

PENGEMBANGAN TES KETERAMPILAN PROSES SAINS DASAR SD/MI

Esti Yuli Widayanti
STAIN Ponorogo

Abstract

The aim of this research is to develop test instrument validly and reliably for measuring the process skill of fundamental science of SD/MI upper class students. This research is research and development (R&D), that research method is used to result appropriate product and exam effectiveness. To reach that's aim, this research was fulfilled by two fases, those are instrument development process fase and psychometric analisis fase. The first fase insist of: identify objective test, specify test content or describe science process skill will be tested, construct instrument lattice include indicators of every skill and questions of each indicators, and make validation to expert judgement. The second fase is psychometric analisis by piloting test. The results of this research are: 1) it was developed the first test instrument by constructing 24 multiple choice questions validly after revising. 2) it was developed the second test instrument that is psychometric analisis by piloting test to get internal reliability and to know discrimination and difficulty grade. The received items are items those have difficulty grade, discriminant index, and answer distribution matching with these criteria. The result of this fase is that there are 10 item received (item number 1, 2, 7, 8, 9, 15, 17, 18, 22, and 23). 13 item was revised (item number 3, 4, 5, 6, 11, 12, 13, 14, 16, 19, 20, 21, 24), and 1 item was not used (item number 10). The reliability was 0.524, that was the medium reliability.

Keywords: Test Development, process Skill of Fundamental Science, Science SD/MI teaching

Abstrak

Tujuan dari penelitian ini adalah untuk mengembangkan instrumen tes yang valid dan reliabel untuk mengukur keterampilan proses sains dasar siswa kelas atas SD/MI. Penelitian ini merupakan penelitian jenis Research and Development (R & D), yaitu metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifannya. Untuk mencapai tujuan penelitian, penelitian dilakukan dalam dua fase, yaitu fase proses pengembangan instrumen dan fase analisis psikometrik.¹ Fase pertama meliputi: mengidentifikasi tes objektif; melakukan spesifikasi isi/konten dari tes atau mendeskripsikan keterampilan proses sains yang akan diteskan; menyusun kisi-kisi instrumen termasuk indikator dari masing-masing keterampilan dan butir soal per indikator; menulis item/butir soal sesuai dengan indikator; dan melakukan expert judgement (untuk meyakinkan face dan content validity). Fase kedua merupakan analisis psikometrik yang dilakukan dengan melakukan tes pilot. Hasil penelitian adalah (1). Dilakukan pengembangan perangkat tes tahap pertama dengan hasil disusun 24 butir soal pilihan ganda yang telah valid setelah dilakukan berbagai revisi sesuai saran ahli. (2) Dilakukan pengembangan perangkat tes tahap kedua yang merupakan analisis psikometrik yang dilakukan dengan melakukan tes pilot untuk menentukan reliabilitas internal sekaligus mengetahui tingkat kesulitan dan diskriminansi. Butir yang diterima adalah butir yang mempunyai tingkat kesulitan, daya beda, serta distribusi jawaban sesuai kriteria

¹ Cohen, Manion, Morrison, *Research Method in Education* (UK: Routledge, 2007)

penelitian ini. Hasil dari tahap ini adalah terdapat 10 butir soal diterima (butir nomor 1, 2, 7, 8, 9, 15, 17, 18, 22 dan 23), 13 butir soal diperbaiki/ direvisi (butir nomor 3, 4, 5, 6, 11, 12, 13, 14, 16, 19, 20, 21, 24), dan 1 soal dibuang / tidak digunakan (nomor 10). Tingkat reliabilitas butir soal tes adalah 0.524, merupakan tingkat reliabilitas dengan klasifikasi sedang.

Kata Kunci: *Pengembangan Tes, Keterampilan Proses Sains Dasar, Pembelajaran sains SD/MI*

PENDAHULUAN

Dalam pelaksanaan pembelajaran materi sains di SD/MI, peserta didik hendaknya dapat berperan sebagai seorang ilmuwan yang menemukan produk sains melalui serangkaian proses ilmiah dengan disertai sikap ilmiah. Melalui pembelajaran sains siswa hendaknya tidak hanya menguasai produk ilmiah dengan pemahaman konsep-konsep sains, tetapi juga menguasai proses ilmiah, dan sikap ilmiah.² Penguasaan proses ilmiah mengacu pada sejauh mana siswa mengalami perubahan dalam kemampuan proses keilmuwan yaitu sejumlah keterampilan untuk mengkaji fenomena alam dengan cara-cara tertentu untuk memperoleh ilmu. Dengan keterampilan proses, siswa dapat mempelajari sains sesuai dengan apa yang para ilmuwan lakukan. Sedangkan penguasaan sikap ilmiah merujuk pada sejauh mana siswa mengalami perubahan dalam sikap dan sistem nilai dalam proses keilmuwan.

Ketiga aspek penguasaan pembelajaran IPA diatas sesuai dengan cakupan ranah dalam kompetensi yang ada pada kurikulum 2013, yakni sikap, pengetahuan, dan keterampilan. Ketiga ranah kompetensi ini terdiri atas empat dimensi yang dijabarkan dalam empat kompetensi inti yang merepresentasikan sikap spiritual, sikap sosial, pengetahuan, dan keterampilan.³ Dalam kurikulum 2013, kompetensi ketiga ranah pembelajaran IPA tersebut telah dijabarkan dengan jelas pada kompetensi inti dan kompetensi dasar IPA kelas IV, V, dan VI.

Pada kurikulum sebelumnya, yaitu KTSP 2006, Pembelajaran IPA di sekolah dasar dengan pendekatan keterampilan proses sudah termaktup dalam standar isi-nya. Sehingga pada praktik pembelajarannya, para guru kelas pun sudah berupaya menggunakan strategi pembelajaran yang menggunakan pendekatan keterampilan proses. Namun, hal ini tidak diimbangi dengan penilaian yang dilakukan. Hasil pengamatan perangkat pembelajaran yang digunakan para guru kelas, meliputi silabus dan RPP, penilaian

² Patta Bundu, *Penilaian Keterampilan Proses dan Sikap Ilmiah dalam Pembelajaran Sains SD*, (Dirjen Dikti Depdiknas, 2006), 18-19.

