

LAPORAN
PENELITIAN DISERTASI DOKTOR



PENGEMBANGAN INSTRUMEN UNTUK MENGIKUR
KEMAMPUAN BERPIKIR TINGKAT TINGGI
DALAM MATA PELAJARAN FISIKA DI SMA

Drs. EDI ISTIYONO, M.Si

NIDN: 0007036802

UNIVERSITAS NEGERI YOGYAKARTA

NOVEMBER 2013

Dibiayai oleh DIPA Universitas Negeri Yogyakarta
dengan Surat Perjanjian Penugasan dalam rangka Pelaksanaan Program
Penelitian Disertasi Doktor Tahun Anggaran 2013
Nomor: 532a/BOPTN/UN34.21/2013 Tanggal 27 Mei 2013

HALAMAN PENGESAHAN

Judul Kegiatan

: Pengembangan Instrumen untuk Mengukur Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Fisika di SMA

Peneliti / Pelaksana

Nama Lengkap : Drs. EDI ISTIYONO M.si.
NIDN : 0007036802
Jabatan Fungsional :
Program Studi : Fisika
Nomor HP : 08121593340
Surel (e-mail) : edi_istiyyono_uny@yahoo.co.id

Institusi Mitra (jika ada)

Nama Institusi Mitra :
Alamat :
Penanggung Jawab :
Tahun Pelaksanaan : Tahun ke 1 dari rencana 1 tahun
Biaya Tahun Berjalan : Rp. 30.000.000,00
Biaya Keseluruhan : Rp. 0,00

Mengetahui
Dekan FMIPA UNY

Yogyakarta, 23 - 11 - 2013,
Ketua Peneliti,



(Dr. Hartono)

(Drs. EDI ISTIYONO M.si.)
NIP/NIK196803071993031001

Menyetujui,
Ketua LPPM UNY



(Prof. Dr. Amik Ghulfron)
NIP/NIK 196211111988031001

PENGEMBANGAN INSTRUMEN UNTUK MENGIKUR KEMAMPUAN BERPIKIR TINGKAT TINGGI DALAM MATA PELAJARAN FISIKA DI SMA

Oleh:

Edi Istiyono

Jurdik Fisika FMIPA Universitas Negeri Yogyakarta

RINGKASAN

Tujuan umum penelitian ini untuk mengadakan perbaikan pembelajaran berkaitan dengan pada model tes dan sistem penskorannya. Tujuan khusus penelitian untuk menghasilkan instrumen penilaian kemampuan berpikir tingkat tinggi (*Higher Order Thinking, HOT*) Fisika pada kelas XI SMA yang memenuhi syarat, meliputi: 1) fit dengan PCM dan 1) tingkat kesulitan item

Responden siswa SMA kelas XI di Daerah Istimewa Yogyakarta terdiri 10 SMA Negeri di DIY yang tersebut di Kota Yoyakarta, Kabupaten Bantul, Kabupaten Kulon Progo, Kabupaten Sleman, dan Kabupaten Gunung Kidul yang masing-masing tiap kabupaten/ kota sebanyak dua SMAN. Langkah-langkah penelitian meliputi; 1) perancangan instrumen, 2) uji coba instrumen, dan 3) perakitan tes. Tahap perancangan instrumen meliputi: a) Penentuan Tujuan Tes, b) Penentuan Kompetensi yang akan diujikan, c) Penentuan Materi yang Diujikan, d) Penyusunan Kisi-kisi Tes, e) Penulisan Item Berdasarkan Prinsip-prinsip Pengembangan Tes *HOT*, f) Validasi Item Tes, g) Perbaikan Item dan Perakitan Tes, dan h) Penyusunan Pedoman Penskoran. Tahap uji coba instrumen terdiri: a) Penetapan Subjek Uji Coba (SMA), b) Pelaksanaan Uji coba Tes, dan c) Analisis Data Hasil Uji Coba dengan teknik: (1) Kecocokan Item Instrumen (*goodness fit*), (2) Reliabilitas, (3) Kurva Karakteristik Item, (4) Indeks Kesukaran, dan (5) Fungsi Informasi dan SEM. Tahap Perakitan Tes untuk Pengukuran.

Hasil penelitian menunjukkan bahwa pada tahap uji coba dan tahap pengukuran berdasarkan kriteria *mean INFIT MNSQ* 1,0 dan simpangan baku 0,0 terbukti tes *fit* dengan PCM. Berdasarkan kriteria batas terendah dan tertinggi INFIT MNSQ 0,77 dan 1,30, semua item sebanyak 44 dari dua set tes semuanya *fit* dengan model. Tingkat kesulitan item antara -0,86 sampai 1,06. Reliabilitas tes tahap uji coba adalah 0,95. Dengan demikian instrumen *PhysTHOT* memenuhi syarat digunakan untuk mengukur kemampuan berpikir tingkat tinggi Fisika Siswa SMA Kelas XI.

Kata kunci: pengembangan instrumen, penilaian, kemampuan berpikir tingkat tinggi, Fisika, SMA

DEVELOPMENT OF PHYSICS TEST FOR HIGHER ORDER THINKING SKILLS (PhysTHOTS) IN SENIOR HIGH SCHOOL

by :
Edi Istiyono

Summary

This research aims to : (1) develop a set of instrument for the measurement of physics higher order thinking skills (PhysTHOTS) for senior high school students, and (2) obtain the characteristics of the PhysHOTS.

This research was conducted through two stages, i.e. the initial development and the instrument tryout. The initial instrument development included writing, reviewing, and validating the learning continuum, blue print of instruments, items, and assessment guidelines. They were validated by the promoters, measurement experts, physics specialists, physics education experts, and practitioners. The instruments consisted of two sets of validated tests tried out to 1001 students. The instruments consisted of two sets of tests each of which consisted of 26 items, including 8 anchor items. The tryout was conducted on 1001 students in 10 senior high schools in Yogyakarta Special Territory (YST). The results of the tryout, four categories of polytomous data, were analyzed according to the Partial Credit Model (PCM) using the QUEST program and Parscale. The QUEST program was used to measure the suitability of the model and the level of difficulty, and Parscale to obtain item characteristic curves, information function, and SEM.

The results of the research show that the tryout based on criteria INFIT MNSQ with mean 1.0 and standard deviation 0.0 was fit to the PCM for four categories of polytomous data. Based on the criteria of the highest and lowest limits INFIT MNSQ 0.77 and 1.30, all 44 items of the two sets of tests were fit to the model. The items' difficulty indexes were between -0.86 and 1.06, which means the items were good because their values were between -2.0 and 2.0. The reliability of the test during the tryout was relatively high, i.e. 0.95. Therefore, PhysTHOTS instruments are qualified for the measurement of higher order thinking skills of physics of Grade XI of senior high school students.

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

PRAKATA

Puji syukur alhamdulillaah kita panjatkan ke hadirat Allah S.W.T yang telah melimpahkan rahmat, taufik, hidayah dan inayahNya sehingga terselesaikannya laporan penelitian yang berjudul ‘Pengembangan Instrumen untuk Mengukur Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Fisika di SMA

Penyelesaian laporan penelitian ini banyak mendapat bantuan dari berbagai pihak, karena itu pada kesempatan baik ini saya mengucapkan terima kasih kepada:

1. Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Negeri Yogyakarta yang telah memberikan jalan pelaksanaan penelitian ini.
2. Direktur Program Pascasarjana beserta staf administrasi, atas segala kebijaksanaan, perhatian, dan bantuan sehingga disertasi ini dapat selesai.
3. Prof. Djemari Mardapi, Ph.D. dan Bapak Suparno, Ph.D selaku promotor disertasi, yang telah memberikan arahan dan bimbingan yang sangat berharga, serta dorongan semangat kepada penulis sehingga disertasi ini dapat terwujud.
4. Para ahli pengukuran, ahli Fisika, ahli Pendidikan Fisika, dan praktisi sebagai reviewer dan validator instrumen yang telah memberikan masukan dalam rangka validasi untuk meningkatkan kualitas instrumen.
5. Kepala sekolah dan Guru Fisika: SMAN 5 Yogyakarta, SMAN 11 Yogyakarta, SMAN 1 Bambanglipuro, SMAN 1 Sedayu, SMAN 1 Wates, SMAN 1 Pengasih, SMAN 1 Gamping, SMAN 1 Minggir, SMAN 1 Wonosari, dan SMAN 1 Patuk, yang telah memberikan izin dan bantuan pada pelaksanaan uji coba.

Semoga bantuan Bapak/Ibu sekalian mendapat pahala dari Allah S.W.T, aamiin. Selanjutnya penulis berharap mudah-mudahan laporan ini bermanfaat khususnya bagi penulis dan umumnya bagi pembaca yang tertarik pada bidang pengukuran dan pengujian.

Yogyakarta, 20 Novemberr 2013

Penulis,

Edi Istiyono

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN PENGESAHAN	ii
RINGKASAN	iii
PRAKATA	v
DAFTAR ISI	vi
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Rumusan Masalah	4
BAB II TINJAUAN PUSTAKA	
1. Hakikat Fisika	6
2. Berpikir Tingkat Tinggi (<i>Higher Order Thinking</i>)	7
3. Penilaian dalam Pembelajaran Fisika	
4. Kemampuan Berpikir Tingkat Tinggi Fisika	13
5. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika	13
BAB III TUJUAN DAN MAFAAT PENELITIAN	
A. Tujuan Penelitian	15
B. Kontribusi Produk yang dikembangkan	15
C. Manfaat Penelitian	15
BAB IV METODE PENELITIAN	
A. Model Pengembangan	16
B. Prosedur Pengembangan	16
BAB V HASIL PENELITIAN	
A. Hasil Pengembangan	21
B. Hasil Uji Coba Produk	24
1. Kegiatan Uji Coba	24
2. Hasil Kegiatan Uji Coba	26
C. Revisi Produk untuk Pengukuran	35
D. Pembahasan	37
BAB VI RENCANA TAHAPAN BERIKUTNYA	39
BAB VII SIMPULAN DAN SARAN	
A. Simpulan	40
B. Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN	45

DAFTAR TABEL

	Halaman
Tabel 1 Sebaran Item Instrumen Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Fisika di SMA Kelas XI	22
Tabel 2 Aspek, Subaspek, dan Indikator Instrumen	23
Tabel 3 SMA yang Digunakan untuk Ujicoba Instrumen	25
Tabel 4 Jumlah Responden dalam Kegiatan Ujicoba	26
Tabel 5 Hasil Estimasi Item dan Estimasi Testi Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajar Fisika SMA pada Kegiatan Uji coba	28
Tabel 6 Tingkat Kesulitan Butir Masing-masing Subaspek untuk Kategori 1, 2, 3, dan 4 pada Kegiatan Ujicoba	30
Tabel 7 Sepuluh Item Termudah pada Kegiatan Uji coba	31
Tabel 8 Sepuluh Item Tersulit pada Kegiatan Uji coba	32
Tabel 9 Persentase Menjawab pada Benar Aspek dan Subaspek Kemampuan Berpikir Tingkat Tinggi Fisika SMA di DIY pada kegiatan Uji coba	33
Tabel 10 Tingkat Kesulitan <i>Anchor Item</i>	35
Tabel 11 Tingkat Kesulitan Item Setiap Perangkat Tes A dan B	36
Tabel 12 Rangkuman Tingkat Kesulitan Item pada Tahap Ujicoba	37

DAFTAR GAMBAR

	Halaman
Gambar 1	11
Gambar 2	12
Gambar 3	20
Gambar 4	27
Gambar 5	29
Gambar 6	30
Gambar 7	34
Gambar 8	34

