



# Jurnal

## Penelitian dan Evaluasi Pendidikan

- Eli Istikom, Dyanari Mandagi, Supriem* Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Endok (PysTHOPE) Peserta Didik SMA
- Gustan, Yuhandiyanto, Dyanari Mandagi* Pengembangan Asesmen Hasil Belajar PENJASORKES Siswa SMA pada Permainan Belah ketupat
- I. W. Gunawan, Budwan Kertanegara, Sri Purwati, Saardianar* Pengembangan Model Evaluasi Program Layanan Pendidikan Anak Usia Dini (PAUD)
- M. Dody, Kasmah, Suryani* Pengembangan Tes Diagnostik Kesulitan Belajar Matematika di SD
- Mubli, Iqbal* Bina Operasi Sekolah Menengah Kejuruan Negeri Kangen SM/RSN di DIY
- Muhammad Gafur, Yematus, Samudro Mulya* Model Evaluasi Reflektif Keterampilan Pendidikan Bahasa Inggris dalam Pengembangan Karakter Bangsa
- Sri Rajadi, Dyanari Mandagi, Komali* Metode *Stowand Satig* untuk Ujian Nasional di Sekolah Dasar
- Susanto, Meliyani, Dyanari Mandagi* Pengembangan Instrumen Evaluasi Uji Kompetensi Keahlian (UKK) Administrasi Perkantoran di SMK
- Yuni Masayu, Dyanari Mandagi* Analisis Metode *Cawitig* pada Tes Berskala Besar
- Zaenal Hamid, Dyanari Mandagi, H. Faridi* Model Asesmen Kepemimpinan Pembelajaran Kepala Sekolah Pendidikan Dasar

# Editorial Team

## ***Director of Publication***

[Djemari Mardapi](#), (Scopus ID: 57170711400) Faculty of Engineering, Universitas Negeri Yogyakarta, Indonesia

## ***Editor in Chief***

[Samsul Hadi](#), (Scopus ID: 57060009200) Faculty of Engineering, Universitas Negeri Yogyakarta, Indonesia

## ***Associate Editors***

[Nur Hidayanto Pancoro Setyo Putro](#), (Scopus ID: 57194874971) Faculty of Languages and Arts, Universitas Negeri Yogyakarta, Indonesia

## ***Editorial Board***

[Heru Widiatmo](#), American College Testing, United States

[Risky Setiawan](#), (Scopus ID: 57195803837) Department of Educational Research and Evaluation, Universitas Negeri Yogyakarta

[Undang Rosidin](#), (Scopus ID: 57201345852) Pendidikan Fisika, Universitas Lampung, Indonesia

[W. Wasis](#), (Scopus ID: 57195518718) Universitas Negeri Surabaya, Indonesia

[Syukrul Hamdi](#), (Scopus ID: 57202629545) Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia

[Edi Istiyono](#), (Scopus ID: 57193458018) Department of Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia

[J. Jailani](#), (Scopus ID: 57204033371) Department of Mathematics Education, Universitas Negeri Yogyakarta, Indonesia

[Badrun Kartowagiran](#), (Scopus ID: 57194710965) Faculty of Engineering, Universitas Negeri Yogyakarta, Indonesia

[S. Sudiyatno](#), (Scopus ID: 57201692859) Faculty of Engineering, Universitas Negeri Yogyakarta, Indonesia

[Alita Arifiana Anisa](#), Lembaga Pendidikan dan Pengembangan Profesi Indonesia, Indonesia

# Vol 18, No 1 (2014)

## Table of Contents

### Articles

<a href="#"><u>PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA (PysTHOTS) PESERTA DIDIK SMA</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 1-12
 <i>Edi Istiyono, Djemari, Mardapi, Suparno, Suparno</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2120">10.21831/pep.v18i1.2120</a>	
<a href="#"><u>PENGEMBANGAN ASESMEN HASIL BELAJAR PENJASORKES SISWA SMA PADA PERMAINAN BOLAVOLI</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 13-29
 <i>Guntur Guntur, Sukadiyanto, Sukadiyanto, Djemari, Mardapi</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2121">10.21831/pep.v18i1.2121</a>	
<a href="#"><u>PENGEMBANGAN MODEL EVALUASI PROGRAM LAYANAN PENDIDIKAN ANAK USIA DINI (PAUD)</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 30-43
 <i>I Wayan Gunartha, Badrun Kartowagiran, Siti Partini, Suardiman</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2122">10.21831/pep.v18i1.2122</a>	
<a href="#"><u>PENGEMBANGAN TES DIAGNOSTIK KESULITAN BELAJAR MATEMATIKA DI SD</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 44-56
 <i>M. Duskri, Kumaidi, Kumaidi, Suryanto, Suryanto</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2123">10.21831/pep.v18i1.2123</a>	
<a href="#"><u>BIAYA OPERASI SEKOLAH MENENGAH KEJURUAN NEGERI KATEGORI SBI/RSBI DI DIY</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 57-71
 <i>Moch Alip</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2124">10.21831/pep.v18i1.2124</a>	
<a href="#"><u>MODEL EVALUASI REFLEKTIF KURIKULUM PENDIDIKAN BAHASA INGGRIS DALAM PENGEMBANGAN KARAKTER BANGSA</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 72-88
 <i>Monika Gultom, Sumarno, Sumarno, Suwarsih, Madya</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2125">10.21831/pep.v18i1.2125</a>	
<a href="#"><u>METODE STANDARD SETTING UNTUK UJIAN NASIONAL DI SEKOLAH DASAR</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 89-97
 <i>Sri Rejeki, Djemari, Mardapi, Kumaidi, Kumaidi</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2126">10.21831/pep.v18i1.2126</a>	
<a href="#"><u>PENGEMBANGAN INSTRUMEN EVALUASI UJI KOMPETENSI KEAHLIAN (UKK) ADMINISTRASI PERKANTORAN DI SMK</u></a>	<a href="#"><u>FULL TEXT PDF</u></a> 98-114
 <i>Suranto Suranto, Muhyadi, Muhyadi, Djemari, Mardapi</i>	
 <a href="https://doi.org/10.21831/pep.v18i1.2127">10.21831/pep.v18i1.2127</a>	

**ANALISIS METODE CHEATING PADA TES BERSKALA BESAR**

 Yance

Manoppo,

Djemari

Mardapi




[10.21831/pep.v18i1.2128](https://doi.org/10.21831/pep.v18i1.2128)

[FULL TEXT PDF](#)

115-128

**MODEL ASESMEN KEPEMIMPINAN PEMBELAJARAN KEPALA SEKOLAH PENDIDIKAN DASAR**

 Zaenal

Fanani,

Djemari

Mardapi,

Wuradji

Wuradji



[10.21831/pep.v18i1.2129](https://doi.org/10.21831/pep.v18i1.2129)

[FULL TEXT PDF](#)

129-145

## PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA (PysTHOTS) PESERTA DIDIK SMA

<sup>1)</sup>Edi Istiyono, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Suparno

<sup>1),3)</sup>Jurusan Pendidikan Fisika FMIPA Universitas Negeri Yogyakarta

<sup>2)</sup>Jurusan Pendidikan Teknik Elektro FT Universitas Negeri Yogyakarta

<sup>1)</sup>edi\_istiyono@uny.ac.id, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>suparno\_mipa@uny.ac.id

