



## **PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA SMA**

**Kistiono**

Fakultas Keguruan dan Ilmu Pendidikan Universitas Sriwijaya  
Jalan Raya Palembang Prabumulih KM.33 Indralaya. Ogan Ilir Sumatera Selatan. Indonesia  
[kistiono.fkip@gmail.com](mailto:kistiono.fkip@gmail.com)

### **ABSTRAK**

Telah dilakukan penelitian pengembangan tes kemampuan berpikir tingkat tinggi fisika SMA. Metode penelitian yang digunakan dalam penelitian ini adalah metode *Risearch* dan *Development* (R and D) dengan merujuk model penelitian pengembangan yang dikembangkan oleh Tiaragajan dengan menggunakan empat langkah, yaitu pendefinisian, perencanaan, pengembangan dan penyebaran. Populasi penelitian seluruh kelas XII yang tersebar diseluruh wilayah kabupaten Ogan Ilir yang berjumlah 35 SMA berstatus negeri dan berstatus swasta yang berjumlah 3225 peserta didik untuk kelas XII. Indikator yang dapat dikembangkan untuk jenjang kemampuan analisis (*analyze*) meliputi indikator: mengelompokkan, menjelaskan, menganalisis, membedakan dan menemukan. Penilaian (*evaluate*) meliputi indikator membandingkan, menafsirkan, menyimpulkan dan membuat urutan. Untuk jenjang kemampuan penciptaan (*create*) indikator yang dikembangkan: menggabungkan merencanakan, memformulasikan dan merealisasikan, dengan jumlah item tes 60 item tes (5 item tes untuk masing-masing indikator). Analisis reliabilitas dan validitas bersumber dari penskoran yang diberikan oleh validator dan alasan jawaban responden melalui empat indikator rubrik kemudian dianalisis dengan menggunakan teknik Cronback Alpha. Hasil analisis dapat disimpulkan: (1) Penilaian ke tiga validator didapat untuk setiap butir tes didapat  $r_{hitung}$  lebih besar dari  $r_{tabel}$ , hal ini menunjukkan bahwa instrumen tes yang dikembangkan valid, dan koefisien reliabilitas untuk setiap jenjang dimensi kemampuan didapat lebih besar dari 0,7 hal ini menunjukkan bahwa butir-butir tes yang dikembangkan pada setiap indikator mempunyai reliabilitas yang baik. (2) Pada tahap uji coba dengan 50 sampel didapat semua item butir tes mempunyai koefisien lebih besar dari 0,279 hal ini menunjukkan bahwa semua butir tes yang dikembangkan valid. Analisis reliabilitas untuk setiap jenjang kemampuan didapat lebih besar 0,7 menunjukkan butir-butir tes yang dikembangkan pada setiap indikator mempunyai reliabilitas yang baik. (3) Pada tahap uji coba luas dengan 312 sampel didapat semua item butir tes mempunyai koefisien lebih besar dari 0,113 hal ini menunjukkan bahwa semua butir tes yang dikembangkan valid. Analisis reliabilitas untuk setiap jenjang kemampuan didapat lebih besar 0,7 maka butir-butir tes yang dikembangkan pada setiap indikator mempunyai reliabilitas yang baik.

**Kata kunci:** Tes Berpikir Tingkat Tinggi, Validitas, Reliabilitas.

**Cara Menulis Sitasi:** Sulistiyono. (2019). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika SMA. *Jurnal Inovasi dan Pembelajaran Fisika*, 6 (1), 70-81.

## **PENDAHULUAN**

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*)

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, 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.

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).

Dalam visi pendidikan nasional tahun 2025 dikehendaki terlahirnya peserta didik yang cerdas dan terampil. Kecerdasan dan keterampilan menjadi prasyarat di kehidupan abad 21. Proses pendidikan yang diterima peserta didik pada hari ini akan menentukan bagaimana ia kelak hidup dan bersaing di dunia nyata. Oleh karena itu implementasi kurikulum 2013 menekankan pada kecerdasan tingkat tinggi yang dibingkai oleh sikap ketuhanan dan nilai-nilai sosial yang terintegrasi dalam proses pembelajaran, peserta didik harus digiring untuk dapat untuk dapat belajar berpikir agar mempunyai kemampuan berpikir tingkat tinggi atau *Higher Order Thinking*..(Kemendikbud 2016).