DAFTAR LAMPIRAN

	Halaman	
Lampiran 1	Standar Kompetensi, Kompetensi Dasar, Matriks HOTS-Fisika, dan Indikator Kemampuan Berpikir Tingkat Tinggi Fisika Kelas XI SMA.....	47
Lampiran 2	Distribusi Item, Tabel masukan dari Penelaah (Validator) Instrumen, dan Surat Keterangan Validitas Istrumen	52
Lampiran 3	Perangkat Tes pada Uji coba	54
Lampiran 4	Estimasi Parameter dan Nilai Mean Infit MNSQ Tes Uji Coba dari Hasil Analisis QUEST	97
Lampiran 5	Nilai Percent Correct, Point Biserial, dan Mean Ability Tiap Item Tes Uji Coba	103
Lampiran 6	Nilai Tingkat Kesulitan (<i>difficulty</i> dan <i>delta</i>) Tes Uji Coba dari Hasil Analisis QUEST	120
Lampiran 7	Peringkat UN Fisika Tahun 2012 di DIY	135
Lampiran 8	Surat Izin Penelitian	178
Lampiran 9	Surat Keterangan Telah Melaksanakan Penelitian	184

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Dewasa ini dunia berada pada era globalisasi. Pada era ini persaingan cukup ketat. Agar dapat menang dalam bersaing, maka negara-negara berlomba-lomba meningkatkan kualitasnya. Kualitas bangsa ditentukan dengan tingkat pendidikan bangsa tersebut. Untuk meningkatkan kualitas pendidikan dapat dimulai dari meningkatkan kualitas pembelajaran. Untuk meningkatkan 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: 160). 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 Indonesia berada pada ranking 40 dari 42 negara (TIMSS & PIRLS International Study Center, 2012:48), sedangkan untuk PISA yang dilaporkan oleh *the Organization for Economic Co-Operation and Development* (OECD) berada pada peringkat 50 dari 57 negara (OECD, 2007:58). Hal senada dinyatakan Ridwan Efendi (2011: 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 dan kemampuan siswa Indonesia rata-rata masih berada pada kemampuan *knowing*; (2) kecenderungan capaian fisika siswa Indonesia selalu menurun pada tiap aspek kognitif sehingga kemampuan fisika siswa Indonesia harus

dingkatkan pada semua aspek, khususnya aspek *reasoning* dengan cara membekalkan pada 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 dengan asesmen yang tepat dapat mendorong siswa untuk belajar dengan berpikir tingkat tinggi.

Berdasarkan teori perkembangan Piaget, tahap operasional formal adalah tahap anak mulai dari sebelas tahun. Pada tahap ini anak sudah mulai dapat mengembangkan kemampuan untuk memanipulasi konsep abstrak melalui penggunaan proposisi dan hipotesis (Reedal, 2010:7). Pada usia ini anak berada pada tingkat sekolah menengah pertama (SMP), sehingga kemampuan berpikir tingkat tinggi siswa SMA tentunya sudah lebih mapan.

Psikolog lain, Vygotsky mengatakan banyak pengaruh konstruksi sosial yang membentuk sisi-sisi kognitif manusia (fungsi psikologis kognisi tinggi). Kalau Piaget mengatakan harus menunggu kematangan dan kesiapan seseorang serta harus cocok antara pengaruh dari luar dan perkembangan di dalam dirinya (*match*), tetapi Vygotsky tidak. 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). Dengan menerapkan konsep ZPD pada pendidikan, maka pembelajaran akan memajukan perkembangan anak. Salah satu wujud konkrit implikasi dari teori Vygotsky adalah dilaksanakannya akselerasi belajar bagi anak berbakat, pendidikan bagi anak-anak yang mempunyai kemampuan intelektual luar biasa dan dalam proses pembelajaran harus selalu meningkatkan kadar mental atau berpikir tingkat tinggi.

Menurut Bloom yang telah direvisi 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 and Krathwohl, 2001:30). Taksonomi Bloom sudah lama diterapkan dalam bidang pendidikan dan sudah lama digunakan. Taksonomi Bloom masih digunakan dalam banyak kurikulum dan bahan pengajaran (Brookhart, 2010: 39). Dengan demikian kemampuan berpikir tingkat tinggi Fisika (*Physics Higher Order Thinking*) meliputi kemampuan fisika dalam menganalisis, mengevaluasi, dan berkreasi.

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 maupun 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.

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 (Nana Sujana, 1990:49). 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 (Nana Sujana, 1990:49).

Berdasarkan informasi dari guru-guru Fisika SMA di Daerah Istimewa Yogyakarta (DIY), 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 peserta didik SMA. Demikian juga untuk mata pelajaran Fisika, tes pilihan

ganda biasa masih menjadi alat andalan untuk mendapatkan hasil belajar peserta didik di 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*).

Berdasarkan wawancara dengan guru Fisika SMA di DIY bahwa penskoran hasil tes dengan model dikotomus, artinya jika item benar diberi skor 1 dan jika salah diberi skor 0. Belum menggunakan model politomus akan lebih adil karena mempertimbangkan langkah-langkah Dengan model skoring 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.

Berdasarkan uraian di atas,. untuk mengukur kemampuan berpikir tingkat tinggi Fisika digunakan tes berbentuk pilihan ganda yang dinamakan Tes Fisika berpikir tingkat tinggi (*Physics Test for Higher Order Thinking (PhysTHOT)*). Untuk itu perlu disusun instrumen penilaian kemampuan berpikir tingkat tinggi fisika (*PhysTHOT*) yang terdiri atas tes dan pedoman penilaian. Agar dihasilkan instrumen yang andal diterapkan prinsip-prinsip pengembangan instrumen. 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 (PhysTHOT)*) dan pedoman penilaian.

B. Rumusan Masalah

Agar penelitian ini lebih terfokus, maka perlu kiranya ditentukan rumusan masalah antara lain:

1. Bagaimanakah konstruksi instrumen *PhysTHOTS* yang dikembangkan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik kelas XI SMA?
2. Bagaimana karakteristik instrumen *PhysTHOT* yang dikembangkan memenuhi syarat digunakan untuk mengukur ditinjau dari: (a) kecocokan dengan PCM, (b) tingkat kesulitan item, dan (c) reliabilitas instrumen.

BAB II

TINJAUAN PUSTAKA

1. Hakekat Fisika

Berkaitan karakteristik Fisika, Sumaji (1990:28) menyatakan bahwa komponen proses pembelajaran Fisika ada 7, antara lain: (1) Fisika dapat melatih siswa berpikir logis, kritis dan memberikan dasar pemahaman terhadap alam sekitar, serta memberikan dasar kepada siswa4melanjutkan studi; (2) Fisika dapat memupuk sikap ilmiah pada diri siswa; dan (3) Fisika memupuk minat siswa terhadap sains dan teknologi. Dengan demikian pelajaran IPA, Fisika Fisika mengembangkan kemampuan berpikir logis, kritis, objektif, memutuskan sesuatu berdasarkan data yang tetap dengan menggunakan metode ilmiah, dan kemampuan untuk komunikasi ilmiah.

Chiappetta dan Koballa (2010:104-114) menyatakan bahwa “sains pada hakekatnya merupakan sebuah kumpulan pengetahuan (“*a body of knowledge*”), cara atau jalan berpikir (“*a way of thinking*”), dan cara untuk penyelidikan (“*a way of investigating*””). Dengan demikian dapat dikatakan sebaliknya bahwa, pernyataan Chiappetta dan Koballa di atas merupakan pandangan yang komprehensif atas hakekat IPA atau sains.

2. Berpikir Tingkat Tinggi (*Higher Order Thinking/HOT*)

Untuk mengetahui yang dimaksud berpikir tingkat tinggi (*higher order thinking*) dapat dilihat definisi menurut Brookhart (2010:5) sebagai berikut:

Higher-order thinking conceived of as the top end of the Bloom’s cognitive taxonomy. The teaching goal behind any of the cognitive taxonomies is equipping students to be able to do transfer. “Being able to think” means students can apply the knowledge and skills they developed during their learning to new contexts. “New” here means applications that the student has not thought of before, not necessarily something universally new. Higher-order thinking is conceived as students being able to relate their learning to other elements beyond those they were taught to associate with it.

Definisi tersebut menyiratkan beberapa hal, sebagai berikut: (1) Berpikir tingkat tinggi merupakan kemampuan pada ujung atas taksonomi kognitif Bloom, (2) Tujuan pengajaran berdasarkan taksonomi kognitif Bloom melengkapi siswa untuk dapat menerapkan pengetahuan dan keterampilan untuk konteks baru. Maksud "Baru" adalah aplikasi konsep yang oleh siswa belum terpikirkan sebelumnya, ini berarti belum tentu sesuatu yang universal baru. (3) Berpikir tingkat tinggi berarti kemampuan siswa untuk menghubungkan pembelajaran mereka untuk hal-hal lain di luar yang pernah diajarkan.

Untuk mengetahui lebih lanjut yang dimaksud dengan kemampuan "ujung atas" taksonomi Bloom dapat disimak pernyataan berikut (Brookhart, 2010:5)

Higher-order thinking is approached as the “top end” of Bloom’s (or any other) taxonomy: Analyze, Evaluate, and Create, or, in the older language, Analysis, Synthesis, and Evaluation

Berdasarkan definisi tersebut berarti, berpikir tingkat tinggi merupakan kemampuan berpikir menurut taksonomi Bloom, yang meliputi: menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*).

3. Penilaian dalam Pembelajaran Fisika

a. Pengertian penilaian dan jenisnya

Penilaian hasil belajar peserta didik pada jenjang pendidikan dasar dan menengah dilaksanakan berdasarkan standar penilaian pendidikan yang berlaku secara nasional. Instrumen penilaian hasil belajar yang digunakan pendidik memenuhi persyaratan (a) substansi, adalah merepresentasikan kompetensi yang dinilai, (b) konstruksi, adalah memenuhi persyaratan teknis sesuai dengan bentuk instrumen yang digunakan, dan (c) bahasa, adalah menggunakan bahasa yang baik dan benar serta komunikatif sesuai dengan taraf perkembangan peserta didik.

b. Karakteristik Tes

Tes dikatakan bermutu, jika jika memenuhi syarat antara lain: valid dan reliabel. Salah satu karakteristik tes adalah validitas tes. Linn (1999: 47) menyatakan bahwa tes yang baik harus memenuhi tiga karakteristik, yaitu: validitas, reliabilitas, dan usabilitas. Adapun validitas dalam model Rasch adalah sesuai atau fit dengan model (Hambleton dan Swaminathan, 1985: 73). Dengan

demikian, validitas untuk instrumen yang analisisnya dengan model Rasch 1 PL dapat dilihat dari kecocokan atau fit terhadap model.

