

PENGEMBANGAN INSTRUMEN INTEGRATED SCIENCE TEST TIPE PILIHAN GANDA BERALASAN UNTUK MENGUKUR HOTS PESERTA DIDIK

Yeni Widiyawati¹, Indri Nurwahidah², Dwi Septiana Sari³

¹Yeni Widiyawati (Program Studi Pendidikan IPA, Fakultas Sains dan teknologi,
Universitas Ivet, Indonesia)

²Indri Nurwahidah (Program Studi Pendidikan IPA, Fakultas Sains dan teknologi,
Universitas Ivet, Indonesia)

³Dwi Septiana Sari (Program Studi Pendidikan IPA, Fakultas Sains dan teknologi,
Universitas Ivet, Indonesia)

Abstract. HOTS is the main purpose of 21st-century science learning. Individu with HOTS will able to make a connection between facts and concepts, apply it into any new situation to solve the problem. The aim of this research was to developed integrated science test instrument to measure HOTS level of junior high school students. Oriondo and Antonio models' was adapted and used in this research method. The test was developed in the reasoning-multiple choice form with three category based on PISA and Marzano HOTS competencies. The item analysis involved content validity, empirical validity, reliability, and difficulty index. Rasch model analysis through Quest and Ministep software output showed (1) Mean of INFIT MNSQ level at 1.00 ± 0.12 that shows fit item with Rasch Model or empirical validity of items; (2) Level of INFIT MNSQ for all item between 0.77 and 1.33; (3) Vary of difficulty index; 34 items were at difficulty index level between -2 and +2; (4) junior high school students' HOTS in SMP IT Multazzam Banyumanik was at low level because they are difficult in giving explanations or arguments about the options that have been selected. Thus, the integrated science test instrument developed can be used to measure students' HOTS.

Keywords: HOTS, integrated science test, Rasch Model, reasoning-multiple choice, Quest.

PENDAHULUAN

Kegiatan pengukuran dan penilaian merupakan hal esensial dalam pembelajaran di sekolah (Sukardiyono & Rosana, 2018) dalam rangka memetakan kemampuan peserta didik dan memantau capaian pembelajaran. Kegiatan tersebut juga dijadikan sebagai landasan pengambilan keputusan para stakeholder yang mendukung lahirnya lulusan yang kompetitif di era Revolusi Industri 4.0 (Setiawan, Dafik, & Lestari, 2014). Proses pengukuran dalam pembelajaran berfungsi mengukur gejala non fisik misalnya kecerdasan yang bersifat laten atau tersembunyi (Bambang & Wahyu, 2014). Instrumen yang baik menjadi salah satu komponen penting dalam kegiatan pengukuran tersebut (Arum & Lestari, 2019; Sukardiyono & Rosana, 2018) untuk mengobservasi gejala laten yang tidak dapat teramati secara empiris berdasarkan indikator-indikator yang telah disusun (Bambang & Wahyu, 2014).

Namun, instrumen pengukuran yang digunakan dalam pembelajaran kurang variatif baik dalam bentuk instrumen maupun kemampuan yang hendak diukur (Arum

¹yeni.widiyawati26@gmail.com

P-ISSN: 1411-5433

E-ISSN: 2502-2768

© 2019 Sainifikita; Jurusan PMIPA, FKIP, Universitas Jember

<http://jurnal.unej.ac.id/index.php/STF>

& Lestari, 2019). Sejumlah pengukuran dalam pembelajaran di kelas baik menggunakan instrumen tes maupun non-tes masih berfokus pada pengukuran kemampuan berpikir tingkat rendah atau *low order thinking skills* (LOTS). Saido, Siraj, Nordin, & Al-Amedy (2015) menyatakan bahwa LOTS hanya membutuhkan kemampuan untuk mengingat dan memahami fakta, gejala maupun konsep yang telah ada sedangkan higher order thinking skills (HOTS) mengharuskan peserta didik untuk mengaplikasikan dan mengevaluasi pengetahuan tersebut. HOTS merupakan kemampuan pemecahan masalah di abad 21 (Retnawati, Djidu, Kartianom, Apino, & Anazifa, 2018). HOTS merupakan atribut yang diasosiasikan pada tiga kemampuan berpikir tingkat tinggi sesuai Taksonomi Bloom revisi (Anderson & Krathwohl, 2001) yaitu menganalisis, mengevaluasi dan mencipta. Lebih jauh, HOTS dapat diartikan sebagai kemampuan berpikir kritis dan kreatif (Heong et al., 2011) yang diperlukan oleh setiap individu dalam konteks pembelajaran di kelas (Yen & Halili, 2015). Thomas & Thorne (Apino & Retnawati, 2017) menyatakan bahwa *higher level thinking* ditekankan pada pengolahan fakta yang mencakup memahami, menghubungkan antar fakta dan konsep, memanipulasi, mengkombinasikan dalam bentuk baru, dan menggunakannya untuk menyelesaikan suatu permasalahan. Keterampilan berfikir merupakan kunci meningkatkan capaian pembelajaran di kelas bagi seluruh peserta didik (Sulaiman, Muniyan, Madhvan, Hasan, & Rahim, 2017).

Definisi lain mengenai HOTS yaitu dari Marzano (Heong et al., 2011; Marzano et al., 1997) yang mengedepankan dimensi belajar sebagai bagian dalam proses berpikir yang kompleks. Indikator HOTS versi Marzano memiliki kelebihan yang dapat membantu peserta didik untuk mencapai pembelajaran yang lebih bermakna (*meaningful learning*). Indikator HOTS Marzano dibagi menjadi 13 yaitu (1) membandingkan, (2) mengelompokkan, (3) penalaran induktif, (4) penalaran deduktif, (5) menganalisis kesalahan berpikir, (6) menyusun pernyataan pendukung, (7) menganalisis perspektif, (8) menyusun abstraksi, (9) membuat keputusan, (10) penyelidikan, (11) pemecahan masalah, (12) penyelidikan eksperimental dan (13) penemuan.

