuk setiap orang. Hal ini disebabkan oleh beberapa faktor seperti umur, pengalaman, pengetahuan dan lingkungan. Fensham (2009) menekankan konteks pada aspek literasi sains. Pada pembahasan ini, aspek literasi sains meliputi pengetahuan sains, penyelidikan sains, cara berpikir sains, dan keterkaitan sains dengan lingkungan, teknologi, dan masyarakat. Hal ini selain mengacu kepada pendapat para ahli tersebut juga sesuai dengan rumusan Chiappetta *et al.*, (1991; 1993), indikator literasi sains yang telah dikembangkan oleh Rusilowati dkk., (2016d), serta PISA.

B. Aspek Literasi Sains

Aspek literasi sains diperinci ke dalam beberapa tujuan, tetapi pada dasarnya mencakup pengetahuan, proses, dan sikap. Chi Lau (2009) mengemukakan aspek penting dalam literasi sains adalah (1) konsep sains dan aplikasinya dalam kehidupan sehari-hari, (2) proses inkuiri sains, (3) memahami hakikat sains, dan (4) memahami hubungan antara sains, teknologi, dan masyarakat. Chiappetta *et al.* (1991) merumuskan aspek sains meliputi *science as a body of knowledge, way of investigating, dan interaction of science, environment, technology, and society*. Rusilowati, dkk.

(2016) menambahkan sub aspek lingkungan dan merumuskan indikator untuk setiap aspek literasi sains sebagai berikut:

1. Sains sebagai batang tubuh pengetahuan (*Science as a body of knowledge*).

Aspek ini biasanya dimaksudkan untuk menampilkan, mendiskusikan atau menanyakan hal-hal untuk mengingat informasi tentang fakta-fakta, konsep-konsep, prinsip-prinsip, hukum-hukum, teori-teori, dan sebagainya. Hal ini akan mencerminkan pemindahan pengetahuan ilmiah manakala siswa menerima informasi. Indikator yang dapat dikembangkan antara lain:

- a. mendefinisikan/menyebutkan fakta-fakta, konsep-konsep, prinsip-prinsip, dan hukum-hukum.
- b. merumuskan hipotesis
- c. menerapkan teori

2. Sains sebagai cara untuk menyelidiki (*Science as way of investigating*)

Aspek ini dimaksudkan untuk menstimulasi berpikir dan melakukan sesuatu dengan menugaskan kepada siswa untuk “menyelidiki”. Hal ini mencerminkan aspek inkuiri dan belajar aktif, melibatkan siswa dalam proses sains seperti melakukan observasi, mengukur, melakukan klasifikasi, menarik kesimpulan, mencatat data, melakukan perhitungan, melakukan percobaan, dan sebagainya. Pembelajarannya dapat menyangkut kegiatan “*hands-on*”. Indikator yang dapat digunakan untuk mengukur ketercapaian aspek ini, antara lain adalah:

- a. merumuskan permasalahan
- b. merumuskan hipotesis
- c. melakukan pengamatan
- d. menyajikan data pengamatan
- e. menganalisis hasil pengamatan

- f. membuat grafik, tabel, dan lain-lain.
 - c. menyimpulkan hasil
3. Sains sebagai Cara Berpikir (*Science as a way of thinking*)
Aspek ini dimaksudkan untuk memberi gambaran sains secara umum dan ilmuwan khususnya dalam melakukan penyelidikan. Hakikat sains mewakili proses berpikir, penalaran (*reasoning*), dan refleksi ketika siswa berbicara tentang berlangsungnya kegiatan ilmiah. Indikatornya antara lain:
- a. menggambarkan bagaimana seorang ilmuwan melakukan eksperimen
 - b. menunjukkan perkembangan historis dari sebuah ide
 - c. menekankan sifat empiris dan objektivitas ilmu sains
 - d. mengilustrasikan penggunaan asumsi-asumsi
 - e. melakukan berpikir induktif dan deduktif
 - f. menganalisis hubungan sebab dan akibat
 - g. menemukan fakta dan bukti
 - h. menyajikan metode ilmiah
 - i. memecahkan masalah
4. Interaksi Sains, Lingkungan Teknologi, dan Masyarakat (*Interaction of Science, Environment, Technology, and Society*) Aspek ini dimaksudkan untuk memberi gambaran tentang pengaruh atau dampak sains terhadap lingkungan, teknologi, dan masyarakat. Aspek melek ilmiah (*scientific literacy*) menyinggung penerapan atau aplikasi sains dan bagaimana teknologi membantu atau mengganggu manusia. Hal ini juga menyinggung soal *issue* sosial dan karir. Siswa menerima informasi tersebut dan umumnya tidak harus menemukan atau menyelidiki. Indikator yang dapat digunakan untuk mengukur aspek ini antara lain:
- a. menjelaskan kegunaan ilmu sains dan teknologi bagi

- masyarakat
- b. menunjukkan efek negatif dari ilmu sains dan teknologi bagi masyarakat
 - c. mengidentifikasi masalah-masalah sosial yang berkaitan dengan ilmu sains atau teknologi
 - d. menyebutkan karir-karir dan pekerjaan-pekerjaan di bidang ilmu dan teknologi

Aspek atau dimensi yang mengkonstruksi kemampuan literasi sains yang dikembangkan PISA 2006 terdiri atas: (1) konteks sains yang terkait kehidupan sehari-hari baik secara individu, sosial maupun global; (2) kompetensi sains mencakup identifikasi isu sains, penjelasan fenomena sains, dan penggunaan fakta-fakta sains; (3) pengetahuan sains mencakup materi dan kerja sains; dan (4) sikap terhadap sains meliputi ketertarikan terhadap sains, dukungan terhadap kegiatan inkuiri, dan respons terhadap sumber daya alam dan lingkungan (Rusilowati, 2013: 6).

Penilaian literasi sains yang dilakukan oleh PISA 2006 (OECD, 2006) memandang kemampuan literasi sains melalui empat aspek yakni aspek konteks, pengetahuan, kecakapan, dan sikap. Aspek konteks seperti siswa mampu mengenali penggunaan sains dan teknologi dalam kehidupan. Aspek pengetahuan seperti siswa memahami gejala alam berdasarkan pengetahuan tentang sains. Aspek kompetensi meliputi siswa mampu memperlihatkan kompetensinya dalam mengidentifikasi masalah sains, menjelaskan fenomena secara ilmiah, dan menjelaskan simpulan berdasarkan fakta. Aspek sikap meliputi siswa tertarik pada sains, mendukung penemuan-penemuan sains, dan bertanggung jawab terhadap ilmu sains.

