

PENGEMBANGAN SOAL BERBASIS *HIGHER ORDER THINKING SKILL* (HOTS) SERTA IMPLEMENTASINYA DI SMA NEGERI 8 PALEMBANG***DEVELOPMENT OF TEST-BASED HIGHER ORDER THINKING SKILL (HOTS) AND ITS IMPLEMENTATION IN SMA NEGERI 8 PALEMBANG***Jhon Riswanda¹, Erie Agusta*², Evitia Lestari¹¹Pendidikan Biologi, Fakultas Ilmu Tarbiyah dan Keguruan, Universitas Islam Negeri Raden Fatah
Jl. Prof. K.H. Zainal Abidin Fikri Km. 3,5 Palembang, Sumatera Selatan 30126²Pendidikan Biologi, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Palembang
Jl. Jendral A. Yani, Plaju, Palembang, Sumatera Selatan 30263

*erieagusta1@gmail.com

Diterima: Maret 2018; Disetujui: Agustus 2018; Diterbitkan: September 2018

Abstrak

Tujuan penelitian ini adalah untuk menghasilkan soal berbasis HOTS lengkap dengan panduan penyusunan soalnya. Metode penelitian ini terdiri dari (1) *defining the test universe, audience and puprose*; (2) *developing a test plan*; (3) *composing the test items*; (4) *writting the administration instruction*; (5) *conduct the piloting test*; (6) *conduct the item analysis*; (7) *revise the test*; (8) *validation of test*; (9) *developing the norms*; (10) *completting the test manual*. Hasil penelitian yang diperoleh 1) keseluruhan item *fit* dengan nilai reliabilitas pada kategori sangat bagus (0,93), akan tetapi item 1 dan item 20 terjangkit DIF yang lebih berpihak terhadap kelompok IPS sehingga keputusan *final* item tersebut perlu dihapuskan. 2) Secara keseluruhan instrumen riset pada penelitian ini menunjukkan bahwa ia mengukur apa yang seharusnya diukur karena telah memenuhi persyaratan unidimensionalitas sebesar 20%. 3) *Manual test* yang disusun berada pada kriteria sangat layak (A). 4) Pada tingkat responden terlihat bahwa 133 siswa memiliki HOTS tinggi sedangkan 64 siswa memiliki HOTS rendah dan 24 responden berada pada kriteria *misfit* yang didalamnya terdapat responden dengan HOTS tinggi. Hal ini mengindikasikan masih banyak siswa yang memiliki kemampuan HOT rendah. Estimasi ini didukung dengan hasil reliabilitas *person* yang menunjukkan kemampuan HOT *person* berada pada kriteria cukup (0,72).

Kata kunci: panduan penyusunan soal, Rasch, soal HOTS**Abstract**

The study aimed to generate the HOTS-based problem with its guidance on preparing the test. Research methods consisted of (1) defining the test universe, audience and puprose; (2) developing a test plan; (3) composing the test items; (4) writting the administration instruction; (5) conduct the piloting test; (6) conduct the item analysis; (7) revise the test; (8) validation of test; (9) developing the norms; (10) completting the test manual. The results obtained 1) the overall fit item with the reliability value were in very good category (0.93), but item 1 and item 20 infected by DIF were more favorable to the IPS group so the final decision of the item needs to be abolished. 2) Overall the research instrument in this study showed that it measures what should be measured because it has met the unidimensionality requirement of 20%. 3) Manual test that arranged was in very feasible (A). 4) At the level of respondents showed that 133 students have high HOTS while 64 students have low HOTS and 24 respondents are on misfit criteria in which there are respondents with high HOTS. This indicated that there are still many students who have low HOT ability. This estimation was supported by the results of the person reliability that shows the HOT person ability was in enough criteria (0.72).

Keywords: *guidance on preparing the test, Rasch, HOTS test*

©Didaktika Biologi: Jurnal Penelitian Pendidikan Biologi
p-ISSN 2549-5267
e-ISSN 2579-7352

Pendahuluan

Pendidikan merupakan sarana untuk mencapai kesuksesan bagi setiap orang. Suatu negara dapat dikatakan maju apabila memperhatikan pendidikan bangsa. Pendidikan berasal dari kata didik, yang berarti memelihara dan membentuk latihan (Duhita, 2015).

Salah satu faktor utama yang menentukan mutu pendidikan adalah guru. Upaya yang dapat dilakukan guru dalam peningkatan kualitas pendidikan, yaitu melalui sistem penilaian (evaluasi) (Sofiana, 2010). Alat evaluasi yang sering digunakan dalam proses belajar mengajar adalah tes (Arifin, 2013). Brookhart (2010) menyatakan bahwa tes dalam evaluasi pembelajaran harus mampu melatih kemampuan berpikir tingkat tinggi siswa sehingga menumbuhkan daya nalar yang baik. Kemampuan berpikir tingkat tinggi merupakan suatu kemampuan berpikir yang tidak hanya membutuhkan kemampuan mengingat saja, namun membutuhkan kemampuan lain yang lebih tinggi, seperti kemampuan menganalisis (C4), mengevaluasi (C5), dan mencipta (C6).