³ Permendikbud No 64 Tahun 2013 tentang Standar Isi

pembelajaran IPA masih berkuat pada penilaian pengetahuan siswa tentang konsep (produk IPA), sebagian besar dalam bentuk tes objektif dan uraian. Guru belum melakukan penilaian secara terstruktur terhadap keterampilan proses sains siswa meskipun pembelajaran yang dilakukan sudah berbasis keterampilan proses. Demikian juga dari RPP yang dibuat oleh para calon guru MI di Prodi PGMI STAIN Ponorogo, instrumen penilaian yang mereka susun belum banyak yang menilai ranah keterampilan proses.⁴ Beberapa instrumen penilaian keterampilan proses yang berbentuk rubrik observasi maupun rubrik kinerja, pada praktiknya membingungkan penggunaannya karena selain belum familiar, pada praktiknya ketika proses penilaian guru harus melakukan pengamatan terhadap masing-masing individu siswa. Kesulitan yang teridentifikasi terletak pada ketidakfamiliaran guru dan calon guru dengan bentuk penilaian keterampilan proses serta kesulitan mengaplikasikan proses penilaian yang berbentuk rubrik pengamatan yang tidak semudah penilaian tertulis (essay atau pilihan ganda). Para praktisi memerlukan instrumen penilaian yang familiar dan mudah diaplikasikan, seperti bentuk soal essay atau pilihan ganda, sama seperti ketika melakukan penilaian untuk ranah pengetahuan.

Keterbatasan instrumen penilaian keterampilan proses di sekolah dasar (baik bentuknya maupun penggunaannya) tidak hanya terjadi pada proses penilaian berbasis kelas, tetapi juga pada penilaian keterampilan proses secara umum (bukan *classroom assessment*) dalam kategori tes diagnostik (yaitu tes untuk melihat kemampuan siswa secara umum pada kompetensi tertentu). Instrumen untuk mengetahui pemahaman siswa terhadap keterampilan proses secara umum belum banyak dikembangkan, apalagi dengan adanya perubahan kurikulum pada tahun 2013. Di Indonesia, beberapa instrumen penilaian keterampilan proses telah dikembangkan oleh Kartika Santi dkk yang mengembangkan keterampilan proses untuk pada mata pelajaran fisika tingkat SMA (Sekolah Menengah Atas). Penelitian ini menghasilkan instrumen penilaian keterampilan proses untuk 11 aspek keterampilan proses sains. Instrumen penilaian divalidasi oleh ahli untuk memberikan penilaian tiap butir yang terdapat di dalam instrumen penilaian sesuai dengan kriteria.⁵ Pada tingkat SMP, Titik Hidayati dkk mengembangkan tes diagnostik untuk mengidentifikasi keterampilan proses sains dengan tema energi pada pembelajaran IPA Terpadu.⁶ Kedua perangkat tes tersebut diperuntukkan bagi siswa sekolah menengah yang

⁴ Hasil pengamatan RPP IPA Kelas Atas di SD/MI kabupaten Ponorogo serta RPP yang disusun mahasiswa Prodi PGMI STAIN Ponorogo pada mata kuliah Pembelajaran IPA MI.

⁵ Kartika Santi, *Pengembangan Instrumen Penilaian Keterampilan Proses Sains pada Mata Pelajaran Fisika*

⁶ Titik Hidayati, Pengembangan Tes Diagnostik untuk mengidentifikasi Keterampilan Proses Sains dengan Tema Energi pada Pembelajaran IPA Terpadu, *Unnes Science Education Journal* 2 (2) 2013, 311-319.

sudah mengaplikasikan keterampilan proses lanjutan. Untuk mendiagnosis keterampilan proses sains dasar diperlukan perangkat tes yang berbeda.

Berdasarkan latar belakang tersebut, peneliti ingin mengembangkan tes keterampilan proses sains dasar yang valid dan reliabel berbentuk soal pilihan ganda untuk kelas atas SD/MI sesuai dengan keterampilan proses yang disyaratkan dalam kompetensi dasar pembelajaran IPA pada kurikulum 2013. Dengan disusunnya perangkat tes ini, guru kelas maupun calon guru dapat menggunakannya untuk tes diagnostik kemampuan siswanya. Tes yang dikembangkan juga dapat digunakan sebagai rujukan untuk menyusun penilaian berbasis kelas (tes formatif maupun sumatif) karena tes disusun berdasarkan kurikulum sekolah, yaitu kurikulum 2013. Jadi tujuan dari penelitian ini adalah untuk mengembangkan instrumen tes yang valid dan reliabel untuk mengukur keterampilan proses sains dasar siswa kelas atas SD/MI.

Dalam kurikulum 2006 disebutkan bahwa Pembelajaran IPA SD/MI sebaiknya dilaksanakan secara inkuiri ilmiah (*scientific inquiry*) untuk menumbuhkan kemampuan berpikir, bekerja dan bersikap ilmiah serta mengkomunikasikannya sebagai aspek penting kecakapan hidup. Oleh karena itu pembelajaran IPA di SD/MI menekankan pada pemberian pengalaman belajar secara langsung melalui penggunaan dan pengembangan keterampilan proses dan sikap ilmiah.

Untuk pembelajaran sains di sekolah dasar, Herlen menyarankan penguasaan hanya pada beberapa jenis keterampilan proses. Keterampilan proses pada tingkat dasar ini disebut keterampilan proses sains dasar. Lima jenis keterampilan yang disarankan Herlen adalah *observing (collecting data, measuring)*, *planning (raising questioning, predicting, devising enquiries)*, *hypothesizing (suggesting, explanation)*, *interpreting (considering evidence, evaluating)*, dan *communicating (presenting report, using secondary sources)*.

Sedangkan Rezba, mengklasifikasikan keterampilan proses sains dasar secara lebih sederhana menjadi 6 jenis keterampilan. Keterampilan ini adalah apa yang orang lakukan ketika mereka mengerjakan ‘sains’, yaitu: mengamati, mengklasifikasi, mengukur, menyimpulkan, memprediksi, dan mengkomunikasikan.⁷ Siswa menggunakan alat indra untuk *mengamati* objek dan peristiwa dan melihat pola dari hasil observasi. Mereka melakukan *klasifikasi* untuk membentuk konsep baru berdasarkan persamaan dan perbedaan. Siswa secara lisan maupun tertulis *mengkomunikasikan* apa yang mereka tahu dan dapat dilakukan. Mereka *mengukur* untuk mengkuantifikasi objek dan peristiwa. Siswa

⁷ Richard J. Rezba, et.al. *Learning and Assessing Science Process Skill*. (Iowa: Kendall/Hunt Publishing Co, 1995), 1.

menyimpulkan penjelasan dan mau merubah kesimpulan jika terdapat informasi baru. Siswa juga *memprediksi* kemungkinan-kemungkinan yang akan terjadi sebelum mereka benar-benar melakukan observasi.