Kemampuan guru dalam mengembangkan dan mengukur HOTS peserta didik menjadi penting bagi keberhasilan pembelajaran di abad kompetitif ini (Retnawati et al., 2018). Pendidik diharapkan dapat mengembangkan instrumen tes yang berkualitas untuk mengukur kemampuan berpikir tingkat tinggi peserta didik. Namun, rendahnya pemahaman pendidik terkait HOTS berakibat pada kurangnya ruang yang diberikan kepada peserta didik untuk mengembangkan kemampuan berpikir tingkat tinggi mereka (Heru & Suparno, 2019; Retnawati et al., 2018). Evaluasi dan pengembangan instrumen penilaian oleh pendidik juga masih berorientasi pada instrumen untuk mengukur LOTS.

Padahal, pengukuran HOTS yang termasuk dalam lingkup gejala kognitif (Bambang & Wahyu, 2015) dan bersifat laten membutuhkan instrumen pengukuran yang terskala baku (Mardapi, Kumaidi, & Kartowagiran, 2011). HOTS dapat diukur dengan menggunakan serangkaian instrumen penugasan yang mencakup kegiatan menganalisis, mengevaluasi, menciptakan, menghubungkan konsep,

menginterpretasikan, memberikan argumen yang tepat, dan mengambil keputusan dalam kegiatan pemecahan masalah (Arum & Lestari, 2019; Retnawati et al., 2018). Instrumen pengukuran HOTS yang berkualitas harus memenuhi kriteria valid (sahih) dan reliabel (dapat dipercaya) (Arum & Lestari, 2019; Bambang & Wahyu, 2015). Diperlukan instrumen yang bukan hanya dapat mengukur tetapi juga dapat digunakan sebagai sarana untuk membiasakan peserta didik dengan butir tes HOTS (Hamdi, Suganda, & Hayati, 2018).

Instrumen pengukuran HOTS standar internasional yang telah banyak digunakan sebagai acuan yaitu Programme for International Student Assessment (PISA) *test* yang dikembangkan oleh OECD (2018). Tes model PISA menuntut peserta didik untuk mampu menalar dan memecahkan masalah (Setiawan et al., 2014). Kurniati, Harimukti, & Jamil (2016) mengungkapkan bahwa kemampuan peserta didik dalam melakukan penalaran, analisis, evaluasi serta kreasi menggunakan soal berstandar PISA berada di level rendah dan sedang.

Fokus PISA Science Framework yaitu pada pengukuran literasi sains. Individu yang melek IPA harus mampu memberikan justifikasi dalam suatu penyelidikan ilmiah berdasarkan penelitian terdahulu yang relevan (OECD, 2009). Konteks literasi sains sangat tepat bagi mata pelajaran IPA yang harus diajarkan secara terpadu dari disiplin ilmu fisika, kimia dan biologi secara koheren untuk mencapai tujuan pembelajaran. Salah satu karakteristik dari *integrated science* yaitu penggunaan tema yang harus menarik, kontekstual dan berkaitan dengan kehidupan nyata (Lathifah & Wilujeng, 2016). Belum banyak instrumen yang menekankan pada konsep *integrated science* tersebut. Sehingga diperlukan instrumen tes dalam bentuk yang tepat untuk mengukur HOTS peserta didik.

Tipe soal dalam tes model PISA sangat beragam, yakni *multiple choice* dan *open ended response*. Beberapa penelitian terdahulu telah mengembangkan instrumen tes model PISA dalam bentuk pilihan ganda (Wibowo & Cholifah, 2018) dan kombinasi pilihan ganda dan uraian (Sinaga, 2015). Tipe soal pilihan ganda lebih sering digunakan dalam pengukuran terkait dengan objektivitas dan lebih mudah dalam pelaksanaan dan pengolahan data. Namun, tipe soal ini tidak dapat mengungkapkan kemampuan peserta didik dalam menalar dan memberikan alasan (Putri, Istiyono, & Nurcahyanto, 2016). Kombinasi antara *constructed respons test* dan *multiple choice test* yaitu pilihan ganda beralasan merupakan alternatif terbaik (Ku, 2009). Namun demikian, belum banyak pengembangan instrumen tes dalam bentuk soal pilihan ganda beralasan yang mengungkapkan HOTS ditinjau dari model PISA yang dikombinasikan dengan model Marzano.

Oleh sebab itu, tujuan dari penelitian ini yaitu untuk mengembangkan instrumen *integrated science test* tipe pilihan ganda beralasan untuk mengukur HOTS peserta didik SMP.

METODE PENELITIAN

Penelitian ini merupakan penelitian pengembangan instrumen tes yang mengadaptasi model pengembangan Oriondo & Antonio (1984). Tahapan pengembangan instrumen tes yaitu (1) perencanaan, (2) uji coba, (3) validasi instrumen, (4) penghitungan reliabilitas, dan (5) interpretasi skor yang diperoleh.

Tahap perencanaan terdiri dari lima langkah utama yaitu (1) penentuan tujuan tes, (2) pengembangan bentuk tes, (3) penyusunan kisi-kisi, (4) penyusunan konstruksi tes dan (5) revisi tes. Tujuan pengembangan instrumen tes yaitu untuk mengukur HOTS peserta didik SMP di kota Semarang pada mata pelajaran IPA. Materi yang dikembangkan menjadi instrumen tes yakni materi IPA kelas VII semester ganjil. Bentuk tes yang dipilih yaitu pilihan ganda beralasan yang dapat menyediakan ruang bagi peserta didik untuk memberikan argumen terkait butir soal yang diujikan.

