

Karakteristik instrumen *integrated assessment* untuk mengukur *critical thinking skills* dan *science process skills*

Anggi Ristiyana Puspita Sari^{1,a*}, Suyanta Suyanta^{2,b}

¹ Universitas Palangka Raya, Jln. Yos Sudarso Palangka Raya, 73111, Kalimantan Tengah, Indonesia

² Universitas Negeri Yogyakarta. Jalan Colombo No. 1, Yogyakarta, 55281, Indonesia

* Corresponding Author. E-mail: ^a airistiyana@gmail.com; ^b suyanta@uny.ac.id

Received: 31 January 2021; Revised: 3 March 2021; Accepted: 17 April 2021

Abstrak: Penelitian ini bertujuan untuk mendeskripsikan karakteristik instrumen *integrated assessment*. Penelitian ini menggunakan model pengembangan instrumen tes dari McIntire. Produk awal ditinjau oleh *peer reviewers* dan divalidasi oleh *experts*. Tahap uji coba instrumen melibatkan 392 peserta didik kelas XI IPA di tiga SMA di Kota Yogyakarta. Hasil uji coba instrumen berupa data politomus yang dianalisis dengan pendekatan PCM 1-PL menggunakan program *Winsteps*. Instrumen pengumpulan data meliputi lembar validasi butir soal, lembar angket respon pengguna instrumen, dan instrumen tes berbentuk uraian. Karakteristik instrumen tes terdiri atas validitas isi, validitas konstruk, tingkat kesukaran butir, fungsi informasi tes, kesalahan pengukuran, dan reliabilitas. Hasil uji coba menunjukkan bahwa instrumen *integrated assessment* dinyatakan valid secara logis dan konstruk, tingkat kesukaran butir berada pada kategori sedang dan fungsi informasi tes sebesar 6,819 pada kemampuan $-0,8$ logit dengan kesalahan pengukuran sebesar 0,383 sehingga instrumen *integrated assessment* dinyatakan reliabel untuk mengukur *critical thinking skills* dan *science process skills* peserta didik.

Kata Kunci: instrumen tes, *integrated assessment*, karakteristik

The characteristic of an integrated assessment instrument to measure critical thinking skills and science process skills

Abstract: The purpose of this study is to describe the characteristics of the integrated assessment instrument. This study used the McIntire development model. The initial product was observed by peer reviewers and validated by experts. The pilot testing involved 392 grade XI students from three senior high schools in Yogyakarta City. The results of the pilot testing were polytomous data which were analyzed by means of the PCM 1-PL approach using *Winsteps*. The data-collecting instrument included a question item validation sheet, questionnaire of instrument user responses, and essay test. The characteristics of the test instrument consist of logic and construct validity, item difficulty, item information function, standard error measurement, and reliability. The integrated assessment instrument is valid in content and construct, item difficulty of integrated assessment instrument has a medium category, and test information function is obtained at 6.819 on the ability -0.8 logit with standard error of estimation at 0.383, and thus the instrument is reliable to measure students' critical thinking skills and science process skills.

Keywords: characteristic, integrated instrument, test instrument

How to Cite: Sari, A., & Suyanta, S. (2021). Karakteristik instrumen integrated assessment untuk mengukur critical thinking skills dan science process skills. *Measurement in Educational Research*, 1(1), 26-38. doi:<http://dx.doi.org/10.33292/meter.v1i1.108>



PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi yang semakin pesat berdampak langsung pada berbagai bidang kehidupan termasuk bidang pendidikan. Setiap sekolah hendaknya mampu membekali anak didiknya agar memiliki kemampuan dalam mengintegrasikan pengetahuan yang diperoleh dengan kehidupan nyata dalam menjawab tantangan global. Bukanlah perkara yang mudah untuk mewujudkan hal tersebut karena dibutuhkan keterampilan dalam mengintegrasikannya. Salah satu keterampilan yang harus dimiliki peserta didik dalam pembelajaran adalah keterampilan proses. Keterampilan proses merupakan kompetensi yang diperlukan untuk menghadapi kompetensi global (*skills dimention*) (Mansilla & Jackson, 2011, p.53).



Model pembelajaran berbasis peningkatan keterampilan proses sains (*science process skills*) adalah model pembelajaran yang mengintegrasikan *science process skills* ke dalam sistem penyajian materi secara terpadu (Beyer, 1991, p.112). Leonor (2015) mengemukakan bahwa *science process skills* harus diterapkan kepada peserta didik karena secara tidak langsung akan melibatkan mereka dalam berbagai kegiatan inkuiri serta mampu mengarahkan mereka untuk menerapkan keterampilan proses sains dasar dan terintegrasi. *Science process skills* dibagi menjadi dua bagian yaitu *basic* dan *integrated science process learning* (Ministry of Education Malaysia, 2002; Rezba, Sprague, & McDonnough, 2007, pp.28-133; Chabalengula, Mumba, & Mbewe, 2011).

Science process skills kimia merupakan elemen penting yang harus dimiliki oleh peserta didik karena melibatkan keterampilan kognitif atau intelektual, manual, dan sosial yang digunakan dalam pemecahan masalah (Karamustafaoğlu, 2011). *Science process skills* dianggap memiliki korelasi dengan keterampilan berpikir kritis (*critical thinking skills*) karena memiliki hubungan erat dengan penguasaan konsep (Chebii, 2012). Peserta didik hendaknya memiliki *science process skills* untuk bisa lebih memahami konsep kimia, selain itu *science process skills* mendukung *life skills* peserta didik untuk menunjang di kehidupan mendatang.

Judge, Jones, dan McCreery (2009, p.2) mengemukakan bahwa *critical thinking skills* melibatkan pengolahan ide dan informasi berdasarkan pandangan objektif. Definisi lain diungkapkan oleh Cottrell (2005, p.2) yang mengemukakan bahwa *critical thinking* merupakan proses diskusi kompleks yang melibatkan keterampilan dan sikap yang mencakup: (1) mengidentifikasi suatu argumen dan kesimpulan milik orang lain; (2) mengevaluasi bukti untuk dijadikan pandangan alternatif; (3) mengemukakan argumen dan bukti dengan adil; (4) mampu membaca poin penting, melihat apa yang tak terlihat, dan mengidentifikasi asumsi yang salah; (5) merefleksikan masalah secara logis; (6) menggambarkan kesimpulan tentang kevalidan dan kebenaran suatu argumen; (7) merepresentasikan pandangan secara terstruktur dan beralasan yang mampu meyakinkan orang lain.