### Abstrak

Penelitian dilakukan untuk mengembangkan instrumen kemampuan berpikir tingkat tinggi fisika (PhysTHOTS) peserta didik SMA dan mendapatkan karakteristik PhysTHOTS. Kisi-kisi instrumen disusun berdasarkan aspek dan subaspek kemampuan berpikir tingkat tinggi, yang selanjutnya digunakan untuk menyusun *item-item*. Instrumen terdiri atas dua perangkat tes yang masing-masing memiliki 26 *item* termasuk delapan *anchor item* dan telah divalidasi oleh ahli pengukuran, ahli pendidikan fisika, ahli fisika, dan praktisi. Instrumen yang telah divalidasi diujicobakan pada 1.001 siswa dari sepuluh SMAN di Daerah Istimewa Yogyakarta. Data politomus dianalisis menggunakan *Partial Credit Model* (PCM). Hasil uji coba menunjukkan bahwa semua *item* sebanyak 44 dan instrumen PhysTHOTS terbukti *fit* dengan PCM, reliabilitas instrumen sebesar 0,95, indeks kesukaran *item* mulai -0,86 sampai 1,06 yang berarti semua *item* dalam kategori baik. Dengan demikian, PhysTHOTS memenuhi syarat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik SMA.

**Kata kunci:** pengembangan instrumen, tes kemampuan berpikir tingkat tinggi, fisika, politomus, dan PCM

## DEVELOPING HIGHER ORDER THINKING SKILL TEST OF PHYSICS (PhysTHOTS) FOR SENIOR HIGH SCHOOL STUDENTS

<sup>1)</sup>Edi Istiyono, <sup>2)</sup>Djemari Mardapi, <sup>3)</sup>Suparno

<sup>1),3)</sup>Jurusan Pendidikan Fisika FMIPA Universitas Negeri Yogyakarta

<sup>2)</sup>Jurusan Pendidikan Teknik Elektro FT Universitas Negeri Yogyakarta

<sup>1)</sup>edi\_istiyono@uny.ac.id, <sup>2)</sup>djemarimardapi@gmail.com, <sup>3)</sup>suparno\_mipa@uny.ac.id

### Abstrak

The objectives of this research were to develop an instrument for measuring senior high school students' physics higher order thinking skills (PhysTHOTS) and to obtain the characteristics of the PhysHOTS. The instrument blue print was developed based on the aspects and sub-aspects of high order thinking skills, then it was used to develop the items. Two sets of instrument consisting of 26 items and each, including eight anchor items were then validated by promotors, measurement experts, physics specialists, physics education experts, and practitioners. The validated instruments were then tried out on 1,001 students of ten senior high schools throughout Special Province of Yogyakarta. The polytomous data were analyzed according to the Partial Credit Model (PCM). The results show that the 44 items and PhysTHOTS were fit to the PCM, the reliability of the test was 0.95, the items' difficulty indexes were between -0.86 and 1.06. Therefore, the PhysTHOTS are qualified to measure senior high school students' physics higher order thinking skills.

**Keywords:** instrument development, physics test of higher order thinking skills, polytomous, and PCM

## Pendahuluan

Dewasa ini dunia berada pada era globalisasi. Pada era ini persaingan cukup ketat, yakni persaingan kualitas sumber daya manusia (SDM). Kualitas SDM bangsa ditentukan oleh tingkat pendidikan bangsa tersebut. Peningkatan kualitas pendidikan dimulai dari peningkatan kualitas pembelajaran. Peningkatan kualitas pembelajaran dapat dimulai dengan menyusun tujuan pembelajaran yang tepat.

Salah satu tujuan Mata Pelajaran Fisika di SMA agar peserta didik memiliki kemampuan mengembangkan kemampuan bernalar dalam berpikir analisis induktif dan deduktif dengan menggunakan konsep dan prinsip fisika untuk menjelaskan berbagai peristiwa alam dan menyelesaikan masalah, baik secara kualitatif maupun kuantitatif (BSNP, 2006, p.160). Hal ini diperkuat dengan anjuran pemerintah dalam kurikulum tingkat satuan pendidikan. Penilaian (asesmen) hendaknya direncanakan untuk mengukur pengetahuan dan konsep, keterampilan proses sains (KPS), dan penalaran tingkat tinggi (Pusat Kurikulum, 2007, pp.23-24). Dengan demikian, melalui pembelajaran fisika diharapkan peserta didik dapat mengembangkan diri dalam berpikir. Peserta didik dituntut tidak hanya memiliki kemampuan berpikir tingkat rendah (*lower order thinking*), tetapi sampai pada kemampuan berpikir tingkat tinggi (*higher order thinking, HOT*).

Berkenaan dengan kemampuan berpikir tingkat tinggi ini, fakta menunjukkan bahwa prestasi fisika yang diukur pada aspek *reasoning* Indonesia berada pada ranking 40 dari 42 negara (TIMSS & PIRLS International Study Center, 2012, p.48). Hal senada dinyatakan Efendi (2010, p.393) bahwa berdasarkan hasil TIMSS dapat ditarik kesimpulan sebagai berikut: (1) rata-rata capaian fisika siswa Indonesia ditinjau dari aspek kognitif (*knowing, applying, reasoning*) masih rendah; (2) kecenderungan capaian fisika siswa Indonesia selalu menurun pada tiap aspek kognitif sehingga kemampuan fisika siswa Indonesia harus ditingkatkan pada semua aspek, khususnya aspek *reason-*

*ing* dengan cara membekali siswa kemampuan berpikir tingkat tinggi. Dengan demikian, prestasi fisika yang menuntut kemampuan berpikir tingkat tinggi siswa sekolah menengah Indonesia, di kancah internasional masih rendah. Prestasi belajar fisika rendah dapat disebabkan karena proses pembelajaran atau model asesmennya yang tidak tepat. Dalam hal ini hanya akan dibahas tentang asesmennya, karena asesmen yang tepat dapat mendorong siswa untuk belajar dengan berpikir tingkat tinggi.

Berdasarkan teori perkembangan Piaget, tahap operasional formal adalah tahap anak mulai berusia sebelas tahun. Pada tahap ini anak sudah mulai dapat mengembangkan kemampuan untuk memanipulasi konsep abstrak melalui penggunaan proposisi dan hipotesis (Piaget, 2005, p.122 dan Reedal, 2010, p.7). Usia siswa SMA antara 15 sampai 18 tahun, sehingga kemampuan berpikir tingkat tinggi mereka sudah mapan.

Piaget mengatakan bahwa kematangan dan kesiapan seseorang harus menunggu serta harus cocok antara pengaruh dari luar dan perkembangan di dalam dirinya (*match*), tetapi tidak demikian menurut Vygotsky. Ada sesuatu di atas tahap perkembangan itu (*plus one matching*). Ada daerah-daerah yang sangat sensitif untuk diaktualisasikan dalam diri anak yang dinamakan Zone Proximal Development (ZPD) (Albert, Corea, & Macadino, 2012, p.14). Dengan menerapkan konsep ZPD pada pendidikan, maka pembelajaran akan memajukan perkembangan anak. Salah satu wujud konkret implikasi dari teori Vygotsky adalah dilaksanakannya akselerasi belajar bagi anak berbakat, pendidikan bagi anak-anak yang mempunyai kemampuan intelektual luar biasa, dalam proses pembelajaran harus selalu meningkatkan kadar mental atau berpikir tingkat tinggi.