Dalam taksonomi Bloom yang direvisi (taksonomi Anderson dan Krathwohl) dimensi kemampuan berpikir yang semula hanya terdapat satu dimensi yaitu dimensi pengetahuan saja, berubah menjadi dua dimensi,yaitu pertama adalah *knowledge dimension* (dimensi pengetahuan) dan *cognitive process dimension* (dimensi proses kognisi). Dimensi proses kognisi terdapat 6 kategori, yaitu kemampuan mengingat, memahami, dan menerapkan yang merupakan kemampuan berpikir tingkat rendah. Selain itu kemampuan menganalisis, mengevaluasi, dan mencipta termasuk kemampuan berpikir tingkat tinggi Kategori kategori dalam dimensi proses kognitif berpikir tingkat tinggi adalah terdiri dari empat kategori, yaitu: (a) Pengetahuan Faktual (PF), (b) pengetahuan konseptual (PK), (c) Pengetahuan ( Prosedural) dan (d) Pengetahuan Metokognisi (PM). Dua dimensi selengkapnya untuk dimensi revisi taksonomi bloom dan contoh kata kerja operasional untuk kemampuan berpikir tingkat tinggi ditunjukkan pada Tabel1.

Tabel 1. Dimensi revisi Taksonomi Bloom dan contoh kata kerja operasional untuk Kemampuan Berpikir Tingkat Tinggi.

Dimensi Pengetahuan (The Knowledge Dimension)	Dimensi Proses Kognisi (The Cognitive Process Dimension)		
	C4	C5	C6
	Analisis (analyze)	Penilaian (evaluate)	Penciptaan (create)
Pengetahuan Faktual (PF)	Membandingkan	Membandingkan, menghubungkan	Menggabungkan
Pengetahuan Konseptual (PK)	Menjelaskan/ Membedakan/ mendekonstruks	Mengkaji, Menafsirkan Menyimpulkan, Meringkas	Merencanakan Mengobinasikan, Memformulasikan
Pengetahuan Metakognisi (PM)	Menemukan	Membuat urutan, Menilai	Merealisasikan

(Anderson and Krathwohl, 2001)

Untuk melihat kemajuan hasil belajar belajar peserta didik secara berkesinambungan diperlukan penilaian. Penilaian hasil belajar adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik (Peraturan Menteri Pendidikan Nasional No 20, 2007). Penilaian hasil belajar dilakukan secara lisan ataupun tertulis. Penilaian hasil belajar 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.

Penilaian hasil belajar tersebut mencerminkan kemampuan berpikir peserta didik. Menurut Van den Berg (2008, p.15) bahwa kurikulum sebagai untuk mengembangkan kemampuan berpikir tingkat tinggi peserta didik. Guru harus merencanakan dengan baik dan melibatkan peserta didik dalam kegiatan pembelajaran untuk mendorong dan mengembangkan kemampuan berpikir tingkat tinggi. Penilaian hasil belajar dapat diimplementasikan untuk memantau kemampuan hasil belajar dan kemampuan berpikir tingkat tinggi mereka. Hal ini didukung pendapat lain yang menyatakan bahwa pertanyaan berpikir tingkat tinggi dapat mendorong peserta didik untuk berpikir secara mendalam tentang materi pelajaran (Barnett & Francis (2012, p.209).

Berdasarkan paparan tersebut dapat disimpulkan bahwa tes dapat memberikan rangsangan kepada peserta didik untuk mengembangkan kemampuan berpikir tingkat tinggi. Nitko & Brookhart (2011, p.223) lebih lanjut dijelaskan bahwa kemampuan penilaian berpikir tingkat tinggi dapat melalui tugas-tugas dan atau dengan menggunakan indikator-indikator yang jelas dan terukur, diantaranya adalah dapat dilakukan dengan menggunakan set-set *item tes* yang bergantung pada konteks, konten dan indikator.

Dalam melaksanakan penilaian hasil belajar pada umumnya menggunakan dua macam teori pengukuran, yakni: (1) teori pengukuran klasik dan, (2) teori pengukuran modern. Teori tes klasik disebut juga dengan *Classical TrueScore Theory*, disebut 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.

Cara lain adalah didasarkan pada tahap-tahap yang dapat diselesaikan peserta didik. Walaupun hanya menyelesaikan tahap awal saja, peserta didik sudah mendapatkan nilai. Nilai tertinggi tentu saja didapatkan ketika peserta didik 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 jawaban 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*“. Analog lain adalah dalam pembelajaran eksak misalnya, alasan memilih option tertentu, *semua* menggunakan kaidah, prinsip, hukum-hukum dengan benar, *sebagian besar* jawaban menggunakan kaidah, prinsip, hukum-hukum dengan benar, *sebagian kecil* jawaban menggunakan kaidah, prinsip, hukum-hukum dengan benar, jawaban *tidak* menggunakan kaidah, prinsip, hukum-hukum dengan benar.