Kegiatan penilaian merupakan bagian integral dari sebuah proses pelaksanaan pembelajaran. Pembelajaran dan penilaian hasil belajar kimia harus memperhatikan karakteristik ilmu kimia sebagai proses dan produk. Adanya asumsi bahwa penguasaan konsep sains berdasarkan memori atau ingatan adalah salah (Aydm, 2011). Oleh karena itu, penilaian hasil belajar kimia harus mencakup berbagai aspek kemampuan peserta didik, sehingga setiap indikator yang merupakan kompetensi dasar spesifik dapat dijabarkan lebih lanjut ke dalam instrumen penilaian. Dengan demikian, proses pembelajaran dan penilaian hasil belajar merupakan dua hal penting yang tidak dapat dipisahkan (National Academy of Sciences, 1996, pp.5-6).

Kimia yang termasuk dalam rumpun IPA tidak terlepas dari kegiatan pembelajaran yang membutuhkan *critical thinking skills* dan *science process skills* peserta didik. Salah satu materi kimia yang terdapat dalam Kurikulum 2013 yaitu asam dan basa. Materi asam dan basa dapat mengasah *science process skills* peserta didik yang membutuhkan pembuktian melalui percobaan. *Science process skills* memastikan apakah peserta didik menguasai materi dengan baik (Askins, Douglas, Angie, Amy, Kamalesh, 2006; Sukarno, Permanasari, & Hamidah, 2013). Materi asam dan basa merupakan salah satu materi kimia yang kompleks karena terhubung dengan konsep kimia yang lain, seperti reaksi kimia, kesetimbangan kimia, stoikiometri (Demircioğlu, Ayas, & Demircioğlu, 2005).

Berdasarkan kompetensi dasar pada Kurikulum 2013, materi asam dan basa merupakan prasyarat untuk materi selanjutnya yaitu larutan penyangga dan hidrolisis. Materi asam dan basa juga erat kaitannya dengan kehidupan sehari-hari. Oleh karena itu, peserta didik diharapkan mampu menguasai konsep asam dan basa dengan mengasah keterampilan proses sains. Berdasarkan karakteristik materi asam dan basa tersebut penilaian yang selayaknya dilakukan mencakup aspek kognitif dan keterampilan proses sains.

Penelitian yang relevan dilakukan oleh Subali (2009) tentang pengukuran keterampilan proses sains pada mata pelajaran biologi menunjukkan bahwa produk instrumen yang dihasilkan berupa tes pengukur keterampilan proses sains pola divergen yang mencakup semua keterampilan proses sains, namun tidak semua aspek dan sub aspek keterampilan proses sains dirumuskan secara eksplisit dalam standar kompetensi dan kompetensi dasar pada standar isi mata pelajaran Biologi SMA.

Penelitian yang serupa dilakukan oleh Sukarno, Permanasari, dan Hamidah (2013) tentang pengukuran *science process skills* peserta didik mengindikasikan bahwa secara umum peserta didik tidak dapat

mencapai skor optimal pada tes *science process skills* dan menunjukkan bahwa pembelajaran sains di Jambi masih belum optimal melatih peserta didik dalam melakukan inferensi, prediksi, dan pengukuran.

Hasil observasi di lapangan yang dilakukan oleh peneliti menunjukkan bahwa masih banyak guru yang melakukan penilaian secara tidak menyeluruh dan lebih menekankan pada aspek kognitif. Tindakan tersebut hendaknya diubah dengan melakukan penilaian mencakup semua aspek, dengan begitu selama proses pembelajaran peserta didik juga tidak hanya memfokuskan dirinya pada aspek kognitif. Permasalahan tersebut dapat terjadi karena padatnnya materi kimia yang harus dipelajari peserta didik dengan waktu pembelajaran yang tidak sesuai, sehingga kebanyakan guru dalam melakukan penilaian lebih memfokuskan pada aspek kognitif.

Berdasarkan hasil wawancara dengan beberapa guru kimia SMA di kota Yogyakarta kendala lain yang muncul adalah masih terbatasnya ketersediaan instrumen tes yang mampu mengukur aspek kognitif dan keterampilan proses peserta didik secara terintegrasi. Beberapa instrumen yang digunakan guru juga belum memenuhi kualitas sebagai instrumen, diantaranya instrumen yang digunakan belum mencakup seluruh indikator pembelajaran. Kondisi tersebut dapat menyebabkan hasil penilaian yang diperoleh masih memiliki beberapa kelemahan. Selain itu, beberapa instrumen yang digunakan untuk mengevaluasi hasil belajar peserta didik juga belum memiliki pedoman penskoran yang jelas, sehingga penilaian yang dilakukan belum bersifat objektif.

Guntur, Sukardiyanto, & Mardapi (2014) mengungkapkan bahwa keberhasilan dalam melaksanakan penilaian hasil belajar akan sangat ditentukan oleh kemampuan guru dalam mengkonstruksi alat ukur dan menggunakan alat ukur tersebut. Apabila keseluruhan keseluruhan kemampuan itu tidak dikuasai oleh guru, maka tidak menutup kemungkinan akan terjadi kesalahan dalam pengukuran hasil belajar yang pada gilirannya akan mengakibatkan kerugian bagi peserta didik. Brookhart (2010, p.18) menegaskan bahwa prinsip dasar penilaian adalah memastikan penilaian agar sesuai dengan tujuan pembelajaran, sehingga jenis penilaian dapat ditentukan.

Integrated assessment dapat mengevaluasi aspek kognitif dan keterampilan proses secara bersamaan. Apabila pembelajaran yang dilakukan lebih ditekankan pada pemanfaatan keterampilan proses sains (*science process skills*), maka akan berdampak positif terhadap hasil belajar peserta didik (Durmaz & Mutlu, 2014). Birenbaum, Breuer, Cascallar, Dochy, Dori, Ridgway, & Wiesemes (2011) mengemukakan bahwa guru harus menyadari bahwa *integrated assessment* fokus pada penilaian pembelajaran bukan penilaian untuk pembelajaran, dengan kata lain penilaian dilakukan untuk menilai kemampuan yang dimiliki peserta didik selama proses pembelajaran. Pernyataan serupa juga diungkapkan oleh Hunter, Roy, & Hazel (2003) bahwa penilaian tersebut dilakukan tidak hanya untuk mengurutkan peringkat peserta didik, tetap tujuan utamanya yaitu meningkatkan pembelajaran peserta didik.

