

ASESMEN LITERASI SAINS: ANALISIS KARAKTERISTIK INSTRUMEN DAN KEMAMPUAN SISWA MENGGUNAKAN TEORI TES MODERN *RASCH MODEL*

Ani Rusilowati*[¶]

Fisika, FMIPA, Universitas Negeri Semarang

*Email korespondensi: rusilowati@mail.unnes.ac.id

ABSTRACT

The low ability of the students in sciences dicipline, especially scientific literacy, is a concern for the government to improve it through curriculum changes. Strengthening the assessment standard is main role in measuring scientific literacy skills. Development of scientific literacy instruments are viewed as solution in training the students to think in accordance of the required aspect for scientific literacy. The aspect measured in the assessments of scientific literacy are: the knowledge of science, the investigative nature of science, science as a way of thinking, and interaction of science, environment, technology, and society. The developed instruments must satisfy with valid and reliable criteria, and have a level of difficulty and differentiation of problems appropriately. One way to analyse the instrument characteristic is by item order test of 1 logistic parameter (1-LP) theory, known well as the Rasch model.

Keywords: assesment, scientific literacy, instrument characteristics, Rasch model

ABSTRAK

Lemahnya kemampuan siswa dalam bidang sains, khususnya literasi sains, menjadi perhatian bagi pemerintah untuk meningkatkannya melalui perubahan kurikulum. Penguatan standar penilaian menjadi kunci utama dalam mengukur kemampuan literasi sains. Pengembangan instrumen literasi sains dipandang menjadi salah satu solusi dalam membiasakan siswa untuk berpikir sesuai dengan aspek yang dituntut dalam literasi sains. Aspek yang diukur dalam asesmen literasi sains mencakup: the knowledge of science (pengetahuan sains), the Investigative nature of science (penyelidikan tentang hakikat sains), science as a way of thinking (sains sebagai cara berfikir), dan interaction of science, environment, technology, and society (interaksi antara sains, lingkungan, teknologi dan masyarakat). Instrumen yang dikembangkan harus memenuhi kriteria valid dan reliabel serta memiliki tingkat kesukaran dan daya pembeda soal yang memadai. Salah satu cara untuk menganalisis karakteristik instrumen adalah dengan teori tes modern 1 parameter logistik (1-PL), yang dikenal dengan Rasch model.

Kata kunci: asesmen, literasi sains, karakteristik instrumen, Rasch model

[¶] Keynote speaker

PENDAHULUAN

Literasi sains merupakan fokus utama dalam pembelajaran abad 21. Sains memiliki tiga komponen yang tidak dapat dipisahkan, yaitu produk, proses ilmiah, dan sikap ilmiah. Oleh sebab itu belajar sains adalah belajar produk, proses, dan sikap. Sains sebagai produk memiliki makna pengorganisasian fakta, konsep, prosedur, prinsip, dan hukum-hukum alam. Sains sebagai proses menjelaskan bahwa temuan sains diperoleh dari proses ilmiah atau kerja ilmiah. Sains sebagai sikap memiliki makna bahwa sikap ilmiah mendasari proses ilmiah yang berguna dalam menghasilkan produk sains. Komponen tersebut menjadi tolok ukur kemampuan literasi sains siswa.

Kondisi saat ini menunjukkan lemahnya kemampuan siswa dalam bidang sains, khususnya literasi sains, dibuktikan oleh hasil asesmen literasi 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.

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

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 buku ajar berbasis literasi sains atau pun soal kemampuan literasi sains sebagai alat evaluasinya. Buku ajar IPA yang digunakan di sekolah-sekolah pun tidak ada yang memuat soal literasi sains seperti 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.

Tabel 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

Hasil wawancara dengan beberapa guru SMP di Jawa Tengah, Sumatera Utara, dan Sulawesi Selatan dan beberapa dosen LPTK bidang Sains menunjukkan bahwa: (1) masih banyak guru yang tidak mengenal istilah literasi sains, (2) mereka tidak mengetahui

program evaluasi terhadap kemampuan sliterasi sains siswa yang dilakukan oleh pihak internasional seperti TIMSS dan PISA, (3) di antara 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. Salah satu cara yang dapat dilakukan untuk meningkatkan kemampuan literasi sains siswa adalah menyediakan instrumen literasi sains yang valid dan reliabel. Instrumen yang sudah valid dapat digunakan untuk menilai hasil pendidikan.

Penilaian pendidikan merupakan cara menempatkan pembelajar dalam konteks yang dapat menyatakan apa yang diketahui, dipahami, dan dilakukan. Penilaian merupakan bukti akuntabilitas kegiatan pengembangan sumber daya manusia di lingkup pendidikan. Teori yang digunakan untuk menganalisis instrumen (tes) dan hasil penilaian, termasuk ujian nasional, selama ini menggunakan teori tes klasik. Tahun 2018 ini pemerintah mencoba menerapkan teori tes modern atau *item response theory* (IRT) sebagai cara menilai hasil ujian nasional. Kemdikbud mengeluarkan indeks integritas bagi sekolah-sekolah yang menyelenggarakan ujian akhir nasional. Analisis dengan IRT ini membantu guru dalam menentukan mana siswa yang memiliki kemampuan tinggi, tidak hati-hati/*careless*, menebak, menyontek/bekerja sama. Kelebihan penggunaan IRT adalah kemampuan siswa tidak bergantung pada butir soal dan tingkat kesulitan butir soal tidak bergantung pada kemampuan siswa. Ada tiga jenis model IRT yaitu 1-PL, 2-PL, dan 3-PL. Pada kesempatan ini akan dibahas tentang model 1-PL atau *Rasch Model*.