Tabel 1
Ciri-ciri Aktivitas Keterampilan Proses Sains Dasar

Keterampilan Proses	Ciri Aktivitas
Observasi	Menggunakan alat indra sebanyak mungkin, Mengumpulkan fakta yang relevan dan memadai
Kuantifikasi	Observasi menggunakan alat ukur, membandingkan dengan menggunakan alat ukur yang sesuai.
Klasifikasi	Mencari perbedaan, mengontraskan, mencari kesamaan, membandingkan, mencari dasar penggolongan
Prediksi	Menggunakan pola, menghubungkan pola yang ada, memperkirakan peristiwa yang akan terjadi
Mengkomunikasikan	Membaca grafik, tabel atau diagram, menjelaskan hasil percobaan, mendiskusikan hasil percobaan, menyampaikan laporan secara sistematis
Inferensi	Menjelaskan hasil observasi, menyimpulkan berdasarkan fakta/bukti dari serangkaian observasi

Dalam kurikulum 2013, penguasaan keterampilan proses sains dijelaskan pada keterampilan dasar ilmu pengetahuan alam dari kompetensi inti ke-4 (kompetensi ketampilan). Kompetensi dasar untuk kelas IV meliputi: melakukan pengamatan, mmembuat laporan tertulis, membaut tabel atau grafik, melakukan dan menyajikan hasil percobaan, membuat karya/model. Kelas V meliputi: pengamatan rantai makanan, membuat bagan, klasifikasi manfaat bagian tumbuhan, merancang rangkaian listrik, membuat kompas dan electromagnet, memprediksi ganngguan keseimbangan alam, melaporkan tentang jenis penyakit organ tubuh. Keterampilan untuk kelas VI meliputi: melakukan percobaan tentang larutan, hantaran, membuat laporan dari percobaan hantaran, mengamati perkembangbiakan tumbuhan dan melaporkan, pengamatan adaptasi makhluk hidup dengan lingkungannya.

Bentuk penilaian sumatif berbeda dengan penilaian formatif dalam beberapa hal. Penialain yang dilakukan pada saat program tersebut berjalan yang ditujukan untuk memepbaiki proses pembelajaran agar tercapai tujuan yang diharapkan disebut penilaian

formatif.⁸ Penilaian formatif memerlukan *feedback* sehingga siswa dapat memperbaiki kekurangannya untuk mencapai tujuan yang diharapkan. Pada penilaian sumatif, penilaian dilakukan sebelum atau sesudah berlangsungnya suatu program dengan tujuan untuk menilai prestasi tertentu dari seseorang.

Contoh bentuk penilaian yang digunakan untuk penilaian sumatif dapat berupa pre dan post tes, tes diagnostik, tes penempatan, tes persiapan, dan tes untuk calon guru. Bentuk-bentuk tes tersebut merupakan tes tertulis berpadu dengan gambar. Untuk dapat menyusun tes sumatif untuk keterampilan proses sains dengan mudah, maka diperlukan penyusunan indikator untuk tiap keterampilan proses sains yang akan dinilai. Menurut Patta Bundu, dalam menyusun tes sumatif berbentuk pilihan ganda dapat menggunakan opsi-opsi pilihan dari proses ketika melakukan praktikum yang diungkapkan dalam bentuk gambar-gambar.⁹

Langkah-langkah dalam menyusun penilaian keterampilan proses secara umum adalah sebagai berikut: 1) menentukan jenis keterampilan proses yang akan dinilai, 2) menentukan indikator-indikator jenis keterampilan proses yang akan dinilai, 3) menentukan dan mengembangkan instrument penilaian yang akan digunakan, dan 4) validasi instrument (validasi ahli atau uji coba lapangan).¹⁰

Bentuk penilaian dapat berupa tes tertulis, tes perbuatan, pemberian tugas, proyek, penilaian sikap, dan penilaian portofolio. Sedangkan Mardapi mengklasifikasikan menjadi: tes lisan, bentuk benar salah, bentuk pilihan ganda, uraian objektif, uraian non-objektif, jawaban singkat, menjodohkan, unjuk kerja, dan portofolio.¹¹ Kemudian, dalam menilai keterampilan proses sains tersebut menurut Winkel dapat dilakukan melalui empat metode, yaitu: 1) membuat daftar pertanyaan (berbentuk pilihan ganda atau skala penilaian), 2) observasi, 3) wawancara, dan 4) laporan tertulis yang dibuat siswa pada akhir suatu program pengajaran.¹²

Untuk keterampilan proses sains, Dillasaw dan Okey menyatakan bahwa meskipun keterampilan proses menginginkan siswa untuk mendemonstrasikan kompetensi keterampilannya, menggunakan prosedur yang bersifat hands-on untuk mengakses akuisis skill pastinya akan sangat merepotkan. Karena itu, *format paper and pencil group testing*

⁸ John A. Glover dan Roger H. Bruning.. *Educational Psychology: Principles and Application*. New York: Harper Collins Publisher, 1990

⁹ Patta Bundu, *Penilaian Keterampilan Proses dan Sikap Ilmiah dalam Pembelajaran Sains SD*. Dirjen Dikti Depdiknas, 2006, 75

¹⁰ Ibid...., 63

¹¹ Djemari Mardapi, *Teknik Penyusunan Instrumen tes dan Nontes* (Yogyakarta: Mitra Cendekia, 2008), 69-85.

¹² W.S. Winkel, *Psikologi Pengajaran*. (Jakarta: Grasindo, 1996),

adalah bentuk yang mudah dan nyaman untuk digunakan dalam meng-ases kompetensi keterampilan proses sains pada jumlah siswa yang besar.¹³

Tes bentuk pilihan ganda adalah tes yang jawabannya dapat diperoleh dengan memilih alternatif jawaban yang telah disediakan. Pedoman utama dalam pembuatan butir soal bentuk pilihan ganda adalah: 1) pokok soal harus jelas, 2) pilihan jawaban homogen dalam arti isi, 3) panjang kalimat pilihan jawaban relatif sama, 4) tidak ada petunjuk jawaban benar, 5) hindari penggunaan pilihan jawab semua benar atau semua salah, 6) pilihan jawaban angka diurutkan, 7) semua pilihan jawaban logis, 8) jangan menggunakan negating ganda, 9) kalimat yang digunakan sesuai dengan tingkat perkembangan peserta tes, 10) bahasa Indonesia yang digunakan sesuai dengan tingkat perkembangan peserta tes, dan 11) letak pilihan jawaban benar ditentukan secara acak.¹⁴

Sesuai *Standar*, 1999, validitas merupakan dukungan bukti dan teori terhadap penafsiran skor tes sesuai dengan tujuan penggunaan tes. Oleh karena itu validitas merupakan fundamen paling dasar dalam mengembangkan dan mengevaluasi suatu tes.¹⁵ Proses validasi yang dilakukan bertujuan melakukan validasi terhadap interpretasi data yang diperoleh melalui prosedur tertentu. Ketepatan interpretasi suatu hasil tes berdasarkan bukti: bukti yang mendukung, yang bersumber dari: bukti berdasarkan isi, berdasarkan proses respon, berdasarkan struktur internal, dan hubungan terhadap variabel lain.