Dorman, Waldrip, & Fisher (2009) menyatakan bahwa *integrated assessment* dapat digunakan untuk mendukung keterlibatan peserta didik dalam kegiatan pembelajaran. *Integrated assessment* merupakan sebuah penilaian terpadu yang mengintegrasikan penilaian penguasaan materi ajar dan keterampilan proses sains. Dalam *integrated assessment*, setiap soal memiliki dua indikator sekaligus, yaitu indikator *critical thinking skills* dan indikator *science process skills*.

Berdasarkan penelitian pendahuluan dan beberapa permasalahan yang telah dipaparkan, penelitian ini berusaha memberikan solusi berupa pengembangan instrumen alternatif yaitu instrumen terintegrasi (*integrated assessment*) yang dapat mengukur keterampilan berpikir kritis (*critical thinking skills*) dan keterampilan proses sains (*science process skills*) peserta didik SMA pada materi asam dan basa. Instrumen yang dikembangkan berupa soal uraian yang dikembangkan menggunakan teori respon butir (*item response theory*, IRT), sehingga diharapkan memiliki validitas dan reliabilitas yang baik dan dapat digunakan untuk skala luas.

Masalah utama yang diangkat pada penelitian ini adalah "Bagaimanakah karakteristik butir soal instrumen *integrated assessment* untuk mengukur *critical thinking skills* dan *science process skills* kimia peserta didik pada materi asam dan basa?"

Sejalan dengan rumusan masalah yang akan diselesaikan, maka tujuan penelitian ini adalah mendeskripsikan karakteristik instrumen *integrated assessment* untuk mengukur *critical thinking skills* dan *science process skills* kimia peserta didik pada materi asam dan basa.

Kegunaan penelitian ini bagi peserta didik adalah memperoleh penilaian yang menyeluruh mencakup *critical thinking skills* dan *science process skills* dengan penerapan *integrated assessment*, bagi guru dapat

memberikan alternatif instrumen penilaian untuk mengukur *critical thinking skills* dan *science process skills* peserta didik SMA yang telah dikembangkan pada pembelajaran kimia pada materi pokok asam dan basa, sedangkan bagi peneliti lain dapat menambah ilmu pengetahuan dan menjadikan inspirasi untuk mengembangkan instrumen penilaian yang ideal sesuai dengan keadaan sehingga tujuan pembelajaran dapat tercapai.

METODE

Penelitian ini merupakan jenis penelitian dan pengembangan atau *Research and Development (R&D)* yang bertujuan untuk menghasilkan produk dan menguji keefektifan produk. Pada penelitian ini produk yang dikembangkan adalah instrumen *integrated assessment* yang dapat mengukur pencapaian *critical thinking skills* dan *science process skills* kimia peserta didik SMA kelas XI pada materi asam dan basa. Pengembangan produk ini mengadaptasi model pengembangan instrumen tes dari McIntire (Mulyatiningsih, 2012, pp.166-180) yang terdiri atas 9 tahapan.

Waktu dan Tempat Penelitian

Penelitian dilakukan pada awal bulan Februari hingga Maret 2016 yang berlokasi di tiga SMA di Kota Yogyakarta meliputi SMA Negeri 1 Yogyakarta, SMA Negeri 2 Yogyakarta, dan SMA Negeri 4 Yogyakarta.

Subjek Penelitian

Besarnya ukuran sampel yang diambil dari populasi menggunakan acuan Linacre (2015) yaitu minimal untuk model 1-PL sebanyak 150-250. Berdasarkan pertimbangan tersebut, subjek coba yang ditetapkan dalam penelitian pengembangan ini adalah 392 peserta didik kelas XI IPA SMA tahun ajaran 2015/2016 di kota Yogyakarta, yaitu SMA Negeri 1 Yogyakarta, SMA Negeri 2 Yogyakarta, dan SMA Negeri 4 Yogyakarta.

Pengambilan sampel dilakukan menggunakan teknik *purposive sampling*. Teknik sampling tersebut digunakan karena peneliti mempunyai pertimbangan-pertimbangan tertentu didalam pengambilan sampelnya berdasarkan informasi yang diperoleh (Fraenkel & Wallen, 1993, p.101). Pertimbangan yang dimaksud seperti sampel yang dipilih dapat mendukung data yang akan diperoleh.

Prosedur

Penelitian ini difokuskan untuk menghasilkan instrumen *integrated assessment*. Prosedur model pengembangan instrumen tes dari McIntire yang terdiri atas 9 tahap meliputi: (1) analisis domain tes, sasaran, dan tujuan; (2) pembuatan rancangan produk; (3) pembuatan butir-butir soal tes; (4) pembuatan petunjuk instrumen; (5) uji coba instrumen; (6) analisis butir secara kuantitatif; (7) revisi instrumen; (8) pengukuran; (9) penyusunan produk akhir.

Uji coba instrumen meliputi tiga tahapan yaitu uji coba tahap I, uji coba tahap II, dan uji coba tahap III. Pada uji coba tahap I produk ditinjau oleh 3 *peer reviewers*, selanjutnya pada uji coba tahap II produk ditinjau dan dikaji oleh *experts judgment*. pada uji coba yang terakhir produk diujicobakan secara luas ke 392 peserta didik kelas XI IPA.

Data, Instrumen, dan Teknik Pengumpulan Data

Data yang diperoleh merupakan data politomus. Instrumen yang digunakan dalam penelitian ini berupa instrumen penilaian *integrated assessment* dan lembar angket yang terdiri atas angket respon pengguna dan lembar validasi.

Lembar angket respon pengguna instrumen digunakan untuk menilai kelayakan instrumen *integrated assessment* yang telah dikembangkan dari aspek substansi, konstruksi, kebahasaan, validitas, dan praktikalitas. Oleh karena itu, lembar angket respon pengguna ini dapat digunakan untuk mengukur tingkat keterbacaan instrumen *integrated assessment*. Lembar angket respon pengguna ditujukan untuk guru kimia SMA. Lembar validasi ditujukan untuk *experts judgement*, yaitu ahli evaluasi dan ahli materi pembelajaran, dan guru kimia SMA untuk menilai kelayakan produk *integrated assessment* yang dikembangkan. Instrumen penilaian *integrated assessment* merupakan instrumen tes yang menjadi produk akhir hasil pengembangan yang digunakan untuk mengukur *critical thinking skills* dan *science process skills* kimia peserta didik SMA kelas XI pada materi asam dan basa. Instrumen penilaian ini berupa soal

uraian yang terdiri atas 11 butir yang diskor menggunakan model politomus sesuai dengan pedoman penskoran.