Konstruksi tes sengaja dipilih berupa *integrated science* agar peserta didik dapat memahami butir tes yang diujikan secara lebih *komprehensif* dan *holistic* baik dari sudut pandang fisika, kimia, maupun biologi. Kisi-kisi tes *integrated science* mengacu pada komponen HOTS Marzano (Heong et al., 2011; Marzano et al., 1997) yang dikombinasikan dengan indikator PISA (OECD, 2018). Kompetensi HOTS Marzano (1997) yang digunakan meliputi (1) mengklasifikasikan, (2) penalaran deduktif, (3) menganalisis kesalahan, (4) mengambil keputusan, (5) penyelidikan, (6) pemecahan masalah, (7) penyelidikan eksperimental dan (8) penemuan. Tiga kompetensi yang diujikan dalam PISA (OECD, 2018) yaitu (1) menjelaskan fenomena ilmiah, (2) mengevaluasi dan mendesain penyelidikan ilmiah, serta (3) menginterpretasikan data dan bukti secara ilmiah selanjutnya dijabarkan menjadi beberapa indikator. Kisi-kisi *integrated science test* disajikan dalam Tabel 1. Sebanyak 35 butir soal tes dikembangkan berdasarkan kisi-kisi yang telah disusun.

Tahap uji coba soal dilakukan terhadap 41 peserta didik SMP di kota Semarang yang berasal dari SMP IT Multazam Banyumanik. Jumlah ini telah sesuai dengan ukuran sampel untuk kestabilan data pada Rasch model (Bond & Fox, 2007). Analisis butir soal dilakukan berdasarkan skor mentah peserta didik dengan Rasch Model. Skala pengukuran dalam Rasch model dinyatakan dalam satuan logit atau log odd unit dalam persamaan 1 (Bambang & Wahyu, 2015).

$$\text{Logit} = \text{Log} (P/(N - P)) \quad (1)$$

dengan

P = skor total (jumlah soal yang dikerjakan dengan benar)

N = jumlah total soal yang disediakan

Fungsi logit tersebut memberikan skala pengukuran dengan interval yang sama sehingga memenuhi kriteria pengukuran. Koefisien validitas, tingkat kecocokan item dengan model, reliabilitas tes, dan tingkat kesukaran soal butir dianalisis berdasarkan fungsi logit tersebut melalui software Quest dan Ministep. Software ini mengestimasi kemampuan peserta didik berdasarkan pendekatan 1-parameter logistic (1PL) yaitu

tingkat kesukaran soal (b). Pendekatan inilah yang biasa disebut Rasch model (Bambang & Wahyu, 2014; Rosana & Sukardiyono, 2015; Widiyawati & Nurwahidah, 2018).

Berdasarkan ukuran sampel yang digunakan, item dikatakan fit dengan model jika memiliki nilai INFIT t antara -2 hingga +2 (Bond & Fox, 2007) jika taraf kesalahan sebesar 5%. Reliabilitas tes secara klasik dalam Quest disajikan dalam indeks alpha Cronbach sebab instrumen tes yang dikembangkan menggunakan penskoran politomus (Subali & Suyata, 2011). Reliabilitas menurut pendekatan IRT dinyatakan dalam *separation index*. Rumus 2 menyatakan pengelompokkan butir dan sampel secara lebih teliti berdasarkan *separation index* (Bambang & Wahyu, 2014).

$$H = \frac{[(4 \times \text{separation index}) + 1]}{3} \quad (2)$$

Tabel 1. Kisi-kisi *Integrated Science Test*

No Butir	Indikator PISA	Indikator HOTS Marzano
1	Mengubah bentuk representasi data.	
2, 3	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	
4	Mengelompokkan benda dalam kategorinya.	
5	Mendesain cara-cara untuk menyelesaikan suatu permasalahan secara ilmiah.	Mengelompokkan benda dalam kategorinya.
6	Menyusun hipotesis penjelas suatu fenomena.	
7	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan	Membuat dan menguji penjelasan suatu fenomena yang diamati.
8	Menyusun hipotesis penjelas suatu fenomena.	
9	Menganalisis dan menginterpretasi data serta menyusun kesimpulan yang tepat.	Membuat dan mengaplikasikan alternatif yang tepat.
10	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	Membuat dan mengaplikasikan alternatif yang tepat.
11	Membuat dan menilai prediksi yang benar.	Menggunakan generalisasi dan prinsip untuk mengambil kesimpulan mengenai informasi atau situasi yang lebih khusus.
12	Mengevaluasi argumen dan bukti ilmiah dari berbagai sumber	Membuat dan mengaplikasikan alternatif yang tepat.
13	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	
14	Mengubah bentuk representasi data.	
15, 17, 18	Mengidentifikasi asumsi, bukti, dan alasan berdasarkan kasus sains yang disajikan.	Menggunakan generalisasi dan prinsip untuk mengambil kesimpulan mengenai informasi atau situasi yang lebih khusus.
16, 20, 27	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	
19	Mengevaluasi argumen dan bukti ilmiah dari berbagai sumber.	Menyelesaikan masalah yang kontradiktif.

21, 24	Mendesain cara-cara untuk menyelesaikan suatu permasalahan secara ilmiah.	Membuat dan menguji penjelasan suatu fenomena yang diamati.
22	Menjelaskan potensi implikasi dari suatu pengetahuan ilmiah untuk masyarakat.	Mengembangkan produk atau proses yang unik untuk memenuhi kebutuhan.
23	Memilih rumusan masalah yang dapat diselidiki secara ilmiah.	Mengatasi masalah atau keterbatasan yang muncul dalam pencapaian suatu tujuan.
25	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	Membuat dan mengaplikasikan alternatif yang tepat.
26	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	Mengelompokkan benda dalam kategorinya.
28	Mengevaluasi argumen dan bukti ilmiah dari berbagai sumber.	Mengidentifikasi dan menyatakan kesalahan suatu pemikiran.
29	Membuat dan menilai prediksi yang benar.	Mengidentifikasi dan menyatakan kesalahan suatu pemikiran.
30, 34, 35	Menganalisis dan menginterpretasi data serta menyusun kesimpulan yang tepat.	Membuat dan mengaplikasikan alternatif yang tepat.
31	Mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk penyelesaian permasalahan.	Mengatasi masalah atau keterbatasan yang muncul dalam pencapaian suatu tujuan.
32, 33	Mengelompokkan argumen berdasarkan bukti dan teori ilmiah dan non ilmiah.	Mengelompokkan benda dalam kategorinya.