Literasi sains menurut PISA (2010) diartikan sebagai “*the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity*”. Tiga kompetensi ilmiah yang diukur

dalam literasi sains diuraikan sebagai berikut: (1) mengidentifikasi isu (masalah) ilmiah yaitu mengenali masalah yang mungkin untuk diselidiki ilmiah, mengidentifikasi kata kunci untuk mencari informasi ilmiah, mengenali fitur kunci dari penyelidikan ilmiah; (2) menjelaskan fenomena ilmiah yaitu menerapkan ilmu pengetahuan dalam situasi tertentu, menggambarkan atau menafsirkan fenomena ilmiah dan memprediksi perubahan, mengidentifikasi deskripsi yang tepat, memberikan penjelasan, dan prediksi; (3) menggunakan bukti ilmiah yaitu menafsirkan bukti ilmiah dan membuat kesimpulan dan mengkomunikasikan, mengidentifikasi asumsi, bukti, dan alasan di balik kesimpulan, berkaca pada implikasi sosial dari ilmu pengetahuan dan perkembangan teknologi (Bybee, 2009).

OECD (2013:100) menyatakan bahwa literasi sains sebagai penggunaan pengetahuan ilmiah untuk mengidentifikasi pertanyaan, memperoleh pengetahuan baru, menjelaskan fenomena ilmiah, dan menarik kesimpulan berdasarkan bukti tentang isu terkait ilmu pengetahuan. Definisi literasi sains oleh Holbrook (2009) adalah mengembangkan kemampuan dan keterampilan sains secara kreatif, yang berhubungan dengan kehidupan sehari-hari dalam memecahkan masalah serta bertanggung jawab terhadap keputusan yang diambil.

Literasi sains menekankan siswa bagaimana menganalisis, memprediksi, dan mengaplikasikan konsep sains dalam kehidupan sehari-hari. Sains dalam pandangan literasi tidak hanya sekedar mengetahui, namun bagaimana proses mereka mengetahui sampai mengaplikasikan dalam kondisi lingkungan sekitar (*how they know and apply in daily life not what they know*) (Ridwan, dkk. 2013:186).

PISA 2015 (OECD, 2016) masih memuat empat aspek literasi sains yaitu konteks, pengetahuan, kompetensi/proses, dan sikap. Uraian berikut dapat memandu kita untuk membedakan aspek literasi sains yang diukur antarperiode.

1. Aspek Konteks

Aspek konteks memandu siswa dalam menentukan ranah mana suatu situasi dialami, apa dampaknya bagi individu dan masyarakat luas. Konteks literasi sains meliputi konteks pribadi, lokal, nasional, dan global.

2. Aspek Pengetahuan

Pengetahuan mengarahkan siswa pada pemahaman tentang fakta, konsep, dan teori yang membentuk dasar konsep pengetahuan ilmiah. Aspek pengetahuan dibedakan menjadi tiga, yaitu pengetahuan konten, prosedural, dan epistemik.

a. Pengetahuan Konten

Konten yang diukur dalam literasi sains PISA 2015 meliputi bidang fisika, kimia, biologi, dan ilmu bumi dan antariksa yang memiliki relevansi tertentu. Relevansi tersebut mencakup kesesuaian dengan situasi kehidupan nyata, merepresentasikan konsep sains yang penting, dan sesuai dengan usia perkembangan siswa (untuk PISA, usia 15 tahun)

b. Pengetahuan Prosedural

Pengetahuan ini terkait dengan bagaimana ide-ide sains diproduksi. Pengetahuan ini diperlukan untuk melakukan penyelidikan ilmiah. Penyelidikan dilakukan secara empiris untuk menghasilkan kejelasan suatu ilmu atau materi yang diuji. Pada pengetahuan prosedural perlu memahami hal-hal sebagai berikut:

(1) Konsep variabel (terikat, bebas, dan kontrol)

(2) Konsep pengukuran secara kuantitatif, observasi (kualitatif), penggunaan skala dan pengelompokan variabel

(3) Mekanisme untuk melakukan replikabilitas (kedekatan antara besaran yang sama dan diukur secara berulang) dan akurasi

- (4) Cara untuk menampilkan data menggunakan tabel, grafik, dan diagram dengan tepat
- (5) Strategi untuk mengontrol variabel dan perannya pada rancangan penelitian, mengujicobakan secara acak untuk menghindari penemuan yang salah, dan mengidentifikasi mekanisme sebab akibat yang dimungkinkan
- (6) Menentukan rancangan penelitian yang sesuai dengan pertanyaan ilmiah, misalnya eksperimental atau hanya melihat pola.

c. Pengetahuan Epistemik

Pengetahuan epistemik merupakan pemahaman tentang alasan mendasar suatu prosedur digunakan. Indikatornya antara lain:

- (1) Menafsirkan data
- (2) Menganalisis kesimpulan
- (3) Menjelaskan bukti yang mendukung hipotesis

3. Aspek Kompetensi

Aspek kompetensi/proses dalam literasi sains PISA memberikan prioritas terhadap kompetensi: menjelaskan fenomena ilmiah, mengevaluasi dan merancang penyelidikan ilmiah, serta menafsirkan data dan bukti ilmiah.

a. Menjelaskan fenomena ilmiah

Indikator kompetensi menjelaskan fenomena ilmiah antara lain:

- (1) Menerapkan pengetahuan ilmiah
- (2) Mengidentifikasi suatu model
- (3) Merepresentasikan suatu model
- (4) Membuat prediksi dengan tepat
- (5) Memaparkan hipotesis dengan jelas
- (6) Menjelaskan implikasi pengetahuan ilmiah bagi masyarakat

b. Merancang dan mengevaluasi penyelidikan ilmiah
Kompetensi ini dapat diukur melalui indikator, antara lain:

- (1) Mengidentifikasi pertanyaan yang dieksplorasi dalam sebuah penelitian
- (2) Mengusulkan cara mengeksplorasi kegiatan penelitian ilmiah
- (3) Menjelaskan berbagai cara untuk memastikan data reliabel dan objektif
- (4) Menggeneralisasikan cara memastikan reliabilitas dan objektivitas data

c. Menafsirkan data dan bukti ilmiah.

Indikator yang dapat digunakan untuk mengungkap kompetensi menafsirkan data dan bukti ilmiah antara lain:

- (1) Memberikan pendapat dalam berbagai bentuk representasi ilmiah
- (2) Mengubah representasi data dalam bentuk yang berbeda-beda
- (3) Menganalisis data
- (4) Menginterpretasikan data
- (5) Membuat simpulan yang tepat
- (6) Mengidentifikasi asumsi, bukti, dan penalaran suatu teks
- (7) Membedakan argumen yang didasarkan pada bukti ilmiah dan teori yang dikemukakan oleh orang lain
- (8) Mengevaluasi argumen ilmiah dan bukti dari sumber yang berbeda

4. Aspek Sikap

Aspek sikap sains menunjukkan minat seseorang pada ilmu pengetahuan, dukungannya terhadap penyelidikan ilmiah, dan motivasinya untuk memperlakukan sumber daya alam dan lingkungan secara bertanggung jawab. Sikap sains dapat diukur melalui indikator antara lain:

- a. Ketertarikan terhadap ilmu pengetahuan dan teknologi
- b. Menilai segala sesuatu dengan pendekatan ilmiah untuk membangun literasi sains
- c. Memahami isu-isu lingkungan
- d. Memiliki kesadaran untuk menjaga lingkungan

Aspek literasi sains yang diukur dalam PISA sebenarnya tidak berbeda dengan aspek sains pada umumnya. Aspek yang dikembangkan PISA sesuai dengan teori yang dikemukakan oleh Chiapetta dan para ahli sains lainnya. Aspek literasi sains yang dituntut dalam PISA dari tahun ke tahun relatif sama, hanya penekanannya yang berbeda. Perubahan terjadi biasanya pada aspek konteks. Oleh sebab itu, pendidikan di Indonesia juga harus mengikuti perkembangan agar tidak tertinggal dari negara lain.