Karakteristik analisis butir soal secara kalsik meliputi analisis tingkat kesukaran, daya beda, dan distribusi jawaban. Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks, berupa proporsi yang besarnya berkisar 0.00 – 1.00. Semakin besar indeks kesukaran berarti semakin mudah soal itu. Jika suatu soal memiliki tingkat kesukaran (TK) 0.00 berarti bahwa tidak ada siswa yang menjawab benar, atau 0% siswa menjawab benar. TK = 1.00, berarti 100% (semua) siswa menjawab benar. Klasifikasi tingkat kesukaran dapat dibagi menjadi tiga kelompok, yaitu mudah, sedang, dan sukar. Sebagai patokan, dapat digunakan klasifikasi tingkat kesukaran sebagai berikut: kategori ‘sukar’ dengan nilai $p = 0.00-0.25$, kategori sedang dengan nilai $p = 0.26-1.00$, dan kategori mudah dengan nilai $p = 0.76-1.00$.

¹³ Kazeni Mungandi Monde Monika, *Development and Validation of a test of Integrated Science Process Skill for the Further Education and Training Learners*, (Disertasi pada University of Pretoria South Africa, 2005), 19.

¹⁴ Djemari Mardapi, *Teknik Penyusunan Instrumen tes dan Nontes* (Yogyakarta: Mitra Cendekia, 2008), 72.

¹⁵ *Ibid*, 16.

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya untuk keperluan ujian semester digunakan butir soal yang memiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi/sukar, dan untuk keperluan diagnostik biasanya digunakan butir soal yang memiliki tingkat kesukaran rendah/mudah.

Kegunaan dari tingkat kesukaran dilihat dari sudut pandang penyusunan tes adalah untuk mengetahui kemungkinan adanya butir soal yang bias dan agar dapat merakit tes yang memiliki ketepatan data soal. Tingkat kesukaran butir soal juga dapat digunakan untuk memprediksi alat ukur itu sendiri (soal) dan kemampuan peserta didik dalam memahami materi yang diajarkan guru.

Jika suatu butir soal termasuk kategori mudah, dapat diprediksi bahwa pengecoh butir soal itu tidak berfungsi serta sebagian besar siswa menjawab benar butir soal itu artinya bahwa sebagian besar siswa telah memahami materi yang ditanyakan. Sedangkan bila suatu butir soal termasuk kategori sukar, dapat dipredikasi bahwa: 1) butir soal itu mungkin salah kunci jawaban, 2) butir soal mempunyai 2 atau lebih jawaban yang benar, 3) materi yang ditanyakan belum diajarkan, 4) materi yang diukur tidak cocok ditanyakan dengan bentuk soal yang diberikan, dan 5) pernyataan atau kalimat soal terlalu kompleks atau panjang.

Daya beda atau daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga belajar yang tidak/kurang/belum menguasai materi yang ditanyakan. Atau antara kelompok yang berprestasi tinggi (*high achiever*) dan kelompok yang berprestasi rendah (*low achiever*).

Manfaat daya pembeda butir adalah untuk: 1) meningkatkan mutu setiap butir soal melalui data empiriknya, dan 2) mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami materi yang diajarkan guru.

Apabila suatu butir soal tidak dapat membedakan kedua kemampuan siswa, maka aptut dicurigai kemungkinan sebagai berikut: 1) kunci jawaban butir soal tidak tepat, 2) butir soal memiliki dua atau lebih kunci jawaban yang benar, 3) kompetensi yang diukur tidak jelas, 4) pengecoh tidak berfungsi, 5) materi yang ditanyakan terlalu sulit, sehingga

banyak siswa yang menebak, dan 6) sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya.¹⁶

Daya beda butir soal yang sering digunakan dalam tes hasil belajar adalah dengan menggunakan indeks korelasi antara skor butir dengan skor totalnya. Daya beda dapat dilihat dari besarnya koefisien korelasi biserial (r -bis) maupun koefisien korelasi point biserial (r -pbis). Kedua korelasi masing-masing mempunyai kelebihan, walaupun para guru/pengambil kebijakan banyak yang suka menggunakan r -pbis.¹⁷ Besarnya korelasi berkisar antar -1.00 s/d 1.00. Klasifikasi daya beda menurut Crocker dan Algina adalah seperti digambarkan pada tabel 7. Klasifikasi Daya Beda Butis Soal Berdasarkan Nilai Korelasi Point Biserial (r -pbis). Kriteria butir soal baik dan dapat digunakan jika nilai r .pbis= 0.40-1.00, kategori sedang dan soal diterima tapi perlu diperbaiki jika nilai r .pbis= 0.3-0.39, kriteria ‘soal dengan eberapa catatan’ dan perlu perbaikan jika nilai r .pbis=0.2-0.29, dan kriteria ‘soal jelek/tidak baik’ dan soal dibuang/tidak dipakai jika nilai r .pbis<0.19.

Sedikit berbeda dengan klasifikasi diatas, menurut Dali S. Naga, untuk kriteria kedua, soal diterima dan tidak perlu revisi/ perbaikan. Dan menurut Azmawi Zainul, kriteria ke-empat, soal masih dapat digunakan dengan revisi. Soal yang dibuang adalah kalau nilai point biserialnya kurang dari 0.00.