HASIL DAN PEMBAHASAN

Validitas Isi

Validitas isi diestimasikan berdasarkan kesesuaian kompetensi HOTS yang hendak di ukur dengan kisi-kisi yang operasional serta butir soal. Validitas ini merupakan sejauh mana tes merepresentasikan kemampuan yang seharusnya diukur. Validitas isi lebih menekankan pada definisi konseptual (Bambang & Wahyu, 2014). Penentuan kompetensi yang akan dikembangkan menjadi instrumen dan penyusunan kisi-kisi tes merupakan langkah dalam memastikan validitas isi instrumen instegrated science yang dikembangkan dalam kriteria baik. *Expert appraisal* oleh empat orang ahli pengukuran di bidang sains dan 2 praktisi membuktikan bahwa komponen HOTS versi Marzano dan PISA telah tercakup dalam 35 butir soal yang dikembangkan. distribusi kompetensi HOTS versi Marzano yang dikombinasikan dengan PISA disajikan dalam Tabel 1.

Kompetensi dan indikator HOTS kombinasi versi Marzano dan PISA direpresentasikan dalam instrumen integrated science test pada materi ajar IPA kelas VII semester ganjil. Materi objek IPA dan pengamatannya; suhu dan kalor; serta energi dan perubahannya terdistribusi secara merata dalam 35 butir soal. Instrumen ini menggali kemampuan peserta didik dalam menghubungkan konsep-konsep dasar yang telah dipelajari secara terintegrasi, bukan lagi fisika, kimia maupun biologi.

Analisis Butir Soal Empiris Secara Matematis

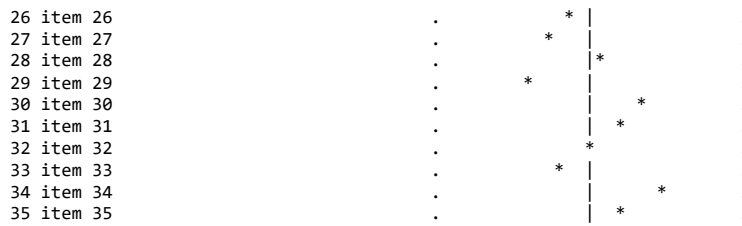
Skor mentah peserta didik dianalisis secara matematis menggunakan bantuan software Quest dan Ministep yang merujuk pada model Rasch dan teori tes klasik

(Rosana & Sukardiyono, 2015). Terdapat tiga parameter yang diukur secara matematis berdasarkan uji empiris butir soal.

Parameter pertama yaitu kecocokan butir dengan model Rasch yang didasarkan pada nilai rata-rata INFIT MNSQ dan simpangan bakunya. Berdasarkan Adams & Khoo (1996), nilai INFIT MNSQ yang disarankan yaitu sekitar 1,0 dengan simpangan baku sebesar 0,0. Nilai rerata INFIT MNSQ pada instrumen pengembangan sebesar $1,00 \pm 0,12$, dan rerata INFIT t sebesar $0,98 \pm 0,15$. Seluruh item fit dengan model Rasch sebab Infit t antara -2,0 dan +2,0, INFIT MNSQ antara -0,77 dan +1,30 (Subali & Suyata, 2011). Distribusi *fit* item dengan model Rasch ditunjukkan dalam output Quest dalam Gambar 1. Kecocokan butir atau *fit item* dengan model Rasch secara matematis dapat dikatakan sebagai validitas butir secara empiris (Bambang & Wahyu, 2015; Rosana & Sukardiyono, 2015). *Fit item* disajikan dalam Gambar 1.

Parameter kedua yaitu reliabilitas yang berdasarkan teori tes klasik. Model penskoran yang dipilih yaitu politomus. Skor yang diberikan yaitu 0 untuk jawaban salah, 1 untuk pilihan atau alasan yang benar, dan 2 untuk pilihan dan alasan yang benar. Butir yang tidak dikerjakan dianggap omit dan tidak dimasukkan dalam analisis. Besarnya *reliability of case estimates* atau *internal consistency* yang menyatakan nilai Alpha cronbach pada output Quest yaitu 0,65. Nilai ini masuk dalam kategori cukup sebab lebih dari 0,6 (Subali & Suyata, 2011). Berdasarkan tinjauan pendekatan IRT, nilai *internal consistency* dalam Quest tidak dapat dikatakan sebagai reliabilitas tes melainkan reliabilitas skor atau jawaban peserta didik sebab sampel yang berbeda akan menghasilkan nilai yang berbeda pula (Bambang & Wahyu, 2014). Perhitungan reliabilitas tes lebih lanjut menggunakan software Ministep yang ditunjukkan dengan nilai *separation index* sebesar 1,65. Analisis pengelompokkan menggunakan rumus 2 diperoleh nilai H sebesar 2,53 dan dibulatkan menjadi 3 yang dapat dimaknai bahwa terdapat tiga kelompok butir yaitu mudah, sedang dan sukar.