ASESMEN LITERASI SAINS

Fokus yang dipentingkan dalam pendidikan sains sekarang adalah literasi sains [1]. Literasi sains merupakan kemampuan siswa menggunakan konsep

sains untuk mengaplikasikannya dalam kehidupan sehari-hari, menjelaskan fenomena ilmiah serta menggambarkan fenomena tersebut berdasarkan bukti-bukti ilmiah [2-8]. Literasi sains menjadi tujuan kurikulum pendidikan sains abad 21 di berbagai negara, hal ini diperkuat lewat program TFCS (*Twenty First Century Science*) di Inggris yang bertujuan untuk mengembangkan kurikulum pendidikan sains [9]. Proyek ini mengembangkan asesmen literasi sains siswa usia 15-16 tahun. Beberapa topik atau bahasan yang dikembangkan bersifat kontekstual, mudah ditemukan siswa dalam kehidupan sehari-hari. Contohnya kualitas udara, radiasi dalam kehidupan, dan lain-lain.

National Science Education Standard [10] menegaskan bahwa Literasi sains didefinisikan sebagai kemampuan mengaplikasikan sains dalam konteks kehidupan sehari-hari (*real life context*). *Science for all American* [1] 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. 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 [11]. Berkaitan dengan literasi sains, beberapa pakar sains membuat tingkat literasi sains dengan alasan literasi sains yang didefinisikan terlalu banyak dan kompleks untuk seseorang. Shwartz *et al* [3] menambahkan bahwa literasi sains ini berbeda untuk setiap orang. Hal ini disebabkan oleh beberapa faktor seperti umur, pengalaman, pengetahuan dan lingkungan. Oleh sebab itu, pada kesempatan ini kemampuan literasi yang akan dikembangkan mencakup empat aspek,

yaitu konteks sains, kompetensi sains, pengetahuan sains, dan sikap terhadap sains.

Literasi sains menurut PISA 2006 diartikan sebagai kemampuan menggunakan pengetahuan sains untuk mengidentifikasi permasalahan dan menarik kesimpulan berdasarkan bukti-bukti dalam rangka memahami serta membuat keputusan tentang sains dan perubahan yang dilakukan terhadap sains melalui aktivitas manusia [2,12,13]. Definisi literasi sains ini memandang literasi sains bersifat multidimensional, bukan hanya pemahaman terhadap pengetahuan sains, melainkan lebih dari itu. PISA juga menilai pemahaman siswa terhadap karakteristik sains sebagai penyelidikan ilmiah, kesadaran akan betapa sains dan teknologi membentuk lingkungan material, intelektual dan budaya, serta keinginan untuk terlibat dalam isu-isu terkait sains, sebagai manusia yang reflektif.

Literasi sains dapat juga diartikan sebagai kemampuan dalam menerapkan pengetahuan dan keterampilan serta untuk menganalisis, bernalar dan berkomunikasi secara efektif apabila dihadapkan pada suatu masalah dan harus menyelesaikan serta menginterpretasi masalah pada berbagai situasi.

ASPEK LITERASI SAINS

Asesmen literasi sains dibangun oleh beberapa aspek yang merupakan hakikat dari sains itu sendiri. Aspek yang diukur dalam asesmen literasi sains mencakup *the knowledge of science* (pengetahuan sains), *the Investigative nature of science* (penyelidikan tentang hakikat sains), *science as a way of thinking* (sains sebagai cara berfikir), dan *interaction of science, technology, and society* (interaksi antara sains, teknologi dan masyarakat) [14,15]. Sesuai dengan tuntutan keterampilan abad 21, Rusilowati, dkk. [16] menambahkan indikator lingkungan pada aspek interaksi. Aspek-aspek ini relevan dengan tuntutan literasi sains oleh PISA 2012. Berikut uraian tiap aspek literasi sains.

Pengetahuan sains

Aspek ini 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 mencerminkan pemindahan pengetahuan sains ketika siswa menerima informasi. Kategori ini merupakan ciri dari materi yang harus dipelajari. Indikator kemampuan literasi sains pada kategori ini adalah:

- (a) Menjelaskan fakta-fakta, konsep-konsep, prinsip-prinsip dan hukum-hukum.
- (b) Menyajikan hipotesis-hipotesis, teori-teori dan model-model.
- (c) Menjawab pertanyaan terkait dengan pengetahuan atau informasi sains.

Penyelidikan tentang hakikat sains

Aspek ini dimaksudkan untuk merangsang kemampuan berpikir dan melakukan sesuatu dengan menugaskan kepada siswa untuk “menyelidiki”. Hal ini mencerminkan aspek inkuiri dan pembelajaran aktif, yang melibatkan siswa dalam metode dan proses sains seperti mengamati, mengukur, melakukan klasifikasi, menarik kesimpulan, mencatat data, melakukan perhitungan, melakukan percobaan, dan sebagainya. Pembelajarannya dapat menyangkut kegiatan “*hands-on*”. Indikator kemampuan siswa pada kategori ini adalah:

- (a) Menjawab pertanyaan kegiatan praktikum melalui penggunaan materi.
- (b) Menjawab pertanyaan melalui penggunaan grafik, tabel, dan sebagainya.
- (c) Membuat kalkulasi/perhitungan.
- (d) Menerangkan langkah prosedural
- (e) Melakukan kegiatan eksperimen atau aktivitas berfikir.

