

**PENGEMBANGAN *HIGHER ORDER THINKING MULTIPLE CHOICE TEST*
UNTUK MENGUKUR KETERAMPILAN BERPIKIR KRITIS IPA
KELAS VII SMP/MTs**

Hartini ¹⁾, Sukardjo ²⁾

SMP Negeri 1 Karangobar Banjarnegara Jawa Tengah ¹⁾, Universitas Negeri Yogyakarta ²⁾
hartiniarif@gmail.com ¹⁾, sukardjo.dikim@yahoo.com ²⁾

Abstrak

Penelitian ini bertujuan untuk: (1) mengembangkan *higher order thinking multiple choice test (HOT-MCT)* untuk mengukur keterampilan berpikir kritis IPA peserta didik kelas VII SMP/MTs, (2) mengetahui kualitas secara teoretis, dan (3) mengetahui kualitas secara empiris. Prosedur penelitian ini terdiri dari tujuh langkah dari 10 langkah model penelitian dan pengembangan Borg & Gall yang meliputi: (1) penelitian pendahuluan, (2) perencanaan, (3) pengembangan, (4) uji coba kelompok kecil, (5) revisi, (6) uji coba lapangan, dan (7) revisi. Uji kualitas secara teoretis dilakukan melalui validasi oleh ahli dan guru IPA. Uji kualitas secara empiris dilakukan melalui uji coba kelompok kecil dan uji coba lapangan. Analisis data menggunakan parameter teori klasik. Penelitian ini menghasilkan soal *HOT-MCT* dengan karakteristik: (1) merupakan soal pilihan ganda dengan empat pilihan jawaban yang terdiri dari 23 butir soal, (2) butir soal secara teoritis valid melalui proses perbaikan, serta (3) butir soal secara empiris valid dengan indeks reliabilitas $\alpha = 0,723$.

Kata Kunci: *higher order thinking multiple choice test*, keterampilan berpikir kritis, IPA SMP

**DEVELOPING A HIGHER ORDER THINKING MULTIPLE CHOICE TEST FOR
MEASURING THE CRITICAL THINKING SKILLS IN SCIENCE OF
GRADE VII OF JUNIOR HIGH SCHOOL**

Abstract

This study aims to: (1) develop higher order thinking multiple choice test (HOT-MCT) to measure the critical thinking skills of students of class VII IPA SMP/MTs, (2) determine the quality of the test theoretically, and (3) determine the quality of the test empirically. The procedure consists of seven-step, which adapted from 10 models of research and development steps Borg & Gall which includes: (1) the preliminary study, (2) planning, (3) development, (4) a small test group, (5) revision, (6) field trials, and (7) revision. The quality of the test theoretically validated by experts and science teachers. Empirically test the quality of trials conducted small groups and field trials. Analysis of the data using the parameters of the classical theory. This research resulted in about HOT-MCT with the following explanation: (1) a multiple choice question with four answer options consisting of 23 items was, (2) the test is valid theoretically valid, and (3) the test is valid empirically with reliability index $\alpha = 0.723$.

Keywords: *higher order thinking multiple choice test, critical thinking skill, science in the junior high school*

PENDAHULUAN

Standar Kompetensi Lulusan Satuan Pendidikan (SKL-SP) menyebutkan bahwa peserta didik SMP/MTs harus memiliki kompetensi kemampuan berpikir logis, kritis, kreatif, dan inovatif (Depdiknas, 2006b, p.342). *National Science Education Standards* menyebutkan bahwa keterampilan berpikir kritis diperlukan untuk melakukan inkuiri ilmiah (NRC, 1996, p.32), Keterampilan berpikir kritis merupakan tujuan penting pendidikan, tetapi keterampilan tersebut belum dikembangkan secara optimal. Kondisi ini dapat dilihat dari prestasi peserta didik berdasarkan pada penelitian tingkat internasional seperti penelitian oleh *Program for International Student Assessment (PISA)* dan *The Tird International Mathematics and Science Study (TIMSS)*. Hasil penelitian menunjukkan peserta didik Indonesia masih jauh di bawah negara anggota *Organization for Economic Co-operation and Development (OECD)*.

Balitbang Dikbud (2011) juga menyatakan bahwa rata-rata skor prestasi literasi IPA, posisi Indonesia masih jauh di bawah rata-rata internasional. Studi PISA tahun 2011 menunjukkan bahwa prestasi belajar IPA dengan rata-rata skor 393, sedangkan negara anggota *OECD* dengan rata-rata skor sebesar 501 (OECD, 2011). Hasil penelitian TIMSS menunjukkan bahwa prestasi IPA peserta didik di Indonesia pada tahun 1999, 2003, dan 2007 secara berturut-turut berada di peringkat ke-32, 37, dan 35 (Balitbang, 2011). Penelitian PISA dan TIMSS ini menggunakan instrumen soal yang memerlukan keterampilan berpikir kritis dalam menyelesaikannya. Rendahnya skor hasil penelitian PISA dan TIMSS menunjukkan rendahnya kemampuan berpikir kritis IPA peserta didik.

Proses pembelajaran merupakan siklus berulang meliputi perencanaan, pelaksanaan, dan penilaian. Proses pembelajaran IPA terlalu banyak memperhatikan pada konten materi. Penilaian pembelajaran IPA cenderung mengukur kemampuan berpikir tingkat rendah, sehingga peserta didik akan belajar menghafalkan materi. Jika kondisi ini berlangsung terus-menerus, maka akan membatasi ruang bagi berkembangnya keterampilan berpikir peserta didik. Sehubungan dengan itu diperlukan penilaian IPA yang tidak hanya terhadap satu aktivitas, seperti selesainya ujian akhir, namun berlanjut sebagai aktivitas simulasi proses pembelajaran (Liu, 2009, p.3). Sebaliknya proses perencanaan dan pelaksanaan pembelajaran berlanjut pada

aktivitas penilaian, karena data penilaian yang memiliki kualitas tinggi dapat digunakan untuk menjawab pertanyaan tentang proses pembelajaran (Liu, 2009, p.3).

Paradigma pembelajaran IPA yang berkembang saat ini yaitu, tujuan pembelajaran IPA tidak lagi hanya berbasis pada konten materi saja, tetapi harus disertai dengan dimensi proses IPA. Pembelajaran yang menekankan dimensi proses IPA peserta didik akan mengembangkan keterampilan proses IPA (*science process skill*) peserta didik. Keterampilan proses IPA ini berkembang beriringan dengan keterampilan berpikir. Peserta didik yang menggunakan keterampilan proses IPA, akan diikuti dengan perkembangan proses berpikir. Salah satu keterampilan berpikir yang perlu diinduksi, yaitu keterampilan berpikir kritis. Langkah ini sesuai dengan harapan Depdiknas (2006b, p.347) bahwa pembelajaran IPA yang baik, yaitu pembelajaran yang dilaksanakan untuk menumbuhkan keterampilan berpikir, bekerja, dan bersikap ilmiah. Tahapan ini menjadi tugas guru IPA yang harus *concern* dalam membantu peserta didik untuk menjadi pemikir yang kritis (Martin *et al.*, 2005, p.233).

IPA merupakan mata pelajaran dengan porsi terbesar hasil belajar pada aspek kognitif, sehingga penilaian hasil belajar aspek kognitif tetap memiliki peran penting dalam proses pembelajaran. Proses penilaian aspek kognitif dilakukan dengan teknik ujian dengan menggunakan instrumen soal. Instrumen soal yang digunakan oleh guru umumnya menggunakan butir soal yang ada di buku teks, lembar kegiatan peserta didik, atau kumpulan soal yang telah diberikan peserta didik. Kondisi ini membuat peserta didik cenderung hanya menggunakan ingatan dalam menyelesaikan butir soal. Hedges (1986, p.65) menyatakan bahwa penilaian dengan menggunakan soal dengan butir soal yang telah diketahui peserta didik akan cenderung menilai aspek kognitif pada dimensi mengingat.

Berdasarkan analisis dari studi pendahuluan yang dilaksanakan di Kabupaten Banjarnegara diperoleh informasi kebutuhan pendidik, yaitu: (1) instrumen penilaian yang dapat mengukur penguasaan pengetahuan (materi) dan sekaligus keterampilan berpikir kritis; (2) instrumen penilaian yang mudah diadministrasikan dalam waktu yang terbatas; serta (3) instrumen penilaian yang memenuhi standar pengukuran. Berdasarkan kebutuhan tersebut perlu dikembangkan instrumen soal pilihan ganda yang dapat mengukur penguasaan pengetahuan dan keterampilan

berpikir kritis. Pengembangan soal pilihan ganda ini menambah variasi butir soal yang dapat digunakan dalam penilaian, sehingga instrumen penilaian yang diperoleh dapat mengakomodasi keterampilan berpikir yang luas.

Berdasarkan penjelasan tersebut, permasalahan yang muncul yaitu instrumen soal terstandar belum banyak, instrumen pengukur keterampilan berpikir kritis menggunakan soal pilihan ganda belum banyak dikembangkan, guru mengajar dengan kelas yang besar, dan perlunya mengembangkan soal pilihan ganda yang berkualitas. Sehubungan dengan itu penelitian ini memiliki tujuan untuk mengembangkan *higher order thinking multiple choice test* untuk mengukur keterampilan berpikir kritis IPA peserta didik kelas VII SMP, serta menentukan kualitas secara teoretis dan empiris.

Keterampilan Berpikir Kritis

Keterampilan merupakan istilah yang mengacu pada kemampuan khusus yang diperoleh melalui pengalaman atau latihan untuk melakukan tugas dengan baik. Berpikir merupakan istilah yang umumnya digunakan untuk mencakup banyak aktivitas berfikir untuk merefeksi dan menganalisis. "*Thinking is a general term used to cover numerous activities, from day dreaming to reflection and analysis*" (Ruggiero, 2012, p.17). *Ministry of Education Malaysia* (2002, p.4) mendefinisikan berpikir merupakan proses mental individu yang memerlukan penggabungan pengetahuan, keterampilan, dan sikap dalam upaya untuk memahami lingkungan. Berpikir merupakan proses kognitif, proses mental yang dapat menghasilkan pengetahuan.

Keterampilan berpikir di kelompokkan menjadi dua kategori, yakni keterampilan berpikir kreatif dan keterampilan berpikir kritis (*Ministry of Education Malaysia*, 2002, p.4). Seseorang yang berpikir kreatif memiliki tingkatan imajinasi yang tinggi yang dapat menurunkan ide-ide asal dan inovatif serta memodifikasi ide, sedangkan seseorang yang berpikir kritis selalu mengevaluasi ide dengan cara sistematis sebelum menerimanya. Ide dievaluasi melalui aktivitas kognitif dengan cara berpikir analitis dan menggunakan proses mental seperti perhatian dan kategorisasi dalam pengambilan keputusan (Nitko & Brookhart, 2011, p.232). Selain berpikir analisis dan evaluatif dalam berpikir kritis juga melibatkan penggunaan keterampilan penalaran, deduktif, dan induktif (Moore & Stanley, 2010, p.10).

Berdasarkan uraian tersebut dapat disimpulkan bahwa keterampilan berpikir kritis merupakan kemampuan proses mental individu yang diperoleh melalui pengalaman, sehingga individu dapat membuat keputusan atau tindakan yang baik. Kemampuan yang dimaksud meliputi kemampuan analisis, evaluatif, dan penalaran yang digunakan secara sistematis. Keterampilan berpikir kritis membuat peserta didik mampu membuat keputusan atau tindakan terhadap permasalahan yang dihadapi. Keterampilan berpikir kritis ini membekali peserta didik dalam menghadapi setiap permasalahan yang dijumpainya dalam kehidupan sehari-hari.

Banyak ahli menyusun indikator berpikir kritis untuk memudahkan dalam mengimplementasikan di berbagai keperluan seperti bidang psikologi ataupun pendidikan. Lumsdaine & Lumsdaine (1995, p.253) menjelaskan karakteristik berpikir kritis di antaranya: (1) merupakan proses bukan hasil, yang meliputi pertanyaan berlanjut pada asumsi; (2) aktivitas yang produktif dan positif; (3) emosi yang terpicu dengan baik; serta (4) ingin tahu, fleksibel, jujur dan *sceptical*. Berpikir kritis tidak bertujuan untuk menemukan solusi, namun untuk mengkonstruksi sebuah gambaran logika pada situasi atau kondisi berdasar pada pendapat dan kejadian yang masuk akal, meskipun model kebenaran yang diperoleh tidak dapat diuji. Keterampilan berpikir kritis merupakan keterampilan yang penting bagi semua orang dalam kehidupan.

Karakteristik keterampilan berpikir kritis yang lebih operasional telah dikembangkan untuk kepentingan pada berbagai bidang keilmuan. Facione (1990, p.6) telah melakukan penelitian kualitatif untuk membuat karakteristik keterampilan berpikir kritis dalam bidang pembelajaran. Penelitian dilakukan berkolaborasi dengan 46 ahli untuk mencari kesepakatan mengenai karakteristik keterampilan berpikir kritis dan mengembangkan cara efektif untuk membelajarkan dan menilai keterampilan berpikir kritis.

Berdasarkan hasil penelitian ini ditetapkan bahwa pemikir kritis harus memiliki dua kategori keterampilan, yaitu keterampilan kognitif dan keterampilan afektif. Keterampilan berpikir kritis untuk aspek keterampilan afektif dibagi menjadi dua kategori, yaitu: (1) pendekatan untuk hidup maupun cara hidup, serta (2) pendekatan pada isu, pertanyaan, apresiasi atau masalah secara spesifik. Penelitian ini membatasi pada keterampilan kognitif.

Karakteristik berpikir kritis menurut Facione (1990, p.6) merupakan aspek berpikir kritis yang dibuat agar dapat digunakan secara umum untuk kepentingan proses pembelajaran, sehingga dapat diterapkan pada semua mata pelajaran. Penggunaan aspek berpikir kritis tersebut pada proses pembelajaran tertentu perlu disesuaikan dengan karakteristik proses pembelajaran. Apabila aspek berpikir kritis akan digunakan untuk kepentingan proses pembelajaran IPA, maka karakteristik berpikir kritis tersebut harus disesuaikan dengan hakikat IPA, yaitu pada dimensi proses IPA yang di dalamnya mencakup proses berpikir.

Keterampilan berpikir kritis untuk aspek (1) interpretasi, (2) analisis, (3) evaluasi, (4) menyimpulkan, dan (5) penjelasan, dapat langsung diimplementasikan dalam proses pembelajaran IPA. Aspek berpikir kritis tersebut merupakan sebagian proses berpikir yang menunjang terhadap keterampilan proses dasar IPA. Aspek *self regulation* dalam pembelajaran IPA masih terlalu umum, sehingga tidak terkait langsung dengan berpikir dalam proses pembelajaran IPA. Meskipun demikian aspek regulasi tetap dapat digunakan dalam proses pembelajaran IPA.

Ministry of Education Malaysia, (2002, p.13) membuat deskripsi keterampilan berpikir kritis dengan tujuan yang lebih spesifik untuk proses pembelajaran IPA setingkat sekolah menengah pertama. Keterampilan berpikir kritis ini disusun dalam kurikulum mata pelajaran IPA untuk sekolah setingkat SMP sebagai acuan berpikir kritis untuk kepentingan pada proses pembelajaran IPA.

Kesamaan aspek berfikir kritis antara *Ministry of Education Malaysia* (2002, p.13) dengan Facione (1990, p.6) terletak pada aspek: (1) menganalisis, (2) mengevaluasi, dan (3) menyimpulkan. Aspek penting pada Facione (1990, p.6) yang tidak terdapat pada *Ministry of Education Malaysia* (2002, p.13), yaitu: (1) interpretasi dan (2) penjelasan, yang keduanya merupakan bagian dari keterampilan proses dasar IPA.

Pada penelitian ini, aspek berpikir kritis yang akan digunakan sebagai indikator keterampilan berpikir kritis, yaitu dengan menggunakan aspek berpikir kritis kombinasi dari Facione dan *Ministry of Education of Malaysia*.

Sebanyak 11 aspek berpikir kritis digunakan sebagai indikator keterampilan berpikir kritis dalam penelitian, yang meliputi: menghubungkan, menginterpretasi, membandingkan dan membedakan, mengelompokkan dan mengklasi-

fikasi, mengurutkan, memprioritas, menganalisis, mendeteksi bias, mengevaluasi, membuat simpulan, dan menjelaskan.

Pengukuran Keterampilan Berpikir kritis

Pengukuran merupakan cara mengumpulkan informasi proses pembelajaran yang dapat dikuantifikasikan atau dinyatakan dengan angka (Sukardjo, 2012, p.5). Pengukuran menunjukkan pada proses dan hasil. Instrumen pengukur yang dipakai untuk mengumpulkan data dapat berbentuk soal maupun nontes. Pengukuran dengan teknik nonujian dapat menggunakan *check lists* dan *rating scales* (Nitko & Brookhart, 2011, p.239).

Check lists merupakan sebuah instrumen yang memuat tingkah laku dan keterampilan berpikir kritis yang dapat diamati dengan memberi daftar cek pada perilaku yang ditunjukkan. *Rating scales* merupakan alat untuk mencatat mengenai tingkat penguasaan keterampilan berpikir kritis peserta didik. Masing-masing cara memiliki kelebihan dalam mengukur keterampilan berpikir kritis, namun pada skala kelas dengan banyak subjek yang diukur akan membuat pengukuran menjadi tidak menyeluruh.

Selain dengan teknik non-ujian, keterampilan berpikir kritis dapat menggunakan teknik ujian dengan instrumen soal. Kneedler (1985, p.276) menjelaskan bahwa terdapat pendekatan tiga bentuk soal yang berbeda, yakni: soal objektif, uraian, dan kosakata keterampilan berpikir kritis untuk menilai berpikir kritis. Penggunaan tipe bentuk soal yang berbeda dapat saling melengkapi dan menutupi kelemahan masing-masing.

Penyusunan soal uraian lebih mudah dibanding dengan soal pilihan ganda. Sebaliknya penyusunan soal pilihan ganda untuk mengukur keterampilan berpikir kritis lebih sulit dan kompleks. Untuk itu penelitian ini berupaya memastikan bahwa butir soal pilihan ganda yang dikembangkan dapat digunakan untuk meningkatkan dan mengukur keterampilan berpikir kritis peserta didik.

Alasan pengembangan instrumen soal pilihan ganda untuk keterampilan berpikir kritis, yaitu: (a) bentuk soal pilihan ganda merupakan bentuk soal yang paling luas penggunaannya, sehingga guru yang mengajar dengan jumlah peserta didik yang banyak dapat menggunakannya; (b) soal pilihan ganda memiliki tingkat objektivitas dan validitas yang tinggi, sehingga dapat digunakan sebagai soal yang standar; serta (c) penyusunan butir soal pilihan ganda untuk

mengukur keterampilan berpikir kritis tidak mudah, memerlukan banyak waktu dan sumber referensi, sehingga diperlukan perhatian khusus untuk menyusunnya.

Pengembangan HOT-MCT

Bentuk soal yang banyak digunakan guru dalam mengevaluasi pembelajaran adalah soal objektif. Soal objektif tipe pilihan ganda (*multiple choice*) diakui sebagai soal yang paling luas digunakan dan diaplikasikan (Gronlund & Kinn, 1985, p.166). Butir soal pilihan ganda merupakan butir soal yang jawabannya diperoleh dengan memilih salah satu alternatif jawaban (Mardapi, 2004, p.14). Butir soal pilihan terdiri atas: pernyataan (*stem*), alternatif jawaban, dan pengecoh (Grunlund, 1998, p.53).

Anderson & Krathwohl (2001, p.31) merevisi dimensi proses kognitif pada taksonomi Bloom sebagai berikut: (1) mengingat (*remember*), (2) memahami (*understand*), (3) mengaplikasi (*apply*), (4) menganalisis (*analyze*), (5) mengevaluasi (*evaluate*), dan (6) mengkreasi (*create*). Dimensi proses kognitif tersebut disusun secara hirarki dari level berpikir rendah ke level berpikir tinggi. Mengingat merupakan level berpikir yang paling rendah, sedangkan mengkreasi merupakan level berpikir yang paling tinggi.

Moore & Standley (2010, p.10) menyebutkan bahwa berpikir tingkat tinggi terdapat pada tiga terakhir tingkat berpikir berdasarkan taksonomi Bloom, yaitu menganalisis, mengevaluasi, dan mengkreasi. Selain Moore & Stanley, Liu (2010, p.54) juga menyatakan hal yang sama. Pendapat lain menyatakan bahwa, berpikir tingkat tinggi terdapat pada empat terakhir tingkat berpikir (Hedges, 1987, p.64). Brookhart (2010, p.17) memiliki pendapat yang berbeda dengan menyebut berpikir tingkat rendah atau mengingat yang berlawanan dengan berpikir tingkat tinggi “...*lower-order thinking or recall versus higher-order thinking*. Pendapat tersebut juga dipertegas dengan penjelasan Nitko & Brookhart (2011, p.223) yang menyatakan:

A basic role for assessing of higher order thinking is to use task that require use of knowledge and skill in new or novel situations. If you only assess student's ability to recall what is in textbook or what you say, you will not know whether they understand or can apply the reasons, explanations, and

interpretations. In short, you must use novel material to assess higher order thinking.

Pernyataan tersebut mengandung maksud aturan utama dalam menilai berpikir tingkat tinggi, yaitu dengan menggunakan materi yang memerlukan penggunaan pengetahuan dan keterampilan dengan situasi atau ide baru. Butir soal yang hanya mengukur kemampuan mengingat yang ada di dalam buku teks dan yang disampaikan guru, tidak dapat mengetahui pemahaman atau aplikasi, penjelasan dan interpretasi.

Haladyna (1997, p.32) menjelaskan tentang karakteristik berpikir tingkat tinggi yang salah satunya, yaitu memahami (*understanding*), “*I characterized the higher forms of mental behavior as (1) understanding, (2) problem solving, (3) critical thinking, and (4) creativity*”. Berdasarkan penjelasan tersebut salah satu bentuk berpikir tingkat tinggi, yaitu *understanding*. Berdasarkan pernyataan Brookhart, Nitko & Brookhart, dan Haladyna tersebut menyamakan berpikir tingkat rendah dengan level berpikir mengingat, dan berpikir tingkat tinggi dimulai pada tingkat memahami, mengaplikasi, menganalisis, mengevaluasi, dan mengkreasi.

Penggunaan batasan tingkat tinggi dapat disesuaikan dengan tingkat sekolah dan karakteristik peserta didik. Berpikir tingkat tinggi untuk peserta didik tingkat dasar, menengah pertama, menengah atas, atau mahasiswa memiliki batasan yang berbeda.

Peserta didik sekolah dasar dan menengah pertama berpikir pada tingkat memahami yang telah dikategorikan sebagai berpikir tingkat tinggi, namun kemampuan itu dikategorikan sebagai berpikir tingkat rendah bagi peserta didik sekolah menengah maupun mahasiswa. Berpikir tingkat tinggi yang digunakan dalam penelitian, yaitu berpikir tingkat tinggi yang didasarkan pada pendapat Brookhart, Nitko & Brookhart, dan Haladyna, yaitu berpikir yang dimulai pada tingkat memahami. Argumen ini digunakan, karena produk yang dihasilkan diperuntukkan bagi peserta didik sekolah menengah pertama yang sebagian besar SK-KD IPA SMP menuntut tingkat berpikir memahami.

HOT MCT merupakan soal pilihan ganda berpikir tingkat tinggi. Kubiszyn & Borich (2003, p.112) menjelaskan tentang soal pilihan ganda tingkat tinggi sebagai berikut:

Good multiple choice items are the most time consuming kind of objectives test items to write. Unfortunately, most multiple choice

items are also written at the knowledge level of the Taxonomy of Educational Objectives. As a new item writer (and, if you're not careful, as an experienced item writer) you will have a tendency to write items at this level. In this section we will provide you with suggestions for writing multiple choice items to measure higher level thinking.

The first step is to write at least some objectives that measure comprehension, application, analysis, synthesis, or evaluation to ensure that your items will be higher than knowledge level if your items match your objectives!

Penjelasan tersebut menekankan bahwa soal pilihan ganda umumnya ditulis dengan tujuan pada level pengetahuan. Langkah pertama dalam menulis soal pilihan ganda untuk mengukur keterampilan tingkat tinggi adalah menentukan tujuan yang mengukur memahami, mengaplikasi, menganalisis, mengevaluasi, dan mengkreasi atau tingkat berpikir yang lebih tinggi dari mengetahui. Pendekatan yang disarankan untuk mengukur berpikir tingkat tinggi, yaitu: (1) menggunakan gambar, grafik tabel; dan (2) menggunakan situasi atau ide baru (Kubiszyn & Borich, 2003, p.113).

Nitko & Brookhart (2011, pp.223-224) menjelaskan cara yang sama seperti yang dikemukakan oleh Kubiszyn & Borich, yaitu butir soal berpikir tingkat tinggi harus menggunakan bahan atau materi baru. Nitko & Brookhart (2011, pp.223-224) menjelaskan cara tersebut dengan lebih rinci. Cara yang disarankan oleh Nitko & Brookhart (2011, pp.223-224) untuk membuat materi baru dalam membuat butir soal berpikir tingkat tinggi, yaitu dengan menggunakan *context dependent item sets* (perlengkapan butir yang bergantung pada konteks). *Context dependent item sets* terdiri atas *the introductory*

materials (materi pengantar) yang diikuti oleh beberapa item. Peserta didik harus berpikir mengenai informasi pada materi pengantar untuk menjawab pertanyaan untuk memecahkan masalah.

Context dependent item sets ini biasanya disebut dengan latihan menafsirkan. Materi pengantar untuk membuat butir soal berpikir tingkat tinggi di antaranya, yaitu: (a) mengutip dari materi bacaan, (b) gambar foto, (c) grafik, (d) gambar, (e) ayat, (f) puisi, formula, (g) tabel data, (h) daftar kata dan simbol, (i) contoh, (j) peta, (k) film, dan (l) suara rekaman. Penggunaan materi pengantar dapat membuat butir soal menjadi butir soal berpikir tingkat tinggi.

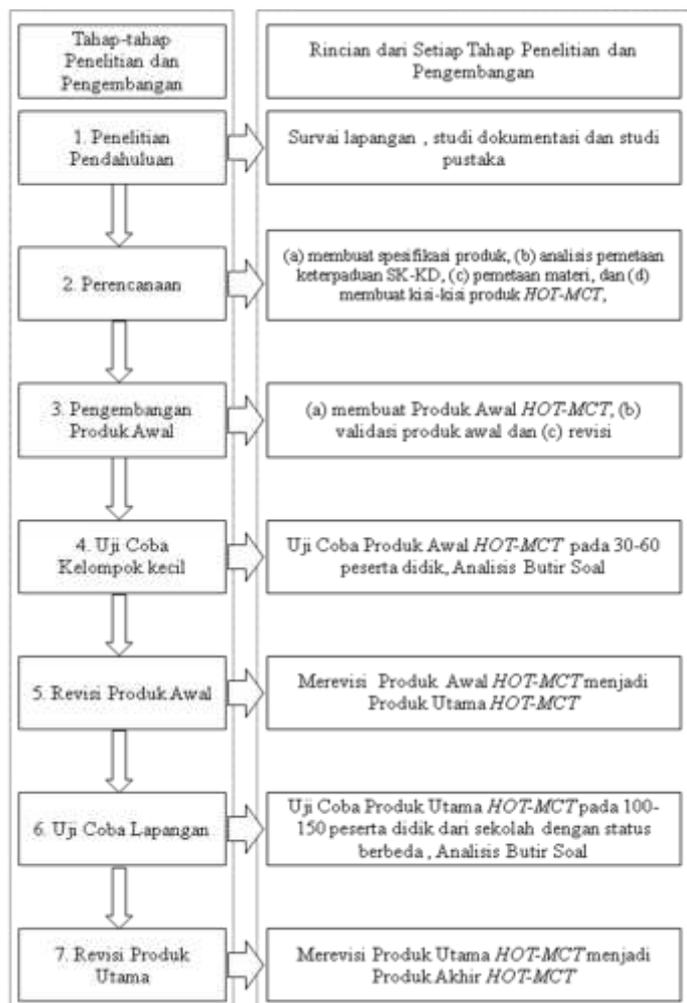
Pembuatan butir soal dengan materi pengantar merupakan langkah yang kompleks. Materi pengantar harus sesuai dengan fakta dan relevan dengan konsep materi pada butir soal tersebut.

METODE

Penelitian ini merupakan penelitian pengembangan (*research and development*, R & D) yang mengacu pada model Borg & Gall (1983). Penggunaan model ini karena relevan untuk mengembangkan instrumen penilaian yang sesuai dengan kebutuhan nyata di lapangan. Pengembangan soal ini juga memiliki keunggulan karena memiliki prosedur kerja yang detail.

Prosedur Pengembangan

Model pengembangan pada penelitian ini diadaptasi dari model pengembangan Borg & Gall (1983) yang terdiri atas 10 langkah. Penelitian ini mengambil 7 langkah dari model tersebut. Langkah pengembangan *HOT MCT* disajikan pada Gambar 2.

Gambar 2. Skema Prosedur Pengembangan *HOT-MCT*

Studi pendahuluan bertujuan untuk mengungkap permasalahan yang ada di lapangan terkait dengan instrumen pengukur keterampilan berpikir kritis. Studi pendahuluan dilakukan dengan cara: (a) survei lapangan dan (b) studi pustaka. Berdasarkan wawancara dan studi dokumentasi diperoleh beberapa informasi bahwa: (1) penilaian keterampilan berpikir kritis peserta didik belum dilakukan oleh pendidik; (2) hanya sebagian kecil butir soal yang menggunakan materi pengantar; serta (3) butir soal yang digunakan pada UAS, UKK, UH, dan UTS sebagian masih mengambil dari buku pegangan dan lembar kegiatan peserta didik.

Studi pustaka dilakukan pengkajian terhadap: (1) SK-KD Kurikulum 2006 kelas VII dan KD pada Kurikulum 2013 sebagai dasar untuk pemetaan keterpaduan, (2) buku referensi sebagai dasar untuk membuat indikator keterampilan berpikir kritis IPA SMP, serta (3) hasil penelitian dan buku referensi yang sesuai untuk pengembangan *HOT-MCT*.

Tahap perencanaan berupa persiapan membuat produk awal *HOT-MCT*. Perencanaan pengembangan produk awal *HOT-MCT* dilakukan melalui langkah: pembuatan produk awal, pemetaan keterpaduan SK-KD, pemetaan materi, serta penyusunan kisi-kisi produk awal *HOT-MCT*.

Kisi-kisi dibuat berdasarkan analisis indikator berpikir kritis, analisis keterpaduan SK-KD, dan pemetaan materi. Format kisi-kisi produk awal menggunakan format yang umum digunakan, hanya ada tambahan untuk indikator berpikir kritis IPA. Penyusunan kisi-kisi ini dibantu dengan matrik konten materi sebagai aspek pengetahuan dan indikator berpikir kritis sebagai aspek kognitif, agar indikator yang dibuat mencakup materi dan berpikir kritis yang proporsional. Kisi-kisi dibuat sebagai langkah awal agar butir soal *HOT-MCT* valid secara teori.

Pengembangan produk awal *HOT-MCT* dilakukan melalui kegiatan, yaitu: (1) penulisan butir soal produk awal, (2) validasi produk awal

dan *review* oleh guru mata pelajaran IPA, serta (3) revisi produk awal *HOT-MCT*.

Butir soal produk awal *HOT-MCT* ditulis berdasarkan kisi-kisi yang telah dibuat, selanjutnya dikembangkan menjadi butir soal. Butir soal ditulis menjadi paket soal produk awal *HOT-MCT*. Jumlah butir soal pada produk awal *HOT-MCT* berjumlah 40 yang dibuat berdasarkan 11 indikator keterampilan berpikir kritis yang menjadi tujuan pengukuran.

Validasi produk awal *HOT-MCT* dilakukan oleh ahli dan *review* oleh guru mata pelajaran IPA untuk mengetahui validitasnya secara teoretis. Validasi dan *review* dilakukan secara terpisah dalam kurun waktu yang sama.

Revisi produk awal *HOT-MCT* dilakukan berdasarkan hasil validasi butir soal dari validator digunakan sebagai dasar untuk memilih butir soal yang diterima, direvisi, dan tidak memenuhi kriteria. Semua masukan dari validator dianalisis dan hasilnya digunakan untuk merevisi butir soal. Rangkuman penilaian dan saran disusun sebagai dasar untuk membuat estimasi mengenai butir soal yang diterima, diterima dengan revisi, atau ditolak.

Uji coba dilakukan sebanyak dua kali, yakni uji coba kelompok kecil dan uji coba lapangan. Setiap akhir uji coba dilakukan revisi. Produk awal *HOT-MCT* diuji coba kelompok kecil. Uji coba dilakukan dengan melibatkan subjek uji coba sebanyak 32 peserta didik (1 kelas) yang dilakukan di SMP Negeri 1 Karangkoobar.

Uji coba kelompok kecil bertujuan untuk mengukur waktu yang perlukan peserta didik dalam mengerjakan soal dan mengetahui kualitas produk secara empiris melalui parameter teori klasik yang dianalisis dengan berbantuan program *MicroCat Iteman*.

Produk awal *HOT-MCT* direvisi berdasarkan karakteristik butir soal yang diketahui dari hasil analisis data dengan berbantuan program *MicroCat Iteman*. Revisi dilakukan pada butir soal dengan kategori diperbaiki, sedangkan butir soal dengan kategori diterima langsung dapat digunakan. Butir soal dengan kategori ditolak, tidak dapat digunakan. Butir soal yang diterima dan yang telah diperbaiki disusun kembali menjadi produk utama *HOT-MCT* yang akan diuji coba lapangan.

Uji coba lapangan terhadap produk utama *HOT-MCT* dilaksanakan di kelas VIII pada awal

semester 1 di tiga sekolah dengan tingkatan yang berbeda. Masing-masing sekolah melibatkan sebanyak 50-60 peserta didik. Penetapan peringkat sekolah didasarkan status sekolah, yaitu satu sekolah eks Rintisan Sekolah Bertaraf Internasional (RSBI), satu sekolah yang termasuk Sekolah Standar Nasional (SSN), serta satu sekolah termasuk Rintisan, yang secara berturut-turut adalah SMP 1 Banjarnegara, SMP Negeri 1 Karangkoobar, dan SMP Negeri 4 Kalibening.

Revisi produk utama *HOT-MCT* untuk menghasilkan produk akhir dilakukan pada penataan butir soal yang telah diketahui kualitasnya. Produk akhir *HOT-MCT* dilengkapi dengan kata pengantar, pendahuluan, analisis keterampilan berpikir kritis, kisi-kisi soal, lembar soal, lembar jawaban, dan kunci jawaban.

Subjek Coba

Subjek uji coba pada penelitian pengembangan ini, yaitu peserta didik yang telah mendapatkan materi kelas VII SMP di Banjarnegara. Subjek uji coba kelompok kecil dipilih pada peserta didik kelas VIII SMP 1 Karangkoobar dalam satu kelas yang terdiri dari 32 peserta didik. Uji coba dilakukan di awal semester 1.

Subjek uji coba lapangan digunakan peserta didik dari tiga sekolah yang berbeda dengan masing-masing subjek uji coba sebanyak 50 peserta didik.

Tabel 1. Subjek Uji Coba Lapangan

No	Nama Sekolah	Jumlah
1	SMP 1 Karangkoobar	50-60
2	SMP N 1 Banjarnegara	50-60
3	SMP N 3 Kalibening	50-60
Jumlah		150-180

Instrumen penelitian ini digunakan untuk menentukan validitas produk awal *HOT-MCT* secara teoretis dan empiris. Teknik dan instrumen pengumpulan data terangkum pada Tabel 2.

Cara yang digunakan untuk menentukan validitas isi, yaitu dengan membuat kisi-kisi soal sebelum penulisan butir soal. Selain penulisan kisi-kisi soal untuk mencapai validitas isi, produk awal *HOT-MCT* juga divalidasi oleh ahli dan di-*review* oleh guru mata pelajaran IPA SMP.

Proses validasi oleh ahli menggunakan teknik angket yang berisi penilaian, pilihan saran dan masukan.

Tabel 2. Teknik dan Instrumen Pengumpulan Data

Validitas	Teknik	Instrumen
Teoretis	Penyusunan kisi-kisi butir soal produk awal HOT-MCT	Lembar kisi-kisi soal produk awal HOT-MCT
	Angket Tertutup	Lembar <i>ceck-list</i>
Empiris	Angket Terbuka	Lembar angket terbuka
	Ujian	Produk awal Produk utama

Untuk melakukan *review* terhadap produk awal *HOT-MCT* digunakan angket terbuka. Instrumen untuk *reviewer* berbeda dengan validator, meskipun tujuan yang diperoleh sama. *Reviewer* diharapkan memberikan saran dan masukan yang lebih luas, sehingga data yang diperoleh lebih bervariasi.

Instrumen yang digunakan untuk validasi empiris berupa satu paket soal produk awal dan satu paket produk utama *HOT-MCT* beserta lembar jawaban. Instrumen digunakan untuk menentukan karakteristik butir soal secara empiris.

Jenis Data

Data kualitatif berasal dari dua sumber, yaitu: (1) hasil penilaian dan masukan dari ahli (validator), dan (2) saran dan masukan dari *reviewer*. Ahli yang menjadi validator, yaitu ahli evaluasi pada penilaian pendidikan IPA dan ahli materi IPA.

Data kuantitatif diperoleh dari hasil uji coba produk awal dan produk utama *HOT-MCT* yang diperoleh dalam bentuk skala nominal. Data diperoleh dari lembar respon peserta didik. Peserta didik yang menjawab benar untuk tiap butir soal mendapat skor 1, sedangkan yang salah mendapat skor 0.

Teknik Analisis Data

Analisis kualitatif dilakukan dengan cara memverifikasi hasil penilaian, saran, dan masukan dari validator dan *reviewer* yang digunakan sebagai dasar untuk merevisi produk awal *HOT-MCT*. Rangkuman hasil penilaian dan saran dianalisis untuk membuat keputusan terhadap perbaikan butir soal dari produk awal.

Saran dan masukan dari *reviewer* berupa revisi kesalahan materi, konstruksi soal, atau kebahasaan dijadikan acuan untuk melakukan telaah terhadap butir soal. Analisis dilakukan dengan memverifikasi masukan untuk setiap butir soal. Estimasi kelayakan butir soal menggunakan kriteria sesuai Tabel 3.

Tabel 3. Kriteria Butir Soal

No	Ket	Penilaian Butir Soal	Keputusan
1	Semua menyatakan valid tanpa revisi	Valid tanpa revisi	Butir soal diterima
2	Minimal salah satu menyatakan valid dengan perbaikan	Valid dengan perbaikan	Butir soal diperbaiki sesuai masukan validator dan reviewer
3	Minimal salah satu menyatakan tidak valid	Tidak valid	Butir soal diganti atau diperbaiki sesuai saran validator dan reviewer

Data hasil uji coba dianalisis dengan program *MicroCat Itean* untuk mengetahui karakteristik butir soal yang meliputi: tingkat kesukaran, daya pembeda, serta penyebaran pilihan jawaban butir soal. Analisis statistik menghasilkan informasi karakteristik mengenai perangkat soal, yaitu rata-rata, standar deviasi, tingkat kesukaran perangkat soal, validitas (daya pembeda) perangkat soal, indeks reliabilitas perangkat soal, dan standar kesalahan pengukuran.

Hasil analisis data hasil uji coba kelompok kecil dan uji coba lapangan digunakan untuk mengambil keputusan yang berbeda

terhadap butir soal. Karakteristik butir soal hasil uji coba kelompok kecil diprioritaskan untuk tujuan perbaikan butir soal produk awal, sedangkan karakteristik butir soal hasil uji coba lapangan digunakan untuk memilih butir soal yang diterima.

Karakteristik butir soal berdasarkan hasil analisis data dari uji coba kelompok kecil digunakan sebagai dasar untuk perbaikan produk awal menjadi produk utama *HOT-MCT*. Kriteria keputusan pemilihan butir soal produk awal *HOT-MCT* dari uji coba kelompok kecil, meliputi: (1) butir soal diterima, apabila butir soal

memenuhi semua kriteria yang ditetapkan. Butir soal yang terlalu mudah atau terlalu sukar dapat diterima apabila memiliki daya beda dan penyebaran pilihan jawaban yang memenuhi kriteria. Butir soal dalam kategori diterima, dapat digunakan secara langsung; (2) butir soal direvisi, apabila salah satu dari segi karakteristik butir soal yang disyaratkan tidak terpenuhi. Butir soal diperbaiki sesuai dengan karakteristik butir soal; (3) butir soal ditolak apabila tidak memenuhi semua karakteristik yang disyaratkan. Butir soal yang ditolak tidak dapat digunakan apabila jumlah butir soal yang diterima dan direvisi jumlahnya masih mencukupi indikator tujuan yang akan dicapai.

Butir soal yang diterima dan yang telah diperbaiki disusun kembali menjadi produk utama *HOT-MCT* yang diujicobakan di lapangan.

Hasil analisis data uji coba lapangan terhadap produk utama *HOT-MCT* digunakan untuk mengambil keputusan pemilihan butir soal yang diterima menjadi produk akhir dan mengetahui reliabilitas perangkat soal. Produk hasil dari tahap ini sebagai produk akhir *HOT-MCT*. Produk akhir *HOT-MCT* yang baik bila memiliki indeks reliabilitas minimal 0,70. Kriteria butir soal yang diterima sebagai produk akhir *HOT-MCT*, yaitu butir soal yang memiliki salah satu dari kriteria: (1) butir soal memenuhi semua kriteria yang dipersyaratkan, yaitu tingkat kesukaran baik, daya pembeda baik, dan semua pengecoh berfungsi baik; (2) butir soal yang terlalu mudah atau terlalu sukar dapat diterima apabila memiliki daya pembeda baik dan semua pengecoh berfungsi baik.

Butir soal yang diterima disusun kembali menjadi produk akhir *HOT-MCT* dan diverifikasi dengan indikator keterampilan berpikir kritis. Verifikasi bertujuan untuk memastikan 11 indikator tujuan keterampilan berpikir kritis terwakili. Butir soal yang diterima dianalisis kembali untuk menghitung reliabilitas produk akhir *HOT-MCT*.

HASIL PENELITIAN

Hasil Pengembangan

Produk awal *HOT-MCT* dibuat berdasarkan: (1) analisis indikator keterampilan berpikir kritis; (2) analisis terpaduan SK-KD dengan acuan keterpaduan KD Kurikulum 2013, dan KD Kurikulum 2006 yang digunakan sebagai dasar penulisan butir soal, yaitu KD pada semester 2 dan semester 1 yang berpotensi untuk dipa-

dukan; serta (3) kisi-kisi penulisan soal produk awal *HOT-MCT* untuk mengukur keterampilan berpikir kritis IPA peserta didik SMP. Kisi-kisi dibuat dengan tujuan agar butir soal yang dibuat memenuhi validitas materi, sehingga butir soal yang disusun mewakili keterampilan berpikir kritis dan materi untuk ke-las VII. Setiap butir soal terdiri atas pokok soal (stem) yang berisi materi pengantar sebagai *context dependent* dan 4 alternatif jawaban.

Produk awal *HOT-MCT* berupa seperangkat soal yang terdiri atas 40 butir soal. Produk awal *HOT-MCT* divalidasi oleh ahli dan *diview* oleh gur IPA SMP.

Hasil penilaian oleh validator dan *reviewer* dari masing-masing butir soal dirang-kum dan diestimasi untuk pengambilan keputusan butir soal yang baik atau diterima, maupun diperbaiki atau diganti. Estimasi butir soal produk awal *HOT-MCT* berdasarkan hasil validasi dan *review* terangkum pada Tabel 4.

Berdasarkan hasil validasi dan *review* dapat disimpulkan bahwa butir soal dengan kategori valid mengalami perbaikan yang didasarkan pada masukan dari ahli dan *reviewer*.

Tabel 4. Estimasi Butir Soal Produk Awal *HOT-MCT*

No	Kategori Penilaian	Jml	Estimasi
1	Valid tanpa revisi	4	Diterima
2	Valid dengan revisi	33	Diperbaiki
3	Tidak valid	3	Diperbaiki atau diganti

Hasil analisis menunjukkan bahwa sebanyak 1 butir soal yang harus diganti, 33 butir soal yang valid perlu diperbaiki, serta 4 butir soal yang valid tanpa revisi (langsung dapat digunakan). Perbaikan dilakukan berdasarkan masukan dan saran dari validator dan *reviewer*.

Hasil Uji Coba Produk

Uji coba produk dilakukan sebanyak dua kali, yaitu: (1) uji coba kelompok kecil; dan (2) uji coba lapangan. Tujuan uji coba dilakukan sebanyak dua kali agar soal yang dihasilkan merupakan perangkat soal *HOT-MCT* yang baik. Berdasarkan uji coba diperoleh data hasil uji coba kelompok kecil dan hasil uji coba lapangan.

Uji coba kelompok kecil merupakan langkah keempat dari penelitian pengembangan ini. Uji coba kelompok kecil dilaksanakan di SMP Negeri 1 Karangobar kelas VIIIB dengan sub-

jek uji coba sebanyak 32 peserta didik, tetapi satu peserta didik tidak masuk, sehingga lembar jawaban yang dianalisis sebanyak 31 buah.

Berdasarkan hasil uji coba kelompok kecil diperoleh informasi yang terkait dengan waktu yang diperlukan peserta didik untuk menyelesaikan soal selama kurang lebih 75 menit. Informasi karakteristik butir soal dan statistik soal diperoleh dari hasil analisis dengan menggunakan program *MicroCat Iteman*. Tingkat kesukaran butir soal produk awal *HOT-MCT* hasil uji coba kelompok kecil terangkum pada Tabel 5.

Tabel 5. Tingkat Kesukaran Butir Soal Produk Awal *HOT-MCT*

No	Kategori	Jumlah	Persentase
1.	$p > 7$ (Mudah)	4	10
2.	$0,3 \leq p \leq 0,7$ (Sedang)	18	45
3.	$p < 0,3$ (Sukar)	18	45

Berdasarkan Tabel 5 dapat diketahui bahwa produk awal *HOT-MCT* terdiri dari 3 butir soal (10%) dengan kategori mudah, 45% butir soal dengan kategori sedang, dan 45% butir soal dengan kategori sukar.

Daya pembeda butir soal diketahui dengan melihat koefisien *point biserial* (r_{pbis}). Karakteristik daya pembeda butir soal terangkum pada Tabel 6.

Tabel 6. Daya Pembeda Butir Soal Produk Awal *HOT-MCT*

No	Kategori	Jml	Persentase (%)
1.	$DP > 0,25$	22	55
2.	$0 < DP \leq 0,25$	15	37,5
3.	$DP \leq 0$	3	7,5

Berdasarkan Tabel 6 dapat diketahui bahwa butir soal dengan daya pembeda dengan kategori diterima sebanyak 22 butir (55%), dengan kategori diperbaiki sebanyak 15 butir (27,5%), sedangkan dengan kategori ditolak sebanyak 3 butir (7,5%).

Hasil analisis penyebaran jawaban (pengecoh) butir soal produk awal *HOT-MCT* berdasarkan data hasil uji coba lapangan seperti ditunjukkan pada Tabel 7.

Tabel 7. Penyebaran Jawaban Butir Soal Produk Awal *HOT-MCT*

No	Kategori	Jumlah	Pesentase
1	Pemilih $\geq 2,5\%$ dan r_{pbis} negatif	9	22,5
2	Pemilih $< 2,5\%$ dan r_{pbis} positif selain kunci jawaban	31	77,5

Tabel 7 menunjukkan bahwa sebagian besar butir soal pengecohnya ada yang tidak berfungsi atau justru mengecoh subjek uji coba kelompok atas. Butir soal dengan pengecoh yang kurang baik sebanyak 77,5% yang terdiri dari pengecoh tidak ada yang memilih, maupun pengecoh dengan r_{pbis} bernilai positif. Butir soal yang memiliki pengecoh dengan kriteria baik sebanyak 9 butir (22,5%).

Berdasarkan karakteristik butir soal yang ditunjukkan pada Tabel 5, Tabel 6, dan Tabel 7 dapat diestimasi jumlah butir soal yang diterima, diperbaiki, dan ditolak seperti ditunjukkan pada Tabel 8.

Tabel 8. Rekap Kategori Butir Soal

No	Kategori	Jumlah (Butir Soal)
1	Diterima	9
2	Diperbaiki	28
3	Ditolak	3

Tabel 8 menggambarkan jumlah butir soal dengan kategori diterima, serta langsung digunakan sebagai produk utama. Butir soal dengan kategori diperbaiki, mengalami perbaikan berdasarkan hasil analisis karakteristik butir soal sebelum digunakan sebagai produk utama. Butir soal dengan kategori ditolak, tidak digunakan sebagai produk utama, karena butir soal masih mewakili indikator keterampilan berpikir kritis.

Berdasarkan statistik soal diperoleh *mean biserial* sebesar 0,308 yang berarti daya beda perangkat soal ini dengan kategori baik. Nilai tersebut menunjukkan bahwa perangkat soal mampu membedakan peserta didik dengan keterampilan berpikir kritis tinggi dan rendah sebesar 30,8%. Skor terendah dicapai 10 dan tertinggi 27, rata-rata 15,774, median 15, simpangan baku sebesar 4,171, dan kemiringan distribusi data (skew) sebesar 0,501. Kemiringan positif menunjukkan sebagian besar skor peserta didik berada pada bagian bawah (skor rendah). Rerata tingkat kesukaran (*mean P*) perangkat soal sebesar 0,394 yang menunjukkan bahwa perangkat soal produk awal *HOT-MCT* dengan tingkat kesukaran dalam kategori sedang.

Uji coba lapangan terhadap produk utama *HOT-MCT* dilaksanakan di 3 SMP negeri dengan tiga status sekolah yang berbeda, yaitu eks RSBI, SSN, dan rintisan SSN yang secara berturut-turut adalah SMP Negeri 1 Banjarnegara, SMP Negeri 1 Karangobar, dan SMP 3 Negeri Kalibening. Jumlah peserta didik sebagai subjek coba yang digunakan pada uji coba lapangan disajikan pada Tabel 9.

Tabel 9. Subjek Uji Coba Lapangan

No	Sekolah	Kelas		Jumlah
		VIII	IX	
1.	SMP Negeri 1 Banjarnegara	49	-	49
2.	SMP Negeri 1 Karangkobar	29	30	59
3.	SMP Negeri 3 Kalibening	40	26	66
Jumlah		118	56	174

Berdasarkan 174 subjek uji coba, lembar jawab yang dapat dianalisis sebanyak 171, karena 3 peserta didik tidak mengikuti tes dikarenakan tidak masuk dan sakit. Hasil tes dari 171 peserta didik dianalisis untuk mengetahui karakteristik butir soal dan statistik perangkat soal produk utama *HOT-MCT*.

Karakteristik butir soal meliputi tingkat kesukaran, daya pembeda dan penyebaran jawaban. Tingkat kesukaran butir soal pada penelitian ini dikelompokkan menjadi tiga kelompok. Sebanyak 23 dari 37 butir soal yang diterima karena semua kriteria yang dipersyaratkan sebagai butir soal yang baik terpenuhi. Selain itu sebanyak 14 butir soal dengan kategori diperbaiki, karena ditemukan salah satu kriteria yang dipersyaratkan tidak terpenuhi, seperti penyebaran jawaban dengan pengecoh nilai biserialnya positif, dan nilai *prop.correct* butir soal kurang dari 0,025. Tingkat kesukaran butir soal untuk produk utama *HOT-MCT* hasil uji lapangan terangkum pada Tabel 10. Tabel 10 menggambarkan bahwa produk utama *HOT-MCT* terdiri dari 5 butir soal dengan kategori mudah, 21 butir soal dengan kategori sedang, serta 11 butir soal dengan kategori sukar.

Tabel 10. Tingkat Kesukaran Butir Soal Produk Utama *HOT-MCT*

No	Kategori	Jumlah	Persentase
1	$p > 7$ atau Mudah	5	14,3
2	$0,3 \leq p \leq 0,7$ atau Sedang	21	56,8
3	$p < 0,3$ atau Sukar	11	29,7

Daya pembeda butir soal diketahui dengan melihat koefisien *point biserial* (r_{pbis}). Karakteristik daya pembeda butir soal terangkum pada Tabel 11. Berdasarkan Tabel 11 dapat diketahui bahwa 81,1% butir soal dengan daya pembeda dengan kategori diterima, 16,2% dengan kategori diperbaiki, dan hanya 0,03% dengan daya pembeda yang ditolak.

Tabel 11. Daya Pembeda Butir Soal Produk Utama *HOT-MCT*

No	Kategori	Jml	Persentase
1.	$DP \geq 0,25$ atau diterima	30	81,1
2.	$0 < DP < 0,25$ atau diperbaiki	6	16,2
3.	$DP < 0$ atau ditolak	1	0,03

Hasil analisis penyebaran jawaban butir soal produk utama *HOT-MCT* seperti ditunjukkan pada Tabel 12 terdiri dari 62,2% butir soal dengan penyebaran yang baik atau memenuhi kriteria, sedangkan 38,2% butir soal dengan pengecoh yang kurang baik.

Tabel 12. Penyebaran Jawaban Butir Soal Produk Utama *HOT-MCT*

No	Kategori	Jml	Persentase
1.	Pemilih $\geq 0,025$ dan r_{bis} negatif	23	62,2
2.	Pemilih $< 0,025$ dan r_{bis} positif selain kunci jawaban	14	37,8

Statistik perangkat soal produk utama *HOT-MCT* seperti ditunjukkan pada Tabel 13.

Tabel 13. Statistik Soal Hasil Analisis *MicroCat* *Iteman* Uji Coba Lapangan

Scale Statistics	
Scale:	0
N of Items	37
N of Examinees	171
Mean	16.205
Variance	25.531
Std. Dev.	5.053
Skew	0.432
Kurtosis	-0.217
Minimum	5.000
Maximum	31.000
Median	16.000
Alpha	0.723
SEM	2.660
Mean P	0.438
Mean Item-Tot.	0.299
Mean Biserial	0.402

Berdasarkan hasil analisis, soal pada Tabel 13 menunjukkan *mean biserial* sebesar 0,402 yang berarti bahwa daya beda perangkat soal ini baik karena nilai *mean biserial* lebih besar dari 0,3. Nilai *mean biserial* menunjukkan bahwa perangkat soal mampu membedakan peserta didik dengan keterampilan berpikir kritis tinggi dan rendah sebesar 40,2%. Skor terendah sebesar 5, sedangkan skor tertinggi sebesar 31, dengan rata-rata 16,205, median 16, dan simpangan baku sebesar 5,053. Skew sebesar 0,432 menyebabkan kemiringan distribusi data sebesar

0,432. Kemiringan positif menunjukkan sebagian besar skor peserta didik berada pada bagian bawah (skor rendah). Rerata tingkat kesukaran (*mean P*) perangkat soal sebesar 0,438 yang berarti tingkat kesukaran produk utama *HOT-MCT* dalam kategori sedang. Nilai *alpha* menunjukkan reliabilitas perangkat soal produk *HOT-MCT*. Berdasarkan analisis nilai *alpha* (α) produk utama *HOT-MCT* sebesar 0,723 yang artinya reliabilitas perangkat soal dengan kategori baik, karena nilai *alpha* tersebut lebih besar dari 0,700.

Revisi Produk

Berdasarkan penilaian dari ahli materi dan ahli evaluasi, serta *review* dari guru mata pelajaran terdapat 4 butir soal yang valid tanpa revisi, 33 butir soal yang valid dengan revisi, serta 3 butir soal dinyatakan tidak valid. Butir soal yang valid tanpa revisi digunakan secara langsung, butir soal yang valid dengan direvisi dan tidak valid, diperbaiki berdasarkan masukan dan saran. Butir soal yang tidak valid dilakukan penggantian, yaitu satu butir soal pada nomor 13, sedangkan dua butir soal lainnya, yaitu nomor 14 dan 15 diperbaiki.

Revisi butir soal didasarkan pada masukan dan *review* yang berupa: (a) perbaikan pada stem di antaranya susunan kalimat, serta kelengkapan informasi bahan pengantar pada stem; (b) perbaikan pada pengecoh di antaranya pengecoh tidak logis dan tidak berfungsi; serta (c) perbaikan pada kunci jawaban, di antaranya kunci jawaban salah. Kunci jawaban yang salah terjadi pada butir nomor 21.

Perbaikan pada indikator, yaitu indikator tidak sesuai dengan butir soal. Ketidaksesuaian indikator dengan butir soal setelah dicermati dikarenakan butir soal tertukar, yaitu indikator untuk butir soal nomor 11 tertukar dengan butir soal nomor 10. Perbaikan terbanyak terdapat pada stem, yaitu dari segi kebahasaan, dan perbaikan pada pengecoh.

Butir soal diperbaiki berdasarkan saran dengan cara menelaah semua saran dan masukan pada butir soal yang diperbaiki. Selanjutnya 40 butir soal yang diperbaiki, ditata, dan diurutkan kembali hingga menjadi produk awal *HOT-MCT* yang memenuhi validitas isi.

Berdasarkan penilaian ahli dan *review* teman sejawat diperoleh hasil rata-rata butir soal pada kategori valid dengan perbaikan, sehingga dapat disimpulkan perangkat soal produk awal *HOT-MCT* dengan kategori valid melalui perbaikan berdasarkan masukan dan saran dari

validator dan *reviewer*. Produk awal *HOT-MCT* hasil pengembangan selanjutnya diuji coba ke peserta didik.

Berdasarkan hasil uji coba kelompok kecil menunjukkan sebanyak 9 butir soal diterima, 28 butir soal harus diperbaiki, dan 3 butir soal ditolak. Butir soal dengan kategori diperbaiki mengalami perbaikan berdasarkan hasil analisis karakteristik butir soal. Sebagian besar perbaikan dilakukan pada penyebaran pilihan jawaban. Perbaikan dilakukan pada pengecoh yang nilai biserialnya positif atau proporsi pemilih kurang dari 2,5% (0,025). Pengecoh dengan nilai proporsi pemilih kurang dari 2,5% menggambarkan pengecoh tidak berfungsi, sehingga harus diperbaiki. Untuk pengecoh dengan nilai biserial positif atau banyak pengecoh kelompok atas juga mengalami perbaikan.

Butir soal yang ditolak karena tidak memenuhi kriteria yang dipersyaratkan salah satunya, yaitu dikarenakan nilai *point biserial* butir tersebut kurang dari nol ($DP < 0$) atau negatif. Kondisi ini disebabkan butir soal tidak dapat membedakan antara peserta didik kelompok atas dengan kelompok bawah. Butir soal yang ditolak tidak digunakan dan tidak diganti dengan soal yang baru, karena butir soal yang ada masih cukup mewakili semua indikator keterampilan berpikir kritis. Butir soal yang ditolak, yaitu nomor 4, 23, dan 25.

Butir soal yang baik dan yang telah diperbaiki disusun kembali menjadi produk utama *HOT-MCT*. Produk utama *HOT-MCT* terdiri dari 37 butir soal yang selanjutnya diuji coba di lapangan.

Berdasarkan hasil analisis uji coba lapangan terhadap produk utama *HOT-MCT*, dari 37 dan 23 butir soal dapat dinyatakan diterima atau baik. Butir soal yang baik diverifikasi kembali dengan indikator keterampilan berpikir kritis untuk mengetahui semua indikator tujuan telah terwakili. Berdasarkan verifikasi seperti pada Tabel 14 menggambarkan bahwa butir soal yang baik telah mencakup semua indikator keterampilan berpikir kritis, meskipun distribusi soal untuk tiap indikator tidak merata. Distribusi butir soal yang baik ditunjukkan pada Tabel 14.

Perbaikan produk akhir *HOT-MCT* hanya dilakukan penataan pada butir soal nomor 23 dan pengemasan menjadi produk yang dapat dimanfaatkan langsung oleh guru IPA SMP kelas VII. Produk akhir *HOT-MCT* ini dilengkapi dengan kata pengantar, pendahuluan, analisis SK-KD, kisi-kisi soal, lembar soal, lembar

jawaban, dan kunci jawaban yang dijilid menjadi sebuah buku.

Kajian Produk Akhir

Validasi oleh ahli dan *review* oleh guru IPA menunjukkan produk akhir *HOT-MCT* telah memenuhi validitas isi, karena masukan yang diberikan menyangkut sisi materi IPA, evaluasi, dan praktisi.

Tabel 14. Distribusi Butir Soal yang Diterima

No	Indikator Keterampilan Berpikir Kritis	Nomor Butir Soal	Jumlah
1.	Menghubungkan	2	1
2.	Menginterpretasi	3, 5	2
3.	Membandingkan dan membedakan	6, 14, 16	3
4.	Mengelompokkan dan mengklasifikasi	18, 19	2
5.	Mengurutkan	11, 37	2
6.	Memprioritaskan	17, 21	2
7.	Menganalisis	22, 25, 27	3
8.	Mendeteksi kerancuan (bias)	23	1
9.	Mengevaluasi	36	1
10.	Membuat kesimpulan	31, 32, 33	3
11.	Menjelaskan	9, 28, 30	3
Jumlah			23

Produk akhir *HOT-MCT* merupakan perangkat soal yang telah valid secara teoretis dan empiris. Valid secara teoretis berdasarkan penilaian ahli dan guru, dan valid secara empiris didasarkan dengan parameter teori tes klasik. Butir soal produk akhir *HOT-MCT* dianalisis kembali untuk mendapatkan reliabilitas perangkat soal *HOT-MCT* dengan kategori baik ($\alpha = 0,723$).

Produk akhir *HOT-MCT* dapat digunakan untuk mengukur keterampilan berpikir kritis peserta didik IPA SMP/MTs yang telah teruji validitas dan reliabilitasnya. Selain untuk mengukur keterampilan berpikir kritis, produk akhir *HOT-MCT* dapat digunakan sebagai soal sumatif untuk mengetahui keterampilan berpikir kritis setelah peserta didik belajar IPA di kelas VII. Produk akhir *HOT-MCT* dapat digunakan sebagai instrumen penelitian yang melibatkan hasil belajar yang berupa keterampilan berpikir kritis peserta didik IPA kelas VII.

Keterbatasan Penelitian

Pengembangan perangkat soal *HOT-MCT* untuk mengukur keterampilan berpikir kritis peserta didik SMP/MTs masih mengandung keterbatasan. Keterbatasan penelitian meliputi perangkat soal *HOT-MCT* ini hanya diujicobakan sebanyak 2 kali. Ruang lingkup sampel pada uji coba lapangan di sekolah dengan peringkat yang berbeda yang masing-masing kelompok hanya diwakili oleh satu sekolah, serta setiap sekolah hanya menggunakan 2-3 kelas. Subjek

uji coba pada kelas VIII dan IX tidak seluruhnya siap untuk mengikuti ujian dengan materi kelas VII, dan banyak konten materi yang telah terlupakan. Sekolah yang digunakan untuk uji coba juga belum menerapkan pembelajaran IPA terpadu. Pada saat mengerjakan perangkat soal *HOT-MCT*, beberapa peserta didik mengerjakan soal *HOT-MCT* dengan tidak bersungguh-sungguh, atau saling bertanya dengan peserta didik lain. Kondisi ini mempengaruhi variasi data yang diperoleh.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan sebagai berikut: (1) telah berhasil dikembangkan *higher order thinking multiple choice test* untuk mengukur keterampilan berpikir kritis peserta didik IPA SMP/MTs kelas VII. Produk akhir *HOT-MCT* berupa seperangkat instrumen soal pilihan ganda dengan 4 pilihan jawaban dengan jumlah 23 butir soal; (2) berdasarkan validasi oleh ahli dan reviewer menggambarkan bahwa produk *HOT-MCT* dapat digunakan untuk mengukur keterampilan berpikir kritis peserta didik IPA kelas VII SMP/MTs berupa soal yang valid dengan perbaikan sesuai dengan masukan dari ahli dan reviewer; (3) kualitas secara empiris produk akhir *HOT-MCT* yang terdiri dari 23 butir soal untuk mengukur keterampilan berpikir kritis peserta didik IPA SMP kelas VII berdasarkan analisis teori klasik yang meliputi: (a) karak-

teristik butir soal, yaitu tingkat kesukaran butir soal, daya pembeda butir soal, dan penyebaran pilihan jawaban butir soal pada kategori diterima atau baik; dan (b) reliabilitas perangkat soal berkategori baik dengan indeks reliabilitas 0,723.

Saran

Produk akhir *HOT-MCT* untuk mengukur keterampilan berpikir kritis telah berhasil dikembangkan, sehingga dapat diajukan beberapa saran pemanfaatannya, yaitu: (1) guru IPA SMP dapat menerapkan produk akhir *HOT-MCT* untuk digunakan sebagai instrumen untuk mengukur penguasaan pengetahuan dan keterampilan berpikir kritis peserta didik, (2) produk akhir *HOT-MCT* ini diharapkan dapat digunakan sebagai contoh oleh guru IPA SMP untuk mengembangkan instrumen pengukur keterampilan berpikir kritis peserta didik, (3) produk akhir *HOT-MCT* untuk mengukur keterampilan berpikir kritis peserta didik belum didesiminasikan pada khalayak sekolah. Untuk itu perlu kiranya dilakukan desiminasi *HOT-MCT* di masa yang akan datang; (4) produk akhir *HOT-MCT* dapat digunakan untuk proses pembelajaran yang menggunakan Kurikulum 2013, karena mengacu pada KD Kurikulum 2013.

DAFTAR PUSTAKA

- Anderson, T.W. & Krathwohl. (2001). *A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives, abridged edition*. New York: Addison Wesley Longman, Inc.
- Balitbang. (2011). *Survai internasional TIMSS (versi elektronik)*. Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan.
- Balitbang. (2011). *Survey internasional PISA (versi elektronik)*. Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan
- Borg & Gall. (1983). *Educational research: an introduction (4th ed.)*. New York: Longman Inc.
- Brookhart S.M. (2010). *How to asses higher order thinking skills in your classroom*. Alexandria: ASCD.
- Facione, P. (1990). *The APA delphi report-critical thinking: A statement of expert consensus for the purposes of educational assessment and instruction*. California: The California Academic Press
- Haladyna, T.M. (1997). *Writing test items to evaluate higher order thinking*. Boston: Allyn and Bacon.
- Hedges, W. D. (1969). *Testing and evaluation for science in the secondary school*. California: Wadsworth Publishing Company, Inc.
- Kneedler, P. (1985). *California assesses critical thinking. Dalam A.L Costa (Eds.), Developing minds, a resource book for teaching thinking. (pp.276-280)*. Alexandria: Association for Supervision and Curriculum Development.
- Kubiszyn, T., & Borich, G. D. (2002). *Educational testing and measurement classroom application and practice (seventh edition.)*. New York: John Wiley & Sons, Inc.
- Liu, F. (2010). *Essentials of science classroom assessment*. Los Angeles: Sage Publications Ltd.
- Lumsdaine, E. & Lumsdaine, M. (1995). *Creative problem solving: thinking skill fo a changing world*. New York: McGraw-Hill, Inc.
- Martin, R., Sexton, C., Franklin, T., & Gerlovich, J., (2005: *Teaching science for all student inquiry methods for constructing understanding*. Boston: Pearson.
- Moore, B. & Standley. (2010). *Critical thinking and formative assessment increasing the rigor in your classroom*. New York: Eye On Education, Inc.
- Ministry of Education Malaysia, (2002). *Integrated curriculum for secondary schools curriculum specifications science form 2*.
- Nitko, A.J. & Brookhart, S.M. (2011). *Educational assessment of student 6th*. Boston: Pearson Education, Inc.
- NRC (1996). *National science education standards*. Washington: National Academy Press.
- OECD. (2011). *Against the odds disadvantaged students who succeed school, PISA*. OECD Publishing

Presiden Republik Indonesia. (2006b). *Peraturan Pemerintah RI Nomor 22 Tahun 2006 tentang Standar Isi*.

Sukardjo. (2012). *Buku pegangan kuliah evaluasi pembelajaran IPA untuk*

mahasiswa S2 program studi sains. (tidak diterbitkan) Yogyakarta: Universitas Negeri Yogyakarta.