Asumsi ini kemudian dikembangkan menjadi *partial credit model* (PCM). Ketika diasumsikan bahwa sebuah *item* mengikuti pola kredit parsial model 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 tingkat tinggi dan pemahaman konseptual dalam konteks eksak (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 hasil belajar termasuk tes fisika yang membutuhkan tahap identifikasi permasalahan hingga solusi akhir.

Menurut 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 dikotomis* menjadi *politomis* 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. Dengan demikian PCM dapat diterapkan pada butir politomis dan dikotomis

Fakta menunjukkan 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 peserta didik dapat dikoreksi dengan mudah dan cepat, (3) jawaban setiap pertanyaan sudah pasti benar atau salah, sehingga penilaian objektif (Sudjana, 1990, p.49). Selain itu kelemahan tes ini, yaitu: (1) memungkinkan peserta didik untuk melakukan tebakan jawaban masih cukup besar dan (2) proses berpikir peserta didik 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 diperlukan mengapa, alasan peserta didik memberikan jawaban/memilih option tertentu.

Hasil studi pendahuluan melalui wawancara dengan beberapa guru-guru fisika SMA di Kabupaten Ogan Ilir diperoleh informasi bahwa; (1) sebagian besar sekolah, baik pada tes tengah semester maupun tes akhir semester dan evaluasi belajar tahap akhir umumnya menggunakan tes pilihan ganda biasa, tes pilihan ganda masih merupakan hal penting dan bahkan andalan dalam mendapatkan data prestasi belajar fisika peserta didik SMA, alasan lain adalah bahwa tes pilihan ganda mempunyai kelebihan yaitu memberikan kemudahan dalam mengoreksi sehingga guru dengan mudah dapat memetakan kemampuan peserta didik berdasarkan jawaban benar dan salah. Kenyataan lain menunjukkan bahwa tes pilihan ganda yang digunakan sebagai 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)*) dan belum mengukur kemampuan berpikir tingkat tinggi (*HOT*), (2) tes yang digunakan guru-guru hampir seluruhnya merupakan soal tes berbentuk pilihan ganda yang mengadopsi dari buku-buku SMA atau yang sederhana, Jumlah skor jawaban (pilihan) yang benar atau menjadi tolak ukur pengukuran keberhasilan peserta didik, (4) soal atau tes yang digunakan belum diketahui tingkat kelayakannya sebagai tes yang baik karena tes yang digunakan tidak dilakukan validitasnya dan reliabelitasnya. Beranjak dari paparan dan permasalahan-permasalahan maka diperlukan *inovasi untuk mengembangkan tes untuk menghasilkan tes untuk mengukur kemampuan berpikir tingkat tinggi yang valid dan reliabel*

### **Metode Penelitian**

Metode penelitian yang digunakan dalam penelitian ini adalah metode *research and development (R and D)* dengan merujuk model penelitian pengembangan yang dikembangkan oleh Tiaragajan dengan menggunakan empat langkah, yaitu pendefinisian, perencanaan, pengembangan dan penyebaran

Penelitian dimulai Juni 2018 sampai dengan Desember 2018. Pengembangan awal tes yang berupa penyiapan tes, validasi, dan penyusunan tes tes dilakukan pada bulan Agustus 2018 sampai dengan September 2018. Uji coba dilaksanakan pada bulan Oktober 2018. Populasi penelitian seluruh kls XII yang tersebar diseluruh wilayah kabupaten Ogan Ilir yang berjumlah 35 SMA berstatus negeri dan berstatus swasta yang berjumlah 3225 peserta didik untuk kls XII

Subjek penelitian 312 peserta didik dengan tingkat kesalahan 5% (Isaac dan Michael dalam Sugiyono, :2007:86-87) Hal ini senada dengan Reckase (2000) sampel ukuran minimum yang baik untuk memperkirakan beberapa parameter (reliabelita, validitas adalah 300 responden.

Langkah-langkah pengembangan instrumen menggunakan model Creswell & Clark, yakni:

#### **(1) Tahap Pendisinisian**

Pada tahap ini melakukan kegiatan antara analisis kebutuhan, antara lain: (a) Mengkaji arah dan tujuan pembelajaran fisika di SMA agar peserta didik mempunyai kemampuan tidak saja (*lower order thinking*), tetapi sampai pada kemampuan berpikir tingkat tinggi (*higher order thinking, HOT*), (b) mengadakan observasi di beberapa SMA yang tersebar di

Wilayah Kab.Ogan Ilir didapattemuan dilapangan tes untuk mengevaluasi keberhasilan belajar masih dalam kategori LOT dan belum ada instrumen tes untuk mengukur kemampuan HOT.