Menurut taksonomi Bloom yang telah direvisi proses kognitif terbagi menjadi kemampuan berpikir tingkat rendah (*Lower Order Thinking*) dan kemampuan berpikir tingkat tinggi (*Higher Order Thinking*). Kemampuan yang termasuk *LOT* adalah kemampuan mengingat (*remember*), memahami (*understand*), dan menerapkan (*apply*), sedangkan

*HOT* meliputi kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*) (Anderson & Krathwohl, 2001, p.30). Taksonomi Bloom sudah lama diterapkan dalam bidang pendidikan dan sudah lama digunakan. Taksonomi Bloom masih digunakan dalam banyak kurikulum dan bahan pengajaran (Brookhart, 2010, p.39, Schraw and Robinson, 2011, pp.158-159). Dengan demikian, kemampuan berpikir tingkat tinggi fisika (*Physics Higher Order Thinking*) meliputi kemampuan fisika dalam menganalisis, mengevaluasi, dan menciptakan.

Menurut Brookhart (2010, p.5) kemampuan berpikir tingkat tinggi (HOTS) adalah (1) berpikir tingkat tinggi berada pada bagian atas taksonomi kognitif Bloom, (2) tujuan pengajaran di balik taksonomi kognitif yang dapat membekali peserta didik untuk melakukan transfer pengetahuan, (3) mampu berpikir artinya peserta didik mampu menerapkan pengetahuan dan keterampilan yang mereka kembangkan selama belajar pada konteks yang baru. Dalam hal ini yang dimaksud “baru” adalah aplikasi konsep yang belum terpikirkan sebelumnya oleh peserta didik, namun konsep tersebut sudah diajarkan, ini berarti belum tentu sesuatu yang universal baru. Berpikir tingkat tinggi berarti kemampuan peserta didik untuk menghubungkan pembelajaran dengan hal-hal lain yang belum pernah diajarkan.

Untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan, diperlukan penilaian. Penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik (Peraturan Menteri Pendidikan Nasional No 20, 2007). Penilaian dapat dilakukan secara lisan ataupun tertulis. Penilaian secara tertulis dilakukan dengan tes tertulis. Secara garis besar ada dua bentuk soal tes tertulis, yaitu: memilih jawaban dan mensuplai jawaban. Soal tes tertulis yang jawabannya dengan memilih jawaban antara lain: pilihan ganda, dua pilihan (benar-salah, ya-tidak), menjodohkan, dan sebab-akibat.

Perlu diketahui bahwa model penilaian juga berpengaruh terhadap kemampuan berpikir siswa. Menurut Van den Berg (2008, p.15) bahwa kurikulum memiliki potensi yang kaya untuk mengembangkan keterampilan berpikir tingkat tinggi peserta didik. Guru harus merencanakan dengan baik dan melibatkan peserta didik dalam kegiatan pembelajaran yang dapat mendorong dan mengembangkan kemampuan berpikir tingkat tinggi tersebut. Penilaian dapat diimplementasikan untuk membantu siswa dalam meningkatkan kemampuan berpikir tingkat tinggi mereka. Hal ini didukung pendapat lain, bahwa pertanyaan berpikir tingkat tinggi dapat mendorong siswa untuk berpikir secara mendalam tentang materi pelajaran (Barnett & Francis (2012, p.209). Berdasarkan dua pendapat ini dapat disimpulkan bahwa tes kemampuan berpikir tingkat tinggi dapat memberikan rangsangan kepada siswa untuk mengembangkan kemampuan berpikir tingkat tinggi juga.

Nitko & Brookhart (2011, p.223) menjelaskan bahwa ketentuan dasar penilaian kemampuan berpikir tingkat tinggi adalah menggunakan tugas-tugas yang memerlukan penggunaan pengetahuan dan keterampilan dalam situasi baru. Untuk melakukan penilaian terhadap kemampuan *higher order thinking* harus menggunakan bahan-bahan baru. Salah satu cara yang dapat dilakukan adalah dengan menggunakan set-set *item* yang bergantung pada konteks.

Kenyataan bahwa tes pilihan ganda lebih banyak digunakan dari pada bentuk tes yang lain. Hal ini karena tes pilihan ganda memiliki kelebihan-kelebihan, antara lain: (1) materi yang diujikan dapat mencakup sebagian besar bahan pembelajaran, (2) jawaban siswa dapat dikoreksi dengan mudah dan cepat, (3) jawaban setiap pertanyaan sudah pasti benar atau salah, sehingga penilaian objektif (Sudjana, 1990, p.49). Walaupun ada juga kelemahan tes ini, yaitu: (1) kemungkinan peserta didik untuk melakukan tebakan jawaban masih cukup besar dan (2) proses berpikir siswa tidak dapat dilihat dengan nyata (Sudjana, 1990, p.49). Di samping itu, kelemahan lain tes objektif yakni:

(1) percaya diri yang tinggi pada testi dan (2) terjadinya kecurangan (*cheating*).

Untuk menghindari kecurangan (*cheating*), misalnya kerja sama dengan peserta didik lain, maka format (set) tes yang dikerjakan peserta didik yang berdekatan sebaiknya berbeda. Oleh karena itu, diperlukan minimal dua perangkat tes.

Berdasarkan hasil survei pendahuluan dengan cara melakukan wawancara dengan guru-guru fisika SMA di Daerah Istimewa Yogyakarta (DIY), diperoleh informasi bahwa sebagian besar di sekolah, baik pada tes tengah semester maupun tes akhir semester umumnya menggunakan tes pilihan ganda biasa. Jadi, tes pilihan ganda masih merupakan primadona dalam mendapatkan data prestasi belajar fisika siswa SMA. Kenyataan bahwa tes pilihan ganda yang digunakan di SMA untuk tes hasil belajar mata pelajaran fisika baru mengukur kemampuan mengingat, memahami, dan menerapkan. Jadi, tes pilihan ganda yang digunakan di SMA baru mengukur kemampuan berpikir tingkat rendah (*Lower Order Thinking (LOT)*) belum mengukur kemampuan berpikir tingkat tinggi Fisika (*PhysHOT*).

Penilaian dalam pendidikan menggunakan dua macam teori pengukuran, yakni: teori pengukuran klasik dan teori pengukuran modern. Teori Tes Klasik disebut juga dengan *Classical True-Score Theory*, dinamakan Teori Tes Klasik karena unsur-unsur teori ini sudah dikembangkan dan diaplikasikan sejak lama, namun tetap bertahan hingga sekarang (Suryabrata, 2002, p.21). Menurut teori pengukuran klasik penskoran hasil tes biasanya dilakukan secara parsial berdasarkan langkah-langkah yang harus ditempuh untuk menjawab benar suatu butir soal. Penskoran dilakukan per langkah dan skor per *item* peserta diperoleh dengan menjumlah skor peserta didik tiap langkah, dan kemampuan diestimasi dengan skor mentah. Model penskoran seperti ini belum tentu tepat, karena tingkat kesulitan tiap langkah tidak diperhitungkan.