Penilaian literasi sains tidak digunakan untuk membedakan seseorang berliterat atau tidak, tetapi merupakan suatu proses yang kontinu dan terus berkembang sepanjang hidup manusia (Shwartz, 2006). Literasi sains terdiri atas pengetahuan dan pemahaman tentang konsep ilmiah dan proses yang dibutuhkan seseorang dalam pengambilan keputusan, partisipasi sosial, budaya, dan produktivitas ekonomi (Dani, 2009). Osborne (2007) menyatakan bahwa argumen pada pembelajaran sains dapat meningkatkan kemampuan berpikir kritis siswa dalam menyampaikan pendapat. Hal ini sangat relevan dengan tuntutan keterampilan abad 21 yang salah satunya adalah keterampilan berpikir kritis.

Evaluasi literasi sains yang dilakukan memberikan perhatian terhadap aspek kognitif dan afektif siswa. Aspek kognitif meliputi pengetahuan siswa dan kapasitasnya untuk menggunakan pengetahuan secara efektif dan melibatkan proses kognitif yang merupakan karakteristik sains dalam bidang personal, sosial, dan global. Aspek afektif berhubungan dengan masalah yang dapat dipecahkan oleh pengetahuan sains dan membentuk siswa yang mampu untuk membuat keputusan pada

saat ini maupun masa depan (PISA, 2010). Kemampuan literasi seseorang dapat berkembang sepanjang hayat (Shwatz, 2006). Kemampuan tersebut pada seseorang dapat sangat tinggi pada bidang tertentu tetapi dapat sangat rendah di bidang lain (Bybee, 2009).

Penilaian literasi sains dalam PISA tidak hanya mengukur tingkat pemahaman terhadap pengetahuan sains, tetapi juga pemahaman terhadap berbagai aspek proses sains, serta kemampuan mengaplikasikan pengetahuan dan proses sains dalam situasi nyata yang dihadapi peserta didik, baik sebagai individu, anggota masyarakat, ataupun warga dunia. Literasi sains sangat penting untuk dikuasai oleh siswa. Hal ini merupakan suatu kompetensi dasar siswa dalam memahami fenomena alam, lingkungan hidup, ekonomi, serta masalah-masalah yang dihadapi oleh manusia saat ini yang sangat bergantung pada kemajuan ilmu pengetahuan dan teknologi.

Chi Lau (2009) menyatakan bahwa kemampuan literasi sains dianggap penting karena:

- 1) Sains merupakan bagian penting bagi manusia untuk mengasah kemampuan berpikirnya.
- 2) Literasi sains memberikan pengalaman laboratorium dalam perkembangan bahasa, logika, dan kemampuan memecahkan masalah di kelas.
- 3) Kehidupan sosial menuntut seseorang membuat keputusan pribadi dan masyarakat tentang situasi yang dihadapi dimana terdapat informasi ilmiah yang berperan penting sehingga seseorang tersebut harus mempunyai pengetahuan tentang ilmu pengetahuan serta pemahaman tentang kemampuan dan metodologi ilmiah. Literasi sains akan melekat seumur hidup bagi siswa dalam berbagai macam situasi dan kondisi.
- 4) Perkembangan zaman dan teknologi bergantung pada kemampuan teknis dan ilmiah kemampuan dan daya saing warganya.

Penguasaan literasi sains diharapkan dapat mempermudah siswa untuk beradaptasi dengan kemajuan ilmu pengetahuan dan teknologi pada masa mendatang. Maka dari itu, faktor yang mempengaruhi tingkat literasi sains menjadi menarik untuk dipelajari. Salah satunya adalah alat evaluasi.

PENGEMBANGAN TES LITERASI SAINS

Penyusunan instrumen evaluasi yang berbasis literasi sains merupakan salah satu upaya untuk mengukur kemampuan literasi siswa terutama dalam bidang sains atau IPA. Untuk mengukur kemampuan literasi sains siswa dapat digunakan beberapa soal dari PISA dan soal yang dikembangkan setipe soal PISA. Indikator soal dapat diturunkan dari muatan sains yang ditentukan oleh PISA dan Chiappetta *et al.* (1991). Muatan sains yang diungkap oleh PISA meliputi pengetahuan (*knowledge*) dan kompetensi/proses. Muatan literasi sains yang diungkap oleh Chiappetta meliputi sains sebagai batang tubuh pengetahuan (*a body of knowledge*), sains sebagai cara untuk berpikir (*a way of thinking*), sains sebagai cara untuk menyelidiki (*a way of investigating*), dan interaksi antara sains, teknologi, dan masyarakat (*interaction between science, technology, and society*).

Pengembangan instrumen meliputi langkah-langkah: (a) menentukan spesifikasi instrumen, (b) menentukan jenis instrumen; (c) menulis instrumen; (d) menentukan sistem penskoran; (e) menelaah instrumen; (f) merakit instrumen; (g) melakukan uji coba; (h) menganalisis hasil uji coba; (i) memperbaiki instrumen; (j) melaksanakan pengukuran; dan (k) menafsirkan hasil pengukuran.

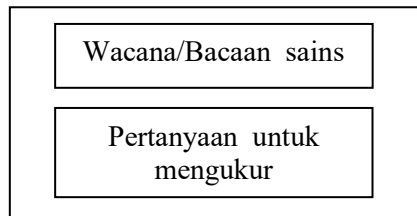
Instrumen yang dikembangkan hendaknya memenuhi kriteria kualitas instrumen seperti reliabilitas, validitas, daya pembeda, dan tingkat kesukaran. Teknik penentuan reliabilitas dapat menggunakan test-retest, tes paralel, tes belah dua. Rumus

yang dapat digunakan antara lain: korelasi *product moment*, Spearman Brown, Alpha, dll. Reliabilitas instrumen ditentukan dari besarnya koefisien korelasi. Pada kesempatan ini digunakan rumus KR-20. Validitas instrumen yang dikembangkan dapat dilihat dari validitas isi atau konstruk. Ketepatan dalam menentukan definisi konseptual, definisi operasional, dan penetapan aspek serta indikator yang direncanakan.