**(2) Tahap Pendesainan**

- 1) Mengkaji materi fisika yang ada di SMA, mengingat begitu luasnya kajian bidang-bidang fisika maka dalam penelitian ini hanya akan memfokuskan pengembangan tes pada kajian dinamika partikel yang terdapat pada mata pelajaran Fisika sekolah menengah atas (SMA), yaitu: (1) hukum newton tentang gerak, (2) gerak jatuh bebas,(3) hukum hooke,(4) rangkaian pegas paralel, (5) osilasi ayunan bandul,(6) osilasi pada pegas,(7) gaya gesekan dan (8) hukum archimedes,
- 2) Setelah mengkaji cakupan materi selanjutnya menentukan draf kisi-kisi dan jumlah item tes untuk setiap kemampuan yang ditunjukkan pada Tabel 1.

Tabel.1. Jumlah item tes dan indikator yang dikembangkan pada pengembangan instrumen tes Kemampuan Berpikir Tingkat Tinggi pada mata pelajaran fisika SMA

Dimensi Pengetahuan (The Knowledge Dimension)	Dimensi Proses Kognisi (The Cognitive Process Dimension)						Jumlah item tes
	C4 Analisis (analyze)	Jumlah item tes	C5 Penilaian (evaluate)	Jumlah item tes	C6 Penciptaan (create)	Jumlah item tes	
Pengetahuan Faktual (PF)	Membandingkan	5	Membandingkan, menghubungkan	5	Menggabungkan	5	15
Pengetahuan Konseptual (PK)	Menjelaskan/	5	Mengkaji, Menafsirkan	5	Merencanakan	5	15
Pengetahuan Prosedural (PP)	Membedakan/ mendekonstruks	5	Menyimpulkan, Meringkas	5	Memformulasikan	5	15
Pengetahuan Metakognisi (PM)	Menemukan	5	Membuat urutan, Menilai	5	Merealisasikan	5	15
	Jumlah	20		20		20	60

- 3) Langkah selanjutnya membuat lembar penilaian validator dan respinden untuk setiap indikator kemampuan berpikir tingkat tinggi, seperti ditunjukkan Tabel 2.

Tabel 2. Cotoh Kisi-Kisi Draft Instrumen Untuk Kemampuan Dimensi Proses C4 Analisis (Analyze)

Dimensi pengetahuan	Dimensi Proses Kognisi	Jumlah item tes	Nomor item tes	Indikator Item tes	Pertanyaan/ pernyataan item tes	Kunci jawaban	Saran/ Rekomendasi Validator	Skor tingkat kesesuaian			
								1	2	3	4
Pengetahuan Faktual (PF)	Membandingkan	5	6-10	Peserta didik dapat memberikan contoh gerak benda yang dipengaruhi gaya gravitasi	1. Manakah dari peristiwa berikut yang merupakan contoh gerak benda yang tidak dipengaruhi oleh gaya gravitasi. a. Gerak koin yang dilemparkan vertikal ke atas dan gerak peluru yang ditembakkan dalam arah horisontal b. Gerak peluru yang ditembakkan dalam arah horizontal dan buah mangga yang jatuh dari pohonnya c. Buah mangga yang jatuh dari pohonnya dan gerak bandul yang berayun d. Gerak bandul yang berayun dan gerak osilasi pegas e. Jawaban a,b,c dan d semuanya benar						
Dst	-	-	-	-	Alasan jawaban.....						