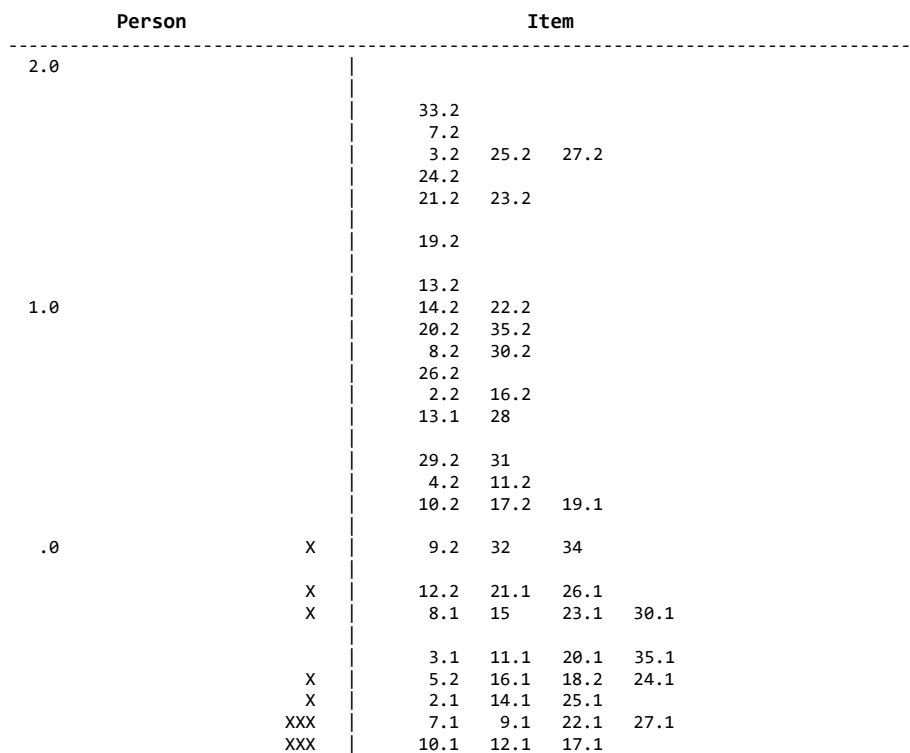
INFIT MNSQ	.56	.63	.71	.83	1.00	1.20	1.40	1.60	1.80
1 item 1				*					
2 item 2			.			*			
3 item 3			.			*			
4 item 4			.		*				
5 item 5			.	*					
6 item 6			.	*					
7 item 7			.			*			
8 item 8			.		*				
9 item 9			.	*					
10 item 10			.	*					
11 item 11			.		*				
12 item 12			.	*					
13 item 13			.		*				
14 item 14			.			*			
15 item 15			.			*			
16 item 16		*							
17 item 17			.	*					
18 item 18		.	*						
19 item 19		.	*						
20 item 20		.	*						
21 item 21		.			*				
22 item 22		.			*				
23 item 23		.			*	*			
24 item 24		.		*					
25 item 25		.		*					

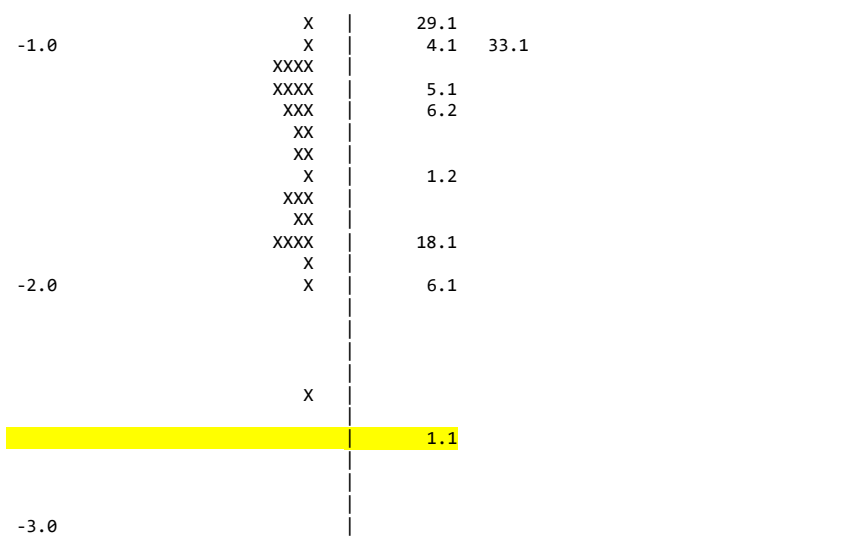


Gambar 1. *Fit item* dengan Model Rasch Berdasarkan Nilai INFIT MNSQ

Parameter ketiga yaitu indeks kesukaran soal yang disajikan sebagai *difficulty index* dalam output Quest. Indeks kesukaran soal yang baik yaitu antara -2,0 dan +2,0 (Putri et al., 2016). Butir soal nomor 1 memiliki *difficulty index* sebesar -2,1 atau lebih kecil dari -2,0. Butir soal ini harus dibuang sebab mengindikasikan bahwa butir tersebut terlalu mudah dan seluruh peserta didik dapat menjawab dengan benar Rerata *index difficulty* instrumen sebesar $0,0 \pm 0,64$. Pengelompokkan butir berdasarkan indeks kesukaran soal dibagi menjadi 3 yaitu mudah, sedang dan sukar dari rerata tersebut yang dapat dilihat dalam *person-item map* dari -3 hingga +3. *Person-item map* disajikan dalam Gambar 2.

Satu tanda X pada *person-item map* mewakili satu peserta didik. Dapat dilihat bahwa terdapat banyak butir yang tidak bisa dijawab dengan benar dan memperoleh skor maksimum (2). Perhatikan butir 33.2, yang menyatakan bahwa butir tersebut paling sukar untuk dijawab sebab tidak ada peserta didik yang dapat mencapai skor maksimum. Butir ini secara kontruksi dan bahasa mungkin membingungkan bagi peserta didik sehingga respon yang diberikan tidak menggambarkan kemampuan mereka. Dengan demikian, butir ini harus direvisi.