Menurut Hambleton dan Swaminathan (1991: 236), penggunaan fungsi informasi tes lebih akurat bila dibandingkan dengan penggunaan reliabilitas karena: (1) bentuknya tergantung hanya pada butir-butir dalam tes, (2) mempunyai estimasi kesalahan pengukuran pada setiap level abilitas. Jadi, untuk estimasi reliabilitas berdasarkan analisis butir soal digantikan dengan menggunakan fungsi informasi karena fungsi informasi jauh lebih akurat dari pada reliabilitas.

c. Tes Pilihan Ganda Beralasan

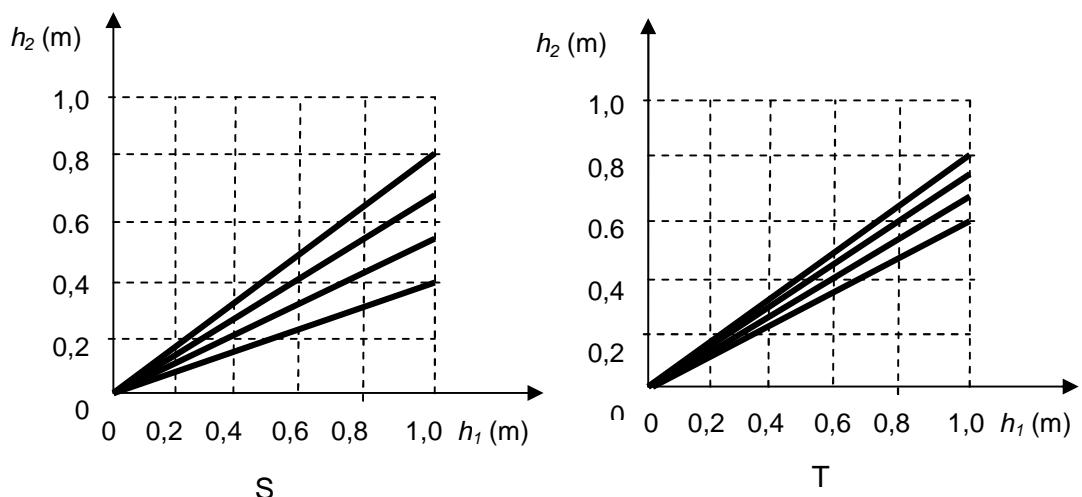
Berbagai alat penilaian tertulis, antara lain: tes memilih jawaban *benar-salah, isian singkat, menjodohkan dan sebab akibat* merupakan alat yang hanya menilai *kemampuan berpikir rendah*, yaitu kemampuan mengingat (pengetahuan). Selain bentuk tes-tes tersebut, tes pilihan ganda merupakan bentuk tes yang sering digunakan. 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 (Nana Sujana, 1990:49). Dengan demikian, maka tes pilihan ganda banyak digunakan untuk tes-tes dalam skala besar, misalnya ujian nasional dan ujian masuk perguruan tinggi. Lebih dari itu, tes pilihan ganda dapat digunakan untuk mengukur berpikir kognitif tingkat tinggi (*Higher Order Thinking (HOT)*) yang lebih praktis dan objektif.

Tes pilihan ganda terdiri atas: pernyataan (pokok soal) atau stem dan alternatif jawaban dan pengcoho. Pokok soal merupakan kalimat yang berisi keterangan atau pemberitahuan tentang suatu materi yang belum lengkap yang harus dilengkapi dengan memilih jawaban yang tersedia. Kunci jawaban adalah salah satu alternatif jawaban yang merupakan pilihan benar, sedangkan pengecoh merupakan alternatif jawaban yang bukan kunci jawaban (Djemari Mardapi, 2004: 74-75).

Pedoman lain dalam pembuatan item soal bentuk pilihan ganda adalah Menurut Haladyna (2004:99), yakni: a) untuk stem: (1) Membuat arah sejelas mungkin, (2) Buatlah pernyataan sesingkat mungkin, (3) Tempatkan gagasan utama dari item dalam stem, bukan pada pilihan dan b) untuk pilhan: (1) Kembangkan pilihan sebanyak mungkin, tapi dua atau tiga mungkin cukup, (2) Variasikan lokasi jawaban yang benar sesuai dengan jumlah pilihan dan menetapkan posisi jawaban yang benar secara acak, (3) Tempatkan pilihan dalam urutan logis atau numerik, (4) Usahakan pilihan independen dan tidak boleh tumpang tindih, (5) Hindari pilihan homogen dalam konten dan struktur gramatikal, (6). Usahakan panjang pilihan hampir sama, (7) Hindari kata-kata negatif seperti tidak atau kecuali, (8) Hindari pilihan yang memberikan petunjuk untuk jawaban yang benar, (9) Membuat distraktor masuk akal, dan (10). Gunakan kesalahan khas siswa untuk menulis distraktor. Dengan pedoman ini, lebih rinci dalam penyusunan pokok soal dan pilihan. Untuk menyusun pokok soal perlu memperhatikan 3 prinsip, sedangkan untuk penyusunan pilihan memperhatikan 10 prinsip tersebut.

Tes pilihan ganda beralasan terdiri dari soal dan alasan yang masing-masing disediakan pilihan jawaban. Berikut adalah contoh item tes pilihan ganda beralasan yang termasuk dalam kemampuan mengevaluasi.

Berdasarkan uji elastisitas bola tenis merk S dan T diperoleh grafik seperti gambar berikut, dengan h_1 merupakan tinggi mula-mula dan h_2 adalah tinggi pantulan bola.



Berdasarkan gambar tersebut, dapat disimpulkan

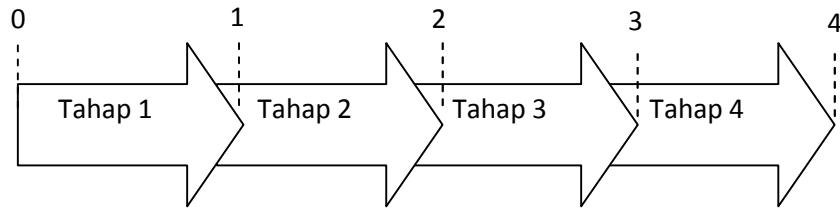
- A. Merk S lebih baik daripada T
- B. Merk T lebih baik daripada S
- C. Tidak dapat disimpulkan mana yang lebih baik antara S dan T
- D. Merk S dan T sama baiknya
- E. Merk T tidak sebaik Merk S

Alasan:

- A. Gradien grafik h_2 terhadap h_1 merupakan koefisien restitusi, maka untuk bola yang baik mestinya garis berimpit
- B. Grafik h_2 terhadap h_1 menggambarkan koefisien restitusi, maka untuk bola yang baik mestinya nilai tetap
- C. Gradien grafik h_2 terhadap h_1 merupakan kuadrat koefisien restitusi, maka semakin baik bola tersebut garisnya semakin mendekati berimpit
- D. Grafik h_1 terhadap h_2 menggambarkan koefisien restitusi, maka untuk bola yang baik mestinya nilainya konstan
- E. Grafik h_2 terhadap h_1 merupakan koefisien restitusi, maka untuk bola yang baik garisnya semakin mendekati berimpit

d. Penskoran dengan *Partial Credit Model* (PCM)

Penilaian ujian didasarkan pada tahap-tahap yang dapat diselesaikan peserta ujian. Prosedur penilaian tersebut sebenarnya sama dengan bagaimana individu merespon butir dalam skala psikologi. Menyelesaikan soal cuma sampai tahap pertama analog dengan kategori ‘*tidak pernah*’ sedangkan kalau sudah sampai tahap akhir, analog dengan kategori ‘*sering*’ yang dinyatakan Gambar 1. Asumsi ini kemudian dikembangkan menjadi Model PCM. Ketika kita mengasumsikan bahwa sebuah butir mengikuti pola kredit parsial maka kemampuan individu lebih tinggi diharapkan memiliki skor yang lebih tinggi daripada individu yang memiliki kemampuan rendah (Wahyu Widhiarsa, 2010: 6). Menurut Masters & Wright, PCM juga sesuai untuk menganalisis respon pada pengukuran berpikir kritis dan pemahaman konseptual dalam sains (van der Linden dan Hambleton, 1997: 101-102).



Gambar 1. Memaknai skala sebagai tahapan penyelesaian butir

PCM dikembangkan untuk menganalisis butir tes yang memerlukan beberapa langkah penyelesaian. Dengan demikian Model PCM cocok untuk dikenakan pada tes prestasi, termasuk soal fisika yang membutuhkan tahap identifikasi permasalahan hingga solusi akhir. PCM merupakan pengembangan dari Model 1-PL dan termasuk keluarga Model Rasch. Model dikotomi dan PCM dapat dikatakan campuran dalam satu analisis (Wu & Adams, 2007). PCM merupakan pengembangan dari Model Rasch butir dikotomi yang diterapkan pada butir politomi. Asumsi PCM yakni setiap butir mempunyai daya yang sama.

Bentuk umum PCM menurut Muraki & Bock (1997:16) dan Ostini & Nering (2006: 28) sebagai berikut.

$$P_{ig}(\theta) = \frac{\exp [\sum_{g=0}^l (\theta - b_{ig})]}{\sum_{h=0}^m \exp [\sum_{g=0}^h (\theta - b_{ig})]} \quad g=1, 2, 3, \dots, m+1 \quad (1)$$

Keterangan :

$P_{ig}(\theta)$: probabilitas peserta berkemampuan θ memperoleh skor kategori g pada item ke-*i*

θ : kemampuan peserta,

$m+1$: banyaknya kategori butir *i*,

b_{ig} : indeks kesukaran kategori *h* butir *i*

Persamaan di atas dapat dijabarkan berdasarkan jumlah kategori di dalam butir. Misalnya sebuah skala memiliki 4 kategori yakni kategodengan skor 1, 2, 3,

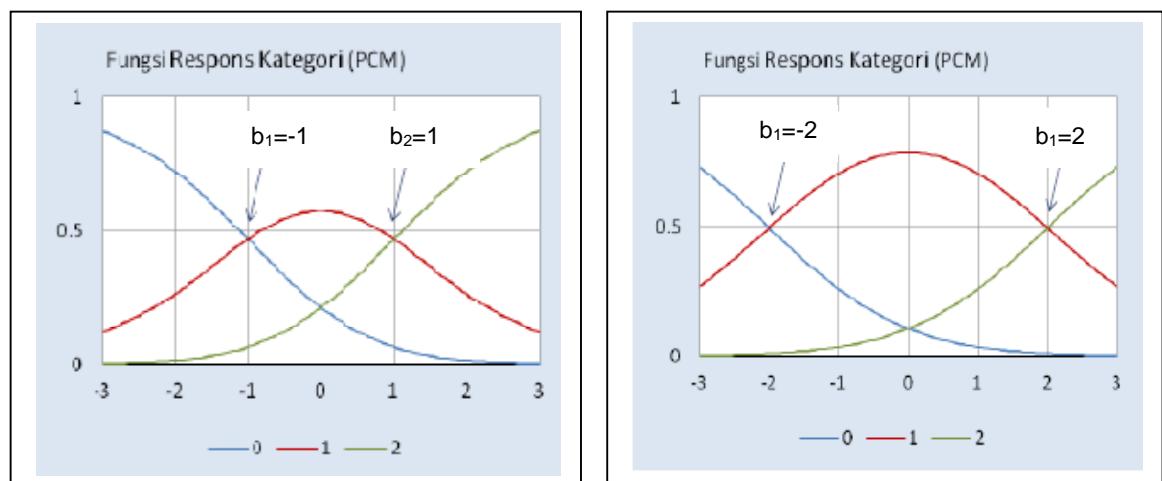
dan 4. Kita dapatkan kategori (g) sebanyak 4 buah persamaan yang probabilitas individu pada tiap kategori. PCM mensyaratkan seperti pada persamaan 1 berikut.

$$\sum_{g=0}^0 (\theta - b_{ig}) \equiv 0 \quad \sum_{g=0}^k (\theta - b_{ig}) \equiv \sum_{g=1}^k (\theta - b_{ig}) \quad (2)$$

Skor kategori PCM menunjukkan banyaknya langkah untuk menyelesaikan dengan benar butir. Skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih besar daripada skor kategori yang lebih rendah.

Partial Credit Model (PCM) merupakan pengembangan dari Model Rasch butir dikotomi yang diterapkan pada butir politomi. Jika i adalah butir politomi dengan kategori skor, $0, 1, 2, \dots, m_i$, maka probabilitas dari individu n skor x pada butir i yang nantinya digambarkan dalam *category response function* (CRF) diwujudkan dalam Persamaan 1 (Ostini and Nering, 2006: 28)

Persamaan di atas dapat dijabarkan berdasarkan jumlah kategori di dalam butir. Misalnya sebuah skala memiliki 5 kategori dengan skor 0, 1, 2, 3, dan 4. Maka kita dapatkan kategori (g) sebanyak 5 buah persamaan yang probabilitas individu pada tiap kategori.