Keterangan

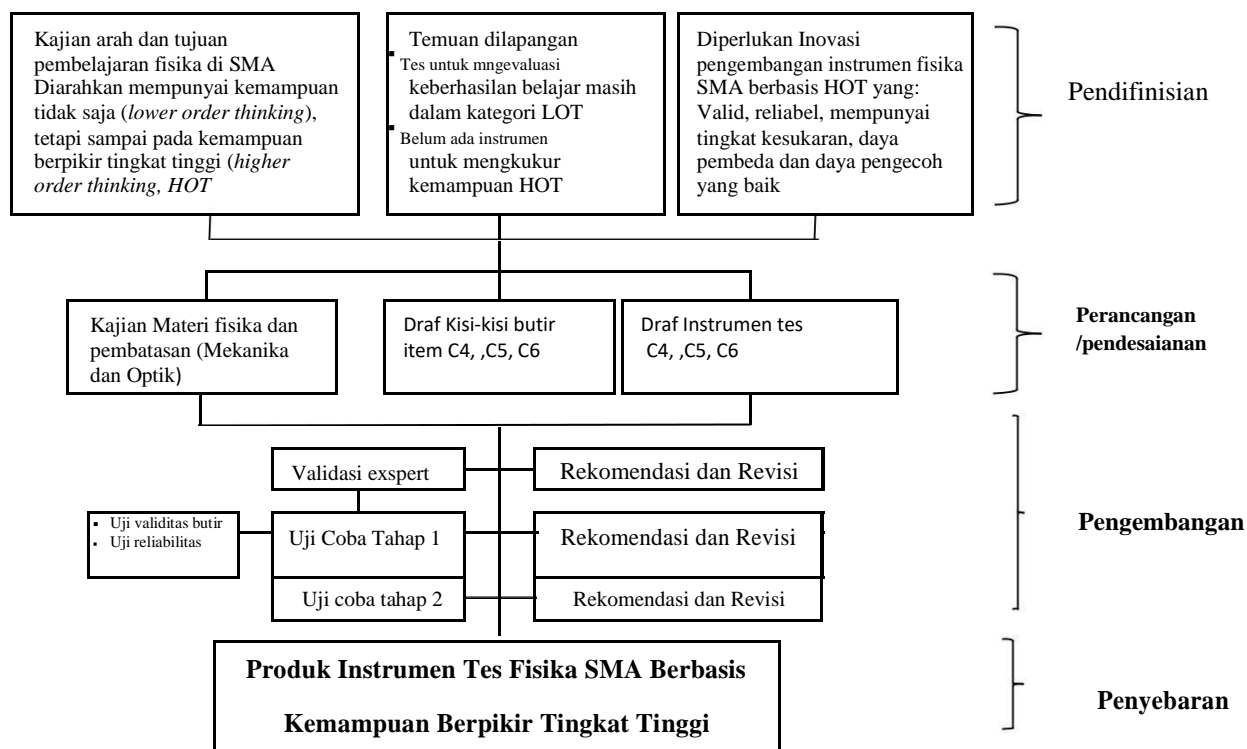
Pedoman Penekoran Validator	
Skor	tingkat kesesuaian
1:	Redaksi, gambar, pertanyaan/ Pernyataan item tes dan jawaban <b>tidak sesuai</b> dengan indikator Item dimensi proses Kognisi
2:	Redaksi, gambar, pertanyaan/ Pernyataan item tes dan jawaban <b>kurang sesuai</b> dengan indikator Item dimensi proses Kognisi
3:	Redaksi, gambar, pertanyaan/ Pernyataan item tes dan jawaban <b>cukup sesuai</b> dengan indikator Item dimensi proses Kognisi
4:	Redaksi, gambar, pertanyaan/ Pernyataan item tes dan jawaban <b>sesuai</b> dengan indikator Item dimensi proses Kognisi

Pedoman Penekoran Jawaban Responden Pada Tahap Uji Coba Terbatas Dan Uji Coba Luas	
Skor	tingkat kesesuaian kaidah, prinsip, hukum-hukum
1:	Jawaban tidak menggunakan kaidah, prinsip, hukum-hukum dengan benar
2:	Jawaban hanya menggunakan sebagian kecil kaidah, prinsip, hukum-hukum dengan benar
3:	Jawaban menggunakan sebagian besar kaidah, prinsip, hukum-hukum dengan benar,
4:	Jawaban menggunakan semua kaidah, prinsip, hukum-hukum dengan benar

### (3) Tahap Pengembangan

- 1) Setelah draf instrumen tes selesai maka dilanjutkan pada uji validasi oleh 3 pakar dari rekomendasi pakar (validator) diadakan revisi-revisi berdasarkan saran –saran yang diberikan dan hasil penilaian kesesuaian oleh tiga validator dianalisis untuk menentukan validitas dan reliabilitas instrumen tes yang dikembangkan
- 2) Melakukan uji coba terbatas (tahap1)
  - Tujuan uji coba terbatas untuk mengetahui apakah instrumen telah memenuhi syarat sebagai instrumen yang baik yaitu di lihat dari (validitas butir dan reliabilitasnya).
  - Subjek uji coba terbatas 143 peserta didik dari yang diambil secara random dari sampel 312 peserta didik (menurut Isaac dan Michael dalam Sugiyono, 86-87:2007).
- 3) Melakukan uji coba luas (tahap2)
  - Tujuan uji coba terbatas untuk mengetahui apakah instrumen telah memenuhi syarat sebagai instrumen yang baik yaitu di lihat dari (validitas butir dan reliabilitasnya).
  - Subjek uji coba luas 312 peserta didik yang diambil secara random dari populasi penelitian seluruh kls XII yang tersebar diseluruh wilayah kabupaten Ogan Ilir yang berjumlah 35 SMA berstatus negeri dan berstatus swasta yang berjumlah 3225 peserta didik untuk kls XII. Menurut Isaac dan Michael (dalam Sugiyono, 86-87: 2007), Subjek penelitian 312 peserta didik dari populasi dengan tingkat kesalahan 5% adalah 312, Hal ini senada dengan Reckase (2000) sampel ukuran minimum yang baik untuk memperkirakan beberapa parameter (reliabilitas, validitas) adalah 300 responden
  - Menganalisis validitas butir dan uji reliabilitasnya terhadap skor-skor jawaban alasan yang diberikan peserta didik (responden) baik pada tahap expert, uji coba terbatas dan uji coba luas.