Namun realitanya kemampuan berpikir tingkat tinggi siswa di Indonesia melalui hasil penelitian PISA (*Program for International Student Assessment*) tahun 2015 menunjukkan siswa Indonesia hanya mampu mengerjakan soal dengan tipe C1-C3 yang merupakan *Low Order Thinking Skill* (LOTS) (Sarnapi, 2016). Dari data tersebut dapat diketahui bahwa kemampuan berpikir tingkat tinggi siswa-siswi Indonesia dalam bidang sains yang termasuk di dalamnya adalah pelajaran Biologi masih sangat memprihatinkan.

Berdasarkan laporan PISA 2015 tersebut dapat menjadi acuan bahwa permasalahan ini sering terjadi pada semua sekolah termasuk SMA Negeri 8 Palembang. Padahal setiap tahun calon siswa yang mendaftar di SMAN 8 Palembang harus mengikuti tes terlebih dahulu.

Berdasarkan hasil studi awal menunjukkan hasil yang masih memprihatinkan. Survei pendahuluan melalui pengambilan sampel soal mata pelajaran Biologi kelas X tahun ajaran 2016/2017 pada tanggal 31 Oktober 2016 di

SMA Negeri 8 Palembang dapat diketahui bahwa soal-soal yang digunakan guru pada kegiatan evaluasi pembelajaran masih terbatas pada tipe soal mengingat (C1), memahami (C2), dan mengaplikasikan (C3) yang bersifat *low order thinking skill*.

Guru berorientasi mengejar target pencapaian materi, tetapi belum mengembangkan kompetensi yang dimiliki siswa secara maksimal. Dewi (2015) menyatakan bahwa salah satu pendekatan pembelajaran yang baik, yaitu dengan menggunakan HOT, baik untuk proses belajar mengajar di kelas maupun dalam evaluasi pembelajaran. Akan tetapi didominasinya soal-soal bertipe C1, C2, dan C3 dari evaluasi pembelajaran siswa mengindikasikan guru kurang melatih siswa dalam berpikir tingkat tinggi yang diwujudkan melalui kemampuan menganalisis (C4), mengevaluasi (C5) dan mencipta (C6). Selanjutnya, berdasarkan hasil wawancara terhadap guru mata pelajaran Biologi kelas X di SMA Negeri 8 Palembang pada tanggal 04 November 2016 dapat diketahui pula bahwa masih rendahnya pengetahuan mengenai *Higher Order Thinking Skill*. Hal ini disebabkan para guru belum terbiasa membuat soal HOTS. Berdasarkan uraian hasil wawancara tersebut maka kebutuhannya adalah panduan mengenai cara pembuatan soal berbasis HOTS untuk para Guru Biologi di SMAN 8 Palembang, dan inilah yang menjadi tujuan dari penelitian ini dilakukan.

Metode Penelitian

Tempat dan Waktu

Penelitian ini dilaksanakan pada bulan Mei 2017 yang dilakukan di SMA Negeri 8 Palembang.

Jenis Penelitian

Penelitian ini merupakan *research and development* terkait pengembangan soal berbasis *Higher Order Thinking Skill* (HOTS) mengikuti alur pengembangan tes menurut model McIntire & Miller (2007).

Populasi dan Sampel

Populasi meliputi seluruh kelas X di SMA Negeri 8 Palembang pada semester genap tahun pelajaran 2016/2017 berjumlah 329 siswa. Pengambilan sampel penelitian

menggunakan teknik *simple random sampling* dengan acuan tabel Krejcie pada taraf signifikan 5%. Subjek yang menjadi uji coba meliputi kelas X MIPA 1, X MIPA 2, X MIPA 3, X MIPA 5, X IPS 1, dan X IPS 4 yang berjumlah 197 siswa.

Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian adalah dokumentasi dan tes tertulis. Dokumentasi digunakan untuk memperoleh sampel soal mata pelajaran Biologi kelas X tahun ajaran 2016/2017 yang biasa digunakan dalam mengukur ketercapaian pembelajaran peserta didik. Setelah itu, penjarangan data dilakukan dengan menganalisis jawaban tes

tertulis siswa setelah mengisi soal berbasis HOTS kelas X SMA. Soal ini terdiri dari 20 butir. Soal ini terdiri dari soal C4 sebanyak 7 butir, soal C5 sebanyak 6 butir, dan soal C6 sebanyak 7 butir.

Teknik Analisis Data

Analisis Hasil Validitas Konten Instrumen Tes

Analisis hasil validasi instrumen tes dilakukan berdasarkan langkah-langkah berikut.

1. Menjumlahkan skor total
2. Mengkonversi skor total menggunakan kategorisasi yang dapat dilihat pada tabel 1 di bawah ini.

Tabel 1. Kategorisasi Skor Penilaian Kelayakan Instrumen Tes oleh Validator

Interval	Nilai	Kategori
$X \geq Y_i + 1.S_{bi}$	A	Sangat Baik
$Y_i + 1.S_{bi} > X \geq Y_i$	B	Baik
$Y_i > X \geq Y_i - 1.S_{bi}$	C	Cukup
$X < Y_i - 1.S_{bi}$	D	Tidak Baik

(Sumber: Mardapi, 2008)

Keterangan :

Y_i (Rata-rata skor ideal) = $1/2$ (S.Maks ideal + S.Min ideal)

S_{bi} (Simpangan baku ideal) = $1/6$ (S.Maks ideal - S.Min ideal)