Teknik pengumpulan data meliputi teknik angket dan tes. Teknik angket dilakukan dengan mengumpulkan data menggunakan instrumen lembar angket yang berupa lembar validasi dan angket pengguna respon instrumen, sedangkan Pengumpulan data kualitas instrumen dilakukan melalui uji coba lapangan kepada peserta didik kelas XI IPA di beberapa SMA di kota Yogyakarta.

Teknik Analisis Data

Teknik analisis data yang digunakan adalah analisis data kuantitatif untuk mengetahui validitas, tingkat kesukaran butir soal, fungsi informasi tes, kesalahan baku pengukuran, dan reliabilitas sebagai berikut.

Validitas Isi

Miller (2008, p.84) menyatakan bahwa validitas isi merupakan validitas yang diestimasi lewat pengujian terhadap isi tes dengan analisis rasional atau lewat *expert judgment*. Analisis tentang validitas instrumen yang digunakan untuk penelitian ini menggunakan formula Aiken untuk menghitung *content-validity coefficient* yang didasarkan pada hasil penilaian dari panel ahli sebanyak n orang terhadap suatu butir dari segi sejauh mana butir tersebut mewakili konstruk yang diukur. Aiken merumuskan formula Aiken's V (Aiken, 1985, p.133) sesuai dengan persamaan sebagai berikut.

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

Keterangan:

s = r - lo

lo = angka penilaian validitas yang terendah (misalnya 1)

c = angka penilaian validitas tertinggi (misalnya 4)

r = angka yang diberikan oleh penilai

Validitas Konstruk

Validitas konstruk adalah validitas yang menunjukkan sejauhmana instrumen mengungkap suatu kemampuan atau konstruk teoretis tertentu yang hendak diukurnya (Nunnally & Brenstein, 1994, p.86; Miller, 2008, p.83). Validitas konstruk diperoleh dari data hasil uji coba lapangan. Data dianalisis dengan menggunakan teknik analisis faktor eksploratori (*exploratory factor analysis*, EFA) yang menggunakan program *SPSS 16*. Interpretasi dari analisis ini akan digunakan sebagai asumsi *undimensi* untuk mengetahui estimasi banyaknya dimensi yang dapat diukur dari instrumen yang telah disusun.

Fungsi Informasi Tes

Fungsi informasi tes dianalisis menggunakan model PCM 1-PL dengan bantuan program *Quest* yang kemudian dihitung secara manual menggunakan *Microsoft Excel*. Untuk fungsi informasi butir dianalisis menggunakan persamaan sebagai berikut.

$$I_i(\theta) = \sum_{x=0}^n \frac{D^2 e^{D(\theta-\delta_i)}}{[1+e^{D(\theta-\delta_i)}]^3}$$

Berdasarkan fungsi informasi butir politomus, dapat ditentukan fungsi informasi tes dengan menggunakan persamaan (Hambleton, Swaminathan, & Rogers, 1991, p.91) sebagai berikut.

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

Sejumlah butir yang memiliki nilai informasi tinggi harus dipilih guna menghasilkan instrumen yang baik.

Kesalahan Baku Pengukuran

Setelah fungsi informasi tes diketahui, kesalahan baku estimasi pengukuran dapat ditentukan dengan menggunakan persamaan sebagai berikut.

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

SE (θ) menyatakan kesalahan baku estimasi dan $I(\theta)$ menyatakan fungsi informasi tes (Hambleton, Swaminathan, & Rogers, 1991, p.94). Kurva hubungan antara TIF dan SE dapat dibuat setelah nilai TIF dan SE diketahui. Kemudian melalui kurva tersebut dapat ditentukan reliabilitas instrumen tes dengan cara melihat titik perpotongan kurva TIF dan SE.

HASIL DAN PEMBAHASAN

Hasil Tahapan Uji Coba Peer Reviewer

Produk awal ditinjau oleh tiga *peer reviewer* yang melakukan penelitian pengembangan serupa. Penilaian dari *peer reviewer* digunakan untuk memperbaiki draf awal sebelum diujicobakan kepada *experts judgment*. Hasil uji coba pada tahap ini berupa saran dan masukan yang selanjutnya dijadikan sebagai bahan pertimbangan. Penentuan saran dan masukan dari *peer-reviewer* yang layak digunakan untuk perbaikan produk tentunya melibatkan *expert judgment*, sehingga tidak semua saran dan masukan dari *peer reviewer* digunakan. Berdasarkan tinjauan *peer reviewer* butir 1-8 perlu diperbaiki, sedangkan butir 9-11 diterima.

Hasil Uji Coba Expert Judgment

Validasi oleh *expert judgment* dilakukan setelah revisi produk awal sesuai dengan hasil uji coba tahap I, yaitu uji coba oleh *peer reviewer*. Produk instrumen yang ditinjau dan dikaji oleh *expert judgment* meliputi dosen ahli evaluasi, dosen ahli materi pembelajaran, dan guru kimia SMA bertujuan untuk memperoleh validitas isi, sehingga dapat digunakan untuk mengukur *critical thinking skills* dan *science process skills* kimia peserta didik. Berdasarkan validasi dari *experts judgment* butir 6, 8, 9 diterima, sedangkan butir yang lainnya perlu diperbaiki. Selain perbaikan dari segi bahasa dan konsep, perbaikan yang disarankan oleh *expert judgment* adalah penggantian nomor butir soal, sehingga kisi-kisi instrumen tes mengalami perubahan.

Selain data kualitatif, pada tahap validasi oleh *expert judgment* juga diperoleh data kuantitatif berdasarkan lembar angket validasi. Hasil indeks Aiken's V untuk tiap butir disajikan pada Tabel 1.