Bagan langkah-langkah penelitian dalam mengembangkan petunjuk praktikum berbasis kerja ilmiah digambarkan oleh Gambar 1.



Gambar 1. Bagan langkah-langkah Penelitian

## Hasil dan Pembahasan

Hasil validasi kontruk dillaksanakan dengan 3 pakar dengan tujuan untuk memperoleh saran dan rekomendasi kesesuaian tiap butir item yang dikembangkan yaitu kesesuaian antara indikator kata kerja dalam dimensi proses kognisi terhadap indikator item tes, indikator item tes terhadap pernyataan/pertanyaan, pernyataan/pertanyaan terhadap pengecoh option jawaban, pengecoh option jawaban terhadap kunci jawaban.

Hasil saran dan rekomendasi ketiga validator untuk ketiga aspek jenjang kemampuan analisis (C4), evaluasi/penilaian (C5) dan kreasi/penciptaan (C6) untuk masing-masing indikator yang dikembangkan validator memberikan rekomendasi sesuai selain itu juga adanya rekomendasi untuk dilakukan revisi terhadap redaksionalnya. Hasil analisis validitas dan reliabelitas dari 3 pakar terhadap kesesuaian pernyataan/pertanyaan, option, jawaban terhadap indikator yang dikembangkan pada setiap dimensi kemampuan berpikir tingkat tinggi ditunjukkan pada Tabel 2.

Tabel 2. Hasil analisis validitas dan reliabelitas oleh 3 pakar terhadap kesesuaian pernyataan/pertanyaan, option, jawaban terhadap indikator yang dikembangkan pada setiap dimensi kemampuan berpikir tingkat tinggi

Dimensi Proses koqnitif	Indikator kata kerja dalam dimensi Proses Kognisi (1)	Rata-rata Validitas (r) untuk 5 item tes	Reliabilitas (R)
C4	Mengelompokkan	0,999	0,987



Analisis ( <i>analyze</i> )	Menjelaskan	1,000	0,877
	Membedakan	0,998	0,887
	Menemukan	1,000	0,798
C5 Penilaian ( <i>evaluate</i> )	Membandingkan	1,000	0,877
	Menafsirkan	1,000	0,834
	Menyimpulkan	0,998	0,851
C6 Penciptaan ( <i>create</i> )	Membuat urutan	1,000	0,862
	Menggabungkan	0,998	0,987
	Merencanakan	1,000	0,877
	Memformulasikan	1,000	0,887
	Merealisasikan	0,999	0,824

Koefisien validitas ( $r$  hitung) didapat 0,998 -1,000, Untuk  $n = 3$   $r_{tabel} = 0,997$  (Dwi Priyatno.2009, h: 216), setelah dibandingkan didapat bahwa semua butir tes mempunyai koefisien lebih besar dari 0,997, ini menunjukkan bahwa semua butir tes yang dikembangkan valid, dalam arti lain bahwa semua redaksi, pertanyaan, jawaban yang dikembangkan yang disusun dalam tes telah sesuai dan tepat untuk menilai setiap indikator C4, C5 dan C6 sehingga instrumen tes yang dikembangkan mampu individu untuk mengambil suatu keputusan terhadap pertanyaan yang dimaksud dalam tes. Menurut Dwi Priyatno (2009, h: 172) koefisien uji reliabelitas kurang 0,6 kurang reliabel, 0,7 dapat reliabelitasnya diterima dan lebih dari 0,8 reliabelitasnya baik. Hasil analisis reliabelitas (Tabel 2) terhadap skor-skor validator didapat koefisien 0,798 – 0,987, hal ini memberikan makna bahwa butir-butir tes yang dikembangkan pada setiap indikator C4, C5 dan C6 ketiga validator memberikan keputusan penilaian yang ajek. Hasil analisis validitas dan reliabelitas terhadap 60 butir terhadap sampel uji coba terbatas dilaksanakan terhadap 132 peserta didik (30% dari 143 peserta didik, Isaac dan Michael dalam Sugiyono, 86-87:2007) ditunjukkan pada Tabel 3.