X = Perolehan skor

S.Maks ideal = Jumlah butir kriteria \times skor tertinggi

S.Min ideal = Jumlah butir kriteria \times skor terendah

Analisis Persamaan Persepsi Ahli terhadap Instrumen Validasi

Analisis persamaan persepsi dosen ahli dilakukan untuk mengetahui besarnya persamaan pendapat para ahli terhadap kelayakan suatu instrumen. Rumus yang digunakan sebagai berikut.

$$R = \left(1 - \frac{A-B}{A+B}\right) \times 100\%$$

Keterangan :

R = Reliabilitas instrumen

A = Frekuensi tertinggi pengamatan

B = Frekuensi terendah pengamatan

Perhitungan dari hasil penilaian oleh *judgement* menghasilkan nilai akhir ($\geq 75\%$) menunjukkan kelayakan/valid terhadap instrumen yang dikembangkan (Borich, 1994).

Analisis Jawaban Tes Tertulis Siswa

Item Fit Order, memberikan informasi mengenai ketidak sesuaian item. *Item Measure*, digunakan untuk mendapatkan informasi mengenai item mana yang paling disetujui oleh responden penelitian dan item mana yang paling sukar disetujui. *Differential Item Functioning/DIF*, mendeteksi bias pada item diperlukan untuk mengetahui apakah item-item yang diberikan memiliki bias dalam kategori responden tertentu atau tidak.

Person Fit Order, cara memeriksa *person* yang *fit* dan *misfit* bisa dengan menggunakan nilai INFIT MNSQ dari tiap *person*, nilai rata-rata dan deviasi standar yang dijumlahkan (jumlah *logit* INFIT MEAN MNSQ + INFIT MNSQ S.D), kemudian dibandingkan dengan nilai *logit* yang ada dalam tiap *person* pada kolom INFIT MNSQ. Nilai *logit* yang lebih besar

dari kriteria INFIT MNSQ maka mengindikasikan *person misfit*. *Person Measure*, memberikan informasi mengenai responden mana yang paling banyak menjawab dengan benar, diurutkan dari atas mulai dari nilai *logit* paling tinggi sampai yang paling rendah (Sumintono & Widhiarso, 2015).

Unidimensionalitas, adalah hal yang penting untuk mengetahui apakah ia mengukur apa yang seharusnya diukur. Menurut Misbach & Sumintono (2014), persyaratan minimal unidimensionalitas sebesar 20%, apabila nilainya lebih dari 40% artinya lebih bagus, apabila lebih dari 60% artinya istimewa. Hal lain yang diinformasikan, yakni varians yang tidak dapat dijelaskan oleh instrumen idealnya tidak melebihi 15% (Bond & Fox, 2007). *Reliabilitas*, diketahui melalui tabel *Summary Statistic*. Menurut Sumintono & Widhiarso (2013), klasifikasi nilai *Alpha Cronbach* dapat dilihat pada Tabel 2. Sedangkan klasifikasi dari nilai *person reliability* dan *item reliability* dapat dilihat pada Tabel 3.

Tabel 2. Klasifikasi Nilai *Alpha Cronbach*

Nilai <i>Alpha Cronbach</i>	Klasifikasi
0,00 $r < 0,50$	Buruk
0,50 $r < 0,60$	Jelek
0,60 $r < 0,70$	Cukup
0,70 $r < 0,80$	Bagus
0,80 $r \leq 1,00$	Bagus Sekali

(Sumber: Sumintono & Widhiarso, 2013)

Tabel 3. Klasifikasi Nilai *Person Reliability* dan *Item Reliability*

Nilai <i>Person Reliability</i> dan <i>Item Reliability</i>	Klasifikasi
0,00 $r < 0,67$	Lemah
0,67 $r < 0,80$	Cukup
0,80 $r < 0,90$	Bagus
0,90 $r < 0,94$	Bagus Sekali
0,94 $r \leq 1,00$	Istimewa

(Sumber: Sumintono & Widhiarso, 2013)

Hasil dan Pembahasan

Menentukan Test Universe, Audiens, dan Tujuan (Defining the Test Universe, Audience, and Purpose)

a. Analisis Kebutuhan dan Karakteristik Siswa

Hasil survei pendahuluan melalui pengambilan sampel soal mata pelajaran Biologi kelas X tahun ajaran 2016/2017

pada tanggal 31 Oktober 2016 di SMA Negeri 8 Palembang dapat diketahui bahwa soal-soal yang digunakan guru pada kegiatan evaluasi pembelajaran terbatas pada tipe soal mengingat (C1), memahami (C2), dan mengaplikasikan (C3) yang bersifat *low order thinking skill*. Temuan lainnya, berdasarkan hasil wawancara terhadap guru mata pelajaran Biologi kelas X di SMA Negeri 8 Palembang pada tanggal 04 November 2016 dapat diketahui pula bahwa masih rendahnya pengetahuan mengenai *Higher Order Thinking Skill* (HOTS). Hal ini disebabkan karena belum adanya panduan mengenai cara pembuatan soal berbasis HOTS.

b. Kompetensi Siswa yang Diharapkan dari Peserta Tes

Kompetensi yang diharapkan peserta tes dapat menghubungkan dan mentransformasi pengetahuan yang dimiliki untuk berpikir secara kritis dan kreatif dalam upaya menentukan keputusan dan memecahkan masalah ketika mengerjakan soal-soal yang menuntut kemampuan berpikir tingkat tinggi siswa.

c. Tujuan Tes HOTS

Tujuan dilaksanakannya tes yaitu untuk melatih dan mengukur kemampuan berpikir tingkat tinggi siswa melalui uji coba kelayakan soal berbasis HOTS yang dibuat mengikuti panduan *manual test* HOTS.