Gambar 2. Person-Item Map

Butir soal IPA yang baik bukan lah butir yang mengukur kemampuan membaca melainkan bagaimana menghubungkan konsep yang telah dipahami dalam situasi yang lebih kompleks. Revisi butir nomor 33 disajikan dalam Gambar 3.

33. Adonan kue termasuk dalam sebab....

- A. Senyawa sebab tidak dapat dipisahkan secara fisika menjadi zat penyusunnya
- B. Campuran sebab tidak dapat dipisahkan secara kimia menjadi zat penyusunnya
- C. Senyawa sebab sudah tidak memiliki sifat yang sama dengan zat penyusunnya
- D. Campuran sebab masih memiliki sifat yang sama dengan zat penyusunnya

Gambar 3. Butir Soal Nomor 33

Butir soal nomor 7.2 (pada skor 2) merupakan butir yang paling sukar. Persentase menjawab butir ini pada skor 0 yaitu 63,4%; pada skor 1 sebesar 34,1 dan pada skor 2 sebesar 2,4. Hal tersebut menandakan bahwa 63% peserta didik tidak dapat menjawab butir tersebut sehingga mendapat skor 0 dan hanya 2,4% yang mampu memberikan alasan yang tepat. Butir soal nomor 7 disajikan dalam Gambar 4. Butir nomor 7 merupakan butir yang membutuhkan perhitungan. Langkah pertama untuk menyelesaikan butir ini yaitu dengan memperhatikan besarnya volume apel dari gambar yang telah disediakan kemudian memasukkan dalam rumus

$$\text{massa jenis } (\rho) = \frac{\text{massa } (m)}{\text{volume } (V)}$$

Diketahui besarnya jenis apel yaitu $0,641 \text{ g/cm}^3$, dan volumenya sebesar 600 ml ($1 \text{ cm}^3 = 1 \text{ ml}$) maka:

$$m = \rho \times V$$

$$m = 0,641 \frac{\text{g}}{\text{ml}} \times 600 \text{ ml}$$

$$m = 384,6 \text{ g}$$

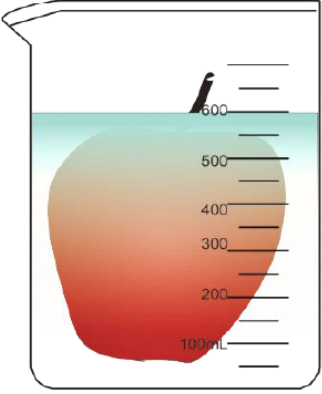
Sebagian besar peserta didik salah dalam menafsirkan gambar. Mereka mengira bahwa angka 600 yang ditunjukkan dalam pengukuran apel merupakan besaran massa.

Angka 600 dijumlahkan dengan massa jenis apel yang telah diketahui yaitu 0,641. Alasan yang mereka kemukakan tersebut tentu saja keliru. Peserta didik yang mendapatkan skor 1 atau memilih option dengan tepat bisa jadi hanya menebak atau menyontek jawaban teman. Konsep pengukuran belum dikuasai oleh peserta didik dengan baik sehingga mereka belum mampu mengingat kembali dan mengaplikasikan pengetahuan ilmiah yang sesuai untuk menyelesaikan masalah. Indikator ini masuk dalam kompetensi menjelaskan fenomena ilmiah.

Selama ini, peserta didik cenderung menghafal rumus yang diberikan tanpa mengetahui maknanya. Mereka akan kesulitan saat diminta untuk menggunakan rumus tersebut dalam konteks yang berbeda atau diubah. Butir nomor 7 yang merupakan butir yang menggabungkan konsep pengukuran (fisika) terhadap benda biotik yaitu apel (biologi) merupakan butir yang terintegrasi. Pendekatan *Integrated science* merupakan salah satu cara untuk mengembangkan keterampilan esensial di abad 21 (Asrizal, Amran, Ananda, & Festiyed, 2018), salah satunya yaitu keterampilan berpikir. Pendekatan pembelajaran secara terintegrasi efektif dalam memberikan pemahaman konten yang lebih mendalam (Drake & Reid, 2018) sehingga dapat melatih peserta didik untuk berpikir tingkat tinggi.

7. Perhatikan gambar di samping. Diketahui massa jenis apel merah adalah $0,641 \text{ g/cm}^3$. Massa apel merah dalam gambar disamping yaitu

A. 38,460 g
B. 384,600 g
C. 600,641 g
D. 606,410 g



Gambar 4. Butir Soal Nomor 7

Butir yang dapat dijawab oleh peserta didik hanya sampai pada skala logit 0,00. Hal tersebut mengindikasikan bahwa banyak butir yang dianggap sukar oleh peserta didik sebagai sampel, yang berasal dari SMP IT Multazzam Banyumanik. Berdasarkan Gambar 2, *person-item map* dan output Quest menunjukkan bahwasannya peserta didik sebagai sampel penelitian memiliki kemampuan yang rendah. Sebanyak 36 orang peserta didik meraih skor logit < SD yaitu skor logit < -0,64 sedangkan 5 orang lainnya berada pada kategori sedang sebab meraih $-0,64 < \text{skor logit} < +0,64$.

Skor maksimum pada sebagian butir soal tidak dapat diraih oleh sebagian besar peserta didik karena mereka tidak mampu menyatakan argumen mereka mengapa memilih option jawaban yang benar. Peserta didik belum terbiasa memberikan alasan yang logis atas pilihan yang mereka anggap benar. Butir-butir tersebut membutuhkan kemampuan analitis yang komprehensif dalam keterampilan berpikir kritis dan memecahkan masalah (Indria, Hindun, Latifatur, Samti, & Azizah, 2019).