Tabel 1. Nilai Aiken's V untuk Tiap Butir

No.	Butir	Nilai Aiken's V
1.	Butir 1	0,83
2.	Butir 2	0,75
3.	Butir 3	0,88
4.	Butir 4	0,83
5.	Butir 5	0,75
6.	Butir 6	0,90
7.	Butir 7	0,80
8.	Butir 8	0,88
9.	Butir 9	0,96
10.	Butir 10	0,92
11.	Butir 11	0,88

Berdasarkan Tabel 1 diperoleh indeks Aiken's V terendah pada butir nomor 2 dan 5 yaitu sebesar 0,75 dan tertinggi pada butir nomor 9 yaitu sebesar 0,96. Perolehan hasil tersebut menyatakan bahwa seluruh butir dapat dinyatakan valid menurut tabel koefisien validitas Aiken (Aiken, 1985, p.134). Pada tabel tersebut dinyatakan bahwa untuk jumlah rater sebanyak 8 dan skala jawaban pada lembar angket menggunakan skala 4, nilai indeks Aiken's V minimal yang harus diperoleh pada taraf signifikansi 1% sebesar 0,83 sedangkan pada taraf signifikansi 5% sebesar 0,75. Oleh karena itu, seluruh butir pada instrumen tes yang dikembangkan dapat dinyatakan valid pada kedua taraf signifikansi.

Hasil Uji Asumsi Pemodelan Rasch

Sebelum melakukan analisis dengan pemodelan Rasch, terdapat tiga asumsi yang harus dipenuhi terlebih dahulu meliputi unidimensi, independensi lokal, dan invariansi parameter termasuk didalamnya parameter butir dan parameter kemampuan.

Uji unidimensi dilakukan dengan analisis faktor eksploratori (*Exploratory Faktor Analysis*, EFA) menggunakan program *SPSS 16*. Hasil analisis faktor menandakan kesahihan suatu konstruk atau validitas secara empiris. Hasil output analisis faktor meliputi statistik *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO), uji Bartlett, *scree plot*, dan *anti-image matrices*. Hasil analisis faktor disajikan pada Tabel 2.

Tabel 2. Hasil Uji KMO dan Bartlett

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0,714
Bartlett's Test of Sphericity	Approx. Chi-Square	446,201
	Df	0,55
	Sig	0,000

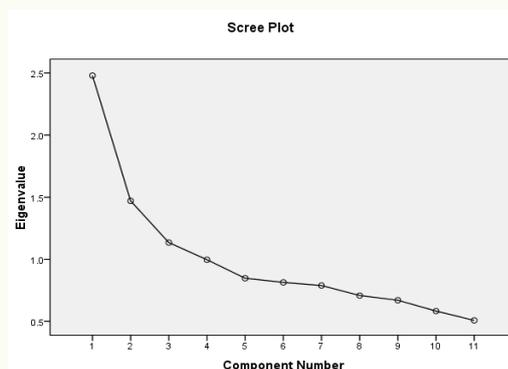
Berdasarkan Tabel 2 nilai KMO yang diperoleh sebesar 0,714 dan hasil analisis faktor menunjukkan nilai *Chi-square* pada uji Bartlett sebesar 446,201 dengan derajat kebebasan 0,55 dan nilai signifikansi 0,000. Hasil tersebut menunjukkan bahwa ukuran sampel sebesar 392 pada penelitian ini telah cukup dan termasuk dalam kategori baik karena nilai KMO yang diperoleh lebih besar dari 0,5 dan uji Bartlett memiliki signifikansi kurang dari 0,05 (Beavers, Lounsbury, Richards, Huck, Skolits, & Esquivel, 2013).

Oleh karena hasil tersebut termasuk dalam kategori baik, maka untuk mengetahui asumsi unidimensi dapat dianalisis lebih lanjut menggunakan analisis faktor. Berdasarkan kriteria Kaiser (Beavers, Lounsbury, Richards, Huck, Skolits, & Esquivel, 2013) dapat dinyatakan bahwa instrumen tes *integrated assessment* memuat tiga faktor. Tabel 3 menunjukkan bagaimana variansi dibagi menurut 3 kemungkinan faktor. Apabila nilai eigen kurang dari satu, maka faktor tersebut tidak dapat dijelaskan lebih lanjut karena kurangnya informasi. Tiga faktor tersebut dapat menjelaskan sekitar 46,217% dari total variansi.

Tabel 3. Nilai Eigen dan Komponen Varians Hasil Analisis Faktor

No. Komponen	Eigen Value	Proporsi (%)	Kumulatif (%)
1	2,479	22,539	22,539
2	1,470	13,365	35,904
3	1,134	10,313	46,217
4	0,996	9,057	55,274
5	0,847	7,702	62,976
6	0,814	7,400	70,375
7	0,789	7,171	77,547
8	0,708	6,4439	83,986
9	0,670	6,095	90,081
10	0,583	5,301	95,381
11	0,508	4,619	100,00

Selain menggunakan tabel *total variance explained*, banyaknya faktor yang terbentuk juga dapat diamati melalui *scree plot* yang disajikan pada Gambar 1. Pada *scree plot* tersebut dapat dicermati bahwa terdapat 3 curaman, sedangkan dari komponen nomor 4 dan seterusnya mulai menunjukkan grafik yang landai dan memiliki nilai eigen yang hampir sama (lebih kecil dari nilai eigen pada faktor yang sebelumnya). Pada praktiknya tidaklah mudah untuk menemukan butir yang murni mengukur hanya satu dimensi pada ciri peserta.



Gambar 1. Scree Plot Hasil Analisis Faktor Eksploratori

Reckase (1979) menegaskan bahwa unidimensi dapat dilihat dari nilai eigen pertama yang tidak kurang dari 20%. Hasil yang diperoleh menunjukkan bahwa nilai eigen yang pertama sebesar 22.539% dan lebih besar dari nilai eigen yang kedua dan seterusnya. Dengan demikian, instrumen tes *integrated assessment* dapat dikatakan memenuhi syarat asumsi unidimensi untuk model IRT. Oleh karena syarat unidimensi telah terpenuhi, maka dapat pula dinyatakan bahwa validasi konstruk untuk instrumen *integrated assessment* yang dikembangkan juga telah terpenuhi.