Mengembangkan Rencana Tes (Developing A Test Plan)

a. Kisi-kisi Soal HOTS

Indikator soal HOTS yang berhasil dikembangkan pada aspek menganalisis (C4), mengevaluasi (C5) dan mencipta (C6) sebanyak 18 indikator.

b. Format Pertanyaan Soal HOTS

Soal HOTS yang diujikan dalam bentuk pilihan ganda dengan lima pilihan jawaban yang mengacu pada 3 dimensi kemampuan berpikir tingkat tinggi dan dihubungkan dengan pengetahuan faktual, konseptual, prosedural serta metakognitif.

Cara Penskoran

Penskoran pada tes ini menggunakan rumus sebagai berikut.

$$\text{Skor} = \left(\frac{B}{N}\right) \times 100$$

Keterangan:

B = Banyaknya butir soal yang dijawab benar

N = Banyaknya butir soal

Menyusun Item Uji (Composing the Test Items)

Hasil *expert judgement* yang dilakukan oleh 2 orang tim ahli Kuratul Aini, M.Pd. (dosen ahli I) dan M. Isnaini (dosen ahli II) selaku dosen Fakultas Ilmu Tarbiyah dan Keguruan UIN Raden Fatah Palembang dapat dilihat pada Tabel 3 di bawah ini.

Tabel 3. Rekapitulasi Data Hasil Validasi Instrumen Tes HOTS

Validator	Total Skor	Kategori
Dosen Ahli I	58	B
Dosen Ahli II	50	B

Tingkat persamaan asumsi dosen ahli terhadap soal HOTS yaitu sebesar 93% yang menunjukkan kelayakan/valid terhadap instrumen HOTS.

Menulis Instruksi Administrasi (Writing the Administration Instructions)

Petunjuk penyelenggaraan berkaitan dengan petunjuk pengerjaan soal untuk peserta tes HOTS, serta tata cara tes HOTS yang dibuat untuk pengawas tes HOTS.

Melakukan Uji Coba (Conduct the Piloting Test)

Perangkat tes diuji coba kepada kelas X MIPA 1, X MIPA 2, X MIPA 3, X MIPA 5, X IPS 1, dan X IPS 4 yang berjumlah 197 siswa.

Melakukan Analisis Item (Conduct the Item Analysis)

a. Item Fit Order

Berdasarkan nilai *point measure correlation negative* untuk semua item yang bernilai positif, ini artinya tidak ada item yang terlalu mudah untuk dijawab oleh siswa. Untuk mengetahui item yang tidak *fit* dapat diketahui nilai *Infit* MNSQ dari setiap item yang tidak boleh lebih dari 1.09, apabila lebih besar hal ini mengindikasikan item tidak sesuai. Berdasarkan kriteria tersebut 2 item dengan nilai *infit* MNSQ yang lebih

besar yakni item nomor 1 dan item nomor 20.

Kriteria lain yang digunakan yaitu dengan melihat nilai dari parameter *outfit* MNSQ, *outfit* ZSTD serta *PT-Measure Correlation*. Nilai *logit PT-Measure Correlation* (x) dari 20 item menunjukkan item nomor 1 dan item 20 tidak memenuhi parameter dari nilai *PT-Measure correlation* (x) yang seharusnya $0.32 < x < 0.8$. Nilai *logit outfit* MNSQ (y) dengan kriteria $0.5 < y < 1.5$ dari 20 item menunjukkan item sesuai dengan kriteria. Pada analisis *outfit* ZSTD (z) dengan kriteria $-2.0 < z < + 2.0$ didapati bahwa item 20 tidak memenuhi kriteria.

b. Item Measure

Item yang paling sukar disetujui oleh 197 responden yakni item nomor 16 dengan nilai *logit* item tertinggi (+0.53) dari nilai *logit* item lainnya, yang termasuk ke dalam aspek HOTS mencipta (C6). Sedangkan item yang paling mudah dijawab oleh siswa yaitu item nomor 1 dengan nilai *logit* item terendah (-2.50) pada aspek HOTS menganalisis (C4).

c. DIF (Differential Item Functioning)

DIF digunakan untuk mendeteksi bias item pada kategori kelompok MIPA dan IPS di SMA Negeri 8 Palembang. DIF diketahui dengan nilai probabilitas dibawah 5% (0.05). Melihat dari hasil analisis DIF dua item terjangkau bias yakni item 1 dan item 20.

d. Person Fit Order

Terdeteksi sebanyak 24 siswa dengan nilai *infit* MNSQ $> + 1.1$. Penjelajahan lebih lanjut yaitu dengan membandingkan hasil *output Scalogram* dan hasil dari *Most Unexpected Response* dari *Winstep*. Item pada *Scalogram* dan *Most Unexpected Response* tersebut terdapat hal menarik dimana 13 *person* dari 133 *person* dengan HOTS tinggi terindikasi *person misfit*. Sedangkan *person* dengan HOTS rendah yang terjangkau *misfit* berjumlah 11 *person* dari 64 *person*.

e. Person Measure

Tabel 4 merupakan penggalan dari hasil analisis 197 responden, yang lebih dititikberatkan kepada *person-person* yang menarik untuk dibahas lebih lanjut.

