

## VALIDITAS DAN RELIABILITAS INSTRUMEN ASESMEN KINERJA LITERASI SAINS PELAJARAN FISIKA BERBASIS STEM

*Khoirul Bashooir<sup>1\*</sup>, Supabar<sup>2</sup>*

<sup>1</sup>SMPIT Ihsanul Fikri, <sup>2</sup>Universitas Negeri Yogyakarta

<sup>1</sup>Jl. Pabelan 1, Pabelan Satu, Pabelan, Mungkid, Magelang, Jawa Tengah 56511, Indonesia

<sup>2</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [khoirul.bashooir@gmail.com](mailto:khoirul.bashooir@gmail.com)

### Abstrak

Penelitian ini merupakan bagian dari penelitian pengembangan asesmen kinerja literasi sains berbasis *STEM* pada pembelajaran fisika. Tujuan dari penelitian ini adalah untuk mengungkapkan validitas isi, validitas empiris, dan reliabilitas instrumen asesmen kinerja literasi sains berbasis *STEM* yang sebelumnya telah disusun. Instrumen yang dikembangkan berupa lembar pengamatan dan tes pilihan ganda. Analisis validitas isi dari lembar pengamatan menggunakan Koefisien V oleh Aiken sedangkan validitas isi instrumen tes dianalisis dengan menggunakan *CVI* (*Content Validity Index*) oleh Lawshe. Validitas empiris reliabilitas instrumen tes diestimasi dengan *IRT* (*Item Response Theory*). Reliabilitas lembar pengamatan ditentukan dengan *ICC* (*Item Correlation Coefficient*). Hasil dari penelitian ini menunjukkan bahwa (1) Lembar pengamatan berupa rubrik penskoran dan penilaian diri terbukti valid dengan koefisien V Aiken 0,75 dan reliabel dengan koefisien Reliabilitas Alfa > 0,8 dan ICC yang *Excellent*. (2) Instrumen tes terbukti reliabel untuk digunakan pada peserta didik dengan kategori sedang sampai dengan tinggi (-0,7 sampai dengan 6,7) dengan *CVI*=1 dan *INFIT MNSQ* sesuai model Rasch. Berdasarkan hasil penelitian tersebut maka asesmen kinerja Literasi Sains berbasis *STEM* layak digunakan.

**Kata kunci:** *validitas isi, validitas empiris, asesmen kinerja, literasi sains, STEM*

## VALIDITY AND RELIABILITY INSTRUMENT OF SCIENTIFIC LITERACY PERFORMANCE ASSESSMENT IN PHYSICS TEACHING BASED ON STEM

### Abstract

This research is part of the development of scientific literacy performance assessment based on STEM in teaching physics. The aim of this research is to reveal the validity (content and also empiric) and reliability of scientific literacy performance assessment instrument based on STEM. The kind of instruments were developed are observational sheet and multiple choice test. The content validity of observational sheet was revealed by used the Aiken's V Coefficient. The content validity of multiple choice tests was revealed by used Content Validity Index (CVI) which proposed by Lawshe. The empirical validity and reliability of multiple choice tests was revealed by used Item Response Theory Analysis. The reliability of observational sheet was revealed by used ICC (Item Correlation Coefficient) Analysis. The results of this study are the validity from the contents and empirical trials from the developed instruments. The observation sheet from scoring rubric and self-assessment has been valid with Aiken's V value that exceeds the standard of 0,75. The reliability of the scoring rubric has Alfa Reliability > 0.8 and Excellent of ICC. Validity values from The written test is shown with CVI of 1 and the MNSQ INFIT value which match to the Rasch model. Based on the TIC and SEM graphs, the written test is stated to be reliable for use in students with moderate to high categories (-0.7 to 6.7). STEM-based Science Literacy performance assessment with caloric material is appropriate to use.

**Keywords:** *content validity, empirical validity, performance assessment, scientific literacy, STEM*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.19590>

## Pendahuluan

Mencerdaskan kehidupan bangsa adalah salah satu tujuan dari Pemerintah Negara Indonesia. Tujuan tersebut termaktub dalam pembukaan Undang-Undang Dasar 1945 dan menjadi salah satu tujuan dari pendidikan nasional negara Indonesia yang diatur dalam undang-undang Republik Indonesia Nomor 20 Tahun 2003 (Presiden Republik Indonesia, 2003). Oleh karena itu, pendidikan yang diterapkan di Indonesia haruslah bertujuan untuk kecerdasan warga negara Indonesia dalam hal ini adalah peserta didik agar mampu bersaing dengan negara-negara yang lain.

Pendidikan yang diselenggarakan melalui proses pembelajaran di sekolah menengah diharapkan dapat menghasilkan lulusan yang mampu mengaplikasikan ilmu pengetahuan yang dimiliki untuk pemecahan masalah dalam kehidupan sehari-hari (Badan Standar Nasional Pendidikan, 2006, p. 443). Berdasarkan acuan tersebut, kegiatan pembelajaran dapat didesain dengan memberikan tugas-tugas sebagai bentuk latihan untuk menyelesaikan permasalahan. Hal ini selaras dengan konsep literasi sains yang dikemukakan oleh *PISA (Programme for International Student Assessment)*. Konteks literasi sains dalam penelitian ini dikhususkan hanya pada pembelajaran fisika.

Sains sebagai ilmu pengetahuan memiliki empat dimensi (Chiappetta & Koballa, 2010, p. 102) yaitu (1) *science as a way of thinking*, (2) *science as a way of investigating*, (3) *science as a body of knowledge*, dan (4) *science and its interactions with technology and society*. Keempat dimensi tersebut diperlukan dalam kegiatan pembelajaran sains.