Tabel 3. Hasil analisis validitas dan reliabelitas pada tahap uji coba (50 peserta didik) terhadap indikator yang dikembangkan pada setiap dimensi kemampuan berpikir tingkat tinggi

Dimensi Proses kognitif	Indikator kata kerja dalam dimensi Proses Kognisi (1)	Rata-	
		rata Validitas ( $r$ ) untuk 5 item tes	Reliabilit as (R)
C4 Analisis ( <i>analyze</i> )	Mengelompokkan	0,701	0,753
	Menjelaskan	0,719	0,642
	Membedakan	0,840	0,882
C5 Penilaian ( <i>evaluate</i> )	Menemukan	0,839	0,882
	Membandingkan	0,746	0,775
	Menafsirkan	0,692	0,718
	Menyimpulkan	0,682	0,867
C6 Penciptaan ( <i>create</i> )	Membuat urutan	0,845	0,735
	Menggabungkan	0,832	0,734
	Merencanakan	0,769	0,802
	Memformulasikan	0,834	0,856
	Merealisasikan	0,762	0,749

Koefisien validitas ( $r$  hitung) didapat 0,701- 0,840, Untuk  $n = 50$   $r_{tabel} = 0,279$  (Dwi Priyatno.2009, h: 216). setelah dibandingkan didapat bahwa semua butir tes mempunyai koefisien lebih besar dari 0,279 ini berarti bahwa semua butir tes yang dikembangkan valid, dalam arti lain bahwa semua redaksi, pertanyaan, jawaban yang dikembangkan yang disusun dalam tes telah sesuai dan tepat untuk menilai setiap indikator C4, C5 dan C6 sehingga

intrumen tes yang dikembangkan individu mampu mengambil suatu keputusan terhadap pertanyaan yang dimaksud dalam tes. Hasil analisis reliabelitas seperti Tabel 2 didapat koefisien 0,642 – 0,882. Menurut Dwi Priyatno (2009, h: 172) bila koefisien reliabelitas kurang 0,6 maka tes kurang reliabel, bila koefisien 0,7 maka intrumen tes dapat diterima, dan bila koefisien reliabelitas lebih dari 0,8 maka intrumen tes mempunyai reliabelitas baik. Hasil analisis mengindikasikan bahwa intrumen tes yang dikembangkan telah mampu mengukur keajegkan jawaban peserta didik pada situasi yang berbeda. Hasil analisis validitas dan reliabelitas terhadap 60 butir terhadap sampel uji coba luas dilaksanakan terhadap 312 peserta didik didapat seperti pada Tabel 4.

Tabel 4. Hasil analisis validitas dan reliabelitas pada tahap uji coba (50 peserta didik) terhadap indikator yang dikembangkan pada setiap dimensi kemampuan berpikir tingkat tinggi

Dimensi Proses kognitif	Indikator kata kerja dalam dimensi Proses Kognisi (1)	Rata-rata Validitas (r) untuk 5 item tes	Reliabilitas (R)
C4 Analisis ( <i>analyze</i> )	Mengelompokkan	0,317	0,712
	Menjelaskan	0,623	0,734
	Membedakan	0,564	0,812
	Menemukan	0,639	0,798
C5 Penilaian ( <i>evaluate</i> )	Membandingkan	0,345	0,871
	Menafsirkan	0,612	0,828
	Menyimpulkan	0,246	0,657
	Membuat urutan	0,289	0,714
C6 Penciptaan ( <i>create</i> )	Menggabungkan	0,431	0,732
	Merencanakan	0,517	0,745
	Memformulasikan	0,631	0,734
	Merealisasikan	0,572	0,767

Koefisien validitas ( $r_{hitung}$ ) didapat 0,701- 0,840, Untuk  $n = 312$   $r_{tabel} = 0,113$  (Sugiyono.2007, h: 333). setelah dibandingkan didapat bahwa semua item butir tes mempunyai koefisien lebih besar dari 0,113 hal ini menunjukkan bahwa semua butir tes yang dikembangkan menunjukkan valid, dalam arti lain bahwa semua redaksi, pertanyaan, jawaban yang dikembangkan yang disusun dalam tes telah sesuai dan tepat untuk menilai setiap indikator C4, C5 dan C6 sehingga intrumen tes yang dikembangkan individu mampu mengambil suatu keputusan terhadap pertanyaan yang dimaksud dalam tes. Menurut Dwi Priyatno, (2009, h: 172) Kriteria koefisien uji reliabelitas: kurang 0,6 maka intrumen tes mempunyai reliabelitas kurang, bila koefisien reliabel, 0,7 maka intrumen tes mempunyai reliabelitas yang dapat diterima dan bila koefisien reliabelitas lebih dari 0,8 maka intrumen tes mempunyai reliabelitas baik. Hasil analisis reliabelitas seperti Tabel 4 didapat koefisien 0,657 – 0,812, ini menunjukkan bahwa butir-butir tes yang dikembangkan pada setiap indikator mempunyai reliabelitas yang baik, dalam arti lain bahwa intrumen tes yang dikembangkan mengindikasikan telah mampu mengukur keajegkan jawaban peserta didik pada situasi yang berbeda.