Tipe butir yang disajikan dalam *integrated science* juga belum biasa mereka kerjakan. Hal tersebut terkait dengan rendahnya kemampuan pendidik dalam memahami

dan menyediakan ruang untuk pembelajaran IPA menggunakan pendekatan terpadu (Rubini, Pusitasari, Ardianto, & Hidayat, 2018) dan HOTS (Apino & Retnawati, 2017).

Padahal pendekatan terintegrasi menciptakan suasana pembelajaran yang dapat memotivasi peserta didik dan lebih baik dibanding kelas tradisional (Drake & Reid, 2018; Raub, Shukor, Arshad, & Rosli, 2015).

Rendahnya kemampuan pendidik dalam menyusun butir soal HOTS (Nurwahidah, 2018) juga menjadi alasan kuat mengapa peserta didik tidak terbiasa mengerjakan soal model PISA. Penting bagi pendidik untuk terampil dan menguasai materi ajar sehingga mereka dapat mengajarkannya kepada peserta didik dalam rangka mencapai kesuksesan pembelajaran (Balakrishnan, Nadarajah, Vellamy, Gnanam, & George, 2016).

Oleh sebab itu, instrumen *integrated science test* ini menjadi salah satu alternatif bagi pendidik untuk mengukur dan melatih HOTS peserta didik menggunakan model PISA (OECD, 2018) yang dikombinasikan dengan versi Marzano (Marzano et al., 1997). Ketika peserta didik menggunakan HOTS maka mereka mampu menghubungkan dan memberikan penjelasan antar fakta maupun konsep, mengkategorikan, serta memanipulasi fakta dan menggunakannya dalam berbagai situasi (Rashika & Salleh, 2019). HOTS sangat penting untuk mewujudkan lulusan yang kompetitif, sukses (Balakrishnan et al., 2016) serta memiliki kesadaran terhadap lingkungan (Perkasa, 2017).

KESIMPULAN

Berdasarkan tujuan dan hasil penelitian dapat disimpulkan bahwa instrumen *integrated science* yang dikembangkan telah memenuhi kriteria valid secara isi maupun empiris, reliabel, dan memiliki tingkat kesukaran soal yang beragam. Dari 35 butir soal yang dikembangkan terdapat 1 butir yang harus dieliminasi sebab memiliki *index difficulty* kurang dari -2,0.

Ucapan Terimakasih

Penelitian ini didanai oleh Kemenristekdikti dalam Hibah Penelitian Skim PDP tahun anggaran 2019. Peneliti mengucapkan terimakasih kepada Kemenristekdikti, validator, guru SMP IT Multazzam Banyumanik, serta pihak-pihak lain yang tidak dapat disebutkan satu-persatu atas bantuan dan kerjasamanya sehingga penelitian ini dapat terselesaikan dengan lancar.

DAFTAR PUSTAKA

- Adams, R. J., & Khoo, S. T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for Learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Apino, E., & Retnawati, H. (2017). Developing instructional design to improve mathematical higher order thinking skills of students. *Journal of Physics: Conference Series*, 812(012100), 1–7. <https://doi.org/10.1088/1742->

6596/755/1/011001

- Arum, S., & Lestari, P. (2019). Pengembangan Instrumen Asesmen Higher Order Thinking SKILL (HOTS) pada Materi Himpunan Kelas VII SMP, 2682, 111–120.
- Asrizal, A., Amran, A., Ananda, A., & Festiyed, F. (2018). Effectiveness of adaptive contextual learning model of integrated science by integrating digital age literacy on grade VIII students. *IOP Conference Series: Materials Science and Engineering PAPER*, 335(012067), 1–8. <https://doi.org/10.1088/1757-899X/335/1/012067>
- Balakrishnan, M., Nadarajah, G. M., Vellasamy, S., Gnanam, E., & George, W. (2016). Enhancement of Higher Order Thinking Skills among Teacher Trainers by Fun Game Learning Approach. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 10(12), 3893–3897.
- Bambang, S., & Wahyu, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial (Revisi)*. Cimahi: Penerbit Trim Komunikata.
- Bambang, S., & Wahyu, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan*. Cimahi: Penerbit Trim Komunikata.
- Bond, T. G., & Fox, C. M. (2007). *Applying the rasch model: Fundamental measurement in the human sciences, 2-nd ed.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Drake, S. M., & Reid, J. L. (2018). Integrated Curriculum as an Effective Way to Teach 21st Century Capabilities.
- Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *REiD (Research and Evaluation in Education)*, 4(2), 126–135.
- Heong, Y. M., Othman, W. B., Yunos, J. Bin, Kiong, T. T., Hassan, R. Bin, Mohaffyza, M., & Mohamad, B. (2011). The Level of Marzano Higher Order Thinking Skills among Technical Education Students, 1(2).
- Heru, M., & Suparno, S. (2019). The Development of Reasoned Multiple Choice Test in Interactive Physics Mobile Learning Media (PMLM) of Work and Energy Material to Measure High School Students ' HOTS. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, 9(2), 141–150. <https://doi.org/http://dx.doi.org/10.30998/formatif.v9i2.3002> The
- Indria, T., Hindun, I., Latifatur, N., Samti, A., & Azizah, N. (2019). JPBI (Jurnal Pendidikan Biologi Indonesia) Critical thinking skills : The academic ability , mastering concepts , and analytical skill of undergraduate students, 5(1), 1–8.
- Ku, K. Y. L. (2009). Assessing students ' critical thinking performance : Urging for measurements using multi-response format, 4, 70–76. <https://doi.org/10.1016/j.tsc.2009.02.001>
- Kurniati, D., Harimukti, R., & Jamil, N. A. (2016). Kemampuan berpikir tingkat tinggi siswa SMP di Kabupaten Jember dalam menyelesaikan soal berstandar PISA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 142–155. <https://doi.org/DOI:http://dx.doi.org/10.21831/pep.v20i2.8058>