Pembelajaran sains yang dilaksanakan harus dapat menjawab tantangan pada abad ke-21 yaitu untuk membentuk manusia yang dapat bertahan dengan perkembangan teknologi. Wagner (2008, pp. 14–38) merumuskan beberapa keterampilan yang dibutuhkan oleh manusia agar dapat bertahan hidup pada abad ke-21. Beberapa keterampilan tersebut adalah: (1) keterampilan berpikir kritis dan menyelesaikan masalah; (2) berkolaborasi dalam komunitas; (3) ketangkasan dan kemam-

puan beradaptasi; (4) berinisiasi dan jiwa bisnis; (5) kemampuan komunikasi secara lisan dan tulis; (6) mengakses dan menganalisis informasi; dan (7) imajinasi dan keingintahuan. Selain itu, literasi sains juga merupakan keterampilan yang penting pada abad ke-21 (P21, 2009, p. 2). Hal tersebut sejalan dengan hasil penelitian Yore & Treagust (2006, p. 304) yang menunjukkan bahwa literasi sains dibutuhkan oleh individu pada abad 21 untuk dapat hidup dan memecahkan permasalahannya terkait dengan fenomena alam.

*PISA* merupakan program asesmen yang diadakan oleh *OECD (Organization for Economic Co-operation and Development)* untuk menilai kemampuan membaca, literasi matematika dan literasi sains peserta yang mengikuti tes tersebut. Pada tahun 2012 (OECD, 2014, p. 19) tes yang dilakukan oleh *PISA* diikuti oleh 510.000 peserta didik dari 65 negara. Indonesia berada pada ranking 64 dari 65 negara. Data tersebut menunjukkan bahwa kemampuan membaca, literasi matematika dan literasi sains peserta didik di Indonesia masih sangat rendah bila dibandingkan dengan 64 negara lain yang mengikuti program tersebut. Peringkat yang diperoleh Indonesia dalam *PISA* mengindikasikan bahwa peserta didik Indonesia memiliki literasi sains yang rendah. Selain itu, peringkat tersebut juga mengindikasikan bahwa Indonesia belum dapat bersaing dengan negara-negara lain.

Ilmu fisika merupakan salah satu konteks dalam sains yang diukur dalam *PISA*. Fisika juga menjadi salah satu ilmu pengetahuan yang mendasari perkembangan teknologi. Dalam fisika terdapat integrasi antara konsep fisika, konsep matematika, cara merangkai peralatan sehingga menghasilkan teknologi yang inovatif.

Berdasarkan observasi yang dilakukan di empat sekolah di kabupaten Magelang yang memiliki peringkat tinggi, sedang dan rendah menurut hasil ujian nasional diketahui bahwa (1) kegiatan pembelajaran masih berpusat kepada guru (*Teacher Oriented*), (2) metode pembelajaran yang digunakan adalah ceramah, dan (3) asesmen yang digunakan

oleh guru untuk mengukur kemampuan peserta didik adalah asesmen tes tertulis. Lebih lanjut diketahui pula bahwa asesmen tes tulis yang digunakan tidak disusun berdasarkan konsep pemecahan masalah dalam kehidupan sehari-hari. Hal tersebut belum mencerminkan tujuan pendidikan yang ditentukan oleh BSNP yang menuntut peserta didik untuk dapat mengaplikasikan pengetahuan yang dimilikinya untuk memecahkan masalah dalam kehidupan sehari-hari.

*STEM (Science Technology Engineering dan Mathematic)* adalah salah satu alternatif solusi bagi pembelajaran abad 21. Pendekatan *STEM* merupakan pembelajaran yang mengintegrasikan *Science Technology Engineering* dan *Mathematic* dalam pembelajaran *Engineering* (Department of Education, 2009, p. 19; Breiner, Harkness, Johnson, & Koehler, 2012, p. 3; Gonzalez & Kuenzi, 2012, p. 1; Reeve, 2013, p. 3; Hernandez et al., 2014, p. 108). *STEM* memenuhi keempat dimensi dalam pembelajaran Sains (Chiappetta & Koballa, 2010, p. 102; Departemen Pendidikan dan Kebudayaan, 2001, p. 1473). *Science* sebagai kerangka utama (*Science as a way of thinking* dan *science as a body of knowledge*). *Science* sebagai *way of investigating*, *Science* dikombinasikan dengan *Engineering* dan *Mathematics*. *Engineering* atau keterampilan teknik digunakan dalam merancang dan melaksanakan percobaan. *Mathematics* digunakan dalam menganalisis data percobaan. *Technology* digunakan untuk membantu dalam menganalisis data serta mengomunikasikan hasil percobaan.

Penggunaan pendekatan *STEM* dimaksudkan agar peserta didik dapat memiliki kemampuan dan pemahaman dalam keempat aspek *STEM* yang saling terkait pada satu pokok bahasan. Peserta didik juga diharapkan dapat mengaplikasikan hal-hal yang dipelajarinya dalam dunia nyata atau lingkungan kerja.

Pengembangan keterampilan literasi sains peserta didik dapat dilakukan melalui pembelajaran *STEM* (Ismail, Permanasari, & Setiawan, 2016, p. 10). Pengembangan literasi sains dan *STEM* peserta didik dapat melalui pendidikan formal dan informal. Pendi-

dikan formal yang dimaksud adalah pendidikan yang dilaksanakan oleh sekolah-sekolah pada umumnya. Pendidikan informal adalah pendidikan yang didapatkan oleh individu karena interaksinya dalam masyarakat di sekitarnya. Penelitian ini membatasi literasi sains dan *STEM* ada pendidikan formal yang dilaksanakan di sekolah, khususnya sekolah menengah pertama.