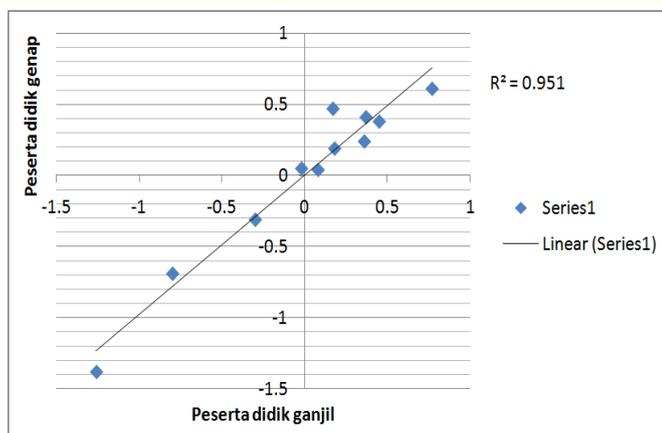
Selain uji asumsi unidimensi, IRT juga mensyaratkan adanya uji asumsi independensi lokal. Uji asumsi independensi lokal akan terpenuhi bila jawaban peserta didik terhadap suatu butir soal tidak mempengaruhi jawaban peserta didik terhadap butir soal yang lain. Dengan demikian, skor dari satu butir tidak boleh ditentukan atau bergantung pada skor butir yang lainnya. Retnawati (2014, p.7) menegaskan bahwa asumsi ini otomatis terbukti setelah dibuktikan dengan unidimensiolitas data respon peserta didik terhadap suatu tes.

Gambar 2 menunjukkan nilai varians-kovarians antarkelompok kemampuan peserta didik. Berdasarkan Gambar 2 terlihat bahwa keseluruhan elemen matriks diluar diagonal utama mendekati nol, sehingga dapat disimpulkan bahwa asumsi independensi lokal terpenuhi. Dengan demikian, setiap butir yang terdapat dalam instrumen tes *integrated assessment* tidak bergantung dengan butir yang lainnya.

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10
Column 1	0.162									
Column 2	0.022	0.004								
Column 3	0.024	0.003	0.005							
Column 4	0.019	0.003	0.003	0.004						
Column 5	0.021	0.003	0.003	0.002	0.004					
Column 6	0.016	0.002	0.003	0.002	0.002	0.004				
Column 7	0.020	0.004	0.003	0.003	0.002	0.002	0.004			
Column 8	0.034	0.004	0.005	0.004	0.005	0.005	0.004	0.009		
Column 9	0.024	0.005	0.003	0.004	0.003	0.002	0.005	0.005	0.006	
Column 10	0.083	0.011	0.014	0.011	0.105	0.013	0.10	0.022	0.012	0.068

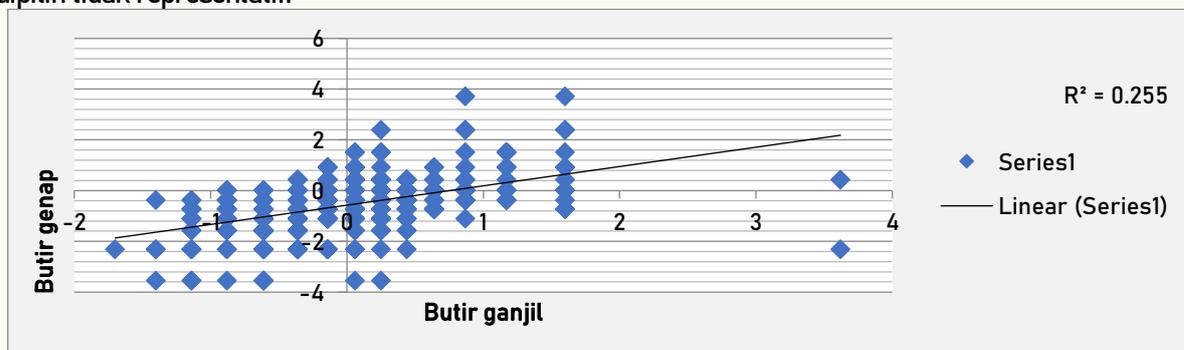
Gambar 2. Hasil Uji Asumsi Independensi Lokal

Uji asumsi yang terakhir adalah uji asumsi invariansi parameter. Invariansi parameter dapat dibuktikan dengan invariansi parameter butir dan parameter kemampuan peserta didik. Gambar 3 merupakan plot invariansi parameter butir. Berdasarkan hasil *scatter plot* dapat dilihat bahwa sebarannya mendekati garis diagonal yang menyatakan bahwa korelasinya tinggi. Selain itu dapat dikatakan pula bahwa parameter butir bersifat invarian. Nilai R^2 yang diperoleh sebesar 0.951 yang berarti nilai R nya sebesar 0.975 dan menandakan bahwa nilai R mendekati 1. Dengan demikian, asumsi invariansi parameter butir telah terpenuhi.



Gambar 3. Scatter Plot Invariansi Parameter Butir

Selain invariansi parameter butir, dilakukan pula analisis invariansi parameter kemampuan. Hasil *scatter plot* invariansi parameter kemampuan dapat dilihat pada Gambar 4. Berdasarkan hasil *scatter plot* dapat dilihat bahwa sebarannya tidak mendekati garis diagonal. Hasil tersebut menunjukkan bahwa invariansi parameter kemampuan tidak terbukti karena kecenderungan pola titik-titik yang tidak mendekati garis, sehingga tidak dapat dikatakan bahwa parameter kemampuan peserta didik bersifat invarian. Kondisi tersebut dapat disebabkan oleh teknik sampling yang tidak *random*, sehingga sampel yang dipilih tidak representatif.



Gambar 4. Scatter Plot Invariansi Parameter Kemampuan

Hasil Uji Coba Instrumen

Kecocokan Butir (Item Fit)

Hasil analisis kecocokan butir (*item fit*) dilihat berdasarkan *output* fit1.txt dari program *Quest* yang ditunjukkan pada Gambar 5.

```

QUEST: The Interactive Test Analysis System
-----
Item Fit                                     31/ 3/16 13:12
all on all (N = 392 L = 11 Probability Level= .50)
-----
INFIT
MNSQ   .63   .71   .83   1.00   1.20   1.40   1.60
-----+-----+-----+-----+-----+-----+-----+
1 item 1      .                               *
2 item 2      .                               *
3 item 3      .                               *
4 item 4      .                               *
5 item 5      .                               *
6 item 6      .                               *
7 item 7      .                               *
8 item 8      .                               *
9 item 9      .                               *
10 item 10    .                               *
11 item 11    .                               *
    
```

Gambar 5. Output Item Fit dari Program Quest

Pada Gambar 5 dapat dilihat bahwa butir nomor 1 memiliki nilai INFIT MNSQ 1,3 meskipun begitu butir nomor 1 tetap dinyatakan *fit* karena masih berada dalam garis vertikal. Dengan demikian, hasil analisis tersebut diperoleh 11 butir soal yang *fit* karena berada dalam garis vertikal atau memiliki nilai INFIT MNSQ diantara 0,77 sampai dengan 1,30. Kondisi tersebut ditunjukkan dengan tanda bintang (*) terletak dalam jalur INFIT MNSQ. Dengan batas penerimaan butir menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,3), maka seluruh butir sebanyak 11 butir seluruhnya dinyatakan *fit* dengan model PCM 1-PL

Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir merupakan salah satu faktor yang mempengaruhi probabilitas jawaban peserta tes dalam merespon butir tertentu. Butir yang memiliki tingkat kesukaran yang relatif tinggi memerlukan kemampuan tinggi untuk dapat dijawab dengan benar, sedangkan butir yang mudah memerlukan tingkat kemampuan yang rendah untuk dijawab dengan benar.