Konstruksi butir soal terdiri dari dua bagian, yaitu pokok soal dan alternatif jawaban. Alternative jawaban juga terdiri dari dua bagian, yaitu kunci jawaban dan pengecoh. Pengecoh dikatakan berfungsi apabila semakin rendah tingkat kemampuan peserta tes semakin banyak memilih pengecoh, atau semakin tinggi tingkat kemampuan peserta tes akan semakin sedikit memilih pengecoh. Apabila peserta tes yang menjawab dengan salah atau memilih pengecoh kurang dari 0.025 maka pengecoh tersebut harus direvisi (*Proporsional endorsing* antara 0.00 s/d 0.025) sedangkan jika pengecoh tidak ada yang memilih (*proporsional endorsing* = 0.00) maka pengecoh tersebut ditolak. *Proporsional endorsing* menunjukkan fungsi daya tarik dari alternatif jawaban.

Selain melihat daya tarik untuk dipilih, pengecoh soal juga perlu memperhatikan daya beda (koefisien korelasi) yang ditunjukkan oleh masing-masing alternatif jawaban. Setiap pengecoh diharapkan memiliki daya beda negatif, artinya suatu pengecoh diharapkan lebih sedikit dipilih oleh kelompok tinggi dibandingkan dengan kelompok

¹⁶ Panduan Analisis Butir Soal, www.dikmenum.go.id, 11

¹⁷ Panduan Analisis Butir Soal, www.dikmenum.go.id, 13

bawah. Atau daya beda pengecoh tidak lebih besar dari daya beda kunci jawaban setiap butir.

Reliabilitas berhubungan dengan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran yang sama. Untuk mengetahui konsistensi dari perangkat tes yang dikembangkan dilakukan dengan 3 cara, yaitu metode konsistensi internal, stabilitas, dan antar penilai. Metode konsistensi internal dilakukan dengan melihat konsistensi antar item atau antar bagian dalam tes itu sendiri. Metode stabilitas dilakukan dengan melihat konsistensi hasil pada kesempatan tes yang berbeda. Sedangkan metode ketiga adalah dengan metode alternatif, yaitu konsistensi hasil dari dua atau lebih tes yang berbeda. Berikut rangkuman berbagai metode untuk mengetahui reliabilitas yang disarikan dari buku “teknik Penyusunan Instrumen Tes dan Non-tes”.¹⁸

Metode konsistensi internal hanya memerlukan satu kali penyajian tes saja, dikenal dengan nama *single trial administration*, dan karena itu masalah-masalah yang timbul akibat penyajian yang berulang dapat dihindari. Pendekatan dalam melakukan estimasi reliabilitas dengan konsistensi internal ada tiga, yaitu klasik parallel, tau-ekivalen, dan konginerik. Masing-masing pendekatan menggunakan persyaratan yang berbeda, yaitu sesuai dengan karakteristik data. Estimasi reliabilitas dengan metode parallel klasik dikenal juga dengan Spearman-Brown dilakukan dengan menghitung korelasi antar skor-tampak antara dua tes yang parallel yang disajikan pada kelompok subjek yang sama. Metode tau-ekivalen dapat dilakukan dengan teknik Rulon, koefisien alpha, formula Kuder-Richardson, dan reliabilitas belah dua. Pendekatan konginerik yaitu pendekatan yang digunakan apabila varians kedua belahan tidak sama dan rerata dua belahan juga tidak sama. Salah satu formula yang termasuk pada pendekatan ini adalah Kristoff untuk tes belah tiga.

Metode kedua yaitu dengan metode stabilitas. Stabilitas hasil pengukuran dapat dilihat dari besarnya korelasi skor hasil pengukuran pertama dan hasil pengukuran kedua. Dua distribusi skor dikorelasikan. Besarnya korelasi ini menyatakan indeks reliabilitas dalam arti stabilitas hasil pengukuran.

Berdasarkan berbagai aspek/kriteria dalam menganalisis butir soal yang telah disebutkan diatas, dapat dirangkum criteria penilaiannya seperti tampak pada tabel 2.

Tabel 2. Kriteria Kualitas Butir Soal

Kategori	Kriteria Penilaian
----------	--------------------

¹⁸ Djemari Mardapi, 30-58.

Baik	<ol style="list-style-type: none"> 1. Tingkat kesukaran $0.25 \leq p \leq 0.75$ 2. Point/korelasi biserial butir soal ≥ 0.40 3. Point/Korelasi biserial alternatif jawaban (distraktor) bernilai negative
Revisi	<ol style="list-style-type: none"> 1. Tingkat kesukaran $p \leq 0.25$ atau $p \geq 0.75$ tetapi korelasi biserial butir ≥ 0.40 dan korelasi biserial distraktor bernilai negatif 2. Tingkat kesukaran $0.25 \leq p \leq 0.75$, point biserial butir soal ≥ 0.40, tetapi ada korelasi biserial yang bernilai positif 3. Tingkat kesukaran $0.25 \leq p \leq 0.75$, point biserial butir soal antara 0.2 sampai 0.3, tetapi korelasi distraktor bernilai negatif selain kunci jawaban
Tidak baik	<ol style="list-style-type: none"> 1. Tingkat kesukaran $p \leq 0.25$ atau $p \geq 0.75$ dan ada point biserial pada distraktor bernilai positif 2. Point/korelasi biserial butir soal < 0.20 3. Point biserial < 0.30 dan point biserial distraktor bernilai positif

Instrumen penilaian keterampilan proses telah dilakukan oleh banyak peneliti dengan masing-masing karakteristik baik untuk sekolah dasar, menengah pertama maupun menengah atas. Aspek keterampilan proses yang dikembangkan juga beragam, dari yang dasar maupun yang sudah terintegrasi.

Untuk keterampilan proses sains dasar untuk siswa sekolah dasar telah dikembangkan oleh Beard pada tahun 1970 dengan istilah *Basic Science Process Test* (BSPT), serta oleh Molitor dan George pada tahun 1976 yang mengembangkan *Science Process Skill Test* (SPST) untuk mengukur keterampilan inkuiri inferensi dan verifikasi siswa kelas 4, 5, dan 6. Smith dan Welliver tahun 1990 mengembangkan *Science Process Assessment* untuk siswa kelas 4 yang mengukur 13 keterampilan proses sains untuk siswa di Amerika Serikat.

Untuk tingkat sekolah menengah pertama, Tannenbaum pada tahun 1969 yang mengembangkan *Test of Science Process* untuk mengukur keterampilan proses siswa kelas 7, 8, dan 9. Kemudian Ong Eng Tek et al mengembangkan perangkat tes keterampilan proses sains untuk pendidikan Malaysia. Instrumen diperuntukkan untuk siswa tingkat menengah pertama sejumlah 60 butir dengan indeks kesukaran antara 0.25 sampai 0.75 serta indeks diskriminasi di atas 0.40.¹⁹ Selain validasi oleh ahli, instrumen juga dianalisis secara psikometrik untuk mengetahui daya beda dan tingkat kesukaran butir.