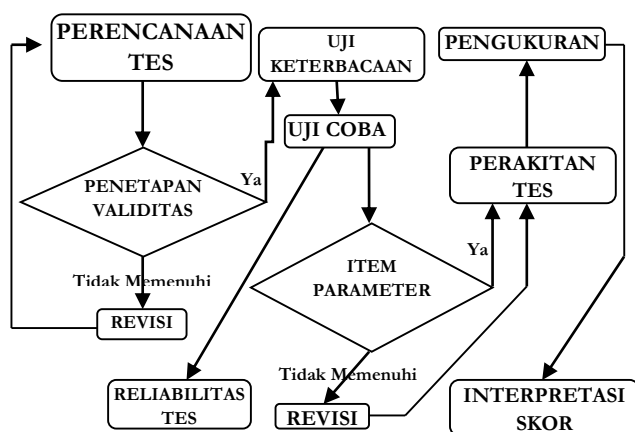
Selain dengan memperbaiki proses pembelajaran yang dilaksanakan, peningkatan kualitas pembelajaran dapat dilakukan melalui perbaikan pada asesmen yang digunakan untuk mengukur hasil belajar peserta didik. Perbaikan dalam aspek penilaian dapat dilakukan dengan memberikan panduan dan pelatihan dalam melakukan penilaian (Kartowagiran & Jaedun, 2016, p. 140). Asesmen yang digunakan untuk memperoleh, menyintesis dan menginterpretasikan informasi yang diperoleh dari peserta didik dengan tujuan untuk menarik keputusan tentang peserta didik di dalam kelas dikenal dengan istilah *classroom assessment*. *Classroom assessment* yang digunakan untuk menilai kinerja peserta didik selama pembelajaran dikenal sebagai *assessment for learning*. Dalam pembelajaran, *assessment for learning* menyatu dengan rencana pelaksanaan pembelajaran yang disusun atau digunakan oleh guru.

Salah satu bentuk *assessment for learning* adalah asesmen kinerja. Asesmen kinerja digunakan untuk mengases kemampuan peserta didik dalam menerjemahkan pemahamannya menjadi suatu kerja nyata, sehingga asesmen kinerja dapat menunjukkan kemampuan peserta didik yang sesungguhnya.

Instrumen yang digunakan dalam asesmen memiliki beberapa syarat agar layak digunakan yaitu valid dan reliabel (Mardapi, 2012, pp. 37–96; Sumintono & Widhiarso, 2015, pp. 7–12). Oleh karena itu, instrumen kinerja berbasis *STEM* yang telah dikembangkan sebelumnya harus memenuhi kedua syarat tersebut. Oleh karena itu, uji validitas dan reliabilitas instrumen asesmen kinerja literasi sains perlu dilakukan.

## Metode Penelitian

Penelitian ini merupakan bagian dari penelitian pengembangan asesmen kinerja literasi sains berbasis *STEM* pada pembelajaran fisika. Instrumen yang telah memenuhi standar kevalidan dan standar kereliatan dapat digunakan untuk tahap pengukuran. Tahapan penelitian secara keseluruhan yang disajikan pada Gambar 1 mengacu pada pembuatan instrumen tes yang dibuat oleh Supahar & Prasetyo (2015, p. 99).



Gambar 1. Skema Prosedur Penelitian Pengembangan Asesmen Kinerja Berbasis *STEM*

Instrumen asesmen dikembangkan berdasarkan aspek dan indikator literasi sains. Penelitian tersebut memperoleh 11 aspek literasi sains (Bashooir & Supahar, 2016, p. 93), yaitu: (1) merumuskan masalah, (2) menyusun variabel fisika, (3) menyusun kajian teori, (4) mendesain eksperimen, (5) mengumpulkan data, (6) menganalisis data, (7) menyusun tabel, (8) menyusun grafik, (9) menyimpulkan, (10) menyusun solusi dan (11) mengomunikasikan solusi. Dalam penelitian ini, instrumen dikembangkan dengan menggunakan 9 aspek yang mengacu pada penelitian tersebut. Ke-9 aspek tersebut adalah 1, 2, 3, 4, 6, 7, 8, 9 dan 10.

Data yang diperlukan untuk penentuan validitas dan reliabilitas instrumen diperoleh melalui beberapa cara. Validitas isi instrumen diperoleh dengan memberikan angket kepada pakar, yaitu pakar asesmen, pakar pembelajaran, pakar ilmu fisika dan pakar pembelajaran fisika. Selanjutnya, uji keter-

bacaan instrumen juga dilakukan dengan memberikan instrumen kepada 6 peserta didik. Hasil dari uji keterbacaan digunakan untuk memperbaiki instrumen.

Validitas empiris dan reliabilitas diperoleh instrumen yang dikembangkan diperoleh melalui uji coba instrumen. Instrumen tes tertulis bentuk pilihan ganda diujicobakan kepada 350 peserta didik di empat sekolah di Kabupaten Magelang. Keempat sekolah tersebut mewakili sekolah dengan peringkat tinggi, sedang dan rendah ditinjau dari hasil ujian nasional di Kabupaten Magelang. Keempat sekolah tersebut adalah SMAIT Ihsanul Fikri, SMA N 1 Muntilan, SMA N 1 Salaman dan SMA Muhammadiyah 1 Muntilan. Selanjutnya, Lembar pengamatan diujicobakan kepada 57 peserta didik yang berasal dari SMAIT Ihsanul Fikri.

### Validitas Isi

Validitas isi suatu butir dapat dibuktikan dengan menggunakan *CVR* dan *CVI* atau koefisien *V* Aiken. *CVR* dan *CVI* diusulkan oleh Lawshe pada tahun 1975 dengan menggunakan 3 skala rating (Lawshe, 1975, p. 567). Pada tahun 1985, Aiken mengusulkan konsep validitas isi dengan lebih rinci. Kerincian tersebut terlihat dari standar kevalidan yang dipengaruhi oleh jumlah rater dan skala rating yang digunakan (Aiken, 1985, p. 133). Penggunaan cara pembuktian harus memperhitungkan asumsi atau syarat yang harus dipenuhi, sedangkan, Pembuktian validitas dengan menggunakan konsep dari Lawshe (1975, p. 567) hanya menggunakan tiga skala rating, yaitu (1) esensial, (2) berguna tetapi tidak esensial, dan (3) tidak perlu. Pembuktian validitas isi oleh Lawshe (Lawshe, 1975, p. 567; Azwar, 2012, p. 114) dapat ditentukan dengan formula:

$$CVR = \frac{2ne}{n} - 1 \quad (1)$$

Keterangan:

*ne* = banyaknya *SME* (*Subject Matter Experts*) yang menilai suatu butir 'esensial'

*n* = banyaknya *SME* yang melakukan penilaian

Nilai *CVR* memiliki rentang antara -1 sampai dengan 1. Jika setengah dari *SME* menyatakan esensial maka nilai *CVR* akan bernilai 0. *CVR* akan bernilai 1 jika seluruh *SME* menyatakan esensial untuk suatu butir. Acuan standar kevalidan item menurut Lawshe (1975, p. 568) ditunjukkan pada Tabel 1.