Indeks tingkat kesukaran butir biasanya berkisar antara kira-kira $-2,0$ logit hingga $+2,0$ logit (Hambleton & Swaminathan, 1985, p.36). Nilai yang semakin mendekati -2 logit menunjukkan karakteristik butir yang semakin mudah dan nilai yang mendekati $+2$ logit menunjukkan karakteristik butir yang semakin sulit bagi peserta ujian.

Instrumen *integrated assessment* yang dikembangkan ini diskor menggunakan penskoran politomus. Penskoran politomus memiliki tingkat kesukaran atau *threshold* lebih dari satu. Pada instrumen yang dikembangkan ini setiap butir memiliki skor yang berbeda, namun data yang digunakan untuk analisis bukanlah data skor melainkan data kategori. Setelah dilakukan penskoran, skor diubah menjadi kategori. Oleh karena tiap butir memiliki skor yang berbeda, maka jumlah kategori tiap butirnya pun juga bervariasi.

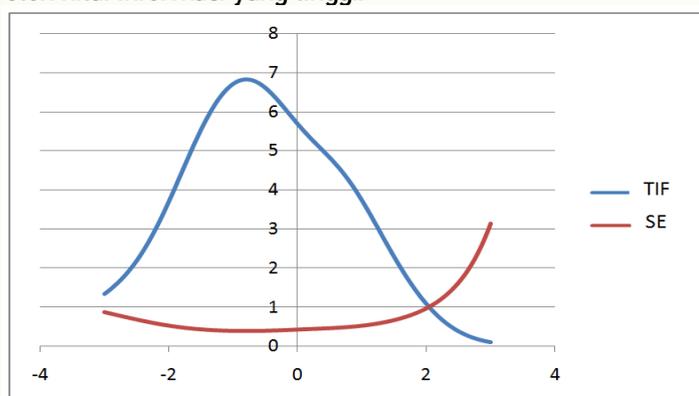
Tabel 4. Tingkat Kesukaran Butir

Butir	Measure (dalam logit)	Keterangan
1	-1,32	Sangat mudah
2	0,32	Sedang
3	0,39	Sedang
4	0,69	Sukar
5	-0,75	Mudah
6	0,19	Sedang
7	0,02	Sedang
8	-0,30	Sedang
9	0,30	Sedang
10	0,41	Sedang
11	0,06	Sedang

Hasil analisis tingkat kesukaran butir dapat dilihat pada Tabel 4. Berdasarkan Tabel 15 dapat diamati bahwa rentang *measure* yang diperoleh dari $-1,32$ logit sampai dengan $+0,69$ logit. Butir yang memiliki tingkat kesukaran paling mudah adalah butir 1, sedangkan butir dengan tingkat kesukaran paling sulit adalah butir 4. Secara keseluruhan, instrumen tes *integrated assessment* untuk mengukur *critical thinking skills* dan *science process skills* memiliki tingkat kesukaran pada kategori sedang atau cukup.

Fungsi Informasi Tes dan Kesalahan Baku Pengukuran

Analisis fungsi informasi dilakukan terhadap tiap butir terlebih dahulu (*Item Information Function*, IIF). Visualisasi analisis fungsi informasi tes dan kesalahan baku pengukuran dapat dilihat pada Gambar 6. Pada visualisasi tersebut terlihat bahwa nilai fungsi informasi tes dan kesalahan baku terletak pada rentang kemampuan -4 logit $\leq \theta \leq +4$ logit. Gambar 6 mengindikasikan bahwa pada kemampuan peserta didik kategori rendah memberikan nilai fungsi informasi yang rendah; sedangkan pada kemampuan peserta didik kategori sedang diperoleh nilai informasi yang tinggi.



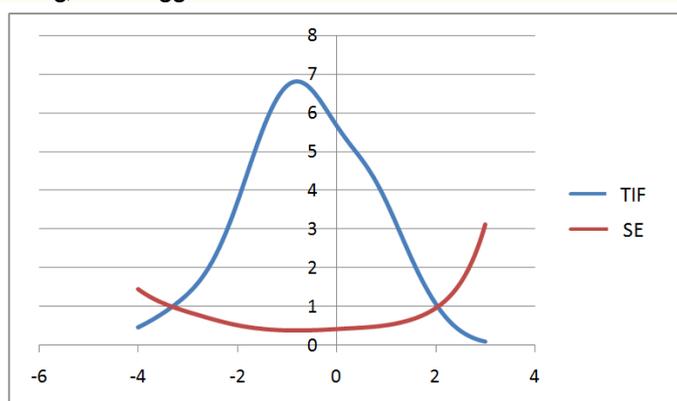
Gambar 6. Fungsi Informasi Tes

Berdasarkan hasil perhitungan dapat diketahui bahwa estimasi fungsi informasi tes diperoleh sebesar 6,819 dengan kesalahan baku pengukuran sebesar 0,383. Kondisi tersebut mengindikasikan bahwa pada saat nilai fungsi informasi tinggi akan diperoleh nilai kesalahan baku pengukuran yang kecil. Semakin besar fungsi informasi tes maka semakin kecil kesalahan baku pengukurannya, sehingga dapat

disimpulkan bahwa besar kecilnya nilai fungsi informasi tes sangat mempengaruhi kesalahan baku pengukuran.