Untuk tingkat sekolah menengah atas, instrumen keterampilan proses sains terintegrasi dikembangkan oleh Burns, Okey and Wise pada tahun 1985. McKenzie dan

¹⁹ Ong Eng Tek et al, The Development and validation of an All Encompassing Malaysian-Based Science Process Skills Test for Secondary School. *Journal of Science and Mathematics Education in South East Asia*, 2011 Vol. 34 No. 2, 203-236.

Padilla tahun 1986 mengembangkan tes pilihan ganda untuk keterampilan membuat grafik untuk siswa kelas 7-12.

Di Indonesia, Kartika Santi mengembangkan instrumen penilaian keterampilan proses sains pada mata pelajaran fisika tingkat SMA (Sekolah Menengah Atas). Penelitian ini menghasilkan instrumen penilaian keterampilan proses untuk 11 aspek keterampilan proses sains, yaitu: mengamati, merumuskan masalah, merumuskan hipotesis, mengajukan pertanyaan, mengidentifikasi variable, mengelompokkan, menginterpretasi data, penarikan kesimpulan, merancang percobaan, berkomunikasi, dan menerapkan konsep. Instrumen penilaian divalidasi oleh ahli untuk memberikan penilaian tiap butir yang terdapat di dalam instrumen penilaian sesuai dengan kriteria.²⁰

Pada tahun 2013, Titik Hidayati dkk mengembangkan tes diagnostik untuk mengidentifikasi keterampilan proses sains dengan tema energi pada pembelajaran IPA Terpadu.²¹

METODE

Penelitian ini merupakan penelitian jenis *Research and Development (R & D)*, yaitu metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifannya. Untuk mencapai tujuan penelitian, penelitian dilakukan dalam dua fase, yaitu fase proses pengembangan instrumen dan fase analisis psikometrik.²² Fase pertama meliputi: 1) mengidentifikasi tes objektif; 2) melakukan spesifikasi isi/konten dari tes atau mendeskripsikan keterampilan proses sains yang akan diteskan; 3) menyusun kisi-kisi instrumen termasuk indikator dari masing-masing keterampilan dan butir soal per indikator; 4) menulis item/butir soal sesuai dengan indikator; dan 5) melakukan *expert judgement* (untuk meyakinkan *face dan content validity*). Fase kedua merupakan analisis psikometrik yang dilakukan dengan melakukan tes pilot untuk menentukan reliabilitas internal sekaligus mengetahui tingkat kesulitan dan diskriminansi. Item akan diterima jika mempunyai tingkat kesulitan 0.25-0.75 dan indeks diskriminasi paling tidak 0.40. item yang tidak memenuhi kualifikasi tersebut akan dimodifikasi atau ditolak.

Pada tahap dilakukan *expert judgement*, instrumen tes keterampilan proses sains akan dinilai oleh validator sebagai *expert/ahli*. Maka dibutuhkan tim validator sebagai

²⁰ Kartika Santi, *Pengembangan Instrumen Penilaian Keterampilan Proses Sains pada Mata Pelajaran Fisika*

²¹ Titik Hidayati, Pengembangan Tes Diagnostik untuk mengidentifikasi Keterampilan Proses Sains dengan Tema Energi pada Pembelajaran IPA Terpadu, *Unnes Science Education Journal* 2 (2) 2013, 311-319.

²² Cohen, Manion, Morrison, *Research Method in Education* (UK: Routledge, 2007)

narasumber yang terdiri dari 2 orang validator . Selanjutnya, para validator memberikan penilaian tiap butir yang terdapat di dalam instrumen penilaian sesuai dengan kriteria.

Penelitian untuk uji keterlaksanaan/tes pilot produk tes keterampilan proses sains kelas atas sekolah dasar dilaksanakan di SDN 1 Tugu. Subjek penelitian adalah siswa kelas atas, yaitu kelas IV dan V. Kelas VI tidak digunakan sebagai subyek tes pilot karena belum menggunakan kurikulum 2013. Tes pilot dilakukan pada semester I tahun pelajaran 2014/2015. Jumlah subyek tes pilot adalah 15 siswa kelas IV, yang terdiri dari 6 siswa perempuan dan 9 siswa laki-laki. Kelas V terdiri dari 14 siswa, terbagi menjadi 7 siswa perempuan dan 6 Siswa laki-laki.

Butir tes terdiri atas pertanyaan/soal, kunci/pedoman penskoran/rubrik, dan pedoman penskorannya. Ada 6 keterampilan proses dasar sains yang diidentifikasi²³, yaitu: mengamati (*observing*), menglompokkan (*classifying*), mengukur (*measuring*), menyimpulkan (*inferring*), meramalkan (*predicting*), dan mengkomunikasikan (*communicating*). Kemudian akan dikembangkan indikator untuk tiap-tiap keterampilan proses dasar ini.

Untuk mengukur keterampilan proses dasar sains, selanjutnya ditulis butir-butir instrumennya berdasarkan indikator tersebut. Tiap indikator dikembangkan menjadi 4 butir instrumen. Jumlah butir ada 24 butir, masing-masing 4 butir untuk indikator keterampilan proses sains dasar. Dalam tes pilot, ke-24 butir instrumen tersebut dikerjakan dalam waktu 40 menit. Waktu pengerjaan ini sudah sesuai dengan yang disarankan oleh Kartowarigan, yaitu bahwa setiap responden sebaiknya tidak mengisi lebih dari 40 butir instrumen (lebih dari 1 jam), namun bila keadaan memaksa harus diberi jeda paling tidak 20 menit.²⁴ Tabel 3. merupakan kisi-kisi Instrumen Keterampilan Proses Sains Dasar untuk Kelas Atas.