Sains sebagai cara berfikir

Aspek ini dapat diidentifikasi dari teks dalam buku sains berupa ilustrasi/gambaran sains secara umum dan ilmuwan dalam melakukan penyelidikan. Aspek sains ini mewakili proses berpikir, penalaran, dan refleksi ketika siswa terlibat dalam kegiatan ilmiah. Indikator pada kategori ini adalah:

- (a) Menggambarkan bagaimana seorang ilmuwan melakukan eksperimen.
- (b) Menunjukkan kemampuan penalaran induktif dan deduktif.
- (c) Menganalisis hubungan sebab dan akibat.
- (d) Menyajikan fakta dan bukti.
- (e) Menyajikan metode ilmiah dan pemecahan masalah.

Interaksi antara Sains, Lingkungan, Teknologi dan Masyarakat

Aspek ini dimaksudkan untuk memberi gambaran tentang pengaruh atau dampak sains terhadap lingkungan dan masyarakat. Aspek literasi sains ini berkaitan dengan aplikasi sains dan teknologi terhadap manusia baik yang berdampak positif maupun negatif. Meskipun demikian, siswa menerima informasi tersebut dan umumnya tidak harus menyelidikinya. Indikator yang dapat diungkap pada kategori ini adalah:

- (a) Menggambarkan kegunaan ilmu sains dan teknologi bagi masyarakat.
- (b) Menunjukkan kepedulian terhadap lingkungan akibat penerapan sains dan teknologi.
- (c) Menunjukkan efek negatif dari ilmu sains dan teknologi bagi masyarakat.
- (d) Mendiskusikan masalah-masalah sosial yang berkaitan dengan ilmu sains atau teknologi, dan
- (e) Menyebutkan karir-karir dan pekerjaan di bidang sains dan teknologi.

INSTRUMEN LITERASI SAINS

Instrumen yang digunakan dalam penilaian pendidikan mencakup tes dan notes. 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 terbanyak pada tingkat pemahaman, aplikasi, dan analisis. Namun hal ini tergantung pada karakteristik bidang studi. Bentuk-bentuk tes yang dapat dikembangkan adalah tes objektif dan uraian.

Tes objektif merupakan tes yang dibuat sedemikian rupa sehingga hasil tes itu 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: soal bentuk melengkapi (*completion*), salah-benar (*true-false*), pilihan ganda (*multiple choice*), soal bentuk klasifikasi atau penggolongan (*classification*), pasangan/menjodohkan (*matching*), sebab-akibat, dan asosiasi pilihan ganda. Untuk soal literasi sains biasanya berbentuk pilihan ganda.

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 Sains. 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 dari pada 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.

Tes berikutnya adalah unjuk kerja yang 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.

Teknik penilaian nontes dapat berupa kegiatan observasi, pemberian kuesioner/skala, melalukan wawancara, pemberian proyek, penerapan portofolio. Dengan demikian, alat penilaian yang sesuai dengan teknik penilaian tersebut adalah: panduan observasi, kuesioner/skala, panduan wawancara, dan 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.

Literasi sains dapat diukur dengan tes objektif (pilihan ganda), tes uraian, tes sikap dan keterampilan. Pengembangan instrumen tes literasi sains bertema Siklustelah dikembangkan oleh Rusilowati, dkk [17] dan tema Interaksi [18].

ANALISIS TERHADAP KARAKTERISTIK INSTRUMEN

Selama ini di Indonesia menggunakan teori tes klasik untuk menentukan kualitas tes dan karakteristik butir soal. Pada tahun 2018 baru dicoba menggunakan teori tes

modern. Pada teori tes klasik, penentuan koefisien reliabilitas tes dapat menggunakan rumus Alpha Cronbach, Kuder Richardson (KR) 20 dan 21, Sperman Brown, dll. Karakteristik butir untuk tingkat kesukaran menggunakan probabilitas menjawab benar responden, dan daya beda butir menggunakan diskriminasi antara probabilitas menjawab benar kelompok atas dan kelompok bawah. Teori tes klasik memiliki kelemahan, yaitu ketergantungan tingkat kesukaran butir soal kepada responden/*testee* yang menjawab tes tersebut. Ketika suatu tes diberikan kepada responden yang pandai, maka tes terasa mudah. Tetapi, bila tes tersebut diberikan kepada responden yang kurang pandai, maka tes terasa sulit. Untuk mengatasi kelemahan tersebut, dikembangkannya teori tes modern.

TEORI TES MODERN (ITEM RESPONSE THEORY/IRT)

Ketika analisis terhadap tes menggunakan IRT, ada asumsi-asumsi yang harus dipenuhi. Asumsi yang mendasari IRT menurut [19,20] meliputi *unidimensional*, *independensi local*, dan *invarians parameter*. Respons terhadap soal dari suatu tes didasari oleh sifat-sifat laten yang jumlahnya tidak lebih dari soal tes tersebut, bahkan dalam berbagai penggunaannya sifat laten ini tunggal (unidimensi). Asumsi independensi lokal menunjukkan unjuk kerja seseorang pada suatu soal tidak mempengaruhi unjuk kerjanya pada soal lain. Artinya apabila kemampuan seseorang konstan, maka respons terhadap soal manapun akan bebas secara statistik. Asumsi invarian parameter menyatakan bahwa karakteristik soal adalah tetap meskipun subjek yang menjawab soal tersebut berubah-ubah. Hasil analisis dengan IRT dapat berupa kurva karakteristik soal, parameter soal dan kemampuan peserta tes. Penganalisisan dapat menggunakan program, di antaranya Winsteps, Bilog, dll. Pemaknaan hasil analisis dengan program

biasanya disertai dengan pentransformasian ke bentuk lain yang lebih mudah diterima oleh orang awam. Program yang digunakan juga dapat mendeteksi kecocokan model analisis yang dipilih dengan jawaban peserta tes.