Penilaian hasil ujian peserta didik didasarkan pada tahap-tahap yang dapat di-

selesaikan peserta didik. Walaupun hanya menyelesaikan tahap awal saja, peserta ujian sudah mendapatkan nilai. Nilai tertinggi tentu saja didapatkan ketika peserta ujian telah menyelesaikan semua tahapan soal ujian dalam butir tersebut. Prosedur penilaian tersebut sebenarnya sama dengan bagaimana individu merespon butir dalam skala psikologi. Misalnya, sebuah butir yang menyediakan empat kategori respons dari '*tidak pernah*', '*jarang*', '*sering*', dan '*selalu*' analog dengan tahap penyelesaian. Menyelesaikan soal hanya sampai tahap pertama analog dengan kategori '*tidak pernah*' sedangkan kalau sudah sampai tahap akhir, analog dengan kategori '*selalu*'. Asumsi ini kemudian dikembangkan menjadi *partial credit model (PCM)*. Ketika diasumsikan bahwa sebuah *item* mengikuti pola kredit parsial maka kemampuan individu lebih tinggi diharapkan memiliki skor yang lebih tinggi daripada individu yang memiliki kemampuan rendah (Widhiarso, 2010, p.6). Menurut Wright & Masters, PCM juga sesuai untuk menganalisis respon pada pengukuran berpikir kritis dan pemahaman konseptual dalam sains (Van der Linden & Hambleton, 1997, pp. 101-102).

PCM dikembangkan untuk menganalisis *item* tes yang memerlukan beberapa langkah penyelesaian. PCM dapat diberikan pada langkah-langkah yang dapat dikerjakan oleh individu. Dengan demikian, PCM cocok untuk dikenakan pada tes prestasi, termasuk soal fisika yang membutuhkan tahap identifikasi permasalahan hingga solusi akhir.

PCM merupakan pengembangan dari Model IRT 1 *parameter of logistic (1-PL)* dan termasuk keluarga Model Rasch. Model dikotomis sederhana dalam Model Rasch merupakan kasus khusus dari PCM. Model dikotomis dan PCM dapat dikatakan campuran dalam satu analisis (Wu & Adams, 2007). PCM merupakan pengembangan dari Model Rasch *item* dikotomis yang diterapkan pada *item* politomis. Model Rasch *item* dikotomis yang hanya berisi satu parameter lokasi *item* (tingkat kesulitan) kemudian dikembangkan dengan menjabarkan lokasi



butir menjadi beberapa kategori. Asumsi pada PCM yakni setiap butir mempunyai daya beda yang sama. PCM mempunyai kemiripan dengan *Graded Response Model* (GRM) pada *item* yang diskor dalam kategori berjenjang, namun indeks kesukaran dalam setiap langkah tidak perlu urut, suatu langkah dapat lebih sukar dibandingkan langkah berikutnya. Ini berarti, PCM merupakan pengembangan Model Rasch dikotomus menjadi politomus dengan satu parameter logistik yaitu tingkat kesulitan.

Skor kategori pada PCM menunjukkan banyaknya langkah untuk menyelesaikan dengan benar butir tersebut. Skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih besar daripada skor kategori yang lebih rendah. Pada PCM, jika suatu butir memiliki dua kategori, maka Persamaan probabilitas menjadi persamaan model Rasch, seperti persamaan yang dinyatakan oleh Hambleton and Swaminathan (1985). Oleh karena itu, PCM dapat diterapkan pada butir politomus dan dikotomus.

Hasil survei pendahuluan menunjukkan bahwa penskoran hasil tes pilihan ganda menggunakan model dikotomus, artinya jika *item* benar diberi skor 1 dan jika salah diberi skor 0. Penskoran belum menggunakan model politomus yang lebih adil karena mempertimbangkan langkah-langkah penyelesaian tes. Model penskoran dikotomus ini belum menghargai tahap-tahap penyelesaian soal, karena dengan tingkat kesalahan yang berbeda mendapatkan skor yang sama yakni 0. Dengan demikian skoring model ini tentu kurang adil.

Tujuan tes untuk mengetahui kemampuan peserta didik dan posisinya dalam kelompok tersebut. Jika tes yang ditempuh peserta didik tidak semuanya sama, tes terdiri atas dua atau lebih perangkat, maka hasil tes tidak dapat dibandingkan antarpeserta didik. Agar hasil tes dapat dibandingkan untuk perangkat tes yang terdiri atas lebih dari satu perangkat, maka perlu ada *anchor item* untuk penyetaraan (*equating*) hasil tes.

Hambleton & Swaminathan (1991, pp.123-143) menyatakan penyetaraan skor tes atau *equating* adalah tindakan mengkon-

versi skor tes yang satu (skor tes X) menjadi skor yang metrik (yang sesuai dengan ukuran) dari tes yang lain (skor tes Y). Menurut Kolen & Brannen (1995, p.2) penyetaraan (*equating*) skor tes adalah proses statistik yang digunakan untuk melakukan penyesuaian skor antara suatu tes dengan tes yang lain. Skor tes yang sudah disesuaikan melalui *equating* bersifat *interchangeable* sehingga dapat diambil keputusan yang adil yang didasarkan pada paket tes yang berbeda. Dengan demikian, berarti agar hasil tes testi yang mengerjakan perangkat tes yang berbeda dari dua atau lebih perangkat tes dapat dibandingkan, maka harus dilakukan penyetaraan tes.

Berdasarkan uraian tersebut, untuk mengukur kemampuan berpikir tingkat tinggi fisika digunakan tes berbentuk pilihan ganda beralasan yang dinamakan Tes kemampuan berpikir tingkat tinggi fisika (*Physics Test for Higher Order Thinking Skills* (*PhysTHOTS*)). Untuk itu perlu disusun instrumen penilaian kemampuan berpikir tingkat tinggi Fisika (*PhysTHOTS*) yang terdiri atas tes dan pedoman penilaian. Dengan demikian, diperlukan adanya penelitian pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi fisika yang terdiri atas tes berpikir tingkat tinggi Fisika (*Physics Test for Higher Order Thinking Skills* (*PhysTHOTS*)) dan pedoman penilaian.

Sejalan dengan permasalahan tersebut, diperlakukan penelitian untuk (1) menghasilkan instrumen untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran fisika, dan (2) mendapatkan karakteristik instrumen penilaian kemampuan berpikir tingkat tinggi dalam mata pelajaran fisika di SMA kelas XI yang meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan

## Metode Penelitian

Penelitian ini termasuk jenis penelitian pengembangan dengan pendekatan kuantitatif. Penelitian pengembangan instrumen ini menggunakan model modifikasi Model Wilson dan Model Oriondo dan Antonio.

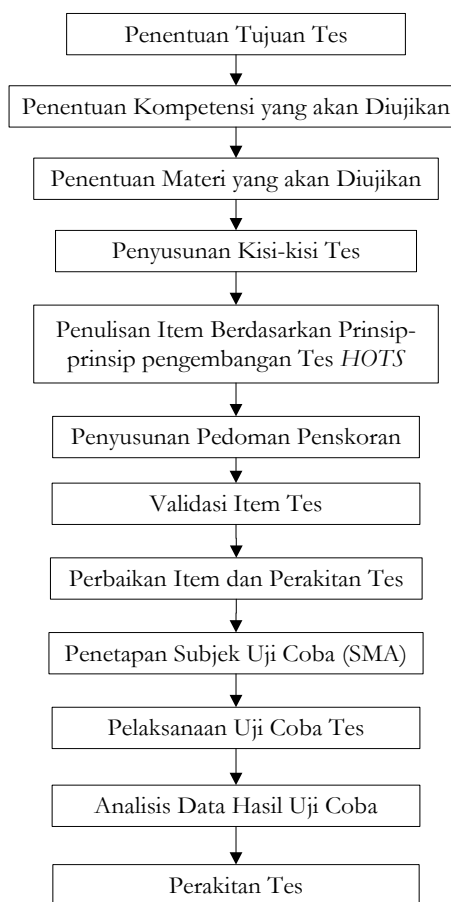
Penelitian dimulai Desember 2012 sampai dengan Mei 2013. Pengembangan awal tes yang berupa penyiapan tes, validasi, dan perakitan tes dilakukan pada bulan Desember 2012 sampai dengan Februari 2013. Uji coba dilangsungkan pada bulan Maret sampai dengan April 2013. Penelitian dilakukan di SMA di wilayah Daerah Istimewa Yogyakarta.