Paling tidak terdapat empat butir soal untuk masing-masing indikator. Empat butir soal ini dirasa sudah mencukupi secara kuantitas untuk nanti diberikan dalam tes pilot. Semakin banyak butir tiap indikatornya akan meningkatkan fitur dalam sebuah pengukuran. Butir instrumen berupa soal tes berbentuk soal pilihan ganda dengan 4 pilihan jawaban. Butir disusun berbentuk *paper and pencil multiple choice test* karena dengan bentuk ini dapat digunakan untuk menilai ke-enam aspek keterampilan proses sains dasar dalam waktu yang singkat, dapat dilakukan dengan mudah (meskipun sampelnya besar),

²³ Rezba, et al. 1995

²⁴ Badrun Kartowagiran. Penyusunan Instrumen Kinerja SMK-SBI. Makalah dalam Workshop Evaluasi Kinerja SMK-SBI P4TK Matematika Yogyakarta, 2009.

dapat di-skor dengan mudah pula, obyektif dan mengurangi kesalahan penilaian.²⁵ Pedoman utama dalam pembuatan butir soal bentuk pilihan ganda adalah: 1) pokok soal harus jelas, 2) pilihan jawaban homogen dalam arti isi, 3) panjang kalimat pilihan jawaban relatif sama, 4) tidak ada petunjuk jawaban benar, 5) hindari penggunaan pilihan jawab semua benar atau semua salah, 6) pilihan jawaban angka diurutkan, 7) semua pilihan jawaban logis, 8) jangan menggunakan negating ganda, 9) kalimat yang digunakan sesuai dengan tingkat perkembangan peserta tes, 10) bahasa Indonesia yang digunakan baku, dan 11) letak pilihan jawaban benar ditentukan secara acak.²⁶

Tabel 3. Kisi-kisi Instrumen Keterampilan Proses Sains Dasar

Keterampilan Proses	Indikator	Nomor Butir
Observasi	Menggunakan alat indra sebanyak mungkin, Mengumpulkan fakta yang relevan dan memadai	1, 8, 9, 20
Kuantifikasi	menggunakan alat ukur, membandingkan dengan menggunakan alat ukur yang sesuai.	2, 21, 22, 24
Klasifikasi	Mencari perbedaan, mengontraskan, mencari kesamaan, membandingkan, mencari dasar penggolongan	3, 5, 17, 18,
Prediksi	Menggunakan pola, menghubungkan pola yang ada, memperkirakan peristiwa yang akan terjadi	10, 11, 12, 14
Mengkomunikasikan	Membaca grafik, tabel atau diagram, menjelaskan hasil percobaan, mendiskusikan hasil percobaan, menyampaikan laporan secara sistematis	15, 16, 19, 23
Inferensi	Menjelaskan hasil observasi, menyimpulkan berdasarkan fakta/bukti dari serangkaian observasi	4, 6, 7, 13

Materi butir soal disesuaikan dengan materi dalam kurikulum 2013 kelas IV, meliputi materi sumber bunyi dan sifat sifat bunyi, energi, macam-macam energi, bagian-bagian telinga dan fungsinya, sifat-sifat cahaya, metamorphosis, sumber daya alam, serta morfologi luar tumbuhan.

Meskipun ditujukan untuk mengetes siswa kelas atas, namun butir soal dikembangkan berdasarkan muatan materi kelas IV sehingga baik siswa kelas IV maupun

²⁵ Ong Eng Tek, et.al. The Development and Validation of an All Encompassing Malaysian-based Science Process Skills Test for Secondary Schools. *Journal of Science and Mathematics Education and Southeast Asia*, 2011 vol. 34. No.2, 2013-236, 212.

²⁶Djemari Mardapi, *Teknik Penyusunan.....*, 72.

kelas V sudah familiar dengan materi tersebut. Butir soal tidak menguji pemahaman materi, tetapi menguji keterampilan proses sains yang dimiliki siswa.

Pengecekan butir soal dilakukan oleh dua orang ahli pendidikan dasar untuk meyakinkan validitas konten dari instrumen ini. Penjelasan lengkap tentang validasi ahli dijelaskan pada sub bab berikut.

Tujuan penelitian adalah untuk memperoleh produk berupa tes keterampilan proses sains dasar untuk kelas atas sekolah dasar. Seperti dijelaskan, penelitian dilakukan melalui dua fase, yaitu proses penyusunan instrument dan analisis psikometrik (perhitungan tingkat kesulitan dan indeks diskriminasi). Pada fase pertama data dikumpulkan dari angket yang diberikan kepada para validator yaitu para ahli di bidang pembelajaran sains dan evaluasi pendidikan. Dari penilaian para validator diperoleh skor untuk menilai validitas muka dan konten. Validator kemudian memutuskan apakah instrumen layak untuk diuji pilot-kan.

Dalam penelitian ini, validasi oleh ahli dilakukan oleh dua orang, yang masing-masing validator memberikan penilaian dalam tiga hal, yaitu materi, konstruksi, dan bahasa. Dengan perangkat instrumen tes yang sudah disusun, validator memberikan penilaian sesuai dengan indikator penilaian yang diberikan. Untuk keperluan ini, lembar validasi berupa penilaian berbentuk ya/tidak dengan memberikan tanda centang pada kolom masing-masing butir soal. Lembar kedua berupa penilaian terbuka yang digunakan untuk menuliskan poin-poin penting yang perlu diperhatikan untuk perbaikan instrumen. Di lembar kesimpulan validasi, validator memberikan keputusan apakah instrumen tersebut dapat digunakan, tidak dapat digunakan, atau dapat digunakan sebagai instrumen dengan perbaikan.

Hasil dari validator pertama, disimpulkan bahwa butir-butir soal dapat digunakan sebagai instrumen keterampilan proses sains dengan beberapa perbaikan yang disarankan. Ada satu soal yang disarankan untuk diganti yaitu butir soal nomor 19. Alasan penggantian karena soal tersebut tidak akan terbaca kalau dicetak hitam-putih. Hal ini akan menyulitkan dalam pengerjaan soal. Perbaikan unsur kebahasaan lain terkait pemilihan kata yang kurang tepat, misalnya: kata “ ‘gambar berikut’ diganti dengan ‘gambar disamping’ karena gambar terletak disamping. Ahli juga menelaah kesesuaian kata dan kalimat dengan usia subyek penelitian (yaitu kelas IV dan V sekolah dasar). Dalam hal ini, semua butir yang diajukan peneliti sudah sesuai dengan tingkat usia siswa.

Validator kedua menyimpulkan bahwa butir soal dapat digunakan sebagai instrumen keterampilan proses sains dasar dengan beberapa catatan perbaikan. Diantaranya

adalah: 1). mengganti kunci jawaban yang bias, sehingga tidak membingungkan siswa dalam memilih jawaban, 2). Menyingkat soal yang terlalu panjang, 3). Mengganti poin jawaban yang terlalu panjang, yang berbeda dengan pilihan jawaban lainnya, serta 4). Mengubah Pilihan jawaban yang berbentuk angka sehingga tersusun berdasarkan urutan besar kecilnya angka. Hasil penilaian validator ahli selengkapnya terdapat di lampiran laporan penelitian ini.