Gambar 2. Contoh OCF dan CRF pada butir dengan tiga kategori

Parameter b_{ig} juga diinterpretasikan sebagai titik pada skala sifat laten dengan dua kategori yang berturutan kurva respons berpotongan sehingga

dinamakan persimpangan kategori (*category response curves intersect*). Parameter b_{ig} merupakan titik dimana dua kategori memiliki probabilitas yang sama untuk dipilih oleh level trait yang terkait (Linacre, 2006). Di sisi lain b_{ig} tidak menunjukkan tingkat kesukaran untuk sukses di tahap kedua atau untuk mencapai skor 2, tetapi lebih menunjukkan tingkat kesulitan butir untuk tahap kedua yang independen dengan tahap-tahap sebelumnya (Wu & Adams, 2007).

4. Kemampuan Berpikir Tingkat Tinggi Fisika (*Physics Higher Order Thinking (PhysHOT)*)

Kemampuan berpikir tingkat tinggi Fisika atau *Physics higher order thinking (PhysHOT)* adalah kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*) pada bidang Fisika. Kemampuan ini ditunjukkan dalam menyelesaikan persoalan Fisika dengan menganalisis, mengevaluasi, dan menciptakan. Kemampuan ini sebenarnya sudah dibiasakan dalam Fisika, karena Fisika sudah melatih mengembangkan kemampuan berpikir logis, kritis, objektif, memutuskan sesuatu berdasarkan data yang tetap dengan menggunakan metode ilmiah, dan kemampuan untuk komunikasi ilmiah.

5. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (*Physics Test of Higher Order Thinking, PhysETHOT*)

Berdasarkan uraian di atas, prinsip penyusunan Tes Kemampuan Fisika Berpikir Tingkat Tinggi (*PhysTHOT*) antara lain: (1) menentukan dengan jelas dan tepat apa yang ingin dinilai; (2) Memilih materi sesuai kurikulum berdasarkan kelasnya (Standar kompetensi dan kompetensi dasar); (3) menggunakan kata-kata yang sesuai dengan tingkat taksonomi Bloom (mengevaluasi, menganalisis, menciptakan); (4) menggunakan bahasa Indonesia yang baku; (5) menggunakan kata-kata yang hanya memiliki arti tunggal, supaya tidak membingungkan; (6) membawa siswa untuk memecahkan suatu kasus, pendek; (7) menggunakan bahan-bahan yang baru (belum pernah dibahas di kelas); (8) memberikan petunjuk cara mengerjakan pada awal soal; (9) membuat kunci jawaban; dan (10) membuat pedoman penskoran.

Pengembangan tes model kedua menurut Oriondo dan Antonio (1998: 34) terdiri lima tahapan, yakni: (1) perencanaan tes (*planning test*), (2) uji coba tes (*trying out the test*), (3) menetapkan validitas tes (*establishing test validity*), (4) menetapkan reliabilitas tes (*establishing test reliability*), dan (5) menafsirkan nilai tes (*interpreting the test scores*). Berdasarkan model ini, maka pengembangan tes dimulai dari persiapan yang meliputi: penentuan tujuan tes, spesifikasi tes, pemilihan format butir, membuat butir soal, dan mengedit butir soal. Setelah butir-butir soal siap, maka dilakukan uji coba untuk mendapatkan karakteristik tes yang meliputi. Setelah dianalisis hasil uji coba, maka dapat diperoleh validitas dan reabilitas tes, Akhirnya bila telah melakukan pengukuran, maka dapat dillakukan penafsiran skor tes.

BAB III

TUJUAN DAN MANFAAT PENELITIAN

A. Tujuan Penelitian

Tujuan umum penelitian ini untuk mengadakan perbaikan pembelajaran berkaitan pada model tes dan sistem penskorannya. Tujuan khusus penelitian: 1) untuk menghasilkan instrumen penilaian kemampuan berpikir tingkat tinggi fisika (*PhysTHOTS*) peserta didik kelas XI SMA yang memenuhi syarat digunakan untuk mengukur dan 2) menentukan karakteristik *PhysTHOTS* yang meliputi: (1) fit dengan PCM, (2) tingkat kesulitan item, dan (3) reliabilitas.

B. Kontribusi Produk yang Dikembangkan

Produk utama yang akan dikembangkan adalah instrumen penilaian. Instrumen ini berupa Tes Fisika Berpikir Tingkat Tinggi atau *Physics Test for Higher Order Thinking (PhysTHOT)*. Tes ini dilengkapi dengan pedoman penilaian. Tes yang dikembangkan ini selanjutnya digunakan untuk melakukan pengukuran kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika pada siswa kelas XI SMA sebagaimana tujuan disertasi yang dilakukan. Penilaian model tersebut akan memberikan kontribusi dalam dunia pendidikan dalam peningkatan penilaian secara adil pada siswa.

C. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian pengembangan ini adalah:

1. Dihasilkan instrumen yang dapat digunakan untuk mengukur kemampuan berpikir tingkat tinggi Fisika (*PhysTHOTS*) pada kelas XI SMA.
2. Instrumen *PhysTHOTS* yang dihasilkan dapat menjadi acuan dalam menerapkan model penskoran politomus pada tes pilihan ganda beralasan.

BAB IV

METODE PENELITIAN

A. Model Pengembangan

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 untuk pengukuran. Langkah-langkah penelitian disertasi secara utuh dan penelitian ini disajikan dalam Gambar 3 berikut.

B. Prosedur Pengembangan Instrumen

Prosedur pengembangan instrumen yang berupa tes terdiri atas: (1) perancangan tes, (2) uji coba tes, dan (3) pengukuran. Langkah-langkah prosedur pengembangan instrumen dapat dijelaskan sebagai berikut.

1. Perancangan Tes

a. Penentuan Tujuan Tes

Tujuan tes untuk mengetahui kemampuan berpikir tingkat tinggi siswa pada mata pelajaran Fisika di kelas XI SMA.

b. Penentuan Kompetensi yang akan Diujikan

Setelah jelas tujuan tes, maka selanjutnya dipilih koperasi yang akan diujikan. Kompetensi ini sesuai dengan standar kompetensi (SK) dan kompetensi dasar (KD) untuk Fisika SMA kelas XI. Berdasarkan standar kompetensi dan kompetensi dasar ditentukan indikator yang bersesuaian.

c. Penentuan Materi yang Diujikan

Berdasarkan standar kompetensi, kompetensi dasar, dan indikator tersebut selanjutnya dideskripsikan materi Fisika kelas XI yang sesuai. Materi Fisika mencakup materi minimal dan materi pengayaan.

d. Penyusunan Kisi-kisi Tes

Untuk dapat membuat item soal yang baik diperlukan kisi-kisi tes. Pada langkah ini perlu diperhatikan penyebaran soal berdasarkan SK, KD, materi, dan kemampuan HOT (menganalisis, mengevaluasi, dan menciptakan).

e. Penulisan Item Berdasarkan Prinsip-prinsip Pengembangan Tes HOT

Sebagaimana telah disebutkan di atas kisi-kisi tes sangat penting perannya dalam pengembangan tes. Berdasarkan kisi-kisi tes, maka baru dapat dibuat item-item soal.

f. Validasi Item Tes

Setelah terbentuk draf item-item soal, langkah yang harus dilakukan adalah validasi terhadap item-item soal tersebut. Pada tahapan dilakukan validasi dengan *expert judgment*.

g. Perbaikan Item dan Perakitan Tes

Atas dasar hasil validasi dan masukan perbaikan dari ahli dan praktisi, dilakukan perbaikan item-item tes. Item-item tes yang telah diperbaiki kemudian dirakit menjadi tes. Hasil perakitan tes ini sudah siap untuk diujicobakan.

h. Penyusunan Pedoman Penskoran

Untuk dapat digunakan suatu tes tersebut perlu dilengkapi dengan pedoman penskoran. Hal ini dengan harapan ada kepastian skor yang diperoleh peserta tes.

2. Uji Coba Tes

Setelah tes berupa set soal terbentuk, maka untuk mengetahui karakteristik dan keterpakaianya perlu dilakukan tahap ke dua adalah uji coba dan revisi.

a. Penetapan Subjek Uji Coba (SMA)

Uji coba dilakukan pada siswa SMA negeri di Provinsi Daerah Istimewa Yogyakarta terdiri atas Kota Yogyakarta, Kabupaten Bantul, Kabupaten Kulon Progo, Kabupaten Sleman, dan Kabupaten Gunung Kidul. Uji coba dilaksanakan masing-masing kota atau kabupaten sebanyak 3 SMA negeri. Berkaitan jumlah sampel, untuk analisis secara IRT beberapa ahli pengukuran sebaiknya 200 sampai dengan 1000 orang (Seon, 2009: 3). Untuk analisis dengan Rasch, sampel yang digunakan antara 30 sampai 300 orang (Bond and Fox, 2007: 43). Tsutakawa dan Johnson (1990) merekomendasikan ukuran sampel sekitar 500 untuk estimasi parameter yang akurat. 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, 2004:206). Jadi dengan model PCM 1PL peserta didik yang dijadikan subjek coba sebanyak 500 orang sudah lebih dari cukup.

b. Pelaksanaan Uji coba Tes

Uji coba instrumen dilakukan dengan desain sebagai berikut: a) instrumen yang berupa: (1) tes pilihan ganda berpikir tingkat tinggi dalam mata pelajaran Fisika (*PhysTHOT*) dan (2) pedoman penilaian (penskoran), dikenakan pada subjek coba.

c. Analisis Data Hasil Uji Coba

Berdasarkan analisis hasil uji coba diketahui: (1) butir-butir soal yang tidak fit dan (2) koefisien reliabilitas. Jika instrumen sudah memenuhi, maka digunakan untuk menguji kemampuan Fisika berpikir tingkat tinggi (*PhysHOT*).

Teknik analisis data terdiri atas beberapa aspek, yakni:

1) Kecocokan Item Instrumen (*goodness of fit*)

Pengujian *goodness of fit* untuk tes secara keseluruhan maupun tiap item dengan program Quest. Pengujian Fit tes keseluruhan dikembangkan Adam dan Khoo (1996:30) 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 dan simpangan bakunya 0,0 atau rerata INFIT t mendekati 0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model.

Pengujian penetapan fit setiap item terhadap model mengikuti kaidah Adam dan Khoo (1996:30), suatu item fit terhadap model jika nilai INFIT MNSQ antara 0,77 smp 1,30. Dengan batas penerimaan item menggunakan INFIT MNSQ atau fit menurut model (antara 0,75 sampai dengan 1,30) dan menggunakan INFIT t dengan batas -2,0 sampai 2,0, maka diperoleh item-item yang cocok memenuhi goodness fit.

2) Reliabilitas

Selain untuk menguji kecocokan, output program Quest juga menampilkan estimasi reliabilitas set instrumen tes. Berdasarkan hasil analisis dengan program Quest juga didapatkan estimasi reliabilitas set instrumen (tes) tersebut.

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

Karakteristik item ditunjukkan dengan kurva karakteristik item (ICC) dan indeks kesukaran. Untuk mendapatkan kurva karakteristik item (ICC) digunakan program Multilog.