Subjek penelitian ini seluruh peserta didik kelas XI dari sepuluh SMAN di Daerah Istimewa Yogyakarta yang terdiri atas dua SMA dari setiap kota/kabupaten. Jumlah subjek penelitian sebanyak 1001 peserta didik.

Langkah-langkah pengembangan instrumen berupa tes menggunakan modifikasi Model Wilson dan Model Oriondo dan Antonio, yakni: (1) perancangan tes, (2) uji coba tes, dan (3) perakitan tes.

Tahap perancangan tes meliputi: (1) penentuan tujuan tes, (2) penentuan kompetensi yang diujikan, (3) penentuan materi yang diujikan, (4) penyusunan kisi-kisi tes, (5) penulisan *item* berdasarkan prinsip-prinsip pengembangan Tes *HOT*, (6) validasi *item* tes, (7) perbaikan *item* dan perakitan tes, dan (8) penyusunan pedoman penskoran. Adapun tahap uji coba tes meliputi: (1) penetapan subjek uji coba (SMA), (2) pelaksanaan uji coba, dan (3) analisis data hasil uji coba. Tahap terakhir dalam pengembangan tes ini adalah perakitan tes. Tahap-tahap pengembangan tes tersebut disajikan pada Gambar 1.

Berkaitan ukuran sampel, untuk analisis secara *IRT* beberapa ahli pengukuran sebaiknya 200 sampai dengan 1000 orang (Seon, 2009, p.3). Untuk analisis dengan Rasch, sampel yang digunakan antara 30 sampai 300 orang (Bond and Fox, 2007, p. 43; Keeves & Masters, 1999, pp.12-13). Reckase (2000) menyimpulkan bahwa sampel ukuran minimum yang baik untuk memperkirakan tiga parameter yang meliputi: diskriminasi (daya beda), tingkat kesulitan, dan *pseudoguessing* adalah 300 (Haladyna, 2004, p.206). Jadi, dengan model PCM 1PL peserta didik yang dijadikan subjek coba sebanyak 500 orang sudah lebih dari cukup.



Gambar 1. Langkah-langkah Pengembangan Instrumen

Dalam kegiatan uji coba ini terlebih dahulu ditentukan SMA di Provinsi DIY yang digunakan untuk uji coba berdasar ranking sekolah berdasar nilai UN Fisika Tahun 2012. Setiap kota/kabupaten dipilih dua SMA yang berada pada dua peringkat UN rendah sedang, atau tinggi. SMA yang digunakan untuk uji coba, antara lain: SMA 5 Yogyakarta, SMA 11 Yogyakarta, SMA 1 Bambanglipuro, SMA 1 Sedayu, SMA 1 Wates, SMA 1 Pengasih, SMA 1 Gamping, SMA 1 Minggir, SMA 1 Wonosari, dan SMA 1 Patuk.

Analisis data penelitian ini menggunakan *Partial Credit Model* 1 PL (PCM 1-PL) untuk pengujian *fit item* tes kemampuan berpikir tingkat tinggi untuk mata pelajaran Fisika SMA. Dasar pertimbangan yang digunakan, yang pertama bahwa PCM sebagai perluasan Rasch Model yang merupakan model 1-PL, dapat menggunakan sampel yang tidak sebesar kalau melakukan

kalibrasi data politomus menggunakan model 2-PL atau 3-PL (Keeves & Masters, 1999, pp.12-13). Kedua, bahwa karakteristik respons terhadap *item* kemampuan berpikir tingkat tinggi mengikuti PCM.

Analisis data dilakukan pada beberapa aspek, yakni: (1) kecocokan *item* instrumen, (2) reliabilitas, (3) kurva karakteristik *item* (ICC), (4) indeks kesukaran, dan (5) fungsi informasi dan SEM.

Pengujian *goodness of fit* untuk tes secara keseluruhan dan testi (*case/person*) secara keseluruhan dikembangkan Adam dan Khoo (1996, p.30) berdasarkan nilai rerata INFIT *Mean of Square (Mean INFITMNSQ)* beserta simpangan bakunya atau mengamati nilai rata INFIT *t (Mean INFIT t)* beserta simpangan bakunya. Jika rerata INFIT MNSQ sekitar 1,0 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0,0 dan simpangan bakunya 1,0, maka keseluruhan tes *fit* dengan model.

Kecocokan butir dengan model diketahui dengan *fit item* dan testi mengikuti kaidah bahwa *Item characteristic curve* (ICC) akan mendatar (*flat*) bila besarnya INFIT MNSQ untuk *item* atau *e* lebih besar dari satuan logit >1,30 atau <0,77. Keadaan ini grafik distribusi membentuk *platokurtic curve*

dan tidak lagi membentuk *leptokurtic curve* (Keeves & Alagumalai 1999, p.36). Oleh karena itu, suatu *item* atau testi/*case/person* dinyatakan *fit* dengan model dengan batas kisaran INFIT MNSQ dari 0,77 sampai 1,30 (Adam & Khoo, 1996, pp.30&90). Dalam hal ini menggunakan kisaran nilai *t* adalah  $\pm 2$  (pembulatan  $\pm 1,96$ ) jika taraf kesalahan atau *alpha* sebesar 5% (Keeves & Alagumalai 1999, pp.34-36; Bond & Fox, 2007, p.43).

*Item* dikatakan baik jika indeks kesukaran lebih dari -2.0 atau kurang dari 2.0 (Hambleton & Swaminathan, 1985, p.36). Berdasarkan fungsi informasi dan SEM, maka dapat diketahui bahwa tes ini cocok untuk siswa dengan kemampuan ( $\theta$ ) rendah, sedang, atau tinggi.

## Hasil Penelitian dan Pembahasan

### Hasil Pengembangan Tes

Instrumen PhystHOTS terdiri atas dua perangkat (set) yang masing-masing meliputi materi: gerak, gaya, usaha dan energi, serta momentum dan impuls dari aspek menganalisis, mengevaluasi, dan menciptakan. Selanjutnya, dengan *expert judgment* PhystHOTS dinyatakan layak digunakan.

Tabel 1. Sebaran Iten Tes Kemampuan Berpikir Tingkat Tinggi Fisika Kelas XI SMA

Aspek	Subaspek	Materi Fisika Kelas XI SMA			
		Gerak	Gaya	Usaha dan Energi	Momentum dan Impuls
Meng-analisis	Membedakan	2 (1A, 9B)	2 (4A, 12B)		1 (21A, 3B)*
	Mengurutkan	1 (19A, 1B)*	2 (5A, 13B)	2 (7A, 15B)	2 (9A, 17B)
	Memberikan ciri khusus	2 (2A, 10B)	2 (6A, 14B)	1 (20A, 2B)*	
Meng-evaluasi	Mengecek		2 (3A, 11B)	2 (8A, 16B)	1 (26A, 8B *)
	Mengkritik	2 (11A, 19B)		1 (25A, 7B)*	2 (10A, 18B)
Men-ciptakan	Memunculkan Ide	1 (22A, 4B)*	4 (13A, 14A, 12B, 22B)	2 (15A, 23B)	
	Merencanakan	2 (12A, 20B)	1 (23A, 5B)*		2 (18A, 26B)
	Menghasilkan		1 (24A, 6B)*	2 (16A, 24B)	2 (17A, 25B)

Keterangan: \*) *anchor item*

PhystHOTS terdiri atas dua set yakni set I berkode A dan set II berkode B. Setiap tes meliputi materi: gerak analisis vektor, gaya dan getaran, usaha energi, momentum dan implus yang meliputi aspek dan sub

aspek menganalisis, mengevaluasi, dan menciptakan dengan sebaran sebagaimana dinyatakan Tabel 1. Kedua perangkat tes tersebut memiliki delapan *item* sebagai *anchor item*.