Tabel 4. Hasil Penggalan *Output Person Measure*

<i>Person Statistics: Measure Order</i>													
<i>Entry Number</i>	<i>Total Score</i>	<i>Total Count</i>	<i>Measure</i>	<i>Model S.E.</i>	<i>Infit</i>		<i>Outfit</i>		<i>Pt-Measure</i>		<i>Exact Match</i>		<i>Pers- on</i>
					<i>MNSQ</i>	<i>ZS-TD</i>	<i>MNSQ</i>	<i>ZS-TD</i>	<i>Corr.</i>	<i>Exp.</i>	<i>Obs%</i>	<i>Exp%</i>	
15	17	20	1.84	.63	.93	.0	.77	-.2	.27	.12	85.0	85.0	15
184	15	18	1.70	.64	.98	.1	.92	.1	.17	.13	83.3	83.3	184
74	15	18	1.68	.64	1.02	.2	.99	.2	.10	.13	83.3	83.3	74
12	16	20	1.48	.57	1.06	.3	1.18	.5	.00	.14	80.0	80.0	12
186	16	20	1.48	.57	1.06	.3	1.06	.3	.04	.14	80.0	80.0	186
156	1	20	-3.22	1.07	1.20	.5	1.09	.5	.07	.28	95.0	95.0	156
Mean	10.2	20.0	.03	.52	.99	.1	1.00	.1			68.4	68.5	
S.D.	3.9	.3	1.01	.10	.11	.7	.27	.8			12.2	9.4	

Responden yang paling mampu untuk menjawab tes dengan baik yaitu responden 15 (ditandai warna biru) menjawab 17 soal dari 20 soal tes HOTS menunjukkan bahwa dirinya cenderung memiliki HOTS yang lebih baik. Sementara responden dengan nilai *logit* terendah yaitu responden 156 (ditandai warna hijau) hanya mampu menjawab 1 soal dengan baik.

Selanjutnya pengelompokkan kemampuan HOTS siswa dibuat dengan memanfaatkan nilai *separation* (1.59) tabel *Summary Statistic* melalui penggunaan formula *person* strata H, sehingga nilai $H = [(4 * separation) + 1] / 3$ maka $H = [(4 * 1.59) + 1] / 3$, $H = 2.45$ nilai *person* strata (H), sebesar 2 menunjukkan bahwa kelompok responden terbagi dalam dua kelompok besar. Berdasarkan hal tersebut siswa yang memiliki kemampuan HOTS dalam kategori tinggi berjumlah 133 siswa sedangkan siswa

yang memiliki kemampuan HOTS rendah berjumlah 64 siswa.

f. Unidimensionalitas

Tabel 5 menunjukkan bahwa hasil pengukuran keragaman (*raw variance*) data 20.0% sama dengan nilai ekpektasinya 20.0% (ditandai dengan warna ungu). Hal ini menunjukkan persyaratan unidimensionalitas 20% terpenuhi. Hal lain yang mendukung, keragaman yang tidak dapat dijelaskan (*unexplained variance*) semua di bawah 10% yang menunjukkan tingkat independensi item dalam instrumen HOTS berada pada kategori baik. Dengan demikian, berdasarkan nilai *raw variance* dan *unexplained variance* ini dapat disimpulkan bahwa instrumen tes HOTS telah benar-benar mampu mengukur HOTS siswa di SMA Negeri 8 Palembang.

Tabel 5. Unidimensionalitas Tes HOTS

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		-- Empirical --	Modeled	
Total raw variance in observations	=	25.0	100.0%	100.0%
Raw variance explained by measures	=	5.0	20.0%	20.0%
Raw variance explained by persons	=	2.3	9.2%	9.2%
Raw Variance explained by items	=	2.7	10.8%	10.8%
Raw unexplained variance (total)	=	20.0	80.0%	100.0% 80.0%
Unexplnd variance in 1st contrast	=	1.9	7.6%	9.4%
Unexplnd variance in 2nd contrast	=	1.7	6.9%	8.7%
Unexplnd variance in 3rd contrast	=	1.5	6.2%	7.7%
Unexplnd variance in 4th contrast	=	1.5	5.9%	7.4%

Merevisi Tes (Revising the Test)

Pengambilan keputusan dengan pertimbangan dari hasil analisis DIF dimana terdapat dua item (item 1 dan 20) yang tidak

fair dalam mengukur HOTS siswa sehingga item tersebut dieliminasi atau dihapuskan.

Validasi Tes (Validation of Test)

Berdasarkan beberapa hasil analisis validitas di atas dapat diketahui semua item dalam tes HOTS berkriteria valid. Reliabilitas instrumen HOTS melalui model Rasch diketahui dari tabel *Summary Statistic*. Beracuan dari pola jawaban responden didapat bahwa nilai *infit* MNSQ adalah 0.99 dan *outfit* MNSQ adalah 1.00 menunjukkan bahwa secara keseluruhan pola jawaban responden pada instrumen berkriteria bagus. Untuk *person* nilai *infit* ZSTD adalah 0.1 dan *outfit* ZSTD 0.1 juga menunjukkan secara keseluruhan pola jawaban responden memiliki kesesuaian dengan model. Akan tetapi nilai reliabilitas responden 0.72 yang artinya responden dalam kategori memiliki kemampuan tingkat tinggi yang cukup.

Nilai pengujian instrumen untuk setiap itemnya, terlihat bahwa nilai *Infit* MNSQ adalah 1.01 dan *Outfit* MNSQ 1.00 sedangkan untuk ZSTD nilai *Infit* ZSTD adalah -0.1 dan *Outfit* ZSTD -0.1. Kedua hal ini menunjukkan bahwa keseluruhan instrumen berada pada kategori bagus sekali mengukur HOTS siswa dengan nilai reliabilitas instrumen yang 0,93. Nilai *alpha cronbach* (KR-20) yang mengukur interaksi antara responden dan item berada pada kriteria bagus yaitu 0,72. Secara keseluruhan ini menunjukkan bahwa data aktual yang diperoleh dalam riset ini mendapat hasil yang sesuai dan baik dengan persyaratan model Rasch.

Mengembangkan Norma (Developing the Norms)

Norma acuan penilaian yang ditetapkan adalah Penilaian Acuan Kriteria (PAK) yang didasarkan pada Kriteria Ketuntasan Minimal (KKM). Nilai KKM mata pelajaran Biologi di SMA Negeri 8 Palembang yaitu 65 berdasarkan hasil uji coba di lapangan dapat diketahui bahwa hanya 63 siswa dari 197 siswa yang dapat mencapai nilai KKM mata pelajaran Biologi di SMA tersebut.

Menyelesaikan Tes Manual (Completing the Manual Test)

Manual test yang dihasilkan berupa buku yang berisi uraian materi tentang cara

pembuatan soal berbasis HOTS. Secara lebih rinci dapat dilihat pada lampiran 1.

Mengacu pada hasil analisis *Item Fit Order* keputusan pertama yang dapat diambil kedua item (1 dan 20) tidak dieliminasi dan tetap dinyatakan layak untuk mengukur HOTS siswa. Hal ini sesuai dengan pandangan dari Sumintono & Widhiarso (2014), apabila satu kriteria dari tiga kriteria (*outfit* MNSQ, *outfit* ZSTD serta *PT-Measure Correlation*) yang dihasilkan dari analisis dengan perangkat lunak menunjukkan bahwa item telah memenuhi syarat, maka item tersebut masih dapat digunakan.

Akan tetapi, menindaklanjuti dari hasil analisis DIF dimana item 1 dan 20 justru lebih berpihak pada kelompok IPS dan menimbulkan ketidakadilan terhadap kelompok MIPA dikarenakan nilai DIF *measure* dari item 1 untuk kelompok MIPA adalah -1.9 sedangkan pada kelompok IPS adalah -3.05, item 20 pada kelompok MIPA memiliki nilai DIF *measure* -0.35 dan pada kelompok IPS adalah -1.38.

Temuan ini, dijadikan landasan untuk menghapus item 1 dan item 20 hal ini diperkuat dengan pandangan dari Safar, & Embong (2015) yang menyatakan bahwa item-item yang terlalu berpihak pada salah satu kelompok perlu diasingkan dalam analisis seterusnya. Selanjutnya dari *Item Fit Order* dapat dilihat lebih lanjut juga bahwa terdapat 6 item yang berasal dari 3 aspek HOTS yang sama (lihat kembali tabel 5, ditandai dengan warna biru) yakni item nomor 6 (*logit* + 1.04) dan 11 (*logit* + 1.04) dari aspek menganalisis (C4), item nomor 10 (*logit* + 0.97), 12 (*logit* + 0.97) dari aspek mengevaluasi (C5) serta item nomor 13 (*logit* + 0.95) dan item nomor 18 (*logit* + 0.95) dari aspek mencipta (C6). Samanya nilai *logit* tersebut menunjukkan bahwa item tersebut telah benar – benar mampu mengukur kemampuan menganalisis, mengevaluasi dan mencipta menurut sebagian besar responden.

Apabila ditinjau dari hasil analisis *item map* dapat diketahui bahwa item yang paling banyak dijawab dengan benar adalah item pada tingkat menganalisis (C4) dengan jumlah 5 item yang berkategori mudah dari 7 item yang dibuat. Hal ini mengindikasikan bahwa kemampuan menganalisis siswa di

SMA Negeri 8 sudah cukup baik, sedangkan kemampuan mengevaluasi (C5) dan mencipta (C6) masih kurang baik, asumsi ini didukung dengan sebaran item C5 dan C6 yang berada pada bagian atas *Item Map*.

Hal lain yang menarik untuk dibahas lebih lanjut yaitu hasil *output person measure* dimana *person 12* (skor 16) dan 186 (skor 16) memiliki skor yang lebih tinggi dibandingkan dengan *person 184* (skor 15) dan 74 (skor 15) (lihat kolom *total score* tabel 9) akan tetapi ia justru berada pada urutan yang lebih tinggi dari *person 12* dan 186. Apabila dilihat dari sisi skor dan nilai, tentu *person 12* dan 186 menduduki peringkat lebih tinggi dari kedua siswa tersebut, tetapi tidak demikian.