4) Indeks Kesukaran (b)

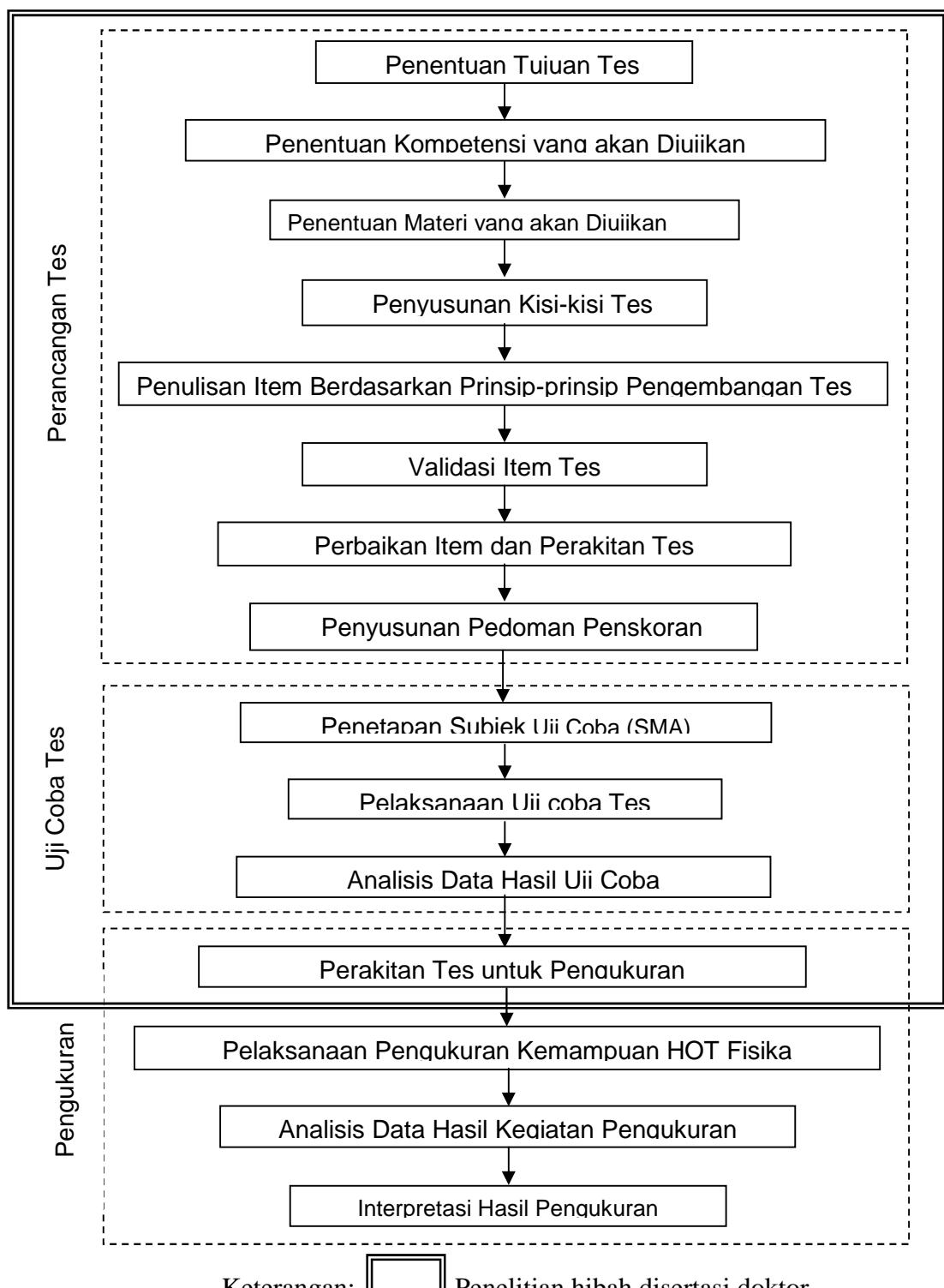
Dengan memanfaatkan analisis berdasarkan dengan Multilog diperoleh indeks kesukaran atau tingkat kesukaran (b) untuk masing-masing item. Item dikatakan baik jika indeks kesukaran lebih dari -2.0 atau kurang dari 2.0 yang dapat dinyatakan dengan $(-2.0 < b < 2.0)$.

5) Fungsi Informasi dan SEM

Berdasarkan analisis dengan Multilog diperoleh fungsi informasi dan standard error measurement (SEM). Berdasarkan fungsi informasi dan SEM, maka tes ini cocok untuk siswa dengan kemampuan (θ) rendah, sedang, atau tinggi.

3. Perakitan Tes Siap untuk Pengukuran

Berdasarkan hasil analisis data kegiatan uji coba dilakukan perbaikan item-item tes. Setelah dilakukan perbaikan item-item tes, maka selanjutnya dilakukan perakitan tes. Tes yang telah dirakit selanjutnya siap digunakan untuk pengukuran kemampuan berpikir tingkat tinggi Fisika.



Gambar 3. Langkah-langkah Pengembangan Instrumen

BAB V

HASIL PENELITIAN DAN PEMBAHASAN

A. Hasil Penelitian Pengembangan

Setelah diperoleh rumusan *learning continuum* kemampuan berpikir tingkat tinggi Fisika materi kelas XI SMA, kisi-kisi instrumen, item-item dan pedoman penilaian kemudian dilakukan ujicoba terbatas, ditelaah, dan divalidasi yang dilakukan oleh ahli pengukuran, ahli Fisika, dan ahli Pendidikan Fisika. Berdasarkan hasil telaah dilakukan revisi dan penyempurnaan item-item.

Ujicoba terbatas instrumen dilakukan di SMA Negeri 1 Kretek Bantul. Instrumen yang digunakan berupa tes kemampuan berpikir tingkat tinggi Fisika (*PhysTHOT*) pada Kelas XI SMA yang terdiri dari dua set tes yang masing-masing berisi 26 item dengan 8 item sebagai *anchor item*. Setiap item tes ini terdiri atas soal dengan 5 pilihan jawaban dan alasan juga disediakan 5 pilihan jawaban. Ujicoba terbatas ini dilakukan dengan tujuan untuk mengetahui keterbacaan item-item tes.

Kisi-kisi instrumen, item-item, dan pedoman penilaian yang telah disusun selanjutnya ditelaah, dan divalidasi yang dilakukan oleh promotor, ahli pengukuran, ahli fisika, ahli Pendidikan Fisika, dan praktisi. Masukan promotor berkenaan dengan kesesuaian dengan tujuan penelitian, sedangkan masukan para ahli berkenaan dengan kebenaran pengukuran, kebenaran materi fisika, dan visibilitas serta efektivitas pada pembelajaran fisika. Masukan dari forum ahli untuk menyempurnakan item, terutama dari segi panjang kalimat baik pada soal maupun dalam pedoman penskoran suatu item, penggantian beberapa istilah, dan panampilan grafik atau gambar agar lebih komunikatif.

Tabel 1
Sebaran Item Instrumen Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Fisika di SMA Kelas XI

Aspek	Sub aspek	Materi Fisika SMA Kelas XI			
		Gerak	Gaya	Usaha dan Energi	Tumbukan
Menganalisis	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) ^{*)}	
	Mengecek		2 (3A, 11B)	2 (8A, 16B)	1 (26A, 8B ^{*)}
	Mengkritik	2 (11A, 19B)		1 (25A, 7B) ^{*)}	2 (10A, 18B)
	Memunculkan Ide	1 (22A, 4B) ^{*)}	4 (13A, 14A, 12B, 22B)	2 (15A, 23B)	
Menciptakan	Merencanakan	2 (12A, 20B)	1 (23A, 5B) ^{*)}		2 (18A, 26B)
	Menghasilkan		1 (24A, 6B) ^{*)}	2 (16A, 24B)	2 (17A, 25B)

Keterangan:

^{*)} *anchor item* (item sama pada set A dan B)

Tes kemampuan berpikir tingkat tinggi Fisika (*PhysTHOT*) untuk Kelas XI SMA 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, implus dan momentum yang meliputi aspek dan sub aspek menganalisis, mengevaluasi, dan menciptakan dengan sebaran sebagaimana Tabel 3 serta indikator pada Tabel 4. Kedua perangkat tes tersebut memiliki delapan item sebagai *anchor item* yang digunakan untuk keperluan penyetaraan (*equating*). Selanjutnya tes dinyatakan layak digunakan oleh beberapa ahli tersebut.

Tabel 2.
Aspek, Subaspek, dan Indikator Instrumen

Aspek/sub aspek		Indikator	No Item
Menganalisis	Membedakan	Membedakan antara kecepatan dan kelajuan dalam gerak lurus beraturan	1A, 9B
		Membedakan gaya normal pada beberapa titik dalam lintasan lingkaran vertikal kaitannya dengan gaya sentripetal	4A, 12B
		Membedakan kecepatan benda sebelum dan sesudah tumbukan pada tumbukan tidak lenting	21A, 3B
	Mengurutkan	Mengurutkan besaran linear dan besaran sudut pada gerak melingkar beraturan	19A, 1B
		Mengurutkan nilai konstanta pegas pada berbagai susunan pegas	5A, 13 B
		Mengurutkan usaha yang dilakukan beberapa gaya yang membentuk berbagai sudut terhadap horizontal	7A, 15B
		Mengurutkan momentum atau kecepatan benda setelah tumbukan lenting sempurna	9A, 17B
	Memberikan ciri khusus	Menentukan sifat posisi atau kecepatan) peluru dalam gerak parabola	2A, 10B
		Menentukan sifat besaran (periode dan frekuensi) pada osilasi pegas	6A, 14B
		Menentukan usaha karena perubahan energi potensial benda	20A, 2B
Mengevaluasi	Mengecek	Mengecek kebenaran gaya gravitasi beberapa benda	3A, 11B
		Mengecek energi kinetik dan kecepatan benda pada berbagai posisi dengan hukum kekekalan energi	8A, 16B
		Mengecek kebenaran jenis tumbukan lenting sempurna berdasarkan kecepatan sebelum dan sesudah tumbukan	26A, 8B
	Mengkritik	Menentukan sudut antara bidang gerak terhadap bidang horizontal dalam melakukan gerak melingkar beraturan yang lebih mudah	11A, 19B
		Memilih kerja yang lebih dalam memindahkan benda secara vertikal dan miring	25A, 7B

		Menentukan yang lebih baik antara dua benda berdasarkan koefisien restitusinya	10A, 18B
Menciptakan	Memeunculkan Ide	Memunculkan hipotesis tentang percepatan gerak luurs berubah beraturan dari grafik S vs t^2	22A, 4B
		Memunculkan hipotesis bahwa adanya perubahan ukuran planet menyebabkan terjadi perubahan sifat gravitasinya	13A, 21B
		Memunculkan cara untuk menentukan usaha dengan perubahan energinya	15A, 23B
	Merencanakan	Merencanakan percobaan gerak parabola	12A, 20B
		Merencanakan kegiatan untuk mengukur gaya sentripetal pada benda yang berputar	23A, 5B
		Merencanakan kegiatan untuk mengukur kelajuan benda pada tumbukan tak lenting	18A, 26B
	Menghasilkan	Menghasilkan karya alat sederhana (neraca pegas) yang bekerja karena gaya pegas	24A, 6B
		Menghasilkan karya alat sederhana untuk mengukur kelajuan benda berdasarkan hukum kekekalan energi	16A, 24B
		Menghasilkan karya alat sederhana (restitusi meter) yang merupakan penerapan tumbukan lenting sebagian	17A, 25B

B. Hasil Uji Coba Instrumen

1. Kegiatan Uji Coba

a. SMA yang Digunakan untuk Kegiatan Ujicoba

Kegiatan ujicoba diawali dengan menentukan SMA di Provinsi DIY yang digunakan untuk uji coba. Penentuan SMA tersebut atas dasar rangking sekolah berdasar nilai ujian nasional (UN) Fisika. Sekolah yang digunakan untuk uji coba sebanyak 2 SMA negeri tiap kabupaten (Bantul, Kulonprogo, Sleman, dan Gunungkidul) atau kota Yogyakarta. Jadi, sekolah yang digunakan ujicoba sebanyak 10 SMA negeri yang tersebar pada peringkat UN Fisika rendah, sedang, dan tinggi sebagaimana dalam Lampiran 13 dinyatakan dalam Tabel 3.