a. Kurva Karakteristik Soal

Ciri utama dari IRT adalah bahwa respons peserta tes terhadap suatu soal yang diujikan akan membentuk sebuah kurva yang disebut kurva karakteristik soal. Kurva ini merupakan formula fungsi dari matematika yang menyatakan hubungan antara peluang menjawab benar $P(\theta)$ dengan kemampuan (θ) [19,20]. Kurva ini berbentuk ogive, yaitu kurva frekuensi kumulatif, sehingga kurva ini disebut kurva monotonik naik. Kurva ini merupakan kurva logistik yang dinyatakan dengan $L(0, 1,7)$, hampir sama dengan kurva normal kumulatif $N(0, 1)$. Selisih taksiran nilai parameter P dari kedua kurva tidak lebih dari 0,01. Jadi hasil taksiran P dari kedua kurva tidak berbeda jauh.

Kurva karakteristik soal dinyatakan dalam fungsi matematis dengan beberapa model, yaitu model satu parameter logistik (1-PL), dua parameter logistik (2-PL), dan tiga parameter logistik (3-PL). Penelitian ini menggunakan model 1-PL. Model ini digunakan dengan asumsi bahwa peserta tidak melakukan penebakan (*guessing*). Model ini cocok untuk data skor soal dikotomus. Persamaan yang digunakan dalam 1-PL adalah sebagai berikut [19,21]:

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

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

$D = 1,7$

$e = 2,718$

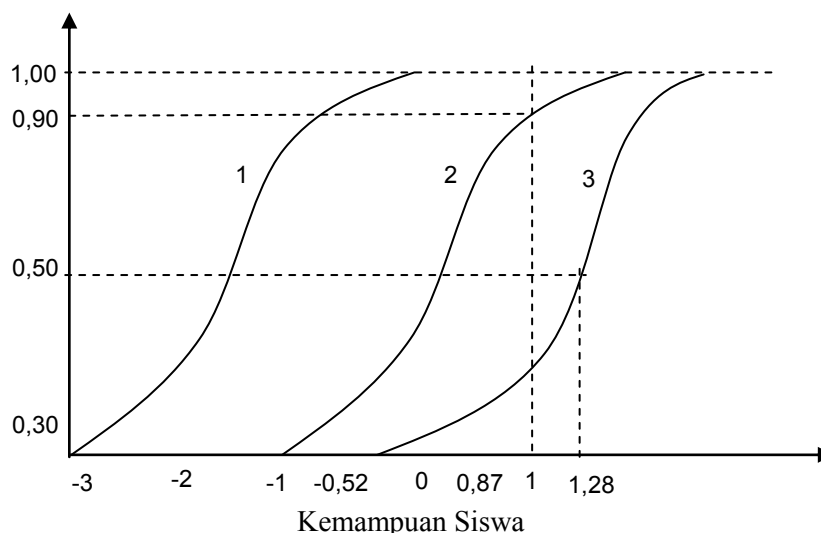
a_i = parameter daya beda soal ke i (*slope*)

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

$i = 1, 2, 3, \dots, n$; n = jumlah soal dalam tes

Bentuk kurva untuk model 1-PL dapat dilihat pada Gambar 1.

Peluang menjawab benar



Gambar 1. Kurva Karakteristik Soal Model 1-PL

b. Parameter Soal dan Kemampuan Peserta

Fungsi karakteristik soal merupakan fungsi dari parameter tingkat kesulitan soal (b_i), daya pembeda soal (a_i) dan parameter kemampuan peserta (θ). Lengkungan kurva

ditentukan oleh parameter soal dan parameter peserta yang merespon soal tersebut. Oleh karena kurva karakteristik soal merupakan kurva monotonik naik, semakin tinggi kemampuan (θ), maka peluang menjawab benar suatu soal semakin besar. Menurut

[10] karakteristik soal adalah tetap dan tidak berubah, sekalipun kelompok peserta yang menjawab soal tersebut berubah-ubah. Begitu pula jika kelompok peserta memiliki ciri sama, karakteristik soal akan tetap, meskipun soal yang dijawab berbeda-beda. Hal ini berarti bahwa kelengkungan respon peserta atau kurva karakteristik soal tidak akan berubah meskipun itu diberikan pada beberapa kelompok peserta dengan distribusi kemampuan berbeda [22].

Skala parameter kemampuan peserta adalah $-\infty < \theta < +\infty$. Biasanya, skala kemampuan peserta dipilih skala $-3 < \theta < +3$. Pada distribusi normal, *range* ini sudah mencakup 99,7% luas daerah di bawah kurva normal. Daya beda soal (a) secara teoretis menurut [19] memiliki skala $(-\infty, +\infty)$. Namun demikian, daya beda yang baik memiliki skala pada interval $(0, 2)$. Daya beda $a = 0$ tidak baik, karena kondisi ini menyebabkan $P_i(\theta)$ menjadi konstan (membentuk garis lurus). Daya beda negatif menyebabkan kemiringan daya beda negatif, sehingga kurva yang terbentuk bukan monotonik naik. Oleh karena itu, soal dengan daya beda negatif merupakan soal yang tidak baik. Kemiringan daya beda merupakan kemiringan garis singgung pada kurva karakteristik soal di titik infleksi (titik belok) yaitu titik pada kurva yang menyebabkan kurva berubah arah.

Tingkat kesukaran soal (b) memiliki skala $-\infty < b < +\infty$. Menurut [19] indeks kesulitan yang baik adalah $-2 < b < 2$, dengan asumsi bahwa kemampuan peserta adalah $-3 < \theta < 3$. Indeks kesulitan soal mendekati -2 merupakan soal mudah, sedangkan indeks kesulitan yang mendekati 2 merupakan soal sukar.