Reliabilitas

Reliabilitas instrumen dapat diketahui dengan melihat titik perpotongan garis TIF dengan SE. Oleh karena pada Gambar 6 hanya terdapat 1 perpotongan garis, maka dilakukan perpanjangan kurva (ekstrapolasi). Ekstrapolasi kurva fungsi informasi tes dan kesalahan baku pengukuran dapat dilihat pada Gambar 7. Titik perpotongan kurva TIF dengan SE merupakan reliabilitas. Berdasarkan Gambar 7 perpotongan kurva TIF dengan SE berada pada nilai $\theta = -3,4$ logit dan $\theta = +2$ logit. Dengan demikian, instrumen tes *integrated assessment* ini bersifat reliabel jika diujikan kepada peserta tes dengan kemampuan rendah, sedang, dan tinggi.



Gambar 7. Ekstrapolasi Kurva TIF dan SE

SIMPULAN

Karakteristik instrumen *integrated assessment* untuk mengukur *critical thinking skills* dan *science process skills* kimia peserta didik SMA kelas XI pada materi asam dan basa dinyatakan bahwa terbukti valid secara logis dan konstruk. Tingkat kesukaran butir pada instrumen menunjukkan bahwa instrumen berada pada kategori sedang. Fungsi informasi tes yang diperoleh sebesar 6,819 pada kemampuan peserta didik - 0,8 logit dengan kesalahan baku pengukuran sebesar 0,383. Perpotongan kurva TIF dengan SE berada pada nilai $\theta = -3,4$ logit dan $\theta = +2$ logit, artinya instrumen tes *integrated assessment* ini bersifat reliabel jika diujikan kepada peserta tes dengan kemampuan rendah, sedang, dan tinggi.

DAFTAR PUSTAKA

- Aiken, L.R. (1985). Three coefficients for analyzing: The reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142.
- Askins, D., Douglas, C., Angie, Amy, & Kamalesh. (2006). Measuring critical thinking. *The Nuts and Bolts Newsletter from Assessment Services*, 4(4), 1-4.
- Aydm, A. (2011). Representation of science process skills in the chemistry curricula for grades 10, 11 and 12 / Turkey. *International Journal of Education and Practice*, 1(5), 51-63.
- Beavers, A.S., Lounsbury, J.W., Richards, J.K., Huck, S.W., Skolits, G.J., & Esquivel, S.L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18(6), 1-13.
- Beyer, B.K. (1991). *Teaching thinking skills: A handbook for elementary school teachers*. New York, USA: Allyn & Bacon.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., & Wiesemes, R. (2011). A learning integrated system. *European Journal of Social Sciences*, 9(1), 1-8.
- Brookhart, S.M. (2010). *How to assess higher-order thinking skills in your classroom*. USA: ASCD.

- Chabalengula, V.M, Mumba, F., & Mbewe, S. (2011). How pre-service teachers' understand and perform science process skills. *Eurasia Journal of Mathematics*, 8(3), 167-176.
- Chebii, R., Wachanga, S., & Kiboss, J. (2012). Effects of science process skills mastery learning approach on students' acquisition of selected chemistry practical skills in school. *Scientific Research*, 3(8), 1291-1296.
- Cottrell, S. (2005). *Critical thinking skills developing effective analysis and argument*. New York: Palgrave Macmillan.
- Demircioğlu, G., Ayas, A., & Demircioğlu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education and Practice*, 6(1), 36-51.
- Dorman, J.P., Waldrip, B.G., & Fisher, D.L. (2009). Using the student perceptions of assessment questionnaire (SPAQ) to develop an assessment typology for science classes. *Journal of Science Education*, 9(1), 13-17.
- Durmaz, H. & Mutlu, S. (2014). The effects of an instructional intervention on 7th grade students' science process skills and science achievement. *Cukurova University Faculty of Education Journal*, 43(2), 155-168.
- Fraenkel, J.R. & Wallen, N.E. (1993). *How to design and evaluate research in education*. New York: McGraw-Hill, Inc.
- Guntur, Sukadiyanto, & Mardapi, D. (2014). Pengembangan asesmen hasil belajar penjasorkes siswa SMA pada permainan bola voli. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 13-29.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston: Kluwer Nijhoff Publishing.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. London: Sage Publications.
- Hunter, C., Roy, MCCOSH, & Hazel, W. (2003). Integrating learning and assessment in laboratory work. *Chemistry Education Research and Practice*, 4(1), 67-75.
- Judge, B., Jones, P., & *thinking skills for education students*. Britain: Learning McCreery, E. (2009). *Critical Matters Ltd*.
- Karamustafaoğlu, S. (2011). Improving the science process skills ability of science student teachers using I diagrams. *Eurasian Journal of Physics and Chemistry Education*, 3(1), 26-38.
- Leonor, J.P. (2015). Exploration of conceptual understanding and science process skills: A basis for differentiated science inquiry curriculum model. *Journal of Information and Education Technology*, 5(4), 255-259.
- Mansilla, V.B. & Jackson, A. (2011). *Education for global competence: preparing our youth to engage the world*. New York: Asia Society.
- Miller, P.W. (2008). *Measurement and teaching*. USA: Patrick W. Miller and Associates.
- Ministry of Education Malaysia. (2002). *Integrated curriculum for secondary schools*. Malaysia: Curriculum Department Centre.
- Mulyatiningsih, E. (2012). *Metode penelitian terapan bidang pendidikan*. Bandung: Alfabeta.
- National Academy of Sciences. (1996). *National science education standards*. Washington, DC: National Academy Press.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory (3rd Edition)*. New York: McGraw-Hill, Inc.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya*. Yogyakarta: Nuh Medika.
- Rezba, R., Sprague, C., McDonnough, J.T. & Matkins, J.J. (2007). *Learning and assessing science process skills fifth edition*. USA: Kendall/Hunt Publishing Company.
- Subali, B. (2009). Pengukuran keterampilan proses sains pola divergen dalam mata pelajaran biologi SMA di provinsi DIY dan Jawa Tengah. *Disertasi Doktor tidak diterbitkan*. Program Studi Penelitian dan Evaluasi Pendidikan: PPs UNY.

Sukarno, Permanasari, A., & Hamidah, I. (2013). The profile of science process skill (SPS) student at secondary high school (Case Study in Jambi). *IJSER*, 1(1), 79-83.

Conflict of Interest Statement: The Author(s) declares that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Copyright: ©Measurement in Educational Research. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Measurement in Educational Research is an open access and peer-reviewed journal published by Research and Social Study Institute, Indonesia

Open Access 