Tabel 3.
SMA yang Digunakan untuk Ujicoba Instrumen

No	Nama SMA	Lokasi	Peringkat di Kota/Kab Berdasar Nilai UN Fisika
1	SMAN 5 Yogyakarta	Kotagede Yogyakarta	Tinggi (4 dari 11)
2	SMAN 11 Yogyakarta	JL AM Sangaji, No,50, Yogyakarta	Rendah (11 dari 11)
3	SMAN 1 Bambanglipuro	Mulyodadi, Bambanglipuro, Bantul	Tinggi (3 dari 19)
4	SMAN 1 Sedayu	Argomulyo, Sedayu, Bantul	Rendah (17 dari 19)
5	SMAN 1 Wates	Terbah, Wates, Kulon Progo	Tinggi (2 dari 11)
6	SMAN 1 Pengasih	Margosari, Pengasih, Kulon Progo	Sedang (5 dari 11)
7	SMAN 1 Gamping	Banyuraden, Gamping, Sleman	Sedang (7 dari 17)
8	SMAN 1 Minggir	Sendangmulyo, Minggir, Sleman	Rendah (16 dari 17)
9	SMAN 1 Wonosari	Wonosari, Gunung Kidul	Tinggi (1 dari 11)
10	SMAN 1 Patuk	Sambipitu, Patuk, Gunung Kidul	Sedang (4 dari 11)

b. Waktu Ujicoba

Ujicoba dilaksanakan bulan 2 Maret 2013 sampai dengan 3 April 2013.

Waktu yang disediakan untuk mengerjakan tes selama 90 menit.

c. Subjek Ujicoba

Jumlah responden dalam pelaksanaan ujicoba atau testi sebanyak 1001 siswa yang berasal dari 10 SMAN tersebut, Adapun rincian jumlah responden masing-masing SMA dinyatakan pada Tabel 4.

Tabel 4.
Jumlah Responden dalam Kegiatan Ujicoba

No	Nama SMA	Set Soal A	Set Soal B	Jumlah
1	SMAN 5 Yogyakarta	60	63	123
2	SMAN 11 Yogyakarta	81	90	171
3	SMAN 1 Bambanglipuro	39	41	80
4	SMAN 1 Sedayu	61	61	122
5	SMAN 1 Wates	56	55	111
6	SMAN 1 Pengasih	34	31	65
7	SMAN 1 Gamping	31	42	73
8	SMAN 1 Minggir	54	50	104
9	SMAN 1 Wonosari	48	48	96
10	SMAN 1 Patuk	29	27	56
Jumlah		493	508	1001

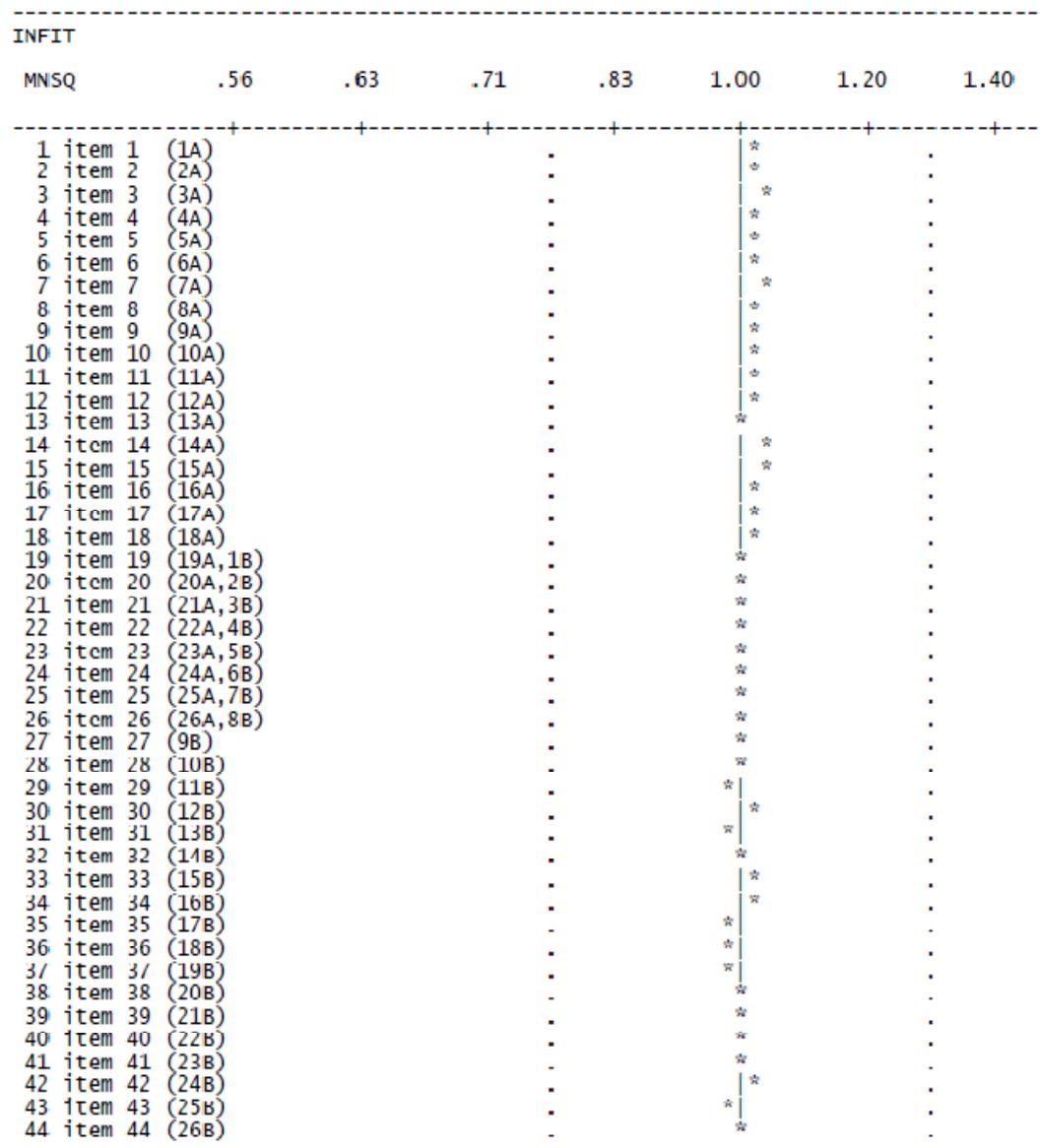
2. Hasil Kegiatan Uji Coba

a. Kecocokan Item Instrumen (*goodness fit*)

Pengujian *goodness of fit* untuk tes secara keseluruhan maupun tiap item dengan program QUEST, Pengujian Fit tes keseluruhan dikembangkan Adam dan Khoo (1996:30) 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 dan simpangan bakunya 0,0 atau rerata INFIT t mendekati 0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model PCM 1 PL, Berdasarkan Lampiran 4, nilai rerata *INFITMNSQ* 1,01 (sekitar 1) dan simpangan baku 0,02 (sekitar 0,0), maka keseluruhan tes fit dengan model PCM 1 PL.

Pengujian penetapan fit setiap item terhadap model mengikuti kaidah Adam dan Khoo (1996:30), yakni suatu item fit terhadap model jika nilai INFIT MNSQ antara 0,77 sampai dengan 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 antara 0,98 sampai dengan 1,05 sebagaimana dinyatakan pada Lampiran 5. 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* semua. Hal ini nampak lebih jelas dilihat pada Gambar 4 yang berasal dari Lampiran 4, bahwa semua item sebanyak 44 semua berada dalam dua garis batas *goodness of fit*, sehingga jelaslah bahwa 44 item semuanya *fit*.



Gambar 4.
Diagram Infit MNSQ Kemampuan Berpikir Tingkat Tinggi Fisika SMA di DIY
pada Kegiatan Ujicoba

b. Hasil estimasi

Berdasarkan analisis data hasil kegiatan ujicoba, didapatkan karakteristik instrumen tes kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika (*PhysTHOT*) untuk kelas XI SMA. Karakteristik tes antara lain: hasil estimasi, tingkat kesulitan item, estimasi reliabilitas, dan *tets optimal ability*.

Hasil estimasi 44 item yang dikerjakan oleh 1001 responden (testi) dengan level peluang 0,5 dan menggunakan penskoran politomus menurut PCM 1PL empat kategori dinyatakan pada Tabel 5, sedangkan perhitungan pada Lampiran 4.

Tabel 5.

Hasil Estimasi Item dan Estimasi Testi Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajar Fisika SMA pada Kegiatan Uji coba

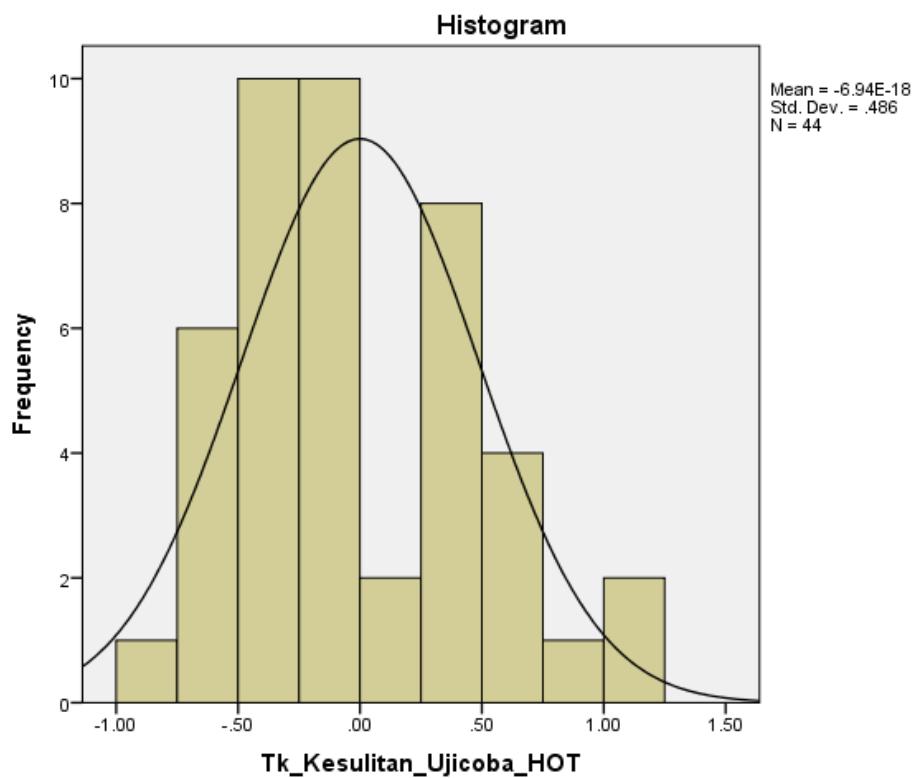
No	Uraian	Estimasi untuk item	Estimasi untuk testi
1	Nilai rata-rata dan simpangan baku	$-0,29 \pm 0,51$	$0,19 \pm 0,01$
2	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 0,50$	$-0,11 \pm 0,00$
3	Reliabilitas	0,95	0,00
4	Nilai rata-rata dan simpangan baku INFIT MNSQ	$1,01 \pm 0,02$	$1,01 \pm 0,07$
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 t	$0,84 \pm 1,47$	$0,25 \pm 1,08$
7	Nilai rata-rata dan simpangan baku OUTFIT t	$0,27 \pm 0,48$	$0,08 \pm 0,61$
8	Item atau testi skor 0	0	0
9	Item atau testi skor perfect	0	0

c. Tingkat Kesulitan

Berdasarkan hasil analisis dengan program QUEST didapat tingkat kesulitan item. Berikut dapat dilihat distribusi tingkat kesulitan item-item.