- Lathifah, I. N., & Wilujeng, I. (2016). Pengembangan perangkat pembelajaran integrated science berbasis kearifan lokal. *Jurnal Penelitian Matematika Dan Sains*, 4(2), 120–129. <https://doi.org/http://dx.doi.org/10.21831/jpms.v4i2.12943>
- Mardapi, D., Kumaidi, K., & Kartowagiran, B. (2011). Pengembangan instrumen pengukur hasil belajar nirbias dan terskala baku. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(2), 326–341.
- Marzano, R. J., Pickering, D. J., Blackburn, G. J., Brandt, R. S., Paynter, D. E., Pollock, J. E., ... Whisler, J. S. (1997). *Dimensions of learning*. Colorado: Mid-continent Research for Education and Learning.
- Nurwahidah, I. (2018). Pengembangan soal penalaran model TIMSS untuk emngukur High Order Thinking (HOT). *Thabiea : Journal of Natural Science Teaching*, 01(01), 20–29.
- OECD. (2009). *Students with Disabilities, Learning Difficulties and Disadvantages in the Baltic States, South Eastern Europe and Malta*. OECD Publication. Retrieved from www.oecd.org/publishing/corrigenda
- OECD. (2018). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. <https://doi.org/https://doi.org/10.1787/b25efab8-en>
- Oriondo, L. L., & Dallo-Antonio, E. M. (1984). *Evaluation educational Outcomes*. Manila: Rex Printing Compagny, Inc.
- Perkasa, M. (2017). Pembelajaran Kimia Berorientasi Sustainable Development untuk Meningkatkan Kesadaran Siswa Terhadap Lingkungan Sustainable Chemistry Oriented on Chemistry Learning to Increase Student Awareness to the Environment, *VI*(2), 63–72.
- Putri, F. S., Istiyono, E., & Nurcahyanto, E. (2016). Pengembangan instrumen tes keterampilan berfikir kritis dalam bentuk pilihan ganda beralasan (politomus) di DIY. *Unnes Physics Education Journal*, 5(2), 76–84.
- Rashika, H., & Salleh, M. (2019). Promoting hots through thinking maps. *International Journal of Education, Psychology and Counseling*, 4(26), 104–112.
- Raub, L. A., Shukor, N. A., Arshad, M. Y., & Rosli, M. S. (2015). An Integrated Model to Implement Contextual Learning with Virtual Learning Environment for Promoting Higher Order Thinking Skills in Malaysian Secondary Schools, 8(13), 41–46. <https://doi.org/10.5539/ies.v8n13p41>
- Retnawati, H., Djidu, H., Kartianom, K., Apino, E., & Anazifa, R. D. (2018). Teachers; Knowledge about higher-order thniking skills and its learning strategy. *Problesms of Education in the 21st Century*, 76(2).
- Rosana, D., & Sukardiyono, S. (2015). The items analysis and the identification of final test score inappropriateness to standardize the assesment. *Jurnal Kependidikan*, 45(2), 130–141.
- Rubini, B., Pusitasari, I. D., Ardianto, D., & Hidayat, A. (2018). Science teachers' understanding on science literacy and integrated science learning" Lesson from teachers training. *Jurnal Pendidikan IPA Indonesia*, 7(3), 259–265. <https://doi.org/10.15294/jpii.v7i3.11443>

- Saido, G. A. M., Siraj, S., Nordin, A. B., & Al-Amedy, O. S. (2015). Teaching strategies for promoting higher order thinking skills: A case of secondary science teachers. *Malaysian Online Journal of Educational Management*, 3(4), 16–30.
- Setiawan, H., Dafik, D., & Lestari, N. D. S. (2014). Soal matematika dalam PISA kaitannya dengan literasi matematika dan keterampilan berpikir tingkat tinggi. In *Prosiding Seminar Nasional Matematika, Universitas Jember* (pp. 244–251). Jember: Universitas Jember.
- Sinaga, T. N. (2015). Pengembangan soal model PISA mata pelajaran ilmu pengetahuan alam terpadu konten fisika untuk mengetahui penalaran siswa kelas IX. *Jurnal Inovasi Dan Pembelajaran Fisika*, 2(2), 194–197.
- Subali, B., & Suyata, P. (2011). *Panduan analisis data pengukuran pendidikan untuk memperoleh bukti empirik kesahihan menggunakan program Quest*.
- Sukardiyono, S., & Rosana, D. (2018). Implementation of integrated science instruction assessment as an alternative to measure science process skills and social attitudes. *Journal of Science Education Research*, 1(1). <https://doi.org/https://doi.org/10.21831/jser.v1i1.16185>
- Sulaiman, T., Muniyan, V., Madhvan, D., Hasan, R., & Rahim, S. S. A. (2017). Implementation of higher order thinking skills in teaching science: A case study in Malaysia. *International Research Journal of Education and Sciences (IRJES)*, 1(1), 1–3.
- Wibowo, A., & Cholifah, T. N. (2018). Pengembangan instrumen tes tematik terpadu kurikulum 2013 berbasis PISA'S literacy di sekolah dasar. *JIPVA (Jurnal Pendidikan IPA Veteran)*, 2(2), 209–221. <https://doi.org/10.31331/jipva.v2i2.726>
- Widiyawati, Y., & Nurwahidah, I. (2018). Elclivis berbasis inquiry untuk meningkatkan penguasaan konsep siswa tuna netra pada materi rangkaian listrik. *Jurnal Inovasi Pendidikan IPA*, 4(2), 212–223. <https://doi.org/https://doi.org/10.21831/jipi.v4i2.21527>
- Yen, T. S., & Halili, S. H. (2015). effective teaching of higher-order thinking (HOT) in education. *The Online Journal of Distance Education and E-Learning*, 3(2), 41–47.