Tabel 1. Standar Minimum *CVR* berdasarkan *SME*

Jumlah <i>SME</i>	Nilai Minimal <i>CVR</i>
5	0,99
6	0,99
7	0,99
8	0,75
9	0,78
10	0,62

Butir yang memperoleh nilai *CVR* negatif merupakan butir yang tidak valid dan harus dieliminasi. Nilai validitas tes secara keseluruhan dapat ditentukan dengan menggunakan *CVI* (*Content Validity Index*). Penentuan *CVI* (Lawshe, 1975, p. 568; Azwar, 2012, p. 115) dilakukan dengan menggunakan formula:

$$CVI = \frac{\sum CVR}{k} \quad (2)$$

Keterangan:

*CVR* = *Content Validity Ratio* dari tiap item  
*k* = banyaknya butir soal

Pada pembuktian validitas isi (Azwar, 2012, p. 113), peneliti dapat menentukan banyaknya kategori rating yang dikehendaki. Banyaknya kategori rating mempengaruhi standar validitas isi yang ditentukan oleh Aiken. Banyaknya kategori rating terkecil yang dirumuskan oleh Aiken adalah 2 dan terbanyak adalah 7 (Aiken, 1985, p. 134). Penelitian ini menggunakan 5 kategori rating dan 7 Rater. Berdasarkan standar yang telah ditetapkan Aiken (Aiken, 1985, p. 134), standar minimal *V* Aiken untuk penelitian ini adalah 0,75 dengan probabilitas 0,041.

Koefisien validitas Aiken didapatkan perhitungan dengan skor mentah dari ahli yang berjumlah *n*. Nilai koefisien *V* Aiken

memiliki rentang antara -1 sampai dengan 1 (Supahar, 2015, p. 66). Validitas isi dengan menggunakan koefisien *V* Aiken (Azwar, 2012, p. 113; Aiken, 1985, p. 133) diperoleh dengan menerapkan formula:

$$V = \frac{\sum(r_i - l_o)}{[n(c-1)]} \quad (3)$$

Keterangan:

*r* = angka yang diberikan oleh penilai  
*l<sub>o</sub>* = angka penilaian validitas terendah  
*c* = angka penilaian validitas yang tertinggi  
*n* = banyaknya ahli & praktisi yang melakukan penilaian  
*i* = bilangan bulan dari 1,2,3 sampai ke *n*  
*n* = banyaknya penilai

Analisis validitas isi dengan menggunakan koefisien Aiken *V* digunakan untuk menguji validitas instrumen lembar pengamatan, sengan validitas isi yang diperkenalkan oleh Lawshe digunakan untuk menguji validitas isi instrumen tes.

#### Validitas Empiris

Validitas empiris diperoleh dari analisis respon terhadap tes yang diberikan kepada responden. Respon didapatkan dari uji coba tes kepada para responden atau testi. Validitas empiris (Sumintono & Widhiarso, 2015, pp. 12-47; Retnawati, 2016, pp. 113–146) dapat ditentukan dengan menggunakan *Classical Test Theory* (*CTT*) atau *Item Response Theory* (*IRT*).

*Rasch model* merupakan bagian dari *IRT* yang dapat dilakukan dengan bantuan program *QUEST*. Butir soal dinyatakan valid apabila nilai *INFITMNSQ* berada pada rentang 0,77 sampai dengan 1,30 (Subali & Suyata, 2012, p. 61). Butir-butir yang berada dalam rentang tersebut dapat digunakan untuk keperluan pengukuran keterampilan literasi sains peserta didik.

#### Koefisien Alfa

Salah satu cara untuk menentukan tingkat reliabilitas instrumen adalah menggunakan koefisien Alfa. Azwar (2015, p. 67) menjelaskan bahwa koefisien alfa dapat digunakan selama masing-masing belahan sama

panjang atau berisi item yang sama banyaknya. Koefisien alfa dapat ditentukan dengan menggunakan persamaan (Azwar, 2015, p. 45):

$$\rho_{XX'} \geq \alpha = \frac{\left[\frac{k}{k-1}\right](\sigma_x^2 - \sum \sigma_{Y_i}^2)}{\sigma_x^2} \quad (4)$$

Keterangan:

$\rho_{XX'}$  = koefisien reliabilitas

$\alpha$  = koefisien alfa

$k$  = banyaknya butir dalam tes

$\sigma_x^2$  = varian skor tes

$\sigma_{Y_i}^2$  = varian skor belahan tes dengan  $i = 1, 2, 3, \text{ dst.}$  (belahan tes)

Dalam penelitian ini estimasi reliabilitas instrumen dengan menggunakan koefisien alfa digunakan untuk mengestimasi reliabilitas instrumen tes. Reliabilitas alfa memiliki rentang nilai antara 0 sampai dengan 1. Berdasarkan (Streiner, 2003, p. 103) sebuah instrumen pada tahap penelitian pendahuluan dinyatakan reliabel apabila nilai reliabilitas alfa 0,7; pada penelitian dasar sebesar 0,8 dan pada penelitian dalam bidang kedokteran dengan tujuan klinis sebesar 0,95.

#### Interclass Correlation Coefficient (ICC)

ICC (*Interclass Correlation Coefficient*) merupakan reliabilitas yang digunakan untuk hasil rating dari pengamatan beberapa rater (Shrout & Fleiss, 1979). Hal yang sama juga disampaikan oleh Mardapi (2012). ICC dapat diestimasi dengan menggunakan persamaan (Mardapi, 2012, p. 89).

$$\rho = \frac{MSrs - MSe}{MSr + (k-1)MSe} \quad (5)$$

Keterangan:

$MSrs$  = rerata kuadrat antar penilai

$MSe$  = varians skor kesalahan

$k$  = jumlah penilai

Analisis ICC pada penelitian ini menggunakan bantuan SPSS 16. Dalam penelitian ini, reliabilitas ICC digunakan untuk mengestimasi reliabilitas instrumen lembar pengamatan.