1) Distribusi Tingkat Kesulitan Item

Indeks kesulitan atau tingkat kesulitan (b) untuk skor 1 (b_1), skor 2 (b_2), skor 3 (b_3), skor 4 (b_4), dan reratanya sebagai *difficulty*. Berdasarkan hasil analisis dengan QUEST, *difficulty* item-item terletak antara antara -0,86 sampai dengan 1,06 dengan rata-rata 0 dan simpangan baku 0,48. Item dikatakan baik jika indeks kesukaran lebih dari -2,0 atau kurang dari 2,0 ($-2,0 < b < 2,0$). Jadi berdasarkan nilai *difficulty*, semua item sebanyak 44 semuanya baik, Adapun histogram distribusi indeks kesulitan hasil dari analisis dengan SPSS 21 dapat dilihat pada Gambar 5. Berdasarkan Gambar 5 distribusi indeks kesulitan mengikuti distribusi normal.



Gambar 5.
Distribusi Tingkat Kesulitan Item Tes pada Kegiatan Ujicoba

2) Tingkat Kesulitan Item Tiap Aspek Instrumen dan Kategori pada Tahap Ujicoba

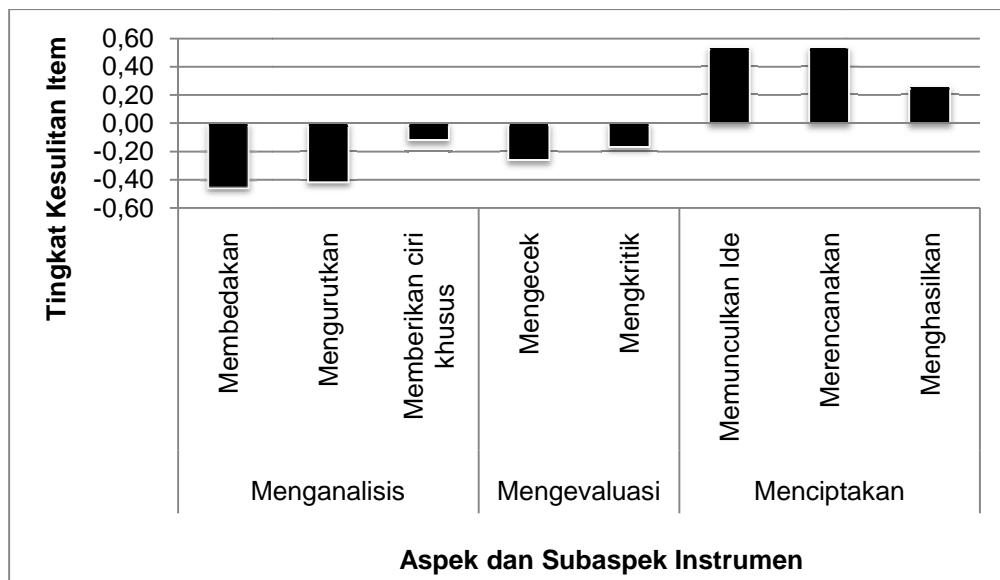
Di depan telah disajikan distribusi tingkat kesulitan item secara keseluruhan. Selanjutnya, berikut dapat jelaskan tingkat kesulitan item pada masing-masing aspek dan subaspek instrumen. Lebih dari itu, dari Tabel 6 dapat diketahui tingkat kesulitan pada masing-masing subaspek dan aspek instrumen untuk masing-masing kategori dalam PCM. Untuk lebih jelasnya, dapat dilihat diagram distribusi tingkat kesulitan item menurut aspek dan subaspek instrumen pada Gambar 9 yang berasal dari Lampiran 7.

Tabel 6.

Tingkat Kesulitan Butir Masing-masing Subaspek untuk Kategori 1, 2, 3, dan 4 pada Kegiatan Ujicoba

No	Aspek	Subaspek	<i>Difficulty</i>	Delta			
				Kategori 1	Kategori 2	Kategori 3	Kategori 4
1	Menganalisis	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	Mengevaluasi	Mengecek	-0,26	-2,00	1,25	0,37	-0,31
		Mengkritik	-0,17	-2,36	1,68	0,19	-0,14
3	Menciptakan	Memunculkan Ide	0,54	0,99	0,32	0,72	-0,23
		Merencanakan	0,53	1,71	-0,28	0,72	-0,16
		Menghasilkan	0,26	-1,10	1,84	0,31	-0,22

Berdasarkan Gambar 6, urutan tingkat kesulitan item masing-masing aspek pada tahap ujicoba berturut-turut adalah menganalisis, mengevaluasi, dan menciptakan.



Gambar 6.
Tingkat Kesulitan Item masing-masing Aspek dan Subaspek Instrumen

3) Item Termudah dan Tersulit

Berdasarkan tingkat kesulitan item, dapat disajikan 10 item termudah dan 10 item tersulit. Item tergolong mudah, jika tingkat kesulitan negatif, sedangkan item yang sulit ditandai dengan tingkat kesulitan positif. Sepuluh item termudah dan sepuluh item tersulit masing-masing disajikan pada Tabel 9 dan Tabel 10.

Tabel 7.

Sepuluh Item Termudah pada Kegiatan Uji coba

No	Item	Tingkat Kesulitan
1	item 5 (5A)	-0,86
2	item 4 (4A)	-0,65
3	item 26 (26A, 8B)	-0,62
4	item 1 (1A)	-0,58
5	item 33 (15B)	-0,58
6	item 10 (10A)	-0,54
7	item 19 (19A, 1B)	-0,51
8	item 37 (19B)	-0,49
9	item 21 (21A)	-0,42
10	item 27 (9B)	-0,37

Tabel 8.
Sepuluh Item Tersulit pada Kegiatan Uji coba

No	Item	Tingkat Kesulitan
1	item 23 (23A, 5B)	1,06
2	item 22 (22A, 4B)	1,03
3	item 24 (24A, 6B)	0,82
4	item 14 (14A)	0,64
5	item 44 (26B)	0,62
6	item 42 (24B)	0,61
7	item 40 (22B)	0,59
8	item 11 (11A)	0,49
9	item 38 (20B)	0,47
10	item 43 (25B)	0,43

4) Persentase Aspek dan Subaspek Tiap Kategori Kemampuan Berpikir Tingkat Tinggi Fiska SMA pada Tahap Ujicoba

Tingkat kesulitan item juga dapat dilacak dari persentase responden menjawab benar untuk masing-masing kategori. Jika persentase responden dapat menjawab kategori 1 paling banyak berarti menggambarkan item tersebut sulit, sebaliknya jika persentase testi paling banyak pada kategori 4 menandakan bahwa item mudah. Persentase responden menjawab benar item pada masing-masing aspek dan suaspek intrumen untuk keempat kategori dinyatakan dalam Tabel 9 yang berasal dari Lampiran 6. Persentase responden dapat menjawab benar untuk kategori 1, 2, 3, dan 4 dapat digunakan sebagai indikator tingkat kesulitan item juga. Jika sebagian besar testi menjawab benar pada kategori 1, maka item tersebut sulit, sebaliknya jika persentase testi paling banyak pada kategori 4 menandakan bahwa item mudah.

Tabel 9.
Percentase Menjawab pada Benar Aspek dan Subaspek Kemampuan Berpikir
Tingkat Tinggi Fiska SMA di DIY pada kegiatan Uji coba

No	Aspek	Subaspek	Kategori 1 (%)	Kategori 2 (%)	Kategori 3 (%)	Kategori 4 (%)
1	Menganalisis	Membedakan	1,81	44,74	36,58	16,87
		Mengurutkan	1,89	46,89	40,93	10,29
		Memberikan ciri khusus	3,31	59,85	29,43	7,41
2	Mengevaluasi	Mengecek	3,25	64,17	21,88	10,70
		Mengkritik	14,52	60,75	12,04	12,68
3	Menciptakan	Memunculkan Ide	54,58	14,83	20,63	9,95
		Merencanakan	56,48	10,38	21,91	11,23
		Menghasilkan	34,67	46,54	10,43	8,36

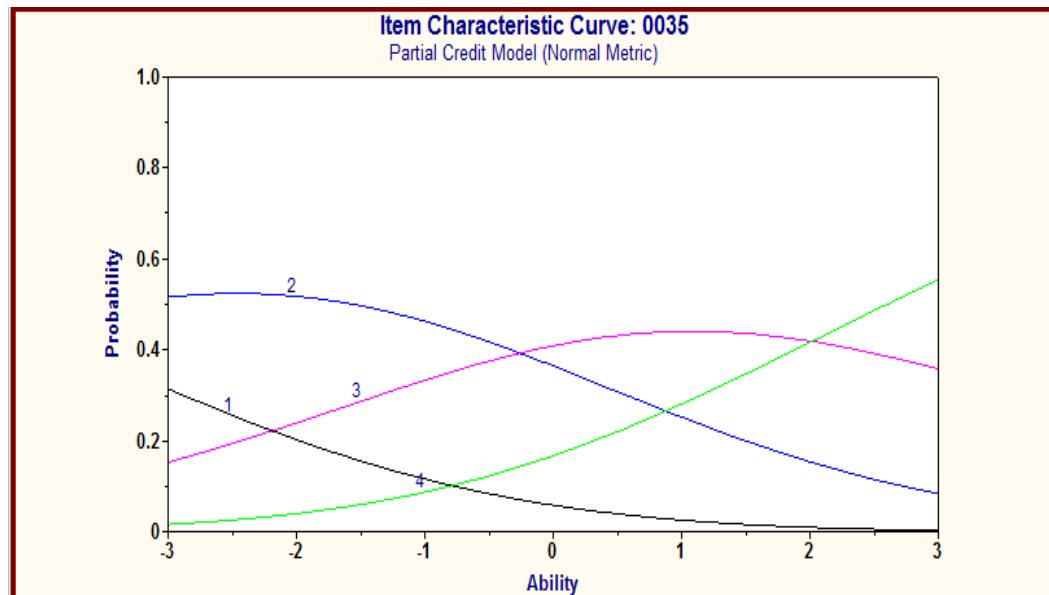
d. Reliabilitas Tes pada Ujicoba

Selain untuk menguji kecocokan, output program QUEST juga menampilkan estimasi reliabilitas set instrumen tes. Berdasarkan hasil analisis dengan program QUEST dalam Lampiran 4 reliabilitas set instrumen (tes) tersebut diestimasikan sebesar 0,95 yang tergolong kategori tinggi.

e. Kurva Karakteristik Item (Item Characteristic Curve)

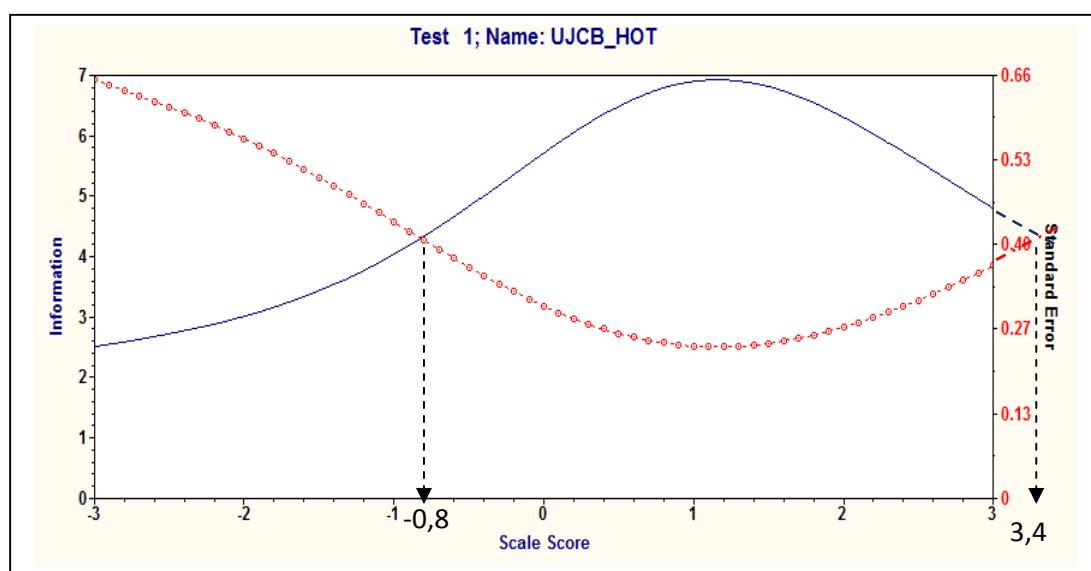
Karakteristik item ditunjukkan dengan kurva karakteristik item (ICC) dan indeks kesukaran. Untuk mendapatkan kurva karakteristik item (ICC) digunakan program Parscale. Berdasarkan analisis dengan Parscale diperoleh kurva karakteristik item (ICC) sebanyak 44 buah. Pada Gambar 7 disajikan contoh ICC untuk item 35 atau 17B, 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 ability tinggi

$(\theta = 1,5)$, dan (d) skor 4 (kategori 4) sebagian besar diperoleh siswa dengan ability sangat tinggi ($\theta = 3$).