Rencana ke depan, soal-soal yang sudah divalidasi dan diuji karakteristiknya dikemas dalam bank soal yang nantinya dapat diakses oleh guru-guru di seluruh wilayah Indonesia. Ketersediaan soal literasi sains dapat terus dikembangkan. Dengan demikian maka ketersediaan soal literasi sains menjadi lengkap, sesuai dengan materi IPA yang dipelajari di tingkat SD, SMP, SMA dan Perguruan Tinggi.

Soal yang ada dapat menjadi contoh bagi guru dan calon guru untuk mengembangkan dan menerapkan pengujian literasi sains. Pemahaman guru dan calon guru terhadap literasi sains perlu ditingkatkan lebih dahulu, agar dapat mengarahkan siswanya dalam mengerjakan soal berbasis literasi sains.

Format soal yang dikembangkan untuk mengukur literasi sains mengacu pada format soal TIMSS atau PISA. Soal diawali dengan sebuah wacana yang secara implisit memuat konsep sains, diikuti dengan pertanyaan. Setiap wacana dapat diikuti satu atau lebih pertanyaan. Setiap pertanyaan yang diajukan jawabannya dapat diperoleh dari konsep yang terkandung di dalam wacana. Secara umum format soal literasi sains dapat digambarkan pada Gambar 8.1.



Gambar 8.1 Format Soal Literasi Sains

Tes literasi sains dapat diterapkan mulai SD, SMP, SMA dan perguruan tinggi khususnya LPTK sebagai pencetak calon guru sains. Aspek dan indikator soal dapat menggunakan teori Chiapetta, et al. (1991), Rusilowati, et al. (2016), atau sesuai dengan tuntutan PISA 2006 dan 2015. Secara umum, aspek literasi sains yang diukur sama, hanya pengelompokannya memiliki nama yang berbeda. Pada bab ini akan diuraikan tentang pengembangan instrumen literasi sains untuk siswa SD, SMP, SMA, dan mahasiswa calon guru IPA. Uraian lebih detail untuk Pengembangan instrumen tingkat perguruan tinggi. Hal ini dimaksudkan, untuk memberi bekal bagi mahasiswa calon guru dalam mengembangkan instrumen literasi sains. Ketika mahasiswa calon guru sudah memahami tentang pengembangan soal literasi sains dan interpretasinya, maka ia akan dapat menerapkan bagi siswanya.

A. Pengembangan Instrumen Literasi Sains SD

Instrumen literasi sains tingkat SD sebaiknya dikembangkan mengacu pada indikator TIMSS. Sejak tahun 1995 hingga 2015, konten TIMSS meliputi *earth science, life science, physical science, dan environment & nature science*. Domain kognitif di tahun 1995 meliputi: (1) pemahaman terhadap informasi sederhana, (2) pemahaman terhadap informasi kompleks, (3) teorisasi, menganalisis, dan pemecahan masalah, (4) menggunakan peraturan, prosedur rutin, dan proses sains, (5) investigasi terhadap alam. TIMSS biasanya dikenakan kepada siswa kelas IV. Tampak bahwa kemampuan yang harus dimiliki oleh siswa SD sudah begitu kompleks. Jadi pendapat bahwa untuk anak SD ranah kognitif yang diukur cukup C1 dan C2 tidaklah benar. Domain kognitif TIMSS tahun 2003 mencakup (1) pengetahuan faktual (*factual knowledge*), (2) pengetahuan konsep (*conceptual knowledge*), (3) penalaran dan analisis (*reasoning & analyzing*). TIMSS tahun 2015 relatif sama dengan TIMSS 2011, tetapi lebih mengacu kepada perkembangan kurikulum sains secara

internasional, seperti *Framework for K-12 Science Education*. Tahun 2015 TIMSS mengukur kemampuan dasar untuk kelas 4 dan 8. TIMSS diselenggarakan empat tahun sekali, dengan domain konten dan domain kognitifnya mengacu pada perkembangan dunia sains. Oleh sebab itu, guru di Indonesia sebaiknya mengikuti perkembangan pendidikan, agar siswanya memiliki kemampuan dan mampu bersaing secara internasional. Bentuk tes biasanya pilihan ganda dan atau esai.

Contoh soal TIMSS untuk kelas 4 dapat dilihat pada Gambar 8.2.

Beberapa hewan sangat bagus. Misalnya, harimau Siberia yang jumlahnya tinggal sedikit. Jika satu-satunya harimau Siberia yang tersisa adalah betina, kemungkinan apa yang akan terjadi?

- a. Harimau betina tersebut akan menemukan jantan jenis hewan lain untuk dikawinkan dan menghasilkan lebih banyak harimau Siberia
- b. Harimau betina akan kawin satu sama lain dan menghasilkan lebih banyak harimau Siberia
- c. Harimau betina hanya akan bisa menghasilkan harimau Siberia betina
- d. Harimau betina tidak akan bisa menghasilkan lebih banyak harimau Siberia, dan mereka akan musnah.

Gambar 8.2. Contoh Soal Literasi Sains untuk SD

Soal tersebut diujikan pada tahun 2011, domain kontennya *life science* (sains dalam kehidupan), topiknya siklus hidup, reproduksi, perkembangan, domain kognitifnya penalaran.

Soal yang dibuat untuk melatih siswa di Indonesia agar melek sains, kondisinya dapat disesuaikan dengan kondisi di Indonesia (kontekstual). Misalnya untuk soal di atas, harimau Siberia, dapat diganti dengan harimau Sumatera. Hal ini tentunya akan mempermudah siswa, karena mereka lebih familier dengan kondisi daerahnya atau negaranya. Sedikit demi sedikit siswa diberikan soal dengan konteks internasional.

B. Pengembangan Instrumen Literasi Sains SMP

Selain TIMSS, penilaian kemampuan dasar termasuk literasi sains, juga dilakukan oleh PISA. Uraian detail tentang konteks dan konten yang dinilai PISA sudah dipaparkan di bab sebelumnya. Berikut diberikan contoh soal dari PISA untuk siswa usia 15 tahun atau setara dengan siswa kelas IX atau X. Contoh soal literasi sains PISA untuk siswa SMP dapat dilihat pada Gambar 8.3.

Bentuk soal PISA biasanya *essay* atau pilihan ganda dan disertai dengan wacana berupa bacaan atau gambar. Tingkat kesulitannya disesuaikan dengan usia perkembangan berpikir siswa setingkat SMP (akhir) atau SMA (awal).

ULTRASONIK

Di beberapa negara, foto/gambar dari janin dalam kandungan dapat diambil dengan pencitraan ultrasonik (echografi). Ultrasonik dinilai aman bagi ibu dan janin dalam kandungannya.



Dokter memegang peralatan ultrasonik dan menggerakkannya melintasi perut ibu hamil tersebut. Gelombang ultrasonik dikirimkan ke dalam perut ibu hamil, di dalam perut gelombang tersebut dipantulkan oleh permukaan janin dalam kandungan. Gelombang pantulan ini ditangkap kembali oleh peralatan ultrasonik yang kini berperan sebagai sensor dan kemudian dikirimkan ke peralatan pemroses lebih lanjut yang dapat menghasilkan gambar.