Hasil penentuan parameter butir dan kemampuan peserta dengan tanda negatif, berskala *logarithm odd unit* (logit) biasanya tidak dapat diterima oleh orang awam. Untuk menghilangkan tanda negatif, dan pecahan perlu dilakukan transformasi ke bentuk lain, Salah satu cara untuk

mentransformasinya menggunakan transformasi WITs.

c. Transformasi Skala

Hasil tingkat kesukaran soal dan kemampuan siswa dengan model logistik berupa skala logit. Bentuk skala ini belum dapat dipahami dengan mudah oleh masyarakat. Oleh sebab itu, skala logit perlu diubah ke dalam bentuk lain. Bentuk ubahan itu dapat dicapai melalui transformasi, baik linier maupun nonlinier. Field [23] menyatakan bahwa apapun bilangan yang digunakan untuk mentransformasi data riil (statistik) tidak akan mengubah substansi dasar dari sebaran data, seperti varian dan tendensi sentral. Jadi model transformasi apapun tidak akan mengubah substansi dasar. Oleh sebab itu, pada penelitian ini dipilih transformasi linier model WITs. Model ini menggunakan logaritma e pada basis 9 yaitu sebesar 0,455. Transformasi WITs menggunakan skala logit setara dengan 45,5 skala WITs dengan mean 500. Wright & Stone [24] merumuskan transformasi WITs sebagai berikut:

$$B = \alpha + \gamma b \quad (\text{untuk kemampuan peserta}) \quad (2)$$

$$D = \alpha + \gamma d \quad (\text{untuk karakteristik soal}) \quad (3)$$

B dan D skala hasil transformasi; α = faktor lokasi yang digunakan untuk membentuk skala baru (= mean = 500 untuk model WITs); γ = faktor pengali untuk skala unit baru (untuk model WITs sebesar 45,5); b dan d skala logit dari kemampuan peserta dan tingkat kesukaran soal.

Dengan transformasi, angka pecahan desimal dan tanda negatif dari skala logit dapat dihilangkan. Pada penelitian ini model transformasi yang digunakan adalah WITs. Persamaan transformasi skala WITs adalah sebagai berikut:

$$W_{\theta} = 500 + 45,5 \theta \quad (\text{untuk kemampuan peserta}) \quad (4)$$

$$W_b = 500 + 45,5 b \text{ (untuk tingkat kesukaran soal)}$$

(5)

d. Kecocokan/Kesesuaian Model

Pencocokan model dimaksudkan untuk mengetahui apakah respons seluruh peserta tester hadap suatu soal fit (cocok) dengan model kurva dari fungsi karakteristik soal yang dipilih. Pengujian kecocokan (*goodness of fit*) model ini menggunakan teknik statistik *Chi Square*. Apabila nilai hitung *Chi Square* lebih kecil dari nilai kritisnya, dengan $db = m - 1$ dan taraf signifikansi 5%, maka dikatakan kurva karakteristik soal cocok dengan model. Simbol db menunjukkan derajat kebebasan dan m menunjukkan banyaknya interval kemampuan yang digunakan dalam perhitungan statistik. Soal yang tidak cocok dengan model merupakan soal yang tidak baik dan tidak berfungsi sebagaimana mestinya. Hal ini disebabkan antara lain oleh konstruksi soal yang jelek, kesalahan kunci jawaban atau berkaitan dengan kebenaran isi soal [25].

RASCH MODEL

Rasch model merupakan salah satu model pengukuran satu parameter logistik. Rasch menyimpulkan bahwa peluang bagi siswa untuk menjawab satu butir soal dengan benar adalah sama dengan ketika kemampuan siswa dibandingkan dengan tingkat kesukaran butir soal. Bond & Fox [26] menyatakan simpulan terkait Rasch model bahwa kesempatan siswa untuk dapat menyelesaikan satu soal dengan benar bergantung pada rasio antara kemampuan siswa dan tingkat kesulitan soal.

Komponen pengukuran yang diperoleh dengan Rasch model meliputi: (1) Analisis butir soal, (2) Kemampuan/Abilitas peserta tes, dan (3) analisis instrumen. Berikut penjelasan setiap komponennya.

1. Analisis Butir Soal

Analisis terhadap butir soal dapat diperoleh informasi tentang parameter butir

soal (*item measure*), kecocokan/kesesuaian butir (*item fit*), dan bias butir (*differential item function/DIF*).

a. Parameter Butir Soal

Analisis terhadap butir soal dengan Rasch model dapat menggunakan program Ministep. Program ini merupakan program komputer yang dapat bekerja di bawah sistem Microsoft Windows yang dikembangkan oleh John Linacre. Paket program ini bersifat *freeware* (dapat digunakan tanpa perlu membeli atau mendapatkan izin pemakaian). Ministep merupakan versi terbatas dari Winsteps. Program ini dapat menganalisis tes bentuk dikotomis (objekti) maupun politomis (uraian). Kemampuan maksimum untuk pengolahan data Ministeps adalah 25 butir soal dan 75 responden. Program ini dapat diunduh di laman <http://www.winsteps.com/ministep/htm> [27]. Jika jumlah soal dan respondennya lebih dari ketentuan di atas maka harus menggunakan Winsteps.