Gambar 7.
Kurva Karakteristik Butir 35 (Butir 17B)

f. Kemampuan yang Sesuai untuk Instrumen yang Dikembangkan



Gambar 11.
Fungsi Informasi dan SEM pada Kegiatan Ujicoba

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

C. Revisi Produk untuk Pengukuran

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 *PhysTHOT* terdiri atas dua set tes A dan B yang masing berisi 26 item dengan 8 *anchor item*.

Tabel 10.

Tingkat Kesulitan *Anchor Item*

No	Item ke-	Tingkat Kesulitan
1	Item 19 (19 A, 1 B)	-0,52
2	Item 20 (20 A, 2 B)	0,20
3	Item 21 (21 A, 3 B)	-0,42
4	Item 22 (22 A, 4 B)	0,31
5	Item 23 (23 A, 5 B)	0,38
6	Item 24 (24 A, 6 B)	0,82
7	Item 25 (25 A, 7 B)	-0,21
8	Item 26 (26 A, 8 B)	-0,62
	Rata-rata	0,01
	SB	0,51

Tingkat kesulitan *Anchor Item* dan tingkat kesulitan seluruh item untuk set tes A dan B masing-masing dinyatakan pada Tabel 12 dan Tabel 13 yang berasal dari Lampiran 7.

Tabel 11.
Tingkat Kesulitan Item Setiap Perangkat Tes A dan B

No	Set A		Set B	
	Item ke-	Tingkat Kesulitan	Item ke-	Tingkat Kesulitan
1	Item 1	-0,59	Item 1*)	-0,52
2	Item 2	-0,29	Item 2*)	0,20
3	Item 3	-0,19	Item 3*)	-0,42
4	Item 4	-0,65	Item 4*)	0,31
5	Item 5	-0,87	Item 5*)	0,38
6	Item 6	-0,20	Item 6*)	0,82
7	Item 7	-0,28	Item 7*)	-0,21
8	Item 8	-0,29	Item 8*)	-0,62
9	Item 9	-0,06	Item 9	-0,42
10	Item 10	-0,54	Item 10	-0,08
11	Item 11	0,49	Item 11	-0,11
12	Item 12	1,06	Item 12	-0,27
13	Item 13	1,03	Item 13	-0,30
14	Item 14	0,63	Item 14	-0,20
15	Item 15	0,38	Item 15	-0,58
16	Item 16	-0,10	Item 16	-0,12
17	Item 17	-0,30	Item 17	-0,33
18	Item 18	0,21	Item 18	-0,09
19	Item 19*)	-0,52	Item 19	-0,49
20	Item 20*)	0,20	Item 20	0,47
21	Item 21*)	-0,42	Item 21	0,38
22	Item 22*)	0,31	Item 22	0,68
23	Item 23*)	0,38	Item 23	0,43
24	Item 24*)	0,82	Item 24	0,63
25	Item 25*)	-0,21	Item 25	0,42
26	Item 26*)	-0,62	Item 26	0,61
	Rata-rata	-0,02	Rata-rata	0,02
	SB	0,51	SB	0,37

Keterangan:

*) : *Anchor Item*

D. Pembahasan

a. Karakteristik Instrumen

Berdasarkan Tabel 7 dan Tabel 16 yang dirangkum pada Tabel 21, parameter instrumen pada kegiatan ujicoba dan kegiatan pengukuran tidak berbeda secara signifikan. Hal ini berarti bahwa *PhysTHOT* yang berupa tes pilihan ganda berasalan dengan lima pilihan jawaban yang menggunakan penskoran PCM 1 PL memiliki tingkat kestabilan yang tinggi. Hal ini juga didukung oleh estimasi reliabilitas yang tinggi 0,95 pada tahap ujicoba dan 0,97 pada tahap pengukuran, 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, 2000: 39-40). Berdasarkan parameter estimasi, instrumen tes ini reliabel dan memiliki kestabilan yang tinggi serta dapat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika siswa SMA.

Menurut Hambleton and Swaminathan (1985:36), skor ability pada kelompok memiliki rata-rata dan simpangan masing-masing 0 dan 1 tingkat kesulitan b bervariasi antara -2,0 sampai dengan 2,0. Item dengan tingkat kesulitan -2 menandakan item yang sangat mudah, sedangkan tingkat kesulitan 2,0 berarti item sangat sulit. Dengan demikian ditinjau dari tingkat kesulitan soal dan kestabilanya maka instrumen ini termasuk kategori baik.

Tabel 12.
Rangkuman Tingkat Kesulitan Item pada Tahap Ujicoba

No	Tingkat kesulitan	Nilai
1	Tertinggi	1,06
2	Terendah	-0,86
3	Rata-rata	0
4	Simpangan Baku	0,48

Atas dasar uraian di atas, maka instrumen *PhysTHOT* yang telah dilakukan validasi melalui *expert judgment* sudah mendapatkan bukti empiris *fit* dengan *Partial Credit Model (PCM)* berdasarkan data politomus empat kategori. Di samping itu, tingkat kesulitan berada pada rentang sebagai instrumen yang baik yakni dari -2,00 sampai 2,00. Lebih dari itu bahwa instrumen *PhysTHOT* terbukti dapat digunakan untuk mengukur kemampuan berpikir tingkat tinggi Fisika pada SMA terpilih menurut *Partial Credit Model (PCM)* berdasar data politomus empat kategori. Dengan demikian instrumen *PhysTHOT* telah memenuhi syarat digunakan untuk mengukur kemampuan berpikir tingkat tinggi dalam Mata Pelajaran Fisika di SMA Kelas XI.

BAB VI

RENCANA TAHAPAN BERIKUTNYA

Penelitian ini adalah penelitian disertasi doktor (PDD) yang bersifat untuk mendukung penulisan disertasi, sehingga jangka waktu penelitian ini hanya satu tahun. Oleh karena itu, rencana tahapan berikutnya menggunakan instrumen PhysTHOTS untuk mengukur kemampuan berpikir tingkat tinggi peserta didik kelas XI SMA di Daerah Istimewa Yogyakarta yang dilakukan pada disertasi. Namun, rencana tahapan berikutnya untuk khusus penelitian judul ini tentu tidak ada.

BAB VII

SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan uraian di atas, simpulan yang dapat ditarik adalah sebagai berikut.

1. Instrumen *PhysTHOTS* dikembangkan 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 tes B yang masing-masing terdapat 26 item dengan 8 *anchor item*.
2. Instrumen *PhysTHOTS* memiliki karakteristik sebagai instrumen yang memenuhi syarat digunakan untuk mengukur, yakni:
 - a. Instrumen *PhysTHOTS* telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris *fit* dengan *Partial Credit Model (PCM)* berdasarkan data politomus empat kategori.
 - b. Seluruh item pada *PhysTHOTS* dalam kriteria baik karena tingkat kesulitannya berada pada rentang antara -2,00 sampai dengan 2,00.
 - c. Reliabilitas *PhysTHOTS* telah memenuhi syarat, bahkan termasuk tinggi (koefisien reliabilitas lebih dari 0,90).
 - d. Berdasarkan fungsi informasi, *PhysTHOTS* sangat tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik yang berkemampuan dari -0,80 sampai 3,40.

B. Saran

Berdasarkan simpulan dan keterbatasan penelitian ini, dapat diberikan saran sebagai berikut.

1. Penelitian ini baru sampai tahap perakitan tes, untuk mengetahui kestabilan instrumen perlu dilakukan uji coba penerapan yang berupa pengukuran kemampuan berpikir tingkat tinggi fisika siswa SMA
2. Dalam penelitian ini baru terbatas pada PhysTHOT kelas XI, agar diperoleh gambaran menyeluruh perlu dilakukan untuk kelas X dan XII.

DAFTAR PUSTAKA

- 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 questionsto 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 rd 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
- Chiappetta, E. L and Koballa, T. R. 2010. *Science Instructiion in the Middle and Secondary Schools*. New York: Allyn & Bacon
- Depdiknas. (2003). *Undang-undang No. 20 tahun 2003 tentang Sistem Pendidikan Nasional*
- (2007). *Peraturan Menteri Pendidikan Nasional No 20 Tahun 2007 tentang Standar Penilaian*
- (2008). *Panduan Penulisan Butir Soal* Jakarta: Direktorat Pembinaan Sekolah Menengah Atas
- 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 diakses tanggal 1 Juni 2012
- Djemari Mardapi. (2004). *Penyusunan Tes Hasil Belajar*. Yogyakarta: PPs Universitas Negeri Yogyakarta
- Haladyna, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*. Boston: Allyn and Bacon
- Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New Jersey: Lawrence Erlbaum Associates, Inc.

Hambleton dan Swaminathan (1991): *Fundamentals of Item Response Theory*. California: SAGE Publications, Inc

Linacre, J. M. (2006). *Winstep : Rasch-model Computer Programs*. Chicago: Winsteps.com.

Linn, R. L. (1999). *Measurement and Evaluation in Teaching* . New York: Macmillan Publishing Company

Madhuri, G. V. Kantamreddi, V. S.S.N and Goteti, P. L. N.S. (2011). *Promoting higher order thinking skills using inquiry-based learning*. European Journal of Engineering Education. <http://www.tandfonline.com/loi/ceee20> Diakses tanggal 20 Novemver 2012

Muraki, E., & Bock, R.D. (1997). *Parscale 3: IRT based test scoring and item analysis for graded items and rating scales*. Chicago: Scintific Software Inc.

Muraki, E. (2003). *Models in PARSCALE*. In M. du Toit (Ed.), *IRT from SSI : BILOGMG, MULTILOG, PARSCALE, TESTFACT*. Chicago: Scientific Software International.

Nana Sudjana. (1990). *Penilaian Hasil Blajar Mengajar*. Bandung: PT Remaja Rosdakarya

Nana Sudjana & Ahmad Rivai. (2003). *Teknologi pengajaran*. Bandung : Sinar Baru Algensindo

OECD. (2007). *PISA 2006: Science competencies for tomorrow world volume 1: Analysis*. Rosewood. Drive: OECD

Oriondo, L.L.and Dallo-Antonio, E.M. (1998). *Evaluation Educational Outcomes*. Manila: Rex Printing Compagny, inc

Ostini, Remo and Nering, M. L. 2006. *Polytomous item response theory models*. California: Sage Publications, Inc.

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/macs/summation>.

Ridwan Efendi. (2010). *Kemampuan Fisika Siswa Indonesia dalam TIMSS*. Prosiding Seminar Nasional Fisika 2010 ISBN : 978-979-98010-6-7

Schraw , G, and Robinson, D.H. (2011). *Assessment of Higher Order Thinking Skills*. New York: Information Age Publishing, Inc

Sumaji (1990). *Faktor Internal dan Eksternal Siswa Sekolah Menengah Atas dan Kemandiriannya Melakukan Percobaan Fisika*. (Disertasi Doktor). Tidak diterbitkan Yogyakrta: IKIP Yogyakarta

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>

Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55, 371–390.

Van der Linden, Wim J and Hambleton, Ronald K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag New York, Inc

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.