1. Untuk membentuk sebuah gambar, mesin pencitra ultrasonik perlu mengukur jarak antara janin dan peralatan ultrasonik. Gelombang ultrasonik merambat melalui perut ibu hamil dengan kecepatan 1540 m/s. Apakah yang harus diukur oleh mesin tersebut untuk dapat menghitung jarak?
2. Gambar dari janin dalam kandungan dapat juga dihasilkan dengan menggunakan sinar-X. Tetapi, ibu hamil disarankan untuk menghindari penyinaran sinar-X pada kandungannya. Mengapa seorang ibu hamil perlu menghindari penyinaran sinar-X di perutnya selama proses kehamilannya?

(PISA 2009)

Gambar 8.3 Contoh Soal Literasi Sains untuk SMP

C. Pengembangan Instrumen Literasi Sains SMA

Literasi sains untuk SMA tentunya berbeda dengan literasi sains SD dan SMP. Pada tingkat SD dan SMP Sains diajarkan secara tematik dan terpadu. Materi yang dibahas meliputi Fisika, Kimia dan Biologi dengan proporsi yang berbeda. Pada tingkat SMA, sains diajarkan per bidang studi meliputi Fisika, Kimia, dan Biologi yang diajarkan secara terpisah. Namun demikian, komponen dan indikator literasi sains yang dapat digunakan untuk mengevaluasi literasi sains relatif sama dengan literasi pada umumnya. Pada hakikatnya, literasi sains mencakup pengetahuan, proses, dan produk. Ranah yang diukur meliputi kognitif, afektif dan psikomotorik.

1. Pengembangan Instrumen Literasi Sains Mahasiswa Calon Guru

Pengukuran terhadap literasi sains calon guru perlu dilakukan agar dapat memetakan kemampuannya dalam mengajarkan sains, khususnya literasi sains. Guru harus paham terhadap tuntutan literasi sains agar dapat mengajarkan sains dengan benar, sehingga mampu mengantar anak didiknya bersaing dengan negara lain. Berikut dicontohkan pengembangan instrumen literasi sains untuk mahasiswa calon guru. Pengembangan ini dilakukan melalui kegiatan penelitian dengan subjek uji coba mahasiswa S1 Pendidikan IPA dan Pendidikan Fisika sebagai calon guru IPA SMP dan atau guru Fisika SMA/SMK. Langkah pengembangan meliputi perencanaan, validasi pakar, uji coba terbatas, dan uji coba skala luas.

a. Perencanaan Pengembangan Instrumen Literasi Sains

Rancangan desain soal dapat berbentuk pilihan ganda atau esai. Kisi-kisi dibuat berdasarkan muatan aspek literasi sains PISA. Aspek yang diukur dalam instrumen literasi sains meliputi pengetahuan (*knowledge*) dan kompetensi/proses. Indikator setiap aspek dapat dilihat pada Tabel 8.1.

Tabel 8.1. Indikator Literasi Sains Berdasarkan PISA

Aspek	Indikator
Pengetahuan	Konten
	Prosedural
	Epistemik
Proses	Menjelaskan fenomena ilmiah
	Mengevaluasi/merancang penyelidikan ilmiah
	Menafsirkan data dan bukti ilmiah

b. Validitas Instrumen Literasi Sains

Validasi terhadap instrumen yang dikembangkan dilakukan oleh ahli materi dan ahli evaluasi. Pada penelitian Rusilowati, dkk (2018) validatornya sebanyak tiga orang. Penelitian dilakukan di Indonesia dan Malaysia, sehingga instrumen yang dibuat dalam bahasa Indonesia juga diterjemahkan ke bahasa Melayu. Validasi dari para pakar meliputi aspek materi, konstruksi dan bahasa. Skala penilaian yang digunakan adalah Gutmann, yaitu pada rentang 0-1, jumlah butir 18. Kriteria penilaian dibagi dalam 5 kategori sebagai berikut:

Interval Skor	Kategori
$13,5 < X \leq 18$	Sangat Valid
$9,0 < X \leq 13,5$	Vaid
$4,5 < X \leq 9,0$	Cukup Valid
$0 < X \leq 4,5$	Tidak Valid

Hasil validasi pakar terhadap instrumen yang dikembangkan disajikan pada Tabel 8.2.

Tabel 8.2. Hasil Validasi Instrumen Literasi Sains

No	Aspek	Rata-Rata (%)	Keterangan
1	Materi	13	Valid
2	Konstruksi	13	Valid
3	Bahasa	17	Sangat Valid

c. Reliabilitas Instrumen Literasi Sains yang Dikembangkan

Hasil analisis menggunakan rumus KR-20 diperoleh besarnya koefisien reliabilitas tes sebesar 0,530. Secara statistik instrumen yang dikembangkan reliabel dengan kategori sedang. Berdasarkan hasil uji coba ini disusun instrumen baru dengan menghilangkan soal-soal yang berdaya beda buruk. Hasil uji coba berikutnya diperoleh koefisien reliabilitas instrumen sebesar 0,71 pada kategori reliabel. Hal ini dapat diartikan bahwa instrumen layak untuk digunakan.

d. Karakteristik Butir Soal Literasi Sains

Hasil analisis terhadap karakteristik butir soal literasi sains meliputi taraf kesukaran dan daya pembeda.

1) Taraf Kesukaran

Taraf kesukaran soal diklasifikasikan menjadi soal yang mudah, sedang, dan sukar. Analisis menggunakan rumus probabilitas menjawab benar. Hasil analisis taraf kesukaran disajikan pada Tabel 8.3.

Tabel 8.3. Taraf Kesukaran Butir Soal Instrumen Literasi Sains

Kategori Taraf Kesukaran	Nomor Butir	Jumlah Butir
Mudah	1, 3, 5, 16, 20, 21, 22, 23, 25, 26, 29, 30, 32, 35	14
Sedang	2, 7, 9, 13, 15, 18, 24, 27, 31, 34, 36, 38, 40, 41, 42, 45, 46, 47	18
Sukar	4, 6, 8, 10, 11, 12, 14, 17, 19, 28, 33, 37, 39, 43, 44, 48, 49, 50	18

Rata-rata indeks taraf kesukaran instrumen literasi sains yang dikembangkan adalah 0,47 pada kategori sedang.

2) Daya Pembeda

Daya pembeda butir soal dianalisis dengan rumus deskriminasi, diklasifikasikan menjadi soal diterima, diterima dengan sedikit revisi, diperbaiki/direvisi, dan ditolak. Hasil analisis daya pembeda butir soal disajikan pada Tabel 8.4.