Data tentang tingkat kesukaran butir soal ditampilkan oleh luaran Ministep **Tabel ke 13** Item Measure, yang memerinci tentang informasi logit dari tiap butir soal. Nilai logit yang tinggi menunjukkan tingkat kesukaran soal tinggi. Hal ini berkorelasi dengan total skor, yang menyatakan jumlah jawaban benar. Informasi yang didapat dari luaran *item measure* adalah:

- 1) Soal mana yang paling sukar, sedang, atau sangat mudah. Misal soal no 1 nilai logitnya +3,50 logit dan hanya ada satu orang yang menjawab dengan benar, berarti soal tersebut merupakan soal paling sukar.
- 2) Tingkat kesukaran butir dapat diperbandingkan, karena skala yang dihasilkan oleh program ini adalah sama. Misalnya, soal no 2 nilai logitnya +2,25 dan soal no 3 nilai logitnya +0,75, ini berarti soal no 2 tingkat kesukarannya 3x lipat dari soal no 3.

- 3) Besarnya deviasi standar, yang dapat dikombinasikan dengan rata-rata nilai logit untuk mengelompokkan soal mana yang termasuk kategori sukar, sedang atau mudah. Misalnya, 0,0 logit + 1 SD merupakan kelompok soal yang sukar; kalau lebih dari +1 SD sangat sukar; 0,0 logit - 1 SD soal yang mudah; lebih dari -1 SD soal sangat mudah.

Contoh bagian luaran *item measure* dapat dilihat pada Tabel 1.

Tabel 1. Contoh Bagian Luar *Item Measure* yang Diperhatikan

Entry Number	Total Score	Total Count	Measure
1	1	25	3.50
2	4	25	2.25
7	7	25	1.36
8	10	25	0.84
3	13	25	0.75
4	15	25	0.55
5	17	25	-1.36
6	21	25	-1.86
9	22	25	-2.86
10	23	25	-2.94

b. Kesesuaian Butir Soal

Informasi lain yang diberikan oleh Rasch model adalah tingkat kesesuaian butir soal (*item fit*). Kesesuaian butir ini menjelaskan apakah butir soal berfungsi normal melakukan pengukuran atau tidak. Jika didapati suatu soal tidak fit, ini merupakan indikasi bahwa terjadi miskonsepsi pada siswa terhadap butir soal tersebut [28]. Informasi ini tentunya sangat bermanfaat bagi guru, agar dapat memperbaiki butir soal yang dibuatnya. Hal ini juga dapat digunakan untuk perbaikan pembelajaran, sehingga miskonsepsi dapat dihindari. Hasil kesesuaian butir soal dapat dilihat dari **Tabel 10 Item Fit Order**. Tabel akan memunculkan secara berurutan butir soal yang mempunyai kriteria tidak fit di bagian atas.

Tingkat kesesuaian butir soal dapat dilihat dari nilai *outfit mean square*, *outfit z_standard*, dan *point measure correlation* [29,26,28]. Kriteria yang digunakan untuk memeriksa kesesuaian adalah sebagai berikut:

- 1) Nilai outfit mean square (MNSQ) yang diterima $0,5 < MNSQ < 1,5$
- 2) Nilai outfit Z-standard (ZSTD) yang diterima $-2,0 < ZSTD < +2,0$
- 3) Nilai point measure correlation (Pt Mean Corr) yang diterima $0,4 < Pt Mean Corr < 0,85$

Jika butir soal pada ketiga kriteria tersebut tidak terpenuhi, dapat dipastikan bahwa butir soalnya kurang bagus sehingga perlu diperbaiki atau diganti.

Tabel 2. Contoh Bagian Luar *Item Fit Order* yang Diperhatikan

OUTFIT MNS	OUTFIT ZSTD	PT-MEASURE CORR
2.28	1.25	.50
1.90	1.05	.25
.96	.5	.16
.71	.6	.84
.63	0	.75
.51	-.5	.55
.17	-.7	.36
.21	-1.05	.85
.22	-1.15	.88
.23	-1.25	.94

Nilai kesesuaian butir sangat dipengaruhi oleh besarnya ukuran sampel. Semakin besar ukuran sampel semakin besar nilai ZSTD nya. Oleh sebab itu, beberapa pakar tidak merekomendasikan menggunakan ZSTD ini sebagai ukuran kesesuaian butir. Besar sampel yang dapat menggunakan ZSTD maksimal 500. Selain ukuran sampel, hal-hal yang mempengaruhi kesesuaian butir adalah kesalahan menentukan kunci jawaban, banyaknya individu yang tidak termotivasi dalam mengerjakan tes, dan soal yang daya pembedanya rendah. Luar *item fit order* yang diperhatikan adalah OUTFIT

MNS, OUTFIT ZSTD dan PT-MEASURE CORR. Contoh bagian luaran item fit ordera dapat dilihat pada Tabel 2.

Selain melalui angka, ilustrasi ketidaksesuaian butir soal juga dapat dilihat melalui grafik, yaitu grafik *expected score ICC*. Prosedurnya, tinggal memilih menu Graphics dan klik box *expected score ICC*. Pendeteksian ketidaksesuaian butir dilihat dari data yang outlier atau berada di luar kurva ruang kepercayaan *outfit*.

c. Deteksi Bias Butir

Pengukuran yang valid ditentukan oleh instrumen yang valid, yang tidak mengandung bias. Indikator bias dapat dilihat dari keberpihakan instrumen pada individu dengan karakteristik tertentu. Misalnya, soal lebih mudah dijawab oleh siswa laki-laki dibanding siswa perempuan. Hal ini menunjukkan bias gender. Pada Rasch model, pendeteksian bias ini disebut dengan DIF (*differential item function*) atau keberfungsian butir diferensial. Pada luaran minsteps dapat dilihat pada **Tabel ke 30. Item DIF**.

Suatu butir soal diidentifikasi bias jika didapati nilai probabilitas butirnya berada di bawah 5% (0,05).