#### Reliabilitas Skala Logit: Item Response Theory (IRT)

Sebagai pendukung, pengujian reliabilitas tiap butir soal dapat ditentukan dengan menggunakan IIC dan reliabilitas perangkat tes dapat diperoleh dari TIC. Grafik IIC dan TIC didapatkan dengan pengolahan data melalui program BILLOG MG (Mathilda du Toit, 2003, pp. 505-508). Melalui grafik ICC dengan 3 parameter dapat diketahui tingkat kehandalan soal. Grafik IIC dan TIC memiliki rentang skala logit antara -3 sampai dengan 3. Sumintono & Widhiarso (2015, p. 39) menjelaskan bahwa kala logit merupakan skala yang menggambarkan abilitas atau kemampuan peserta didik. Skala minus menunjukkan abilitas yang rendah dan semakin positif menunjukkan abilitas yang tinggi.

#### Hasil Penelitian dan Pembahasan

##### Lembar Pengamatan: Rubrik Penskoran dan Penilaian Diri

Skor yang diperoleh untuk perangkat asesmen dalam bentuk rubrik penskoran dianalisis dengan menggunakan analisis validitas V Aiken. Hasil analisis dapat dikategorikan valid apabila memenuhi batas koefisien V Aiken. Syarat batas koefisien V Aiken untuk 5 skala rating dan 7 rater adalah 0,75 dengan probabilitas 0,41. Tiap aspek penilaian dalam rubrik penskoran aspek psikomotorik, LKPD (lembar Kerja Peserta Didik), laporan dan penilaian diri mendapatkan skor Aiken di atas 0,75. Oleh karena itu, ketiga rubrik penskoran tersebut dinyatakan valid sehingga layak digunakan dalam eksperimen. Skor rata-rata V Aiken disajikan dalam Tabel 2.

Tabel 2. Skor Rata-rata V Aiken untuk Rubrik Penskoran dan Penilaian Diri

Perangkat Asesmen	Rata-rata V Aiken
Penilaian Diri	0,90
Rubrik Penskoran Aspek Psikomotorik	0,88
Rubrik Penskoran LKPD	0,88
Rubrik Penskoran Laporan	0,89

### Perangkat Asesmen: Tes Tertulis

Hasil dalam validasi isi untuk perangkat asesmen tes tertulis dianalisis dengan menggunakan Validitas isi Lawshe dimana standar kevalidan *CVR* tergantung pada jumlah *SME*. Nilai *CVR* harus memenuhi 0,99 agar butir dapat dinyatakan valid. Hal tersebut berlaku untuk validasi isi dengan menggunakan 7 *SME* (Lawshe, 1975, p. 568). Nilai *CVR* yang diperoleh dari tiap butir adalah 1 dan secara lengkap disajikan dalam lampiran. Nilai *CVI* yang diperoleh dari rata-rata *CVR* sebesar 1. Berdasarkan nilai *CVR* yang melebihi 0,99 maka seluruh butir dinyatakan valid (Lawshe, 1975, p. 568) dan layak digunakan untuk penelitian lebih lanjut.

### Uji Coba Tes Tulis

Butir soal yang telah terbukti validitasnya kemudian disusun ke dalam 2 paket soal (Paket A dan Paket B). Setiap paket terdiri dari 34 butir soal dengan 7 butir yang digunakan sebagai angkor, sehingga terdapat 7 butir soal yang termasuk dalam Paket A dan Paket B. Uji coba keterbacaan soal dilakukan dengan melibatkan 6 peserta didik. Hasil uji coba keterbacaan butir soal digunakan sebagai bahan untuk memperbaiki instrumen tes tulis sebelum diujicobakan pada responden dengan jumlah lebih besar.

Uji coba ke dua dilakukan terhadap 350 peserta didik yang berada di Kabupaten Magelang. Respon peserta didik dianalisis dengan menggunakan *IRT* model *Rasch* berbantuan program *QUEST* dan *BILOG MG*. Berikut ini adalah hasil analisis yang diperoleh:

### *Kecocokan Butir Instrumen Tes Tulis dengan Model Rasch*

Butir dinyatakan valid apabila memiliki nilai *INFIT MNSQ* dalam rentang 0,77 sampai dengan 1,30 (Subali & Suyata, 2012, p. 61; Supahar, 2014, p. 141). Analisis yang dilakukan dengan menggunakan bantuan program *QUEST* menunjukkan bahwa ke 61 butir memiliki *INFIT MNSQ* yang berada dalam rentang 0,77 sampai dengan 1,30. Oleh karena itu, seluruh butir dinyatakan valid. Ilustrasi secara lengkap kevalidan butir soal disajikan dalam Gambar 2.

### *Hasil Estimasi*

Hasil estimasi 61 butir tes yang dikerjakan oleh 350 responden dengan tingkat peluang 0,5 dan menggunakan penskoran dikotomus berdasarkan Model *Rasch* dinyatakan dalam Tabel 3. Dalam Tabel 3 dapat diketahui estimasi untuk butir dan estimasi untuk testi.