Tabel 8.4. Daya Pembeda Butir Soal Instrumen Literasi Sains

Kategori Daya Pembeda Soal	Nomor Butir	Jumlah Butir
Diterima	21, 34, 38, 41, 42, 43, 44, 45	8
Diterima dengan sedikit perbaikan	11, 22, 27, 31, 46, 47, 48, 49	8
Diperbaiki	10, 13, 14, 15, 16, 17, 25, 30, 36, 37, 39, 40, 50	13
Ditolak	1-9, 12, 18, 19, 20, 23, 24, 26, 28, 29, 32, 33, 35	21

Rata-rata indeks daya beda instrumen literasi sains yang dikembangkan adalah 0,28 pada kategori buruk atau banyak yang harus diperbaiki. Setelah dilakukan perbaikan, diperoleh 30 soal yang berdaya beda baik pada kategori diterima.

Contoh bentuk soal literasi sains untuk mahasiswa calon guru dapat dilihat pada Gambar 8.4.

Wacana 1



Seorang astronot asal Indonesia bernama Rizman akan mewakili Indonesia untuk pergi ke bulan. Sebelum berangkat ke bulan, ia menimbang tubuhnya beserta seragamnya. Skala pada timbangan menunjukkan angka 120. Saat berada di bulan ia menggunakan timbangan yang sama, menimbang kembali tubuhnya. Gravitasi di bulan yaitu $\frac{1}{6}$ kali gravitasi bumi.

Wacana 1 digunakan untuk menjawab soal nomor 1, 2, 3.

1. Pernyataan yang tepat berdasarkan wacana di atas adalah...
 - a. berat tubuh Rizman saat berada di bumi adalah 120 kg
 - b. massa tubuh Rizman saat berada di bumi adalah 120 N
 - c. gaya berat tubuh Rizman saat berada di bumi adalah 120 kg
 - d. massa tubuh Rizman saat berada di bumi adalah 120 kg
 - e. massa dan berat tubuh Rizman adalah 120 kg
2. Massa tubuh Rizman saat berada di bulan adalah
 - a. 12 kg
 - b. 20 kg
 - c. 20 N
 - d. 120 kg
 - e. 120 N
3.

Gambar 8.4. Bentuk Soal Literasi Sains untuk Calon Guru

Pada contoh soal tersebut tampak bahwa informasi yang ada di dalam bacaan tidak menunjukkan secara eksplisit tentang massa dan berat badan. Ketika mahasiswa tidak memahami tentang konsep massa dan berat, maka tidak dapat menjawab soal dengan benar. Kedua soal tersebut merupakan soal untuk mengukur aspek **pengetahuan** pada indikator **konten**. Hasil pengembangan kisi-kisi soal dan tes berbasis literasi sains dapat dilihat di lampiran.

DAFTAR PUSTAKA

- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*. 45, 131-142.
- American Association for the Advancement of Science [AAAS]. 1993. *Benchmarks for science literacy: A Project 2061 report*. New York: Oxford University Press.
- Aryani, A. K., H. Suwono., Parno. 2016. Profil Kemampuan Literasi Sains Siswa SMPN 3 Batu. *Prosiding Seminar Nasional Pendidikan IPA Pascasarjana UM*. 1, 847-855.
- Azwar, S. (2012). *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar.
- Bell, R. & Lederman, N. 2003. Understandings of The Nature of Science and Decision Making on Science and Technology based Issues. *Science Education*, 87(3): 352-377.
- Boujaude. (2002) Balance of scientific literacy themes in science curricula: the case of Lebanon. *International Journal of Science Education* 24(2), 139-156.
- Bybee,R., McCrae, B. & Laurie, R. (2009). PISA 2006: An Assessment of Scientific Literacy. *Journal of Reseach in Science Teaching* 46(8), 865-883.
- Bybee, R. (2008).Scientific Environmental Issue and PISA 2006: The 2008 Paul-F Brand wein Lecture. *Journal of Science Education and Technology* 17. 566-583.
- Bybee, R. (2009). PISA 2006 and Scientific Literacy: A Perspective for Science Education Leader . *Science Educator* 18 (2).
- Bybee, R. (2010). *Teaching Science: 21st century perspective* .USA: National Science Teacher Association Press.
- Bybee, R. (2011). *PISA 2006 : Implication for science teaching and learning*. USA: National Science Teacher Association Press.

- Chiappetta, E.L, Fillman, D.A, dan Sethna, G.H. 1991. A Method to Quantify Major Themes of Scientific Literacy in Science Textbooks. *Journal of Research in Science Teaching*, 28 (8): 713-725.
- Chiappetta, E.L, Fillman, D.A, dan Sethna, G.H. 1993. Do Middle School Life Science Textbooks Provide a Balance of Scientific Literacy Themes?. *Journal of Research in Science Teaching*, 30 (2): 787-797.
- Chi-lau, K. (2009). A critical examination of PISA 's Assesment on scientific literacy. *International Journal of Mathematics and Science Education* 7, 1061-1088.
- Dani, D. 2009. Scientific Literacy and Purposes for Teaching Science: A Case Study of Lebanese Private School Teachers. *International Journal of Environmental & Science Education*, 4(3): 289-299.
- Djanuarsih, E. 2012. Validitas dan Reliabilitas Butir Soal. *E-Jurnal Dinas Pendidikan Kota Surabaya*, 1: 1-12.
- Fang, Z & Wei. 2010. Improving Middle School Student's Literacy Through Reading Infusion. *The Journal of Educational Research*, 103(4): 262-273.
- Fensham, P. J. (2009). Real world contexts in PISA science: implications for context-based science education. *Journal of Research in Science Teaching*, 46(8), 884-896.
- Fu'adah, H., Rusilowati, R. & Hartono. 2017. Pengembangan Alat Evaluasi Literasi Sains untuk Mengukur Kemampuan Literasi Sains Siswa Bertema Perpindahan Kalor dalam Kehidupan. *Jurnal Lembaran Ilmu Kependidikan*, 46(1): 51-59.
- Garson, D. (2006). *Structural equation modeling*. Diambil pada tanggal 20 Agustus 2006, dari <http://www2.chass.ncsu.edu/garson/pa765/structur.htm>.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariat data analysis (5th ed.)*. London: Prentice-Hall International.

- Holbrook, J & Rannikmae. (2007). The nature of science Education for enhancing scientific literacy. *International Journal of Science Education*, 29, (11)1347-1362.
- Husain, H., Baisb, B., Hussainb, A. & Samadb, S.A. 2012. How to Construct Open Ended Questions. *Procedia Social and Behavioral Sciences* 60:456-462.
- Lawshe, C.H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*. 28(4), 563-575.
- Mardapi, D. 2008. *Teknik Penyusunan an Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia.
- Mardhiyyah, L.A., Rusilowati, A. & Linuwih, S. 2016. Pengembangan Instrumen Asesmen Literasi Sains Tema Energi. *Journal of Primary Education*, 5(2): 147-154.
- Maturradiyah, N. & Rusilowati, A. (2014). Analisis Buku Fisika kelas XI yang Beredar di Pati Berdasarkan Literasi sains. *Makalah*. Seminar Nasional Pendidikan IPA Unnes tanggal 26 April 2014 (*Accepted*).
- Millar (2006) Twenty First century Science : Insight from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education* 28 (13), 1499-1521.
- NRC (National Research Council). (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- OECD. (2007). PISA 2006. Science competencies for tomorrow's world. Volume I: Analysis. Paris:OECD.
- OECD. (2009). *PISA 2006 technical report*. Paris: OECD.
- OECD. (2012). *PISA 2006 Result in Focus*. Paris: OECD.
- OECD. (2016). *PISA 2015 Result in Focus*. Paris: OECD.
- Osborne, J. (2007). Science Education for twenty first Century. *Eurasia Journal of Mathematics and Science Education* 3(3), 173-184.