Hasil Uji Coba

*Kecocokan Item Instrumen (goodness fit)*

Pengujian *goodness of fit* dilakukan untuk tes secara keseluruhan ataupun tiap *item*. Pengujian *fit* tes secara keseluruhan menggunakan kaidah yang dikembangkan oleh Adam dan Khoo (1996, p.30) yakni berdasarkan nilai rerata INFIT *Mean of Square (Mean INFITMNSQ)* beserta simpangan bakunya atau mengamati nilai rata INFIT *t (Mean INFIT t)* beserta simpangan bakunya. Jika rerata INFIT MNSQ sekitar 1,0 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0,0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model PCM 1 PL. Berdsarkan Tabel 2, nilai rerata *INFITMNSQ* 1,01 (sekitar 1) dan simpangan baku 0,02 (sekitar 0,0), maka keseluruhan tes *fit* dengan model PCM 1 PL

Tabel 2. Hasil Estimasi *Item* dan Testi HOTS Fisika menurut PCM 1-PL

No	Uraian	Estimasi untuk <i>item</i>	Estimasi untuk testi
1	Nilai rata-rata dan simpangan baku	$-0,29 \pm 0,51$	$0,19 \pm 0,01$
2	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,50$	$-0,11 \pm 0,00$
3	Reliabilitas	0,95	
4	Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,01 \pm 0,02$	$1,01 \pm 0,07$
5	Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$1,01 \pm 0,02$	$1,01 \pm 0,12$
6	Nilai rata-rata dan simpangan baku INFIT <i>t</i>	$0,84 \pm 1,47$	$0,25 \pm 1,08$
7	Nilai rata-rata dan simpangan baku OUTFIT <i>t</i>	$0,27 \pm 0,48$	$0,08 \pm 0,61$

Pengujian penetapan *fit* setiap *item* pada model mengikuti kaidah Adam dan Khoo (1996, p.30), yakni suatu *item fit* pada model jika nilai INFIT MNSQ antara 0,77 sampai 1,30. Dengan batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30) dan menggunakan INFIT *t* dengan

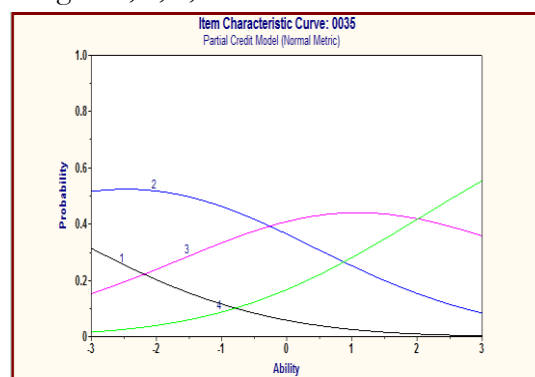
batas -2,0 sampai 2,0, maka diperoleh *item-item* yang cocok memenuhi *goodness of fit*. Nilai INFIT MNSQ PhysTHOTS antara 0,98 sampai dengan 1,05. Dengan batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30), maka semua *item* sebanyak 44 *item* fit dengan PCM semua.

*Reliabilitas*

Berdasarkan hasil analisis diperoleh reliabilitas set instrumen (PhysTHOTS) tersebut sebesar 0,95. Nilai reliabilitas ini tergolong kategori tinggi.

*Kurva Karakteristik Item (Item Characteristic Curva, ICC)*

Karakteristik *item* ditunjukkan dengan kurva karakteritik *item* (ICC) dan indeks kesukaran. Berdasarkan analisis diperoleh kurva karakteristik *item* (ICC) sebanyak 44 buah. Pada Gambar 2 disajikan contoh ICC untuk *item* 35 (*item* 17 B), yang dapat dijelaskan bahwa: (a) skor 1 (kategori 1) sebagian besar diperoleh siswa dengan *ability* sangat rendah ( $\theta = -3$ ), (b) skor 2 (kategori 2) sebagian besar diperoleh siswa dengan *ability* rendah ( $\theta = -2$ ), (c) skor 3 (kategori 3) sebagian besar diperoleh siswa dengan abilitas tinggi ( $\theta = 1,5$ ), dan (d) skor 4 (kategori 4) sebagian besar diperoleh siswa dengan *ability* sangat tinggi ( $\theta = 3$ ). Tingkat kesulitan dari kecil ke besar berurutan kategori 1, 2, 3, dan 4.



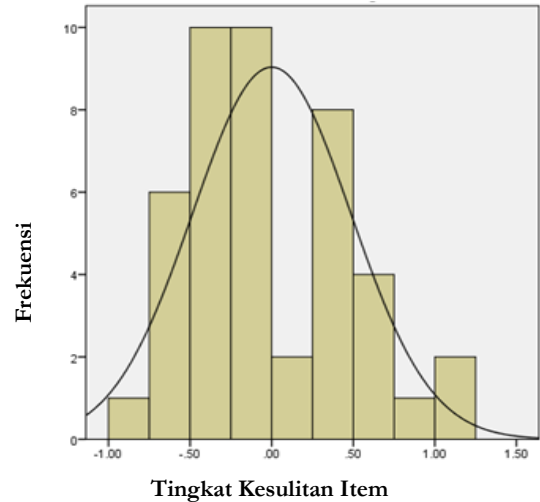
Gambar 2. Kurva Karakteristik Butir 35 (Butir 17B)

*Indeks Kesukaran (b)*

Indeks kesukaran atau tingkat kesukaran (*b*) untuk skor 0 ( $b_0$ ), skor 1 ( $b_1$ ), skor 2 ( $b_2$ ), skor 3 ( $b_3$ ), skor 4 ( $b_4$ ), dan reratanya

sebagai *difficulty*. Berdasarkan hasil analisis, *difficulty item-item* terletak antara -0,86 sampai dengan 1,06 dengan rata-rata 0 dan simpangan baku 0,42. *Item* dikatakan baik jika indeks kesukaran lebih dari -2,0 atau kurang dari 2,0 ( $-2,0 < b < 2,0$ ) (Hambleton & Swaminathan, 1985, p.36). Jadi, berdasarkan *difficulty*, semua *item* sebanyak 44 semuanya baik.