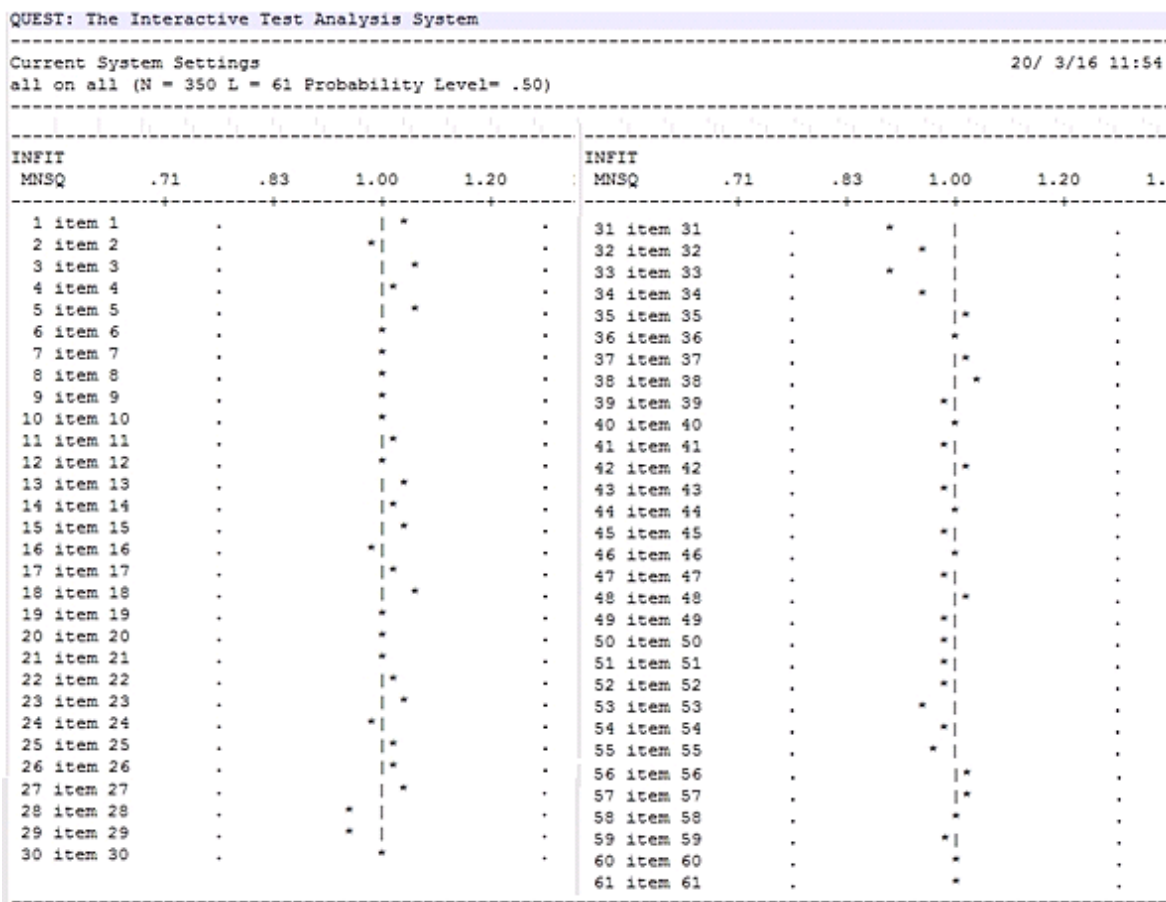
### *Tingkat Kesulitan*

Tingkat kesulitan tiap aspek dalam penelitian ini diperoleh dari rerata tingkat kesulitan tiap butir dalam satu aspek. Semakin tinggi nilai tingkat kesulitan maka butir soal dinilai semakin sulit (Subali & Suyata, 2012, p.56). Hal yang sebaliknya, jika nilai tingkat kesulitan semakin rendah maka butir soal dinilai semakin mudah.

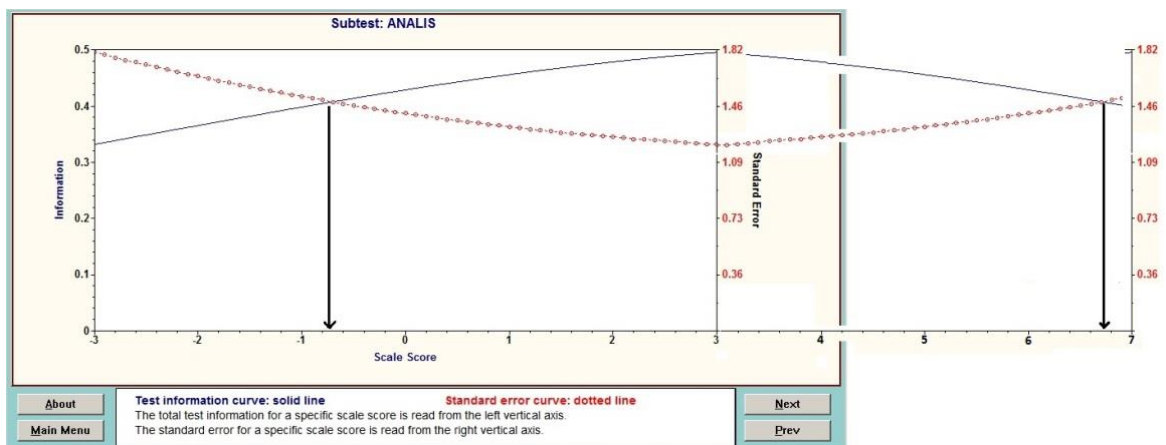
Indikator dengan tingkat kesulitan terendah adalah indikator 2 pada aspek menyusun tabel. Indikator tersulit adalah indikator 3 pada aspek menyusun kesimpulan. Ilustrasi perbandingan tingkat kesulitan antar aspek disajikan dalam Gambar 3.

Tabel 3. Hasil Estimasi Butir dan Estimasi Testi Keterampilan Literasi Sains Peserta Didik SMA pada Kegiatan Uji Coba

No.	Uraian	Estimasi untuk butir	Estimasi untuk testi
1.	Nilai rata-rata dan simpangan baku	101,74 ± 64,74	17,73 ± 3,88
2.	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	0,00 ± 1,13	-1,11 ± 0,41
3.	Reliabilitas	0,98	
4.	Nilai rata-rata dan simpangan baku <i>INFIT MNSQ</i>	1,00 ± 0,03	1,00 ± 0,10
5.	Nilai rata-rata dan simpangan baku <i>OUTFIT MNSQ</i>	1,01 ± 0,10	1,01 ± 0,32