- Ozdem, Y., Cavas, P., Cavas, B., Eylul, D., Cakiroglu, J., & Ertepinar H. 2010. *An Investigation Of Elementary Student's Scientific Literacy Levels*. Journal of Baltic Science Education, 9(1): 6-19.
- Plummer, D. M. & Kuhlman, W. 2008. Literacy and Science Connections in the Classroom. *Reading Horizons*, 48 (2): 95-110.
- Popham, W. J. (1995). *Classroom assessment: What teachers need to know*. USA: Allyn & Bacon.
- Purwantari, T. & Kartono. *Ilmu Pengetahuan 5 untuk Sekolah dasar & Madrasah Ibtidaiyah Kelas V*. Jakarta: Pusat Perbukuan, Departemen Pendidikan Nasional.
- Ridwan, M. S., Mardhiyyah, L.A., & Rusilowati, A. 2013. Pengembangan Instrumen Asesmen Dengan Pendekatan Kontekstual Untuk Mengukur Level Literasi Sains Siswa. *Prosiding Seminar Nasional Evaluasi Pendidikan Tahun 2013*, ISBN 978-602-14215-0-5.
- Ridwan, M. S. & Rusilowati, A. (2012). Literasi Sains Sebagai Kerangka Asesmen Pembelajaran Sains Abad 21. *Prosiding Seminar Nasional FMIPA Unesa Surabaya*.
- Rusilowati, A. (2012). *Pengembangan Tes Kemampuan Dasar Membaca Sains Berdasarkan Psikologi Kognitif*. Salatiga: Griya Media.
- Rusilowati, A. (2013). Peningkatan Literasi Sains Siswa Melalui Pengembangan Instrumen Penilaian. *Naskah Pidato Pengukuhan Profesor*. Semarang: Unnes.
- Rusilowati, A. (2014). Analisis Buku IPA yang Digunakan di Semarang berdasarkan Muatan Literasi Sains. *Makalah Seminar Nasional pada tanggal 22 Maret 2014 di Unnes Semarang*.

- Rusilowati, A., Sri Mulyani, E.S., Sunyoto, E.N. (2015). Developing of Science Textbook Based on Scientific Literacy for Seventh Grade of Secondary School. *Proceeding International Conference Mathematic Education and Science*. Semarang 5-6 September 2015.
- Rusilowati, A., Lina, K., Nugroho, S. E., & Widiyatmoko, A. (2016a). Scientific Literacy Evaluation Instrument to Measure The Student's Scientific Literacy Skill on The Cycle Theme. *International Journal of Environment al and Science Education*, 11 (12), 5718-5727.
- Rusilowati, A., Prabowo, H., & Nugroho, S. E. (2016b). Scientific Literacy Assessment Instrument For Measuring The Students' Scientific Literacy Capability Of Interaction Theme. *Proceeding ICMSE 2016. Seminar Internasional tanggal 3 September 2016*.
- Rusilowati, A. (2016). Pembelajaran Inovatif Mengantar Calon Guru Mencetak Generasi yang Literate. Dalam Zaenuri, Cahyono, E., Sudarmin, Kurniawan, C. (Eds). *Rumah Ilmu: Inovatif dan Membumi*. (pp 118-133). Semarang: FMIPA Unnes.
- Rusilowati, A., Nugroho, S. E., & Susilowati, S. M. (2016c). Development Of Science Textbook Based On Scientific Literacy For Secondary School. *Jurnal Pendidikan Fisika Indonesia*, 12(2), 98-105.
- Rusilowati, A., Nugroho, S. E., & Susilowati, S. M., Mustika, T., Harfiyani, N., Prabowo, T. H. (2017). The Development of Scientific Literacy Assessment to Measure Student's Scientific Literacy Skills in Energy Themed. *Makalah Seminar Internasional ICMSE4*. 18 September 2017.
- Shute, V. J. & Becker, B. J.. (2010). *Innovative Assessment for the 21st Century* (Eds). London: Springer Science+Business Media, tersedia dalam www.springer.com.

- Shwartz, Y., Ruth, B. & Hofstein, A. (2006). The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students. *Chemistry Education Research and Practice*, **7** (4), 203-225.
- Wibowo, M. E. (2012). Menyiapkan Bangkitnya Generasi Emas Indonesia. *Makalah Populer* di Universitas Muria Kudus.
- Wulandari, N. & Sholihin, H. (2016). Analisis Kemampuan Literasi Sains pada Aspek pengetahuan dan Kompetensi Sains Siswa SMP pada Materi Kalor. *Research Article Center for Science Education EDUSAINS*. **8** (1), 66-73.
- Yuliyanti, T. E. & Rusilowati, A. (2014). Analisis Buku Fisika SMA Kelas X yang Digunakan di Tegal Berdasarkan Literasi Sains. *Makalah*. Seminar Nasional Prndidikan IPA Unnes tanggal 26 April 2014 di Semarang.

GLOSARIUM

- Accuracy soal : keakuratan soal, biasanya dilihat dari kunci jawaban
- Asesmen : sekelompok pengertian yang mengacu pada pengumpulan data dan informasi untuk tujuan menjelaskan tingkat pengetahuan, penampilan/kinerja, dan prestasi dari individu.
- Daya pembeda : kemampuan suatu butir soal dapat membedakan antara siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan.
- Distraktor (Pengecoh) : Alternatif jawaban yang benar disebut kunci jawaban dan yang salah atau kurang benar disebut).
- Evaluasi : penilaian yang dikaitkan dengan keberhasilan suatu program, termasuk program pendidikan, dimana hasil evaluasi dapat digunakan untuk memberikan rekomendasi kepada pihak pengambil keputusan.
- Holistik : sebuah cara pandang terhadap sesuatu yang dilakukan dengan konsep engakuan bahwa hal keseluruhan merupakan sebuah kesatuan yang lebih penting daripada bagian-bagian yang membentuknya.
- Integratif : pembauran hingga menjadi kesatuan yang utuh atau bulat