2. Analisis Kemampuan/Abilitas Siswa

Analisis terhadap abilitas individu dapat diperoleh informasi tentang kemampuan individu (*person measure*) dan kecocokan/kesesuaian individu (*person fit*),

a. Kemampuan Individu (*Person Measure*)

Data luaran abilitas siswa dapat diklik di Tabel 17. *Person Measure* dari Ministep. Seperti halnya analisis butir soal, luaran *person measure* ini memberikan beberapa informasi sebagai berikut.

- 1) Siswa mana yang paling pintar, sedang, atau kurang. Misal siswa no 1 nilai logitnya +3,50 logit dan hanya ada satu siswa yang memiliki nilai tersebut, berarti siswa tersebut tersebut paling pintar.

- 2) Tingkat kemampuan siswa dapat diperbandingkan, karena skala yang dihasilkan oleh program ini adalah sama. Misalnya, siswa no2 nilai logitnya -0,50 dan siswa no 3 nilai logitnya +0,57, ini berarti siswa no 2 tingkat abilitasnya setengahnya siswa no 3.

- 3) Besarnya deviasi standar, yang dapat dikombinasikan dengan rata-rata nilai logit untuk mengelompokkan soal mana yang termasuk kategori tinggi, sedang atau rendah. Titik tolaknya dari nilai rata-rata logit person. Pengelompokan ini mirip dengan aplikasi pebilai acuan norma.

d. Kesesuaian Individu (*Person Fit*)

Rasch model juga dapat mendeteksi individu yang pola responsnya tidak sesuai. Ketidaksesuaian pola respons ini disebabkan oleh ketidaksesuaian jawaban yang diberikan berdasarkan abilitasnya yang dibandingkan dengan model ideal. Hal ini dapat digunakan guru untuk mengetahui konsistensi berpikir siswa dan kecurangan yang mungkin dilakukan siswa. Hasil kesesuaian individu dapat dilihat dari Tabel 6 *Person Fit Order*. Tabel akan memunculkan secara berurutan individu yang mempunyai kriteria tidak fit di bagian atas. Kriteria untuk menentukan ketidaksesuaian individu sama dengan kriteria yang digunakan untuk menentukan ketidaksesuaian butir soal, yaitu dilihat dari nilai *outfit mean square*, *outfit z_standard*, dan *point measure correlation*.

3. Analisis Instrumen

Selain dapat menganalisis tingkat kesukaran butir soal dan kemampuan individu, Rasch model juga dapat menganalisis instrumen. Analisis dilakukan secara rinci, disajikan dalam bentuk ringkasan statistik dan fungsi informasi pengukuran.

a. Ringkasan Statistik

Informasi ringkasan statistik instrumen ditampilkan oleh output **Tabel 3.1 Summary Statistics**. Informasi yang ditampilkan meliputi skala logit untuk person, nilai reliabilitas (person dan tes), skala logit untuk butir soal.

- 1) Skala logit untuk person, jika nilai rata-ratanya $< 0,0$ logit menunjukkan kecenderungan abilitas siswa lebih kecil dari tingkat kesukaran soal
- 2) Nilai reliabilitas person, semakin tinggi nilai koefisien reliabilitasnya semakin tinggi konsistensi jawaban siswa.
- 3) Nilai reliabilitas tes, semakin tinggi nilai koefisien reliabilitasnya semakin tinggi bagus kualitas tes.
- 4) Skala logit untuk butir soal, semakin tinggi rata-rata logit, semakin sukar tes.

b. Fungsi Informasi Pengukuran

Setiap pengukuran selalu menghasilkan informasi pengukuran, yang menunjukkan hubungan antara tes dengan individu yang diukur. Informasi pengukuran dipengaruhi oleh variasi hasil yang didapatkan (misalnya jawaban benar, jawaban salah, alternatif jawaban lebih dari satu cara, dll). Fungsi informasi ditunjukkan dengan grafik berbentuk kurva normal. Sumbu X menunjukkan abilitas siswa, sumbu Y sebagai fungsi informasi. Apabila abilitas siswa rendah, maka fungsi informasinya juga rendah. Pada abilitas sedang fungsi informasi yang diberikan paling tinggi. Apabila grafik condong ke kiri, berarti set soal memiliki tingkat kesulitan rendah, dan bila condong ke kanan set soal merupakan tes yang sukar. Ada kalanya grafik yang terbentuk memiliki dua puncak, set soal yang mudah dan set soal yang sulit.

Selain informasi-informasi tersebut, Rasch model juga dapat untuk (1) mengidentifikasi *Respons error*, (2) memprediksi skor data hilang (*missing data*), (3) membuktikan bahwa abilitas tidak hanya bergantung pada jawaban benar, dan (4) mengidentifikasi terhadap tebakan.

1. Identifikasi *Respons error*

Ketika siswa dengan abilitas tinggi dan mampu mengerjakan soal sukar dengan benar, tetapi salah ketika menjawab soal yang tingkat kesukarannya rendah, maka siswa tersebut diindikasikan salah merespon. Seharusnya soal yang mudah tidak menjadi masalah baginya untuk menyelesaikan.

2. Memprediksi Skor Data Hilang (*Missing Data*)

Ketika siswa dapat mengerjakan soal pada tingkat sukar, tetapi tidak menjawab nomor soal yang tingkat kesukarannya lebih rendah, maka dapat diprediksikan skor untuk soal yang tidak dikerjakan adalah benar.