Penyebab *person 184* memiliki skor yang lebih rendah tetapi berada pada tingkat yang lebih tinggi dari *person 12* dan 186 dikarenakan terdapat dua item yang tidak mampu dijawabnya, yaitu item 3 dan 7, padahal *person 184* mampu menjawab item pada tingkat yang lebih sulit yakni item 13, 12, dan 11 bahkan 16. Hasil ini dapat dilihat pada *output Scalogram*. Asumsi lain yang menyebabkan *person 184* lebih tinggi nilai *logitnya* (+ 1.70) daripada *person 12* dan 186 yakni adanya jawaban yang tidak terduga dari *person* ini yang terdeteksi dari *Most Unexpected Response* pada item 2 dengan adanya angka 0, sebagai petunjuk bahwa sesungguhnya *person* ini mampu untuk menjawabnya dengan tepat. Sehingga skor yang seharusnya dimiliki oleh *person 184* adalah 17 bukan 15. Hal ini merupakan suatu bentuk kurang cermatan (*careless*) *person 184* dan bukan diakibatkan ketidakmampuannya.

Selaras dengan *person 184*, kasus pada *person 74* juga demikian, *person 74* berkedudukan lebih tinggi dari *person 12* dan 186 dengan nilai *logit* (+ 1.68) namun melalui hasil *output Scalogram* diketahui bahwa terdapat dua item yang tidak mampu dijawabnya, yaitu item 14 dan 16, sesungguhnya *person 74* mampu menjawab item 14 dan 16 tersebut asumsi ini didasarkan pada soal dengan tingkat kesulitan lebih tinggipun dapat diselesaikan dengan benar (*person 74* mampu menjawab item 18). Sementara munculnya jawaban tidak terduga yang ditandai adanya angka 0 yang muncul pada hasil *output Most*

Unexpected Person untuk item 2 dan 6 merupakan suatu bentuk kurang cermatan (*careless*) *person 74* dan bukan diakibatkan ketidakmampuannya. Dari sisi abilitas, *person 74* sebenarnya mampu menyelesaikan soal ini dengan mudah dan betul. Hal inilah yang menyebabkan *person 74* meskipun memiliki skor yang lebih rendah dari *person 12* dan 186 tetapi memiliki HOTS yang lebih baik dari *person* tersebut.

Berdasarkan analisis *person misfit* dan *person measure* HOTS siswa di SMA Negeri 8 Palembang dapat ditarik *point – point* penting bahwa :

1. *Person* dengan HOTS tinggi belum tentu menjamin *person* tidak terjangkit *misfit*.
2. Skor tidak dapat dijadikan patokan kemampuan seseorang ketika tes yang diujikan dalam konteks pilihan ganda.
3. Data hilang dan ketidakmampuan siswa menjawab suatu item tidak dapat langsung dijadikan acuan jika siswa tidak mampu menjawabnya tanpa analisis lebih lanjut.

Alasan adanya *point-point* penting di atas dikarenakan abilitas tidak bergantung pada jawaban benar yang ditunjukkan oleh responden. Inilah yang menyebabkan skor mentah pada tes ini tidak dapat dijadikan skor *final* tanpa analisis yang lebih mendalam lagi. Dikarenakan responden dalam suatu penelitian juga dapat mengalami beberapa kejadian-kejadian seperti di atas yaitu adanya jawaban tebakan (*lucky guess*), kurang cermat (*careless*) serta adanya data yang hilang (tidak dijawab oleh siswa).

Banyaknya siswa yang terjangkit *misfit*, munculnya jawaban tidak terduga, kemampuan siswa yang mayoritas berada pada aspek menganalisis (C4) menunjukkan bahwa masih banyak siswa yang memiliki kemampuan HOT rendah meskipun *input* siswa di SMA Negeri 8 Palembang baik. Estimasi ini didukung juga dengan hasil reliabilitas *person* yang menunjukkan kemampuan HOT *person* berada pada kriteria cukup (0,72).

Inti dari tujuan penelitian pengembangan ini adalah menghasilkan soal berbasis HOTS yang layak lengkap dengan panduan penyusunan soalnya (*manual test*) di SMA Negeri 8 Palembang dalam rangka

menyempurnakan atau memperbaiki kualitas soal yang kurang baik pada kegiatan evaluasi pembelajaran.

Uji coba keefektifan *manual test* sendiri dilakukan terhadap 6 kelas di SMA Negeri 8 Palembang dengan jumlah 197 siswa melalui penggunaan instrumen tes HOTS yang cara pembuatan soalnya menggunakan alur yang terkandung dalam *manual test*. Berdasarkan hasil penilaian *manual test* dari para ahli, diperoleh hasil bahwa nilai R melebihi 75% hal tersebut menunjukkan kelayakan terhadap *manual test* dan siap untuk digunakan bagi kalangan pendidik maupun calon tenaga pendidik.