- ITEMAN** : program komputer yang digunakan untuk menganalisis butir soal secara klasik
- Kisi-kisi tes** : matrik yang berisi spesifikasi soal, merupakan pedoman bagi pembuat soal
- Kuis** : sejumlah pertanyaan singkat (lisan atau tertulis) yang menanyakan hal-hal yang prinsip (pengetahuan dan pemahaman)
- Literasi sains**:kemampuan menggunakan konsep sains untuk mengaplikasikannya dalam kehidupan sehari-hari, menjelaskan fenomena ilmiah serta menggambarkan fenomena tersebut berdasarkan bukti-bukti ilmiah
- Pengukuran** : proses mendapatkan informasi dan biasanya diwujudkan dengan angka untuk individu atau menentukan karakteristik individu menurut aturan tertentu.
- Penilaian** : proses sistematis untuk mengumpulkan, menganalisis, menginterpretasikan informasi, dan menentukan tingkat keberhasilan siswa terhadap tujuan pembelajaran.
- Penilaian kelas**: suatu kegiatan yang dilakukan guru berkaitan dengan pengambilan keputusan tentang pencapaian kompetensi dasar setelah mengikuti proses pembelajaran
- Portofolio** : kumpulan dokumen seseorang, kelompok, lembaga, organisasi, perusahaan atau sejenisnya yang bertujuan untuk mendokumentasikan perkembangan suatu proses dalam mencapai tujuan yang telah ditetapkan

- Reliabilitas tes : keajegan skor yang diperoleh oleh para subjek yang diukur dengan alat yang sama
- Stem : butir soal bentuk pilihan ganda berupa satu pernyataan yang belum lengkap
- Tes : suatu alat atau prosedur yang dipakai dalam rangka kegiatan pengukuran dan penilaian
- Tes objektif : tes yang dibuat sedemikian rupa sehingga hasil tes tersebut dapat dinilai secara objektif, dinilai oleh siapapun akan menghasilkan skor yang sama
- Validitas tes : derajat kecermatan ukur suatu tes
- Verifikasi : proses menentukan kebenaran dari suatu pernyataan dengan menggunakan sebuah metode yang empirik

INDEKS

- Accuracy* 86,
- Aiken V 28,
- Adjusted goodness of fit index* 38,
- Akaike information criterion* 39,
- Alat Evaluasi 3
- Alat Penilaian 23,
- Alfa Cronbach 50,
- Analisis Butir 4
- Analisis kualitatif 14
- Analisis Kuantitatif 14
- Asesmen 1,6,25
- Atribut tes 81
- Autentik 7,
- Communicability* 86,
- Competency based assessment* 9,
- Comparative fit index* 39,
- Conformity* 86,
- Consistent Akaike information criterion* 39,
- Characteristics curve theory* 72,
- Chi square 38,
- Cut of value* 39,
- CVR (*Content Validity Ratio*) 33,
- Daya Pembeda 61,62,121
- Domain hakikat sains 97
- Domain personal 97
- Domain sosial 97,
- Eigenvalue* 36,
- EXCEL 71,
- Evaluasi 6
- Evaluasi hasil belajar 6
- Goodness of Fit Index* 38,
- Holistik 3
- Incremental fit index* 39,
- Item response theory* 72,

Integratif 3
Instrumen 4,11,14,15
ITEMAN 64,
Kisi-Kisi 10
Koefisien reliabilitas 43,
Kuis 15
Latent trait theory 72,
Literasi Sains 1,3,,4,98,122
Lifescience 116,
Model Rasch 76,
Non Tes 23,78
Normed fit index 39,
Observasi 15, 16, 23
Parsimonius goodness of fit index 39,
Pengembangan Tes 4
Pengukuran 6,7,19,26, 118
Penilaian 7
Proyek 15,
Portofolio 15, 18
Rater 29,
Reinforcement 17,
Relative fit index 39,
Rigth Tail Probability 31,
Root mean square error of approximation 38,
Science as a Body of knowledge 99
Science as Interaction of science, environment, technology and society 99,101,112
Science as a way of thinking 99, 100,112
Spearman-Brown 46,48
Suitability 86
Standar Evaluasi 4
Structural equation modeling 38,
Teknik Penilaian 15
Tes 6,10,16,78
Tes Objektif 19,93
Tes Lisan 24,
Tes standar 81,

Tes Tertulis 24,
Tes Uraian 22,90,
Test retest 43,
Tes paralel 43
Tingkat ketepatan 43,
Tingkat Kesukaran 59,121,
Tingkat konsistensi 43,
Tucker Lewis index 38,
Uji kecocokan absolut 38,
Uji kecocokan inkremental
38,
Uji kecocokan Parsimoni 39,
Validitas Instrumen
10,,36,119,
Validitas Tes 26,
Validitas Isi 27
Validitas Konstruk 27,35
Validitas Berdasar Kriteria
27,
Vaiditas muka 28
Validitas Logik 28,
Validitas konkuren 42,
Validitas Prediktif 42,
Way of investigating 99
Wawancara 15

BIODATA PENULIS



Ani Rusilowati lahir di kabupaten Semarang Jawa Tengah. Gelar Sarjana Pendidikan Fisika diperoleh tahun 1984 dari IKIP Semarang. Pada Tahun 1999 melanjutkan studi pendidikan S2 dan memperoleh gelar Magister Pendidikan bidang Evaluasi dari Universitas Negeri Yogyakarta. Gelar Doktor diperoleh dari Universitas yang sama pada tahun 2008. Menjadi staf pengajar IKIP Semarang (kini Universitas Negeri Semarang) sejak tahun 1985 hingga sekarang. Mata kuliah yang diampu adalah Evaluasi Pendidikan, Dasar-dasar Proses Pembelajaran dan Metode Penelitian. Sejak tahun 2013 mencapai jenjang Profesor bidang Evaluasi Pendidikan. Beberapa penelitian bertema pembelajaran dan evaluasi telah dilakukan dan beberapa judul artikel telah disajikan pada jurnal dan seminar baik nasional maupun internasional.



Budi Astuti lahir di kabupaten Jawa Tengah. Ia menyelesaikan pendidikan sekolah tingkat dasar, tingkat menengah pertama dan atas diselesaikan di Kabupaten Banjarnegara tepatnya di MI Cokroaminoto Wanakarsa, SMPN 2 Wanadadi dan SMA N 1 Bawang. Pada tahun 2001, Ia meraih gelar Sarjana Pendidikan Fisika di Universitas Negeri Semarang. Pada tahun 2003, Ia melanjutkan studi S2 dan memperoleh gelar Magister Fisika Material dari Universiti Teknologi Malaysia dua tahun kemudian. Gelar Doktor diperoleh dari Universitas yang sama pada tahun 2013 pada bidang Fisika Material Elektronik. Sejak tahun 2005 sampai sekarang, Ia menjadi staf pengajar di Universitas Negeri Semarang. Selain bidang Fisika

Material yang ditekuni, Ia juga berkolaborasi dengan pakar pendidikan sebagai kontribusi nyata sebagai staf pengajar di Jurusan Fisika FMIPA Unnes. Hal tersebut karena Jurusan Fisika FMIPA Unnes memiliki dua prodi yaitu Pendidikan Fisika dan Fisika. Hasil kolaborasi dan penelitian dibidang pendidikan maupun non pendidikan telah dipublikasikan melalui jurnal maupun seminar, baik nasional maupun Internasional.