3. Abilitas Tidak Hanya Bergantung Pada Jawaban Benar

Ketika dua orang siswa (A dan B) memiliki skor total yang sama, tetapi siswa A dapat mengerjakan soal dengan tingkat kesukaran yang sukar, sedangkan siswa B hanya dapat mengerjakan soal yang mudah, maka dapat dipastikan siswa A kemampuannya lebih tinggi dibanding B. Hal ini tidak akan dapat dilakukan ketika menggunakan teori tes tradisional.

4. Mengidentifikasi Tebakan

Seseorang yang dapat mengerjakan soal dengan tingkat kesukaran rendah, dan tidak dapat mengerjakan soal sedang, tetapi dapat menjawab dengan benar soal yang paling sukar. Ini dapat dipastikan jawaban yang diberikan siswa tersebut pada soal yang sukar adalah tebakan yang kebetulan benar (*lucky guess*).

KESIMPULAN

Asesmen literasi perlu dilakukan, agar siswa terbiasa dengan instrumen yang ini sehingga tidak kagok ketika dievaluasi oleh pihak internasional. Instrumen literasi sains dikembangkan berdasarkan aspek pengukur literasi yains yang terdiri atas 4 aspek, yaitu *the knowledge of science* (pengetahuan sains), *the Investigative nature of science* (penyelidikan tentang hakikat sains),

science as a way of thinking (sains sebagai cara berfikir), dan *interaction of science, environment, technology, and society* (interaksi antara sains, lingkungan, teknologi dan masyarakat). Analisis terhadap instrumen literasi sains dapat dilakukan dengan teori tes klasik dan teori tes modern. Teori tes klasik memiliki kelemahan, yaitu bergantung pada responden/testee, sehingga dapat digunakan teori respon butir. Teori tes modern, khususnya Rasch model, sangat praktis untuk melakukan analisis, penjelasan, dan simpulan dari data suatu ujian. Pola respons yang diberikan menunjukkan ketepatan respons dari setiap responden (*person*) kepada setiap butir soal (*item*). Kelebihan ini yang tidak dimiliki oleh teori tes klasik (tradisional). Dengan Rasch model dapat dilakukan analisis butir soal, kemampuan/abilitas peserta tes, dan analisis instrumen. Selain itu, Rasch model juga dapat untuk mengidentifikasi *Respons error*, memprediksi skor data hilang (*missing data*), membuktikan bahwa abilitas tidak hanya bergantung pada jawaban benar, dan mengidentifikasi terhadap tebakan. Teori tes ini dapat diterapkan pada instrumen tes prestasi dan tes lainnya.

REFERENSI

1. AAAS (American Association for the Advancement of Science). 1993. *Benchmarks for science literacy: A Project 2061 report*. New York: Oxford University Press.
2. OECD. (2007). PISA 2006. Science competencies for tomorrow's world. Volume I: Analysis. Paris: OECD.
3. 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.
4. Bybee, R., McCrae, B. & Laurie, R. (2009). PISA 2006: An Assessment of Scientific Literacy. *Journal of Research in Science Teaching* 46(8), 865-883.
5. Bybee, R. (2008). Scientific Environmental Issue and PISA 2006 : The 2008 Paul-F Brandwein Lecture. *Journal of Science Education and Technology* 17. 566-583.
6. Bybee, R. (2009). PISA 2006 and Scientific Literacy : A Perspective for Science Education Leader . *Science Educator* 18 (2).
7. Bybee, R. (2010). *Teaching Science : 21st century perspective*. USA : National Science Teacher Association Press.
8. Bybee, R. (2011). *PISA 2006 : Implication for science teaching and learning*. USA : National Science Teacher Association Press.
9. 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.
10. NRC (National Research Council). (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
11. Chi-lau, K. (2009). A critical examination of PISA 's Assessment on scientific literacy. *International Journal of Mathematics and Science Education* 7, 1061-1088.
12. OECD. (2009). *PISA 2006 technical report*. Paris: OECD.
13. OECD. (2012). *PISA 2006 Result in Focus*: OECD.
14. 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.
15. 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.
16. 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.
 17. Rusilowati, A., Nugroho, S. E., & Susilowati, S. M. (2016b). Development Of Science Textbook Based On Scientific Literacy For Secondary School. *Jurnal Pendidikan Fisika Indonesia*, 12(2), 98-105.
 18. Rusilowati, A., Prabowo, H., & Nugroho, S. E. (2018). Scientific Literacy Assessment Instrument For Measuring The Students' Scientific Literacy Capability Of Interaction Theme. *Journal of Physics Conference Series*. Vol 983.
 19. Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications.
 20. Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
 21. Alen, M. J., & Yen, W. J. (1979). *Introduction to measurement theory*. Monterey, California: Brooks/Cole Publishing Company.
 22. Rusilowati, A. 2012. *Pengembangan Tes Kemampuan Membaca Sains Berdasarkan Psikologi Kognitif*. Salatiga: Griya Media.
 23. Field, A. (2000). *Discovering statistic using SPSS for windows: Advanced techniques for beginner*. London: SAGE Publications Ltd.
 24. Wright, B. D., & Stone, M. H. (1992). *Best test design*. Chicago: Mesa Pres.
 25. Bahrul Hayat. (1994). *Pengujian dan penilaian pendidikan*. Jakarta: Puslitbang Sisjian balitbang Dikbud.
 26. Bond, T. G. & Fox, C. M. (2015). *Applaying the Rasch Model, Fundamental Measurement in the Human Science (3rd edition)*. New York: Routledge.
 27. <http://www.winsteps.com/ministep/htm>
 28. Sumintono, B. & Widhiarso, W. (2015). Apilkasi Pemodelan Rasch pada Assessment Pendidikan. Cimahi: Trim Komunikata.
 29. Boone, W. J., Scraver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.