Hasil-hasil analisis reliabelitas intrumen tes untuk mengukur kemampuan siswa pada jenjang berpikir tingkat tinggi terhadap 60 butir yang dikembangkan baik pada tahap expert dan uji coba terbatas dengan sampel 132 dan 312 responden pada tahap uji coba luas didapat reliabelitas dalam kategori baik, ini menunjukkan bahwa tingkat keajegkan intrumen

tes yang dikembangkan dalam dimensi kemampuan berpikir tingkat tinggi (HOT) mampu memberikan jawaban yang ajeg ketika digunakan pada waktu dan siswa yang berbeda, kesimpulan ini sesuai dengan pendapat Azwar (2013: 176) yang menyebutkan bahwa suatu pengukuran dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama, diperoleh hasil pengukuran yang relatif sama, selama aspek yang diukur dalam diri subyek belum berubah, senada dengan hal tersebut Arifin (2011: 122) juga menyatakan bahwa suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda, kemudia Sudjana (2014: 16) juga menyatakan bahwa reliabilitas alat penilaian adalah ketepatan atau keajegan alat tersebut dalam menilai apa yang dinilainya. Artinya, kapanpun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama.

Selain analisis reliabelitas, juga dilakukan analisis validitas tes dalam kategori baik hal ini ditunjukkan dari rata-rata dari 5 item butir untuk setiap kemampuan berpikir tingkat tinggi (20 butir untuk setiap dimensi proses koqnitip) atau 60 butir item tes C4, C5 dan C6, setelah dibandingkan didapat koefisien validitas ( $r_{hitung}$ ) lebih besar dari koefisien ini menindikasikan bahwa semua butir tes yang dikembangkan valid, dalam arti lain bahwa semua redaksi, pertanyaan, jawaban yang dikembangkan yang disusun dalam tes telah sesuai dan tepat untuk menilai setiap indikator C4, C5 dan C6 sehingga instrumen tes yang dikembangkan mampu individu untuk mengambil suatu keputusan terhadap pertanyaan yang dimaksud dalam tes. Kesimpulan ini senada dengan pendapat Azwar (2013: 173) menyatakan bahwa suatu tes dikatakan memiliki validitas yang tinggi apabila alat tersebut menjalankan fungsi ukur secara tepat atau memberikan hasil ukur yang sesuai dengan maksud dilakukannya pengukuran tersebut. Artinya hasil ukur dari pengukuran tersebut merupakan besaran yang mencerminkan secara tepat fakta atau keadaan sesungguhnya dari apa yang diukur, senada dengan hal tersebut Suryabrata (2000: 41) juga menyatakan bahwa validitas tes pada dasarnya menunjuk kepada derajat fungsi pengukurannya suatu tes, atau derajat kecermatan ukurnya sesuatu tes, Sudjana (2004: 12) menyatakan bahwa validitas berkenaan dengan ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai. Untuk mengetahui koefisien validitas instrumen tes dapat digunakan teknik *corrected item total correlation* (korelasi produk momen) yaitu dengan membandingkan hasil korelasi  $r_{hitung}$  dibandingkan dengan  $r_{tabel}$ . Menurut Dwi Priyanto (2009, 172) bila harga  $r_{hitung}$  lebih besar dari  $r_{tabel}$  dan bernilai positif maka item objek tes yang dianalisis valid

## 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.
- Azwar, Saifudin. *Sikap Manusia Teror dan Pengukurannya*. Yogyakarta: Pustaka Pelajar. Badan.
- 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 September 2018].
- Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Class-room*. Alexandria: ASCD.
- BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP-Depdiknas.
- Kemendikbud. 2016. *Silabus Materi Fisika SMA*. Kemendikbud. Jakarta.
- Nitko, A.J & Brookhart, S. M. (2011). *Educational assessment of students. (6th ed)*. Boston: Pearson Education, Inc.
- Priyatno, Duwi. 2009. *5 Jam Belajar Olah Data dengan SPSS 17*. Andi Yogyakarta. Pusat Kurikulum. (2007). *Naskah Akademik Kajian Kebijakan Kurikulum Mata Pelajaran IPA*. Jakarta: Balitbang Depdiknas.
- Sudjana, Nana. (1990). *Penilaian Hasil Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Sudjana, Nana. 2004. *Penilaian Hasil Proses Belajar Mengajar*, Bandung: Remaja Rosdakarya.
- Suryabrata, S. (2002). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset
- 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 Agustus 2018 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
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions