

## **Pengembangan Instrumen Penilaian IPA untuk Memetakan *Critical Thinking* dan *Practical Skill* Peserta Didik SMP**

**Naomi Dias Laksita Dewi<sup>1</sup>\*, Zuhdan Kun Prasetyo<sup>2</sup>**

<sup>1,2</sup> Program Studi Pendidikan Sains, Program Pascasarjana, Universitas Negeri Yogyakarta. Jalan Colombo No. 1, Karangmalang, Yogyakarta, 55281, Indonesia

\* Korespondensi Penulis. Email: [naomidiyas@yahoo.com](mailto:naomidiyas@yahoo.com), Telp: +62274-550836

### **Abstrak**

Penelitian ini bertujuan untuk: (1) mengembangkan instrumen penilaian berupa tes pilihan ganda *critical thinking* dan *practical skill* yang valid dan reliabel (2) memetakan *critical thinking* dan *practical skill* peserta didik SMP. Prosedur pengembangan dalam penelitian ini mengacu pada langkah-langkah model prosedural dari Model 4D Thiagarajan. Hasil penelitian adalah sebagai berikut. (1) Instrumen penilaian IPA berupa tes pilihan ganda terdiri atas 36 item yang fit. Instrumen penilaian IPA berupa tes pilihan ganda telah memenuhi validitas isi dengan *expert judgement* yang dibuktikan dengan perolehan indeks kesepakatan (indeks Aiken) yang berada pada rentang angka 0,8 hingga 1,00. Kegiatan uji coba (uji empiris) dilakukan pada 121 peserta didik. Reliabilitas instrumen penilaian IPA mengalami peningkatan dari 0,95 menjadi 0,97. (2) Secara umum, terdapat 92 dari 220 didik atau 41,82% yang memiliki kemampuan *critical thinking* dan *practical skill* di atas rata-rata. Pada kemampuan *critical thinking* sebagian besar berada pada kategori rendah dan sangat rendah mencapai 49,55% sedangkan pada kemampuan *practical skill* sebagian besar berada pada kategori sedang 43,64%.

**Kata Kunci:** tes, *critical thinking* dan *practical skill*

## ***Developing Science Assessment Instrument to Map Critical Thinking and Practical Skill of Junior High School Students***

### **Abstract**

*The study aims to (1) to developing a valid and reliable assessment instrument in the form of multiple choice test of critical thinking and practical skill and (2) to mapping critical thinking and practical skill of Junior High School Students. This study was research and development (R & D) that refers to Thiagarajan 4D Model. The result of this study are (1) assessment instrument in the form of multiple choice test consist of 36 fit items. The assessment instrument in the form of multiple choice test is already comply with content validity that be proven with Aiken's index from 0.8 to 1.00. 121 students is be responden in small-scale tryout. Item reliability increases from 0.95 to 0.97. (2) Mapping result image shows that there are 92 out of 220 students (41.82%) who have critical thinking and practical skill in over average. Almost the Pilot Project Junior High School students have critical thinking ability in low and very low category that is 49.55% whereas they have practical skill ability in average category that is 43.64%.*

**Keywords:** test, *critical thinking* and *practical skill*

**How to Cite:** Dewi, N., & Prasetyo, Z. (2016). Pengembangan instrumen penilaian IPA untuk memetakan *critical thinking* dan *practical skill* peserta didik SMP. *Jurnal Inovasi Pendidikan IPA*, 2(2), 213-222. doi:<http://dx.doi.org/10.21831/jipi.v2i2.11963>

**Permalink/DOI:** <http://dx.doi.org/10.21831/jipi.v2i2.11963>

## PENDAHULUAN

Mutu pendidikan suatu bangsa menentukan kualitas bangsa tersebut. Untuk meningkatkan mutu pendidikan dapat diawali dari memperbaiki dan meningkatkan mutu pembelajaran. Untuk meningkatkan mutu pembelajaran dapat diawali dengan menyusun tujuan pembelajaran yang tepat. Tujuan pembelajaran merupakan salah satu komponen penting dalam pembelajaran.

Komponen pembelajaran terdiri atas tujuan pembelajaran, proses pembelajaran, dan evaluasi pembelajaran. Penjelasan Kementerian Pendidikan dan Kebudayaan bahwa ketiga hal ini merupakan kesatuan yang tidak dapat dipisahkan. Kegiatan belajar mengajar berorientasi kepada tujuan pembelajaran. Pencapaian tujuan pembelajaran akan terjawab apabila dilakukan evaluasi yang memperhatikan tujuan pembelajaran dan materi pembelajaran. Kegiatan evaluasi merupakan proses yang dilaksanakan evaluator pada suatu kejadian. Kegiatan ini menggunakan tes sebagai sarana menentukan nilai. Pada kegiatan pembelajaran, biasanya tes dilakukan untuk mengukur dan menilai hasil belajar peserta didik.

Untuk mengetahui kualitas peserta didik pada tingkat sekolah menengah pertamanya bidang IPA dapat kita lihat dari prestasi di tingkat internasional maupun nasional. Prestasi IPA di tingkat internasional dapat dilihat pada peringkat *Trends in International Mathematics and Science Study (TIMSS)* dan *Program for International Student Assessment (PISA)*. Menurut hasil TIMSS tahun 2011, kemampuan IPA Indonesia berada pada peringkat 40 dari 42 negara (Martin, Mullis, Foy & Staco, 2011, p.48) sedangkan hasil PISA pada tahun 2012 berada di peringkat 64 dari 65 pada tahun (Gurria, 2007, p.5). Untuk prestasi belajar IPA SMP di Indonesia khususnya di Daerah Istimewa Yogyakarta dapat dilihat dari hasil ujian nasional tahun 2011/2012 nilai IPA terendah 3,78 dan tertinggi 9,38 serta rata-rata 6,08, hasil UN tahun 2012/2013 nilai IPA terendah 3,27 dan tertinggi 8,74 serta rata-rata 5,89 dan hasil UN tahun 2013/2014 nilai IPA terendah 3,67 dan tertinggi 9,37 serta rata-rata 6,36. Berdasarkan hasil ujian nasional dari tahun 2011 hingga 2013, prestasi belajar peserta didik dirasa masih belum memuaskan. Berdasarkan hasil TIMSS, PISA dan Ujian Nasional 2011 hingga 2013, prestasi belajar IPA peserta didik masih rendah.

TIMSS dan PISA memiliki esensi mengukur domain pengetahuan, penerapan dan

penalaran sehingga dapat dikatakan kemampuan berpikir peserta didik Indonesia belum optimal dan dibutuhkan solusi untuk menangani permasalahan tersebut. Seperti yang dituliskan dalam *PISA Draft Science Framework* (2015, p.7), esensi dari penilaian PISA salah satunya adalah literasi sains. Komponen utama PISA untuk literasi sains antara lain (1) Kompetensi; menjelaskan fenomena sains, mengevaluasi dan merancang inkuiri sains, menginterpretasikan data dan bukti sains, (2) Pengetahuan; pengetahuan materi sains mengenai fisika, biologi dan ilmu kebumihan dan antariksa, pengetahuan prosedural, pengetahuan epistemik dan (3) Sikap; tertarik dalam sains, menghargai pendekatan saintifik dan kesadaran lingkungan. Menggaris bawahi pada komponen pengetahuan bahwa di dalam penilaian PISA peserta didik dapat memiliki kemampuan untuk menjelaskan konsep dan ide ilmiah (pengetahuan materi), langkah dan strategi yang digunakan dalam proses inkuiri sains (pengetahuan prosedural) dan cara untuk mendapatkan hasil yang benar dan pertanggungjawaban sains. Hal ini berarti bahwa penilaian PISA salah satunya mengukur kemampuan berpikir kritis peserta didik. Didukung Dani (2009, p.291) dalam jurnalnya disebutkan bahwa salah satu aspek dalam literasi sains adalah investigasi hakikat sains (*the investigative nature of science*) dan sains sebagai sebuah jalan untuk mengetahui (*science as a way of knowing*). Komponen di dalam investigasi hakikat sains (*the investigative nature of science*) antara lain menggunakan metode dan proses sains seperti observasi, mengukur, mengklasifikasi, inferensi, merekam dan menganalisis data, mengkomunikasikan data menggunakan berbagai cara seperti laporan, presentasi, membuat grafik dan tabel. Di sisi lain komponen sains sebagai sebuah jalan untuk mengetahui (*science as a way of knowing*) adalah fokus pada kemampuan berpikir, memberi alasan tepat dan merefleksikan hakikat sains dan kerja ilmiah. Dengan demikian penilaian PISA dapat dijadikan dasar penilaian *critical thinking* dan *practical skill* karena esensi di dalamnya.

Berdasarkan arahan Mendikbud, Kurikulum 2013 memiliki empat elemen perubahan antara lain Standar Kompetensi Lulusan, Standar Isi, Standar Proses, dan Standar Penilaian. Hal ini tentu saja membutuhkan instrumen penilaian yang tepat untuk mengukur ketercapaian peserta didik pada ketiga aspek tersebut. Standar proses menjelaskan bahwa dengan menerapkan Kurikulum 2013 salah satunya

diharapkan adanya peningkatan keseimbangan antara keterampilan fisik dan keterampilan mental. Selain itu, standar penilaian merupakan salah satu elemen perubahan yang menjadi perhatian. Standar penilaian hasil belajar peserta didik mencakup kompetensi sikap, pengetahuan, dan keterampilan yang dilakukan secara berimbang sehingga dapat digunakan untuk menentukan posisi relatif setiap peserta didik terhadap standar yang telah ditetapkan. Melihat kenyataan seperti itu menuntut ketiga kompetensi tersebut untuk dapat dinilai secara baik. Pada umumnya penilaian yang dilakukan masih menitikberatkan pada penilaian aspek pengetahuan. Hal ini dikarenakan, guru masih mengalami kesulitan dalam mengukur kompetensi peserta didik pada aspek keterampilan dan sikap.

Penilaian pengetahuan (*knowledge*) dilakukan menggunakan teknik tes. Penilaian sikap (*attitude*) dilakukan secara non tes sedangkan penilaian keterampilan (*skill*), dilakukan secara non tes atau unjuk kerja (*performance assessment*). Penilaian proses dan hasil belajar saling berkaitan. Penilaian dapat dilakukan secara tertulis maupun lisan. Penilaian secara tertulis dilakukan dengan tes tertulis seperti memilih jawaban (pilihan ganda, jawaban benar-salah, menjodohkan (*matching*) dan sebab-akibat) dan mengisi jawaban (isian singkat dan uraian).

Tes jenis pilihan ganda lebih banyak digunakan daripada jenis tes lainnya. Hal ini dikarenakan tes pilihan ganda memiliki kelebihan antara lain: (1) tes pilihan ganda merupakan salah satu tes yang dapat diterapkan secara luas untuk mengukur pencapaian, (2) tes ini dapat mengukur berbagai tipe pengetahuan dan hasil belajar kompleks secara efektif, (3) tes pilihan ganda bebas dari berbagai karakteristik singkat dari tes tipe lain, (4) keambiguan yang kerap terjadi pada tes jawaban singkat dapat dihindari disebabkan struktur pilihan yang berdasarkan pada situasi yang lebih baik, (5) tes pilihan ganda membantu dalam menentukan skor jawaban salah, (6) peserta didik harus mengetahui jawaban yang paling tepat untuk mendapatkan skor, (7) tes pilihan ganda memberikan reliabilitas yang lebih baik, (8) tes pilihan ganda melalui pelatihan menjodohkan menghindari keseragaman materi sebab tiap tes mengukur ide tunggal, dan (9) penerapan tes pilihan ganda beserta keunggulannya memudahkan dalam penyusunan tes tes berkualitas tinggi (Marmiyanah, 2011, p.4). Karena beberapa kelebihan tersebut maka tes jenis ini sering ditemui dalam proses pengukuran.

Sementara itu tes jenis unjuk kerja (*performance assessment*) yang idealnya digunakan untuk menilai keterampilan (*skill*), Reynolds (2009, p.267) mendaftar beberapa kelemahan dari *performance assessment*, yaitu: (1) penyeoran *performance assessment* dalam cara yang reliabel sulit dilakukan, (2) biasanya hanya memberikan sampling dari domain konten yang terbatas sehingga sulit membuat generalisasi atas keahlian dan pengetahuan peserta didik, (3) membutuhkan banyak waktu dan sulit disusun, diadministrasi dan diskor; dan (4) ada beberapa praktek yang membatasi penggunaan *performance assessment*. Menurut beberapa guru IPA SMP, tidak cukup waktu untuk menilai unjuk kerja setiap peserta didik dalam kegiatan praktikum. Oleh karena itu menjadi kendala tersendiri bagi guru sehingga terkadang penilaian unjuk kerja diabaikan dan dilakukan sebatas mengerjakan lembar kerja peserta didik.

Hasil pengamatan di sekolah menunjukkan kenyataan pada Lembar Kerja Peserta Didik yang biasa digunakan di sekolah masih banyak ditemui tes yang jawabannya berupa hafalan atau level C1 (pengetahuan/*knowledge*). Tes yang demikian tentu akan menghasilkan penilaian yang tidak utuh, membatasi peserta didik dengan aktivitas menghafal, kurangnya keterampilan berfikir dan tidak terasahnya kemampuan praktek IPA. Selama ini guru melakukan penilaian menggunakan tes yang mengukur aspek pengetahuan saja sementara aspek sikap dan keterampilan belum dijadikan tujuan yang dicapai peserta didik melalui pembelajaran IPA. Hal ini tentu saja karena terdapat beberapa kendala terhadap penilaian keterampilan. Kendala tersebut berkaitan dengan banyaknya jumlah peserta didik yang diampu dan kompetensi yang harus dicapai. Rata-rata guru IPA mengampu sebanyak 6 kelas dengan jumlah peserta bervariasi antara 30 sampai 34 peserta didik per kelas.

Temuan lainnya adalah penilaian keterampilan peserta didik juga membutuhkan waktu yang cukup lama, karena guru dituntut untuk menyelesaikan sejumlah kompetensi yang dirumuskan dalam SKL (Standar Kompetensi Lulusan). Ketika guru melakukan penilaian keterampilan, maka kompetensi yang harus diselesaikan dalam waktu satu semester tidak dapat dicapai tepat waktu. Sehubungan dengan hal itu, guru menilai peserta didik dengan cara perkiraan yang subjektif, dimana menilai baik keterampilan proses peserta didik yang dilihat aktif dalam melakukan praktek. Hal ini dapat disimpulkan

bahwa guru belum melakukan penilaian sesuai dengan fungsi dan sasaran pengukuran menggunakan instrumen yang valid dan reliabel. Selain itu, penilaian praktek peserta didik ketika melakukan praktikum masih sebatas mengisi lembar kerja atau petunjuk praktikum saja yang dikerjakan secara berkelompok. Hal demikian tentunya belum secara tepat dapat mengukur kemampuan tiap peserta didik khususnya dalam mengukur keterampilan praktek (*Practical Skill*).

Kemampuan berpikir kritis merupakan salah satu modal dasar atau modal intelektual yang sangat penting bagi setiap orang. Berpikir kritis merupakan kegiatan menganalisis ide atau gagasan ke arah yang lebih spesifik, membedakan secara tajam, memilih, mengidentifikasi, mengkaji dan mengembangkannya ke arah yang lebih sempurna. Proses mental ini menganalisis ide dan informasi yang diperoleh dari hasil pengamatan, pengalaman, akal sehat atau komunikasi.

Costa (1988, p.276) berpandangan bahwa terdapat tiga pendekatan yang dapat digunakan untuk menilai keterampilan berpikir kritis peserta didik.

*three different approaches will be used to assess students' critical thinking skills: objective question, a student essay, and CTS vocabulary. Although the major part of the assessment consists of objective question, the optional essay portion which will be score locally, is also integral. The third part of the tet, the CTS vocabulary, requires student to demonstrate their understanding of terms associated with critical thinking, such as generalization, hypothesis, inference and premise.*

Tiga pendekatan tersebut adalah pertanyaan obyektif, sebuah esai, dan kosa kata keterampilan berpikir kritis. Meskipun sebagian besar penilaian terdiri dari pernyataan objektif, bagian esai juga akan dinilai tidak terpisahkan. Bagian ketiga yaitu kosakata keterampilan berpikir kritis meminta peserta didik menunjukkan tentang pemahaman peserta didik terhadap keterampilan berpikir kritis yang terkait dengan pemikiran kritis, seperti generalisasi, hipotesis, inferensi dan premis.

Berpikir kritis menuntut upaya keras untuk memeriksa setiap keyakinan atau pengetahuan asumsi berdasarkan bukti pendukungnya dan kesimpulan lanjutan yang diakibatkan. Guru harus mengembangkan keterampilan berpikir

pada peserta didik dalam proses pembelajaran maupun penilaian menuju tingkat yang lebih tinggi. Peserta didik berpotensi besar untuk mengembangkan kemampuan berpikir sejak awal sehingga akan tumbuh menjadi seorang pemikir yang kritis pada suatu masalah atau keadaan pengambilan keputusan. Selanjutnya, *Ministry of Education of Malaysia* (2002, pp.5-6) mendeskripsikan kemampuan berpikir kritis sebagai berikut: (a) *Attributing*, mengidentifikasi kriteria seperti sifat, ciri, kualitas dan elemen dari suatu konsep atau sebuah objek, (b) *Comparing and Contrasting*, mencari kesamaan dan ketidaksamaan (perbedaan) berdasarkan kriteria seperti sifat, ciri, kualitas dan elemen dari suatu konsep atau sebuah objek, (c) *Grouping and Classifying*, memisahkan dan mengelompokkan objek atau fenomena dalam suatu kategori berdasarkan criteria yang pasti seperti sifat dan ciri yang umum, (d) *Sequencing*, menyusun objek dan informasi berdasarkan sifat dan ciri umum seperti ukuran, waktu, bentuk atau jumlah, (e) *Prioritizing*, menyusun objek dan informasi berdasarkan kepentingannya atau prioritasnya, (f) *Analyzing*, menguji informasi secara rinci dengan memerinci menjadi bagian yang lebih kecil atau sederhana untuk mencari makna sebenarnya dan hubungannya, (g) *Detecting Bias*, mengidentifikasi pendapat dan opini yang memiliki tendensi untuk mendukung atau menentang suatu hal dalam sebuah cara yang tidak wajar atau menyesatkan, (h) *Evaluating*, membuat keputusan berdasarkan fakta dan alasan yang valid, dan (i) *Making Conclusions*, membuat pernyataan tentang hasil investigasi yang berdasar hipotesis

Fahim & Masouleh (2012, p.1371) dalam jurnalnya menuliskan "*critical thinking refers to the use of cognitive skills or strategies that increase the probability of a desirable outcome. Critical thinking is considered to be as reasonable reflective thinking focused on deciding what to believe or to do.*" Berpikir kritis mengarah pada penggunaan keterampilan berpikir atau strategi berpikir yang meningkatkan probabilitas hasil yang diinginkan. Berpikir kritis dipertimbangkan sebagai cara berpikir yang reflektif beralasan yang fokus pada memutuskan apa yang dipercaya atau dilakukan.

Lebih lanjut dituliskan dalam Friedel (2008, p.76) "*critical thinking only measures the critical thinking skills constructs analysis, evaluation and inference as Facione found that measuring only these three constructs gives an accurate measure of critical thinking skill*

*abilities*". *Critical Thinking* hanya mengukur kemampuan berpikir konstruktif analisis, evaluasi dan inferensi seperti yang ditemukan Facione bahwa pengukuran hanya pada ketiga konstruk yang memberikan pengukuran akurat pada kemampuan *Critical Thinking*.

Dari beberapa pendapat tersebut, tampak adanya persamaan hal dalam sistematika berpikir yang berproses. Berpikir kritis merupakan proses menarik kesimpulan. Proses menilai atau mengevaluasi masih merupakan tahapan sebelum menarik kesimpulan karena yang dinilai adalah sejumlah alternatif sebagai dasar penyimpulan.

Berdasarkan beberapa definisi yang telah dikemukakan maka untuk mencapai hasil belajar peserta didik yang baik dibutuhkan kemampuan berpikir kritis yang tinggi dalam proses pembelajaran. Dalam hal ini kemampuan berpikir kritis peserta didik akan ditingkatkan dengan bantuan instrumen penilaian IPA berupa tes. Kemampuan berpikir kritis seorang peserta didik akan sangat membantu mengambil keputusan secara tepat, cermat, sistematis, benar dan logis, dengan mempertimbangkan berbagai sudut pandang atau aspek sehingga keputusan inilah yang dijadikan sebagai solusi dari suatu permasalahan. Kemampuan berpikir kritis (*Critical Thinking*) merupakan proses kemampuan berpikir untuk menghasilkan keputusan yang rasional dengan cara melaksanakan proses menggali, mengenali, dan menilai segala hal yang terkait untuk menjadi bahan pertimbangan dalam mengambil keputusan. Proses tersebut antara lain melibatkan kemampuan untuk menganalisis, mensintesis, mengenal dan memecahkan masalah, menyimpulkan dan mengevaluasi sehingga dapat digunakan sebagai hasil belajar IPA peserta didik.

Salim (2012, p.548) berpendapat keterampilan praktis peserta didik di laboratorium terkait dengan domain psikomotorik. Domain ini berfokus pada tugas manual yang memerlukan manipulasi benda-benda dan aktivitas fisik. Tabel 1 memaparkan level dan deskripsi dari masing-masing kemampuan psikomotor.

Istilah psikomotorik dikemukakan Arikunto (2013, p.135) berhubungan dengan kata "*motor, sensorymotor atau perceptual motor*". Jadi ranah psikomotor merupakan pengetahuan yang lebih banyak didasarkan dari pengembangan proses mental melalui aspek-aspek otot dan membentuk keterampilan peserta didik. Di dalam istilah psikomotorik terdapat dua hal yang berkaitan yakni keterampilan dan

kemampuan. Peserta didik yang mampu mengolah kemampuannya akan menjadikannya terampil. Di dalam pembelajaran IPA, keterampilan psikomotorik meliputi keterampilan yang ditemui ketika melakukan metode ilmiah. Istilah psikomotor dikhususkan sebagai *practical skills* dengan memperhatikan keselamatan kerja.

Tabel 1. Kategori dan Deskripsi Kemampuan Psikomotorik

No	Kategori	Deskripsi
1	<i>Recognition of tools and materials</i>	Kemampuan mengenali alat dan bahan
2	<i>Handling of tools and materials</i>	Kemampuan untuk menangani (memilih, menggerakkan dan meletakkan) alat dan bahan dan menangani benda tanpa merusak baik obyek atau benda lainnya dalam lingkungannya atau membahayakan untuk orang lain
3	<i>Basic operation of tools</i>	Kemampuan untuk melakukan kemampuan dasar seperti memegang alat tepat untuk digunakan, untuk mengatur alat dalam suatu rangkaian
4	<i>Competent operation of tools</i>	Mampu menggunakan alat kit atau alat rancangan untuk melakukan berbagai praktek sesuai dengan alat yang telah dirancang
5	<i>Expert operation of tools</i>	Kemampuan untuk menggunakan alat secara cepat, efisien, efektif dan aman dalam kegiatan praktek

Lock (1989, p.224) menyebutkan penilaian *practical skill* dapat dilihat dari kemampuan observasi (*observation*), manipulasi (*manipulation*), interpretasi (*interpretation*), merencanakan (*planning*), melaporkan (*report*) dan percaya diri (*self reliance*). Kerja laboratorium sangat potensial mengajak peserta didik dalam suatu penyelidikan yang otentik dimana peserta didik dapat mengidentifikasi masalah mereka sendiri untuk diselidiki, merancang prosedur, dan membuat kesimpulan. Pada umumnya kerja laboratorium dapat digunakan untuk mengenalkan peserta didik hasil belajar berikut ini: (1) sikap terhadap ilmu pengetahuan (*attitudes toward science*), (2) sikap ilmiah (*scientific attitudes*), (3) inkuiri ilmiah (*scientific inquiry*), (4) pengembangan konsep (*conceptual development*), (5) kemampuan praktek (*technical skills*) dan (6) kemampuan bekerja sama (*teamwork skills*) (Chiappetta & Koballa, 2010, p.213). Oleh karena itu, dengan melakukan percobaan

laboratorium, peserta didik belajar mempraktikkan keterampilan yang tidak bisa dipelajari secara teoritis.

Lebih lanjut Chijioke & Okoye (2012, p.782) dalam jurnalnya membagi kategori Practical Skill antara lain (1) Designing Skill, (2) Planning Skill, (3) Constructing Skill, (4) Assembling Skill, (5) Finishing Skill dan (6) Testing Skill. Practical skills memiliki kategori antara lain: keterampilan prosedural dan keterampilan manipulatif (*Procedural and Manipulative Skills*); keterampilan observasi (*Observational Skills*); keterampilan menggambar (*Drawing Skills*); keterampilan melaporkan dan menginterpretasi (*Reporting and Interpretative Skills*).

Berdasarkan pertimbangan-pertimbangan tersebut di atas, maka dikembangkan instrumen penilaian IPA SMP dengan harapan mampu memetakan kemampuan berpikir kritis (*Critical Thinking*) dan kemampuan praktek (*Practical Skill*) peserta didik SMP Pilot Project.

Terkait *critical thinking* dan *practical skill*, diperlukan suatu pengembangan tes berindikator *critical thinking* dan *practical skill* untuk memetakan *critical thinking* dan *practical skill* peserta didik SMP Pilot Project. Tujuan penelitian adalah mengetahui validitas instrumen penilaian IPA yang berupa tes pilihan ganda untuk memetakan *critical thinking* dan *practical skill*, mengetahui reliabilitas instrumen penilaian IPA yang berupa tes pilihan ganda untuk memetakan *critical thinking* dan *practical skill* dan mendapatkan hasil pemetaan *critical thinking* dan *practical skill* peserta didik SMP Pilot Project Kota Yogyakarta.

## METODE

Penelitian dilaksanakan dengan metode penelitian dan pengembangan (*Research and Development*) yang diarahkan untuk menghasilkan suatu produk. Penelitian uji coba terbatas (uji instrumen) dilaksanakan di SMP Negeri 5 Yogyakarta pada 13 hingga 18 Februari 2015. Pelaksanaan tes dalam rangka uji lapangan (pengukuran) berlangsung pada 2 hingga 20 Maret 2015.

Subjek penelitian untuk uji coba butir soal yang dikembangkan dalam penelitian ini adalah peserta didik SMP kelas VII Semester Genap Tahun Ajaran 2014/2015 sejumlah 121 peserta didik. Subjek penelitian uji lapangan sebanyak 220 peserta didik.

Prosedur pengembangan dalam penelitian mengacu pada langkah-langkah model prose-

dural yang diadaptasi dari Four-D (Thiagarajan, 1975, p.5) yang terdiri dari tahap pendefinisian (*Define*), tahap perencanaan (*Design*), tahap pengembangan (*Develop*), dan tahap penyebaran (*Disseminate*).

Data yang diperoleh dari penelitian ini merupakan data primer. Data yang diperoleh berupa hasil validasi ahli materi dan ahli evaluasi, data respon jawaban peserta didik dan wawancara guru IPA. Data hasil validasi ahli diperoleh melalui lembar validasi, data respon jawaban peserta didik diperoleh ketika peserta didik menjawab butir soal. Selain itu dilakukan wawancara menggunakan pedoman wawancara bagi guru IPA di sekolah yang dilakukan penelitian.

Analisis validitas tes dilakukan secara kualitatif oleh ahli materi dan evaluasi sedangkan analisis kuantitatif dilakukan perhitungan indeks Aiken. Azwar (2015, p.113) menyatakan Aiken telah merumuskan formula Aiken's V untuk menghitung koefisien validitas isi berdasarkan penilaian ahli terhadap item. Statistik Aiken's V dirumuskan sebagai:

$$V = \frac{\sum s}{[n(c-1)]}$$

keterangan:

s = r-lo

lo = angka penilaian validitas yang terendah (dalam hal ini =1)

c = angka penilaian validitas yang tertinggi (dalam hal ini =3)

r = angka yang diberikan oleh seorang penilai

Analisis kuantitatif butir soal yang dilakukan adalah sebagai berikut: (1) Uji Unidimensionalitas instrumen. Uji ini merupakan ukuran penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur. Apabila nilai unidimensionalitas lebih dari 20% maka persyaratan unidimensionalitas terpenuhi (Sumintono & Widhiarso, 2014, p.122). (2) Kecocokan butir soal menggunakan kriteria a) menggunakan nilai *Infit MNSQ* dari setiap butir soal, nilai rata-rata dijumlahkan dengan nilai deviasi standar kemudian dibandingkan. Nilai logit yang lebih besar dari nilai tersebut maka mengindikasikan butir soal tidak fit (misfit), b) nilai outfit nilai *Outfit Means Square (MNSQ)* yang diterima  $0.5 < MNSQ < 1.5$ . c) nilai *Outfit Z-Standard (ZSTD)* yang diterima  $-2.0 < ZSTD < +2.0$  dan d) nilai *Point Measure Correlation (Pt Mean Corr)* yaitu  $0.4 < Pt Measure Corr < 0.85$  (Sumintono & Widhiarso, 2014, p.115). (3) Reliabilitas

menggunakan fungsi informasi tes dan lima kriteria sebagai berikut: a)  $< 0,67$  berarti reliabilitas lemah, b)  $0,67 - 0,80$  berarti reliabilitas cukup, c)  $0,81 - 0,90$  berarti reliabilitas bagus, d)  $0,91 - 0,94$  berarti reliabilitas bagus sekali dan e)  $> 0,94$  berarti reliabilitas istimewa. (4) Peta butir dan individu, menunjukkan peta hubungan antara kemampuan individu dan jenis ujian yang sudah diurutkan dalam skala logit. Makin positif nilai logit pada *person* berarti abilitas makin tinggi. Makin positif nilai logit pada butir soal berarti butir soal memiliki tingkat kesulitan yang tinggi. (5) Deteksi bias butir soal dapat dianalisis dengan *Rasch* yang dapat ditampilkan dalam keberfungsian butir soal diferensial. Hal ini diperlukan untuk mengetahui apakah butir soal yang diberikan mempunyai bias dalam kategori responden tertentu atau tidak. (6) Menghitung presentase kemampuan peserta didik dalam tiap subaspek dapat dicari dengan cara sebagai berikut:

% tiap item dalam 1 subaspek

$$= \frac{\text{total responden yang menjawab benar}}{\text{total semua responden} \times \text{jumlah item dalam 1 subaspek}} \times 100\%$$

(7) Mengkategorikan presentase kemampuan *critical thinking* dan *practical skill* sebagai berikut:

Tabel 2. Kategori Presentase Kemampuan

Kategori	Presentase Rerata
Tinggi	$X > 66,7\%$
Sedang	$33,3\% > x \geq 66,7\%$
Rendah	$X > 0\%$

(Sukarno, Permanasari, & Hamidah, 2013, p.80)

## HASIL DAN PEMBAHASAN

Produk penelitian divalidasi ahli materi dan ahli evaluasi untuk penilaian kelayakan. Perhitungan indeks kesepakatan (indeks Aiken) diperoleh rentang angka 0,8 hingga 1,00. Hal ini diinterpretasikan bahwa seluruh item memiliki validitas isi yang baik dan mendukung validitas isi secara keseluruhan.

Berdasarkan Tabel 3 pada uji coba dan uji lapangan nilai infit rata-rata dan outfit rata-rata MNSQ mendekati 1,00. Menurut Sumintono & Widhiarso (2014, p.112), nilai infit rata-rata dan outfit rata-rata MNSQ yang makin mendekati 1,00 berarti makin baik. Hal ini berarti bahwa secara keseluruhan pola jawaban respon pada instrumen adalah bagus. Selain itu didukung nilai infit rata-rata dan outfit rata-rata ZSTD pada tahap uji coba dan uji lapangan yang mendekati 0,00. Lebih lanjut menurut Sumintono &

Widhiarso (2014, p.112) menjelaskan nilai rata-rata infit dan rata-rata outfit ZSTD makin mendekati 0,00 maka kualitas makin baik. Hal ini berarti keseluruhan instrumen adalah sangat bagus. Konsistensi jawaban responden cukup bagus. Hal ini dibuktikan dari nilai reliabilitas *person* yang mengalami peningkatan diuji lapangan. Reliabilitas butir soal juga mengalami peningkatan dari 0,95 menjadi 0,97.

Tabel 3. Rangkuman *Summary Statistic* pada Tahap Uji Coba dan Uji Lapangan

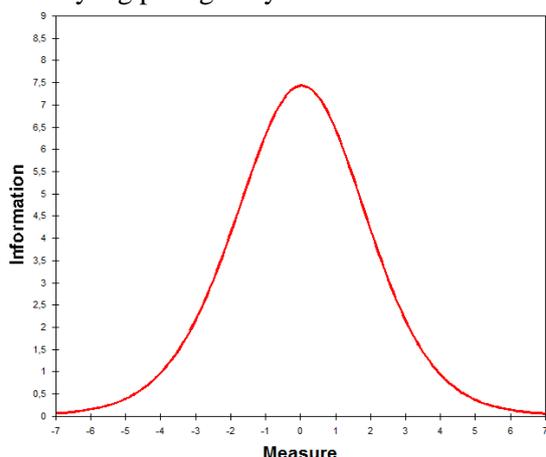
No.	Uraian	Uji Coba		Uji Lapangan	
		<i>Person</i>	Item	<i>Person</i>	Item
1.	Nilai rata-rata dan standar deviasi	0,94	0,00	0,02	0,00
		dan	dan	dan	dan
		0,72	1,04	0,72	1,01
2.	Reliabilitas	0,69	0,95	0,70	0,97
3.	Infit rata-rata MNSQ dan Outfit MNSQ	1,00	1,00	0,99	1,00
		dan	dan	dan	dan
		1,02	1,02	1,03	1,03
4.	Infit rata-rata ZSTD dan Outfit ZSTD	0,0 dan	-0,2	0,0 dan	-0,2
		0,0	dan -	0,1	dan
			0,1		0,0

Setelah dilakukan analisis pada item fit order menggunakan empat kriteria (*infit MNSQ*, *outfit MNSQ*, *Outfit ZSTD* dan *Point Measure Correlation*) ternyata terdapat 1 butir soal yang misfit yaitu butir soal 3 (1,17). Butir soal 5, 7, 8 dan 19 memang memiliki nilai infit MNSQ dan standar deviasi yang lebih besar. Namun, apabila dilihat nilai *Point Measure Correlation* hanya item 3 yang memiliki negatif. Hal ini berarti hanya butir soal nomor 3 yang misfit dan keempat item lainnya fit namun memiliki kecenderungan untuk tidak fit.

Atas uraian ini berarti terdapat 35 butir soal fit dengan model, hal ini disebabkan beberapa hal antara lain: (1) butir soal dikembangkan sesuai prosedur pengembangan dengan benar, (2) butir soal dikembangkan sesuai dengan indikator *Critical Thinking* atau *Practical Skill* dan materi IPA SMP, (3) tes telah melalui uji validitas isi oleh expert judgement dan (4) peserta didik yang melakukan tes mengerjakan tes dengan sungguh-sungguh karena guru dilibatkan untuk mengawasi.

Fungsi informasi yang menunjukkan reliabilitas pengukuran yang dilakukan ditunjukkan pada Gambar 1. Dari Gambar 1 terlihat bahwa kurva fungsi informasi menunjukkan item yang dikembangkan dan dikerjakan peserta didik tidak terlalu mudah dan tidak terlalu sulit. Fungsi informasi menunjukkan bahwa instrumen tes reliabel dan nilai ekstrim tidak terlalu

banyak. Data mendekati nilai rata-rata memiliki jumlah yang paling banyak.



Gambar 1. Fungsi Informasi Tes

Berdasarkan analisis deteksi bias, butir soal nomor 21, 22, 24, 26, 28, 29, 30, 34, 35 dan 36 yang merupakan item berindikator *Critical Thinking* dirasa lebih mudah dikerjakan oleh peserta didik yang berada di sekolah negeri. Butir soal nomor 2, 3, 5, 7, 10, 11 dan 19 yang hampir semuanya merupakan butir soal dengan indikator *Practical Skill* (kecuali item 7) dirasa lebih mudah dikerjakan oleh peserta didik yang berada di sekolah swasta. Didukung dengan peta distribusi respon butir dan responden memberikan informasi bahwa peserta didik sekolah negeri memiliki kemampuan yang merata. Hal ini dapat dilihat pada sebaran respon peserta didik yang cenderung merata mampu menjawab butir soal sulit, sedang maupun mudah (menyebar dari logit positif ke negatif) dengan nilai rata-rata logit *person* berada di tengah pola respon peserta didik.

Tabel 4. Rangkuman Tingkat Kesulitan Butir Soal pada Tahap Uji Coba dan Uji Lapangan

No	Tingkat Kesulitan	Uji Coba	Uji Lapangan
1	Tertinggi	1,85	2,18
2	Terendah	-2,06	-2,60
3	Rata-Rata	0	0
4	Standar Deviasi	1,04	1,01

Berdasarkan Tabel 4 tingkat kesulitan butir soal pada tahap uji coba dan uji lapangan hampir sama. Tingkat kesulitan uji coba dari -2,06 sampai dengan 1,85 sedangkan tahap pengukuran adalah antara -2,60 sampai dengan 2,18, dengan rata-rata 0 dan simpangan baku 1,04 (uji coba) dan 1,01 (uji lapangan).

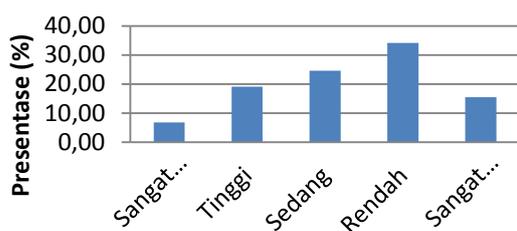
Apabila dilihat pada tiap butir terdapat 5 butir soal *Critical Thinking* yang pola tingkat

kesulitannya berubah yaitu pada butir soal sub-aspek membedakan, meramalkan, menghitung (1), menghitung (2) dan menyimpulkan (1). 13 butir soal yang lain memiliki pola tingkat kesulitan item yang tidak berubah pada tahap uji coba dan uji lapangan. Butir soal yang pola tingkat kesulitannya berubah dimungkinkan karena tiap sekolah memiliki kedalaman pembahasan materi IPA yang berbeda-beda atau guru sama sekali tidak menerangkan secara mendalam materi tersebut.

Pada kemampuan *Practical Skill* terdapat 2 butir soal *Practical Skill* yang pola tingkat kesulitannya berubah yaitu pada butir soal subaspek A “menyeleksi instrumen/piranti untuk melakukan eksperimen” (1) dan item C “memasang atau merangkai instrumen eksperimen secara tepat” (1). 16 butir soal yang lain memiliki pola tingkat kesulitan butir soal yang tidak berubah pada tahap uji coba dan uji lapangan. Butir soal yang pola tingkat kesulitannya berubah dimungkinkan karena di dalam pembelajaran tidak pernah diberi tahu bagaimana meletakkan termometer dengan benar dan dimungkinkan karena materi yang disampaikan tidak dikaitkan dalam kehidupan sehari-hari.

Kemampuan *Critical Thinking* dan *Practical Skill* pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 9,09%, 21,82%, 35,91%, 30,45% dan 5,00%. Keadaan ini belum memuaskan karena hanya terdapat 92 dari 220 peserta didik atau 41,82% yang berada di atas rata-rata. Dengan demikian peserta didik kelas VII SMP Pilot Project Kota Yogyakarta memiliki kemampuan *Critical Thinking* dan *Practical Skill* sebagian besar berada pada kategori sedang 39,55%, disusul dengan kategori rendah dan sangat rendah mencapai 35,45%, dan kategori tinggi dan sangat tinggi mencapai 30,91%.

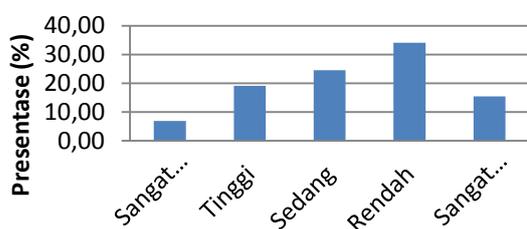
Apabila dilihat dari sisi kemampuan *Critical Thinking* pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 6,82%, 19,09%, 24,55%, 34,09% dan 15,45%. Keadaan ini berarti peserta didik kelas VII SMP Pilot Project Kota Yogyakarta memiliki kemampuan *Critical Thinking* sebagian besar berada pada kategori rendah dan sangat rendah mencapai 49,55%, disusul kategori tinggi dan sangat tinggi 25,91% dan kategori sedang 24,55%. Grafik kemampuan *Critical Thinking* pada tiap kategori disajikan pada Gambar 2.



Kategori *Critical Thinking* Peserta Didik

Gambar 2. Presentase Kategori Kemampuan *Critical Thinking* Peserta Didik

Apabila dilihat dari sisi kemampuan *Practical Skill* pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 9,09%, 16,36%, 43,64%, 23,18% dan 7,73%. Keadaan ini berarti peserta didik kelas VII SMP Pilot Project Kota Yogyakarta memiliki kemampuan *Practical Skill* sebagian besar berada pada kategori sedang 43,64%, disusul dengan kategori rendah dan sangat rendah mencapai 30,92%, dan kategori tinggi dan sangat tinggi mencapai 25,45% seperti yang ditunjukkan Gambar 3.



Kategori *Practical Skill* Peserta Didik

Gambar 3. Presentase Kategori Kemampuan *Practical Skill* Peserta Didik

## SIMPULAN DAN SARAN

### Simpulan

Instrumen penilaian IPA berupa tes pilihan ganda yang dikembangkan layak dan telah memenuhi persyaratan untuk memetakan *critical thinking* dan *practical skill* peserta didik dengan rincian sebagai berikut: (a) instrumen penilaian IPA berupa tes pilihan ganda terdiri atas 36 item yang fit. Masing-masing terdiri dari 18 butir soal *critical thinking* dan 18 butir soal *practical skill*, (b) instrumen penilaian IPA berupa tes pilihan ganda telah memenuhi validitas isi dengan expert judgement dan telah mendapatkan bukti empiris. Pemeriksaan oleh *expert judgement* sehingga dinyatakan layak dibuktikan dengan indeks kesepakatan (indeks Aiken) yang diperoleh berada di rentang angka 0,8 hingga

1,00. Bukti empiris telah dilakukan uji coba di SMP Negeri 5 Yogyakarta dengan jumlah responden 121 peserta didik dan (c) reliabilitas instrumen penilaian IPA berupa tes pilihan ganda telah memenuhi syarat, termasuk tinggi dengan koefisien reliabilitas lebih dari 0,90.

Hasil pemetaan kemampuan *critical thinking* dan *practical skill* peserta didik kelas VII SMP Pilot Project Kota Yogyakarta sebagai berikut: (a) secara umum terdapat 92 dari 220 didik atau 41,82% yang memiliki kemampuan *critical thinking* dan *practical skill* di atas rata-rata. Kemampuan *critical thinking* dan *practical skill* peserta didik kelas VII SMP Pilot Project Kota Yogyakarta pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 9,09%, 21,82%, 35,91%, 30,45% dan 5,00% dan (b) kemampuan *critical thinking* peserta didik SMP Pilot Project Kota Yogyakarta pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 6,82%, 19,09%, 24,55%, 34,09% dan 15,45%. Keadaan ini berarti peserta didik SMP Pilot Project Kota Yogyakarta memiliki kemampuan *critical thinking* sebagian besar berada pada kategori rendah dan sangat rendah mencapai 49,55%, disusul kategori tinggi dan sangat tinggi 25,91% dan kategori sedang 24,55%. Apabila dilihat dari sisi kemampuan *practical skill* pada kategori sangat tinggi, tinggi, sedang, rendah dan sangat rendah berturut-turut 9,09%, 16,36%, 43,64%, 23,18% dan 7,73%. Keadaan ini berarti peserta didik kelas VII SMP Pilot Project Kota Yogyakarta memiliki kemampuan *practical skill* sebagian besar berada pada kategori sedang 43,64%, disusul dengan kategori rendah dan sangat rendah mencapai 30,92%, dan kategori tinggi dan sangat tinggi mencapai 25,45%

### Saran

Berdasarkan kesimpulan penelitian dapat dikemukakan saran-saran yang perlu ditindaklanjuti sebagai berikut produk akhir tes *critical thinking* dan *practical skill* berupa soal pilihan ganda ini dapat dimanfaatkan guru IPA memetakan *critical thinking* dan *practical skill* peserta didik, produk akhir soal diharapkan dapat digunakan sebagai contoh oleh guru IPA SMP untuk mengembangkan instrumen pengukur dan *critical thinking* dan *practical skill* peserta didik, perlu dikembangkan sistem analisis secara komputerisasi dengan menggunakan lembar jawab komputer (LJK) yang siap dianalisis dengan Winstep 3.73 sehingga tidak menimbulkan masalah apabila akan dilakukan dalam

skala yang lebih besar. Penelitian ini masih dikerjakan secara manual. Berdasar hasil penelitian R & D, maka produk harus dilakukan uji coba II sebelum didesiminasikan. Mengingat kemampuan *critical thinking* dan *practical skill* peserta didik belum tinggi maka perlu adanya perbaikan pembelajaran yang dapat melatih kemampuan *critical thinking* dan *practical skill* IPA. Peningkatan kualitas pembelajaran diharapkan mampu menjadikan peserta didik untuk meningkatkan kemampuan *critical thinking* dan *practical skill*.

#### DAFTAR PUSTAKA

- Arikunto, S. (2013). *Dasar-dasar evaluasi pendidikan edisi 2*. Jakarta: PT Bumi Aksara
- Azwar, S. (2015). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Chijioko, P. O., & Okoye, R. K. (2012). Development and validation of instrument for assessing practical skills in building electronics systems in nigerian technical colleges. *Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) Scholarlink Research Institute Journals*, 779-785.
- Costa, A. L. (1988). *Developing minds a resource book for teaching thinking*. Virginia: Association for Supervision and Curriculum Development (ASCD).
- Dani, D. (2009). Scientific literacy and purpose for teaching science: A case study of lebanese private school teachers. *Internastional Journal of Environmental & Science Education Vol. 4, No. 3*, 289-299.
- Gurria, A. (2014). *PISA 2012 Results in focus: What 15-year-olds know and what they can do with what they can do. Programme for International Student Assessment*. OECD .
- Fahim, M., & Masouleh, S. N. (2012). Critical thinking in higher education: a pedagogical look. *Theory and Practice in Language Studies Vol.2 No.7 July 2012*, 1370-1375.
- Friedel, C., Irani, T., Rudd, R., Gallo, M., Eckhardt, E., & Ricketts, J. (2008). Overtly teaching critical thinking and inquiry-based learning: a comparison of two undergraduate biotechnology classes. *Journal of Agricultural Education Volume 49 Number 1*, 72-84.
- Kemendikbud. (2013). *Permendikbud No 65 Tahun 2013 tentang Standar Proses Pendidikan Dasar dan Menengah*.
- Lock, R. (1990). Assessment of practical skill part 1: The relationships between component skills. *Reaseacrh in Science and Technology Education Vol. 7 No.1.*, 221-233.
- Marmiyannah. (2011). *Menyusun soal tes objektif: bentuk pilihan ganda (constructing objective test items: multiple-choice form)*. Program Pascasarjana Universitas Sriwijaya Program Studi Pendidikan Bahasa Indonesia .
- Martin, O. M., Mullis, V. I., Foy, P., & Stanco, G. M. (2011). *Trends in International Mathematics and Science Study (TIMSS) 2011 International Result in Science*. Boston: TIMSS & PIRLS International Study Center
- Ministry of Education of Malaysia. (2002). *Integrate curriculum for secondary schools: curriculum specification science form 2*. Malaysia: Curriculum Development Centre Ministry of Education Malaysia.
- PISA. (2013). *PISA 2015 draft science framework*
- Reynolds, C. L. (2009). *Measurement and assessment in education, Second Edition*. New Jersey: Pearson.
- Salim, R., M., P., & S, M. D. (2012). Assessing students' practical skills in basic electronic laboratory based on psychomotor domain model. *International Conference on Teaching and Learning in Higher Education (ICTLHE 2012)*. Malaysia: Procedia - Social and Behavioral Sciences 56 ( 2012 ) 546 – 555.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial*. cimahi: Trim Komunikata Publishing House.