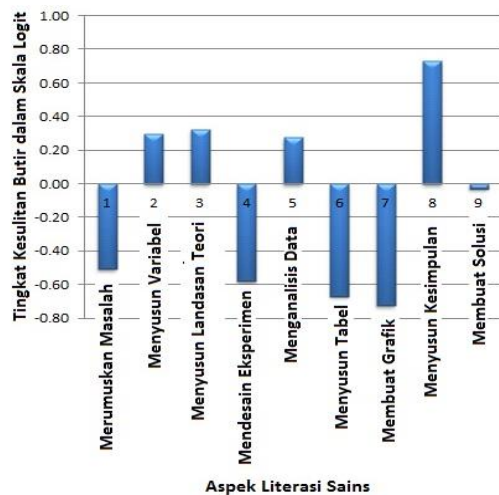


Gambar 2. Distribusi Nilai *INFI* *MNSQ* tiap butir Soal dalam *Fit Model*



Gambar 3. *Total Information Curve (TIC)* untuk Instrumen Tes Tulis





Gambar 4. Tingkat Kesulitan Butir Tiap Aspek Keterampilan Literasi Sains

Gambar 4 menunjukkan perbandingan tingkat kesulitan butir soal dari tiap indikator dan tiap aspek keterampilan literasi sains. Sumintono & Widhiarso (2015, p. 70) mengategorikan tingkat kesulitan skala logit dalam 4 kategori, yaitu sangat sukar, sukar, mudah dan sangat mudah. Kategori sangat sukar memiliki kriteria skala logit lebih dari 1. Kategori sukar berada dalam kriteria 0 sampai 1. Kategori mudah dan kurang dari -1 merupakan kategori sangat mudah. Secara keseluruhan, tingkat kesulitan tiap aspek berada dalam rentang -1 sampai dengan 1. Hal tersebut berarti bahwa tingkat kesulitan rata-rata dari butir pada tiap aspeknya memiliki kategori mudah dan sukar.

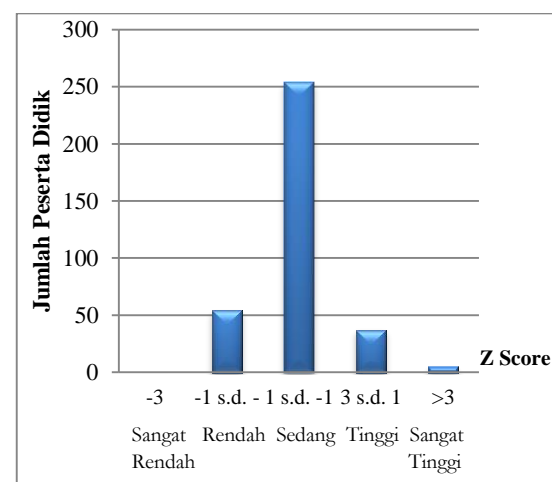
#### Reliabilitas Tes

Reliabilitas perangkat tes tulis dapat diperoleh dengan menggunakan bantuan program *BILOG MG*. Reliabilitas tersebut dapat dilihat pada grafik *Total Information Curve (TIC)* dan *Standard Error of Measurement (SEM)*. Instrumen tes dinyatakan reliabel untuk digunakan kepada peserta didik dengan kategori sedang (-0,7) sampai dengan tinggi (6,7). Hal tersebut memiliki arti bahwa instrumen tes akan tepat diujikan kepada responden dengan kemampuan sedang (-0,7) sampai dengan tinggi (6,7) (Sumintono & Widhiarso, 2015, pp. 7-12; Supahar &

Prasetyo, 2015, p. 104). Hasil analisis reliabilitas tes dapat dilihat pada Gambar 2.

#### Kemampuan Peserta Didik

Kemampuan peserta didik dapat diketahui dari *output* program *BILOG MG* yaitu *file* dengan format *PH3* dan *SCO*. Kemampuan peserta didik dalam literasi sains disajikan dalam kolom *ability* dengan skala logit. Nilai *ability* merupakan  $\theta$  (*tetha*). Nilai  $\theta$  dalam uji coba memiliki rentang antara -2,03 sampai dengan 1,19. Nilai tersebut jika dikonversikan dalam  $z$  score adalah -3,82 sampai dengan 3,92. Keterampilan literasi sains diinterpretasikan sebagai dalam 5 skala yaitu (1) sangat tinggi, (2) tinggi, (3) sedang, (3) rendah, dan (3) sangat rendah. Ilustrasi tingkat keterampilan literasi sains peserta didik disajikan dalam Gambar 5.



Gambar 5. Tingkat Keterampilan Literasi Sains Peserta Didik Uji Coba

#### Uji Coba Rubrik Penskoran

Pada saat mengujicobakan rubrik penskoran, peserta didik diberikan tugas-tugas dalam bentuk LKPD. Pengamat atau rater menggunakan rubrik penskoran untuk menilai ketuntasan kinerja yang dilakukan peserta didik. Skor yang diperoleh peserta didik dianalisis untuk ditentukan reliabilitas Alfa dan *ICC*. Analisis dilakukan dengan menggunakan bantuan program *SPSS 16*. Hasil analisis dirangkum dalam Tabel 4.

Interpretasi berdasarkan analisis skor peserta didik dari pengamatan adalah sebagai berikut (1) Reliabilitas Alfa yang diperoleh mendapatkan nilai di atas 0,8 (Streiner, 2003, p.103) yang merupakan batas minimal pada penelitian dasar. (2) Skor ICC menunjukkan tingkat reliabilitas sangat baik (*Excellent*) yang berada dalam rentang 0,75 sampai dengan 1,00 (Cicchetti, 1994, p. 286).

Tabel 4. Hasil Analisis Reliabilitas Alfa dan *Interclass Coefficient Correlation (ICC)*

No	Statistik	Reliabilitas Alfa	ICC
1.	Rubrik Penskoran Aspek Psikomotorik	0,983	0,936
2.	Rubrik Penskoran LKPD	0,993	0,971
3.	Rubrik Penskoran Laporan	0,999	0,995

### Simpulan

Lembar pengamatan berupa rubrik penskoran dan penilaian diri dinyatakan layak berdasarkan nilai validitas isi V. Rubrik penskoran juga dinyatakan reliabel dengan nilai Reliabilitas Alfa melebihi 0,8 dan ICC dengan kriteria *Excellent*. Perangkat tes tulis terbukti valid dilihat dari sudut pandang *content* atau isi dengan nilai *CVI* sebesar 1 dan valid secara empirik karena sesuai dengan nilai *INFIT MNSQ* model *Rasch*. Berdasarkan Grafik *TIC* dan *SEM*, perangkat tes tulis dinyatakan reliabel untuk digunakan pada peserta didik dengan kategori sedang sampai dengan tinggi (-0,7 sampai dengan 6,7). Oleh karena itu, Asesmen kinerja Literasi Sains berbasis *STEM* layak digunakan karena telah memenuhi syarat dalam aspek validitas isi, validitas empiris dan reliabilitas.

Saran bagi peneliti yang hendak melakukan penelitian yang sejenis adalah sebagai berikut (a) Semakin banyak pakar yang menelaah instrument maka akan semakin baik kualitas instrumen dalam aspek validitas isi, (b) penting untuk mempersiapkan lebih dari satu untuk tiap indikator agar dapat menghindari ketidakterwakilan suatu indikator jika butir tertentugugur dalam prses validasi (c) penting untuk memilih res-

ponden yang mewakili peserta didik dengan kemampuan rendah, sedang dan tinggi dalam uji coba empiris sebagai upaya untuk mendapatkan validitas yang baik.

### Daftar Pustaka

- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2015). *Metode penelitian*. Yogyakarta: Pustaka Pelajar.
- Badan Standar Nasional Pendidikan. (2006). *Panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah*. Jakarta: BSNP. Retrieved from [http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/Panduan\\_Umum\\_KTSP.pdf](http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/Panduan_Umum_KTSP.pdf)
- Bashooir, K., & Supahar, S. (2016). Analisis aspek kinerja literasi sains pada materi kalor Fisika. *UPEJ Unnes Physics Education Journal*, 5(1). Retrieved from <https://journal.unnes.ac.id/sju/index.php/upej/article/view/12711>
- Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics*, 112(1), 3–11. <https://doi.org/10.1111/j.1949-8594.2011.00109.x>
- Chiappetta, E. L., & Koballa, T. R. (2010). *Science instruction in the middle and secondary schools: developing fundamental knowledge and skills* (7th ed.). USA: Pearson Education, Inc.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological*

- Assessment*, 6(4), 284–290.  
<https://doi.org/10.1037/1040-3590.6.4.284>
- Departemen Pendidikan dan Kebudayaan. (2001). *Kamus besar babasa Indonesia* (3rd ed.). Jakarta: Balai Pustaka.
- Department of Education. (2009). *Report of the STEM review*. Retrieved from [https://www.education-ni.gov.uk/sites/default/files/publications/de/Report of the STEM Review 2009\\_1.PDF](https://www.education-ni.gov.uk/sites/default/files/publications/de/Report%20of%20the%20STEM%20Review%202009_1.PDF)
- Gonzalez, H. B., & Kuenzi, J. J. (2012). Science, technology, engineering, and mathematics (STEM) education: a primer. Retrieved from <https://fas.org/sgp/crs/misc/R42642.pdf>
- Hernandez, P. R., Bodin, R., Elliott, J. W., Ibrahim, B., Rambo-Hernandez, K. E., Chen, T. W., & de Miranda, M. A. (2014). Connecting the STEM dots: measuring the effect of an integrated engineering design intervention. *International Journal of Technology and Design Education*, 24(1), 107–120. <https://doi.org/10.1007/s10798-013-9241-0>
- Ismail, I., Permanasari, A., & Setiawan, W. (2016). Efektivitas virtual lab berbasis STEM dalam meningkatkan literasi sains siswa dengan perbedaan gender. *Jurnal Inovasi Pendidikan IPA*, 2(2), 190. <https://doi.org/10.21831/jipi.v2i2.8570>
- Kartowagiran, B., & Jaedun, A. (2016). Model asesmen autentik untuk menilai hasil belajar siswa sekolah menengah pertama (SMP): implementasi asesmen autentik di SMP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 131. <https://doi.org/10.21831/pep.v20i2.10063>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- OECD. (2014). *PISA 2012 results: what students know and can do student performance in mathematics, reading and science volume I*. Paris: OECD Publishing.
- P21. (2009). 21st century skills map. Retrieved from [http://www.p21.org/storage/documents/21st\\_century\\_skills\\_english\\_map.pdf](http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf)
- Presiden Republik Indonesia. Undang-Undang Republik Indonesia nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional (2003). Indonesia.
- Reeve, E. M. (2013). Implementing science, technology, mathematics, and engineering (STEM) education in Thailand and in ASEAN. Retrieved from [http://dpst-apply.ipst.ac.th/specialproject/images/IPST\\_Global/document/Implementing STEM in ASEAN - IPST May 7 2013 - Final.pdf](http://dpst-apply.ipst.ac.th/specialproject/images/IPST_Global/document/Implementing%20STEM%20in%20ASEAN%20-%20IPST%20May%207%202013%20-%20Final.pdf)
- Retnawati, H. (2016). *Validitas reliabilitas dan karakteristik butir*. Yogyakarta: Parama Publishing.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18839484>
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. [https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18)
- Subali, B., & Suyata, P. (2012). *Pengembangan item tes konvergen dan divergen: penyelidikan*

- validitasnya Secara empiris*. Yogyakarta: Diandra Pustaka Indonesia.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan RASCH pada assessment pendidikan*. Cimahi: Tim Komunikata Publishing House.
- Supahar, & Prasetyo, Z. K. (2015). Pengembangan instrumen penilaian kinerja kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 96–108. Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/4560>
- Supahar, S. (2014). The estimation of inquiry performance test items of high school physics subject with quest program. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Supahar, S. (2015). Applying content validity ratios (CVR) to the quantitative content validity of physics learning achievement tests. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Wagner, T. (2008). *The global achievement gap*. New York: Basic Book.
- Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, 28(2–3), 291–314. <https://doi.org/10.1080/09500690500336973>