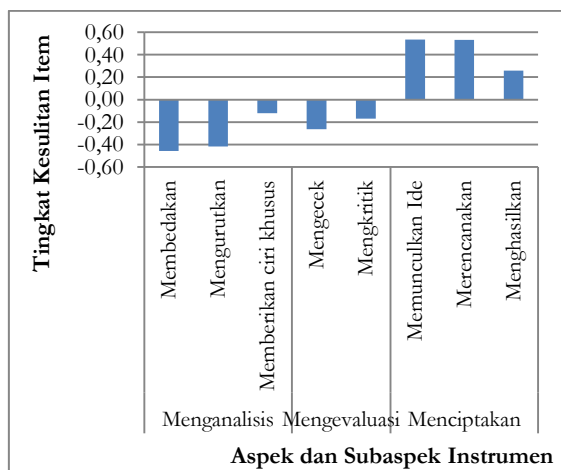
Adapun histogram distribusi indeks kesulitan dapat dilihat pada Gambar 3. Berdasarkan Gambar 3 distribusi indeks kesulitan mengikuti distribusi mendekati normal. Hal ini tidak menjadikan masalah, karena tidak ada persyaratan bahwa distribusi indeks kesulitan *item* harus normal.



Gambar 3. Distribusi Tingkat Kesulitan *Item* Tes

Tabel 3. Tingkat Kesulitan Butir setiap Subaspek

No	Aspek	Subaspek	Difficulty	Kesulitan Tahap			
				Kategori 1	Kategori 2	Kategori 3	Kategori 4
1	Meng-analisis	Membedakan	-0,46	-2,96	0,45	0,92	-0,16
		Mengurutkan	-0,42	-3,09	0,32	1,38	-0,21
		Memberikan ciri khusus	-0,12	-2,69	0,95	1,41	0,00
2	Meng-evaluasi	Mengecek	-0,26	-2,00	1,25	0,37	-0,31
		Meng kritik	-0,17	-2,36	1,68	0,19	-0,14
3	Men-ciptakan	Memunculkan Ide	0,54	0,99	0,32	0,72	-0,23
		Merencana kan	0,53	1,71	-0,28	0,72	-0,16
		Menghasil kan	0,26	-1,10	1,84	0,31	-0,22



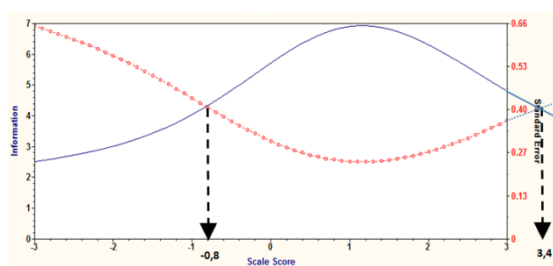
Gambar 4. Tingkat Kesulitan *Item* masing-masing Aspek dan Subaspek Instrumen

Tabel 3 menjelaskan besar tingkat kesulitan pada setiap subaspek dan aspek instrumen untuk masing-masing kategori da-

lam PCM. Berdasarkan Tabel 3 juga dapat diketahui tingkat kesulitan pada setiap subaspek dan aspek instrumen untuk setiap kategori dalam PCM. Untuk lebih jelasnya, dapat dilihat diagram distribusi tingkat kesulitan *item* menurut aspek dan subaspek instrumen yang disajikan pada Gambar 4.

Berdasarkan Gambar 4, dapat dilihat bahwa urutan tingkat kesulitan *item* setiap aspek pada tahap ujicoba berturut-turut adalah menganalisis, mengevaluasi, dan menciptakan. Hal ini berarti bahwa aspek kemampuan secara gradasi dari rendah ke tinggi secara berturut-turut adalah aspek kemampuan menganalisis, mengevaluasi, dan menciptakan. Kondisi ini sudah sesuai dengan pendapat Anderson & Krathwohl (2001, p.30).

### Fungsi Informasi dan SEM



Gambar 5. Fungsi Informasi dan SEM

Hasil analisis diperoleh fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan Gambar 5, maka tes ini cocok untuk peserta didik dengan kemampuan berpikir tingkat tinggi Fisika ( $\theta$ ) dalam kategori tinggi, yakni  $-0,8 \leq \theta \leq 3,4$ .

#### Revisi Instrumen

Berdasarkan hasil uji coba, semua *item* adalah *fit*. Dua *anchor item* yang tingkat kesulitannya lebih dari satu, yakni *item* ke-22 (1,03) dan ke-23 (1,06) ditukar dengan *item* ke-13 (0,31) dan ke-12 (0,38). Penukaran *item* didasarkan pada kesesuaian sub aspek dan sub materi fisika. Instrumen *PhysTHOTS* terdiri atas dua set tes A dan B yang masing berisi 26 *item* dengan 8 *anchor item*.

#### Pembahasan

Reliabilitas *PhysTHOTS* cukup tinggi yakni 0,95 yang menegaskan bahwa hasil pengukuran dengan instrumen ini reliabel. Lebih dari itu, tes yang memiliki koefisien reliabilitas sekurang-kurangnya 0,90 hasil testing dengan tes tersebut dapat digunakan untuk membuat keputusan tentang individu (Suryabrata, 2002, pp.39-40). Selain itu, fungsi informasi relatif tinggi untuk kemampuan antara -0,8 sampai 3,4. Ini berarti bahwa instrumen ini memiliki kekuatan dan reliabel yang tinggi karena tersusun dari *item-item* yang memiliki fungsi informasi yang tinggi (Hambleton dan Swaminathan, 1985, p.94). Hal ini dapat terjadi karena tes ini sesuai dengan kemampuan peserta didik yang dites. Berdasarkan koefisien reliabilitas, fungsi informasi tes, dan parameter estimasi, berarti *PhysTHOTS* ini reliabel dan memiliki kestabilan yang tinggi.

Validitas isi tes telah dibuktikan dengan *expert judgment*. Validitas konstruk secara empiris dibuktikan dengan *goodness of fit* pada *partial credit model* (PCM). Berdasarkan Tabel 2 nilai rata-rata dan simpangan baku INFIT MNSQ yakni masing-masing 1,01 (sekitar 1) dan 0,02 (sekitar 0), maka tes *fit* dengan PCM 1 PL. Hal ini berarti bahwa secara empiris *PhysTHOTS* tersebut valid. Tes yang berisi pertanyaan berpikir tingkat tinggi yang valid mendorong peserta didik untuk berpikir secara mendalam tentang materi pelajaran (Barnett & Francis, 2012, p.209). Validitas tes tersebut ini didukung oleh semua *item* memiliki nilai INFIT MNSQ antara 0,98 sampai dengan 1,05 yang terletak antara batas penerimaan *item* menggunakan INFIT MNSQ atau *fit* menurut model (antara 0,77 sampai dengan 1,30) berarti semua *item* sebanyak 44 *item fit* semua. Hal ini disebabkan karena beberapa hal, antara lain: (1) *item-item* dikembangkan sesuai prosedur pengembangan *item* instrumen yang benar, (2) *item-item* dikembangkan dari indikator yang diturunkan dari aspek kemampuan berpikir tingkat tinggi dan materi fisika, (3) tes yang terdiri atas 44 *item* telah melalui uji validitas isi dengan *expert judgment*, dan (4) responden (peserta didik) yang diuji sungguh-sungguh dalam mengerjakan karena melibatkan pengawas dari guru fisika mereka. Dengan demikian, karena *PhysTHOTS* yang dikembangkan valid akan mengukur kemampuan berpikir tingkat tinggi fisika dengan hasil yang valid pula serta mendorong peserta didik untuk berpikir secara mendalam tentang materi pelajaran fisika.

Menurut Hambleton & Swaminathan (1985, p.36), tingkat kesulitan *b* untuk *item* yang baik bervariasi antara -2,00 sampai dengan 2,00. *Item* dengan tingkat kesulitan -2,00 menandakan *item* tersebut sangat mudah, sedangkan tingkat kesulitan 2,00 berarti *item* tersebut sangat sulit. Dengan demikian, ditinjau dari tingkat kesulitan *item* dan kestabilannya maka instrumen ini termasuk kategori baik.