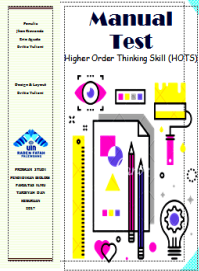
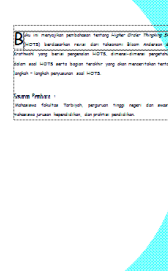

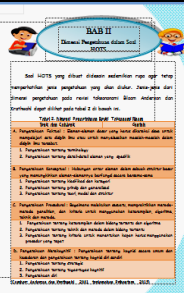
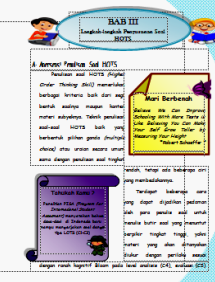
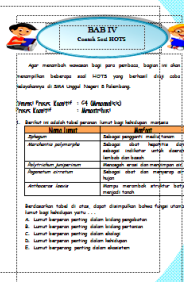
Simpulan

Berdasarkan analisis data dari uji coba lapangan panduan penyusunan soal HOTS (*manual test*) yang dikembangkan di SMA Negeri 8 Palembang menunjukkan kelayakan pada kriteria sangat layak (A). Hasil uji coba menunjukkan semua item HOTS *fit* dengan nilai reliabilitas yang berada pada kategori sangat bagus (0,93) meskipun 2 item terdeteksi DIF dan dihapuskan. Sementara itu, hasil reliabilitas *person* menunjukkan kemampuan HOT *person* berada pada kriteria cukup (0,72).

Daftar Pustaka

- Arifin, Z. (2013). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Bond, T.G. & Fox, C.M. (2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences (2nd ed)*. London: Lawrence Erlbaum Associates, Publishers.
- Borich, G.D. (1994). *Observation Skill for Effective Teaching (2nd ed)*. New York: McMillan Publishing Company.
- Brookhart, S.M. (2010). *How to Assess Higher Order Thinking Skill in Your Classroom*. Alexandria: ASCD Publisher.
- Dewi, N.D.L. (2015). Pengembangan Instrumen Penilaian IPA untuk Memetakan *Critical Thinking* dan *Practical Skill* Peserta Didik SMP. *Tesis*, tidak dipublikasikan. Universitas Negeri Yogyakarta.
- Duhita, R. (2015). Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Pengantar Akuntansi Kelas X Akuntansi SMK Negeri 1 Godean Tahun Ajaran 2014/2015. *Skripsi*. Universitas Negeri Yogyakarta.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendikia Press.
- McIntire, S.A. dan Miller, L.A. (2007). *Foundation of Psychological Testing A Practical Approach*. California: Sage Production.
- Misbach, I.H. & Sumintono, B. (2014). *Pengembangan dan Validasi Instrumen Persepsi Siswa terhadap Karakter Moral Guru di Indonesia dengan Model Rasch*. Makalah diseminarkan pada Seminar Nasional Psikometri, Universitas Muhammadiyah Surakarta.
- Safar, J.H. & Embong. (2015). The Construction Of Peer-Based Behavioural Assesment Instrument For Secondary School Students. *Jurnal Persatuan Saintis Muslim Sriwijaya*, 5 (2), 46-61.
- Sarnapi. (2016). *Peringkat Pendidikan Indonesia Masih Rendah*. Diakses dari <http://www.pikiranrakyat.com/pendidikan/2016/06/18/peringkatpendidikan-indonesia-masih-rendah372187> pada 18 Juni 2016.
- Sofiana, S. (2010). Analisis Butir Soal Ulangan Kenaikan Kelas Mata Pelajaran Kimia Kelas X SMA Negeri 8 Surakarta Tahun Ajaran 2009/2010. *Skripsi*. Universitas Negeri Sebelas Maret.
- Sumintono, B. & Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: TrimKom Publishing House.
- Sumintono, B. & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assesment Pendidikan*. Cimahi: TrimKom Publishing House.

Lampiran 1. Bagian-bagian Pedoman Penyusunan Soal HOTS (Manual Test)

Bagian Buku	Keterangan	Bagian Buku	Keterangan
	<p>Cover didesain dengan warna, gambar dan tulisan yang menarik. Cover depan terdiri dari nama penulis (Jhon Riswanda, Erie Agusta dan Evitia Yuliani) juga terdapat judul buku (<i>Manual Test Higher Order Thinking Skill</i>)</p>		<p>Cover belakang didesain dengan lebih sederhana yang memuat gambaran secara umum isi <i>manual test</i> dan sasaran pembaca.</p>
<p>Bagian cover depan</p>		<p>Bagian cover belakang</p>	
	<p>Bagian ini menjelaskan mengenai pengertian HOTS yang sebelumnya telah dikemukakan secara ringkas pada bagian apersepsi. Bagian ini juga menjelaskan proses-proses kognitif HOTS serta KKO (Kata Kerja Operasional) yang ada pada setiap aspek HOTS.</p>		<p>Bagian ini menjelaskan macam-macam dimensi pengetahuan yang terlibat dalam soal HOTS, setiap dimensi pengetahuan tersebut memiliki aspeknya masing-masing yang disertai dengan contoh dalam konteks yang kehidupan sehari - hari.</p>
<p>Bab I Berisi Pengenalan HOTS</p>		<p>Bab II Berisi Dimensi Pengetahuan dalam Soal HOTS</p>	
	<p>Bab ini dibuat dengan menjelaskan langkah-langkah penyusunan soal HOTS, pada bagian ini pembaca akan dilatih untuk mengaplikasikan kemampuan HOT yang dimilikinya dalam membuat soal-soal.</p>		<p>Bagian ini memuat bentuk-bentuk soal HOTS yang telah diuji kelayakannya di SMA Negeri 8 Palembang.</p>
<p>Bab III Berisi Langkah-langkah Penyusunan Soal HOTS</p>		<p>Bab IV Contoh Soal HOTS</p>	