Berdasarkan fungsi informasi dan SEM, *PhysTHOTS* akan tepat digunakan

pada peserta didik yang memiliki kemampuan dari -0,8 sampai dengan 3,4 (termasuk dalam level tinggi). Dengan demikian, instrumen ini diyakini dapat digunakan untuk menggali informasi mengenai kemampuan berpikir tingkat tinggi fisika dengan tepat pada peserta didik SMA yang memiliki abilitas tinggi. Hal ini karena peserta didik yang berabilitas tinggi memiliki pemahaman konsep yang mendalam sehingga lebih mungkin dapat menerapkan pengetahuan itu untuk memecahkan masalah baru. (Ramos, Dolipas, & Villamor, 2013, p.57).

## Simpulan dan Saran

### Simpulan

Berdasarkan uraian tersebut, simpulan yang dapat ditarik adalah sebagai berikut: (1) instrumen *PhysTHOTS* dikembangkan dalam bentuk pilihan ganda ber-alasan pada kemampuan menganalisis, mengevaluasi, dan menciptakan untuk materi fisika gerak, gaya, usaha dan energi, serta momentum dan impuls yang terdiri atas perangkat tes A dan perangkat tes B yang masing-masing terdapat 26 *item* dengan 8 *anchor item*; (2) instrumen *PhysTHOTS* telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris validitas konstruk *fit* pada *Partial Credit Model (PCM)* berdasarkan data politomus empat kategori; (3) seluruh *item* pada *PhysTHOTS* dalam kriteria baik karena tingkat kesulitannya berada pada rentang antara -2,00 sampai dengan 2,00.

Reliabilitas *PhysTHOTS* telah memenuhi syarat, bahkan termasuk tinggi (koefisien reliabilitas lebih dari 0,90).

Berdasarkan fungsi informasi, *PhysTHOTS* sangat tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik yang berkemampuan dari -0,80 sampai 3,40.

### Saran

Berdasarkan hasil analisis, disarankan: (1) agar para guru menerapkan tes kemampuan berpikir tingkat tinggi fisika di SMA; (2) diadakan pelatihan penyusunan tes kemampuan berpikir tingkat tinggi fisika bagi guru; (3) dilakukan penelitian

lebih lanjut dengan menggunakan analisis data politomus menurut *generalized partial credit model (GPCM 3PL)*.

## 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.
- Albert, L.R, Corea, D & Macadino, V. (2012). *Rhetorical Ways of Thinking Vygotskian Theory and Mathematical Learning*. New York: Springer
- Anderson, L.W., and Krathwohl, D.R. (2001). *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Barnett, J. E and Francis, A.L. (2012). *Using higher order thinking questions to foster critical thinking: a classroom study*. Educational Psychology: An International Journal of Experimental Educational Psychology. <http://www.tandfonline.com/loi/cedp20>. Diakses tanggal 10 Desember 2012
- Bloom, B.S., et al. (1979). *Taxonomy of Educational Objectives: Handbook I Cognitive Domain*. London: Longmans Group Ltd.
- Bond, TG and Fox, CM. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences (2<sup>rd</sup> edition)*. Mahwah, NJ: Lawrence Erlbaum. Bonk.
- Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom*. Alexandria: ASCD
- BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP-Depdiknas
- Depdiknas. (2007). *Peraturan Menteri Pendidikan Nasional No 20 Tahun 2007 tentang Standar Penilaian*
- Dikpora DIY. (2012). *Hasil Nilai UN 2012 SMA/MA/SMK di DIY (27 Mei 2012)*. Yogyakarta: Dikpora DIY [http://www.pendidikan-diy.go.id/?view=v\\_berita&id\\_sub=2692](http://www.pendidikan-diy.go.id/?view=v_berita&id_sub=2692) diakses tanggal 1 Juni 2012.

- Efendi, Ridwan. (2010). *Kemampuan Fisika Siswa Indonesia dalam TIMSS*. Prosiding Seminar Nasional Fisika 2010 ISBN: 978-979-98010-6-7
- Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton & Swaminathan. (1985). *Item Response Theory Principles and Applications*. Boston: Kluwer Nijhoff Publishing.
- Hambleton dan Swaminathan (1991): *Fundamentals of Item Response Theory*. California: SAGE Publications, Inc
- Keeves, J. P. & Alagumalai. (1999). New Approach to measurement. Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in Measurement in Educational Research and Assessment* (pp.23-42). Amsterdam: Pergamon, An imprint of Elsevier Science
- Keeves, J.P. & Masters, G.N. (1999). Introduction. Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in Measurement in Educational Research and Assessment* (pp.1-22). Amsterdam: Pergamon, An imprint of Elsevier Science
- Kolen, M.J. & Brannen, R.L. (1995). *Test equating: Methods and practices*. New York: Sprinegr-Velag New York, Inc.
- Nitko, A.J & Brookhart, S. M. (2011). *Educational assessment of students*. (6<sup>th</sup> ed). Boston: Pearson Education, Inc.
- Oriondo, L.L.and Dallo-Antonio, E.M. (1998). *Evaluation Educational Outcomes*. Manila: Rex Printing Compagny, inc
- Piaget, J. (2005). *The psychology of intellegence* [Versi elektronik]. Taylor & Francis e-Library.
- Pusat Kurikulum. (2007). *Naskah Akademik Kajian Kebijakan Kurikulum Mata Pelajaran IPA*. Jakarta: Balitbang Depdiknas
- Ramos, J. L. S., Dolipas, B. B., & Villamor B. B. (2013). Higher Order Thinking Skills and Academic Performance in Physics of College Students: A Regression Analysis. *International Journal of Innovative Interdisciplinary Research*, ISSN 1839-9053l , p: 48-60. Diambil tanggal 20 Desember 2013 dari <http://www.auamii.com/jiir/Vol-01/issue-04/5ramos.pdf>
- Reedal, K.E. (2010). *Jean Piaget's Cognitive Development Theory in Mathematics Education*. Department of Mathematics and Computer Science – Ripon College. Summation, May 2010, pp. 16-20 <http://ripon.edu/mac/s/summation>.
- Schraw, G, and Robinson, D.H. (2011). *Assessment of Higher Order Thinking Skills*. New York: Information Age Publishing, Inc
- Seon, Hi Sin. (2009). How to tread omitted respons in Rasch model based equating [Versi elektronik] *Practical Assessment, Research & Evaluation*. ISSN 1531-7714, Volume 14, Number 1, p: 1-8
- Sudjana, Nana. (1990). *Penilaian Hasil Blajar Mengajar*. Bandung: PT Remaja Rosdakarya
- Suryabrata, S. (2002). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset.
- TIMSS & PIRLS International Study Center. (2012). *TIMSS 2011 international results in science*. Boston: The TIMSS & PIRLS International Study Center, Boston College. Diambil tanggal 5 Januari 2013, dari <http://timss.bc.edu/timss2011/release.html>
- Van den Berg, G. 2008. The use of assessment in the development of higher-order thinking skills. *Africa Education Review*, 1:2, 279-294. Diambil Tanggal 24 Oktober 2012 dari <http://dx.doi.org/10.1080/18146620408566285>.
- Van der Linden, W. J & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag New York, Inc
- Widhiarso, Wahyu. (2010). *Model politomi dalam teori respons butir*. Yogyakarta: Psikologi UGM
- Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.