Setelah instrumen tes dinyatakan valid oleh para validator, kemudian dilakukan dilakukan tes pilot di sekolah yang telah ditetapkan. Hasil dari tes ini kemudian diskor sesuai pedoman pedoman penskoran. Hasilnya dikumpulkan untuk kemudian dilakukan dianalisis secara psikometrik.

Subyek tes pilot yang dilakukan adalah siswa kelas IV dan V di SDN 1 Tugu. Jumlah responden adalah 15 siswa kelas IV dan 13 siswa kelas bawah. Tes dilaksanakan secara bersamaan di ruang kelas masing-masing. Siswa mengerjakan soal uji coba sebanyak 24 butir soal dalam waktu maksimal 40 menit. Soal untuk tes diagnostik disarankan tidak lebih dari 40 soal yang dikerjakan tidak lebih dari 60 menit. Jawaban siswa kemudian direkam tidak hanya benar salah tiap butirnya, tetapi semua pilihan jawaban yang dipilih siswa, apakah itu pilihan (a), (b), (c), atau (d).

Dari jawaban siswa tersebut, akan dipertimbangkan berbagai aspek sesuai kriteria analisis butir soal klasik, yaitu dengan mempertimbangkan tingkat kesukaran, daya beda, distraktor, serta reliabilitas tes. Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks, berupa proporsi yang besarnya berkisar 0.00 – 1.00. Semakin besar indeks kesukaran berarti semakin mudah soal itu. Jika suatu soal memiliki tingkat kesukaran (TK) 0.00 berarti bahwa tidak ada siswa yang menjawab benar, atau 0% siswa menjawab benar. $TK = 1.00$, berarti 100% (semua) siswa menjawab benar. Klasifikasi tingkat kesukaran dapat dibagi menjadi tiga kelompok, yaitu mudah (p bernilai 0.76-1.00), sedang (p bernilai 0.26-0.75), dan sukar (p bernilai 0.00-0.25).

Kriteria kedua adalah dengan memperhatikan daya beda atau daya pembeda soal, yaitu kemampuan suatu butir soal dapat membedakan dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga belajar yang tidak/kurang/belum menguasai materi yang ditanyakan. Atau antara kelompok yang berprestasi tinggi (*high achiever*) dan kelompok yang berprestasi rendah (*low achiever*).

Daya beda butir soal yang sering digunakan dalam tes hasil belajar adalah dengan menggunakan indeks korelasi antara skor butir dengan skor totalnya. Daya beda dapat

dilihat dari besarnya koefisien korelasi biserial (r -bis) maupun koefisien korelasi point biserial (r -pbis). Kedua korelasi masing-masing mempunyai kelebihan, walaupun para guru/pengambil kebijakan banyak yang suka menggunakan r -pbis.²⁷ Besarnya korelasi berkisar antar -1.00 s/d 1.00. Klasifikasi daya beda yang diaplikasikan dalam penelitian ini adalah kesimpulan dari kriteria oleh beberapa ahli. (Dali S. Naga dan Azmawi Zainul). Butir soal dikatakan baik, kalau mempunyai korelasi point biserial lebih atau sama dengan 4.00. Soal sedang, korelasi 0.3 – 0.39, soal dengan beberapa catatan mempunyai korelasi 0.2-0.29, soal jelek dan dapat diperbaiki kalau korelasinya ≤ 0.19 , dan soal jelek dan tidak dapat diperbaiki kalau korelasinya ≤ 0.00 .

Kriteria berikutnya adalah terkait distribusi jawaban, yaitu kunci jawaban dan pengecoh. Pengecoh dikatakan berfungsi apabila semakin rendah tingkat kemampuan peserta tes semakin banyak memilih pengecoh, atau semakin tinggi tingkat kemampuan peserta tes akan semakin sedikit memilih pengecoh. Apabila peserta tes yang menjawab dengan salah atau memilih pengecoh kurang dari 0.025 maka pengecoh tersebut harus direvisi (*Proporsional endorsing* antara 0.00 s/d 0.025) sedangkan jika pengecoh tidak ada yang memilih (*proporsional endorsing* = 0.00) maka pengecoh tersebut ditolak. *Proporsional endorsing* menunjukkan fungsi daya tarik dari alternatif jawaban.

Selain melihat daya tarik untuk dipilih, pengecoh soal juga perlu memperhatikan daya beda (koefisien korelasi) yang ditunjukkan oleh masing-masing alternatif jawaban. Setiap pengecoh diharapkan memiliki daya beda negatif, artinya suatu pengecoh diharapkan lebih sedikit dipilih oleh kelompok tinggi dibandingkan dengan kelompok bawah. Atau daya beda pengecoh tidak lebih besar dari daya beda kunci jawaban setiap butir.

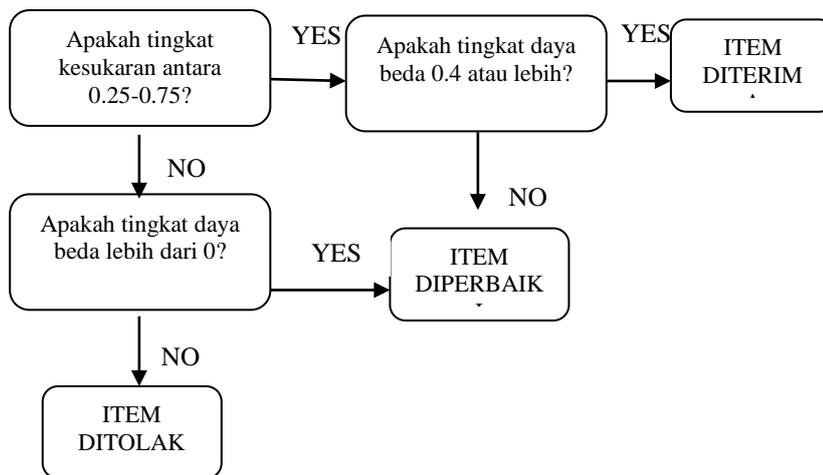
Selanjutnya, untuk melihat konsistensi soal, perlu dilakukan tes untuk mengetahui tingkat reliabilitas. Pada penelitian ini tes reliabilitas menggunakan koefisien alpha. Kriteria reliabilitas perangkat tes adalah sangat tinggi (0.800- 1.00), tinggi (0.600-0.799), cukup (0.400-0.599), rendah (0.200-0.399), dan sangat rendah (0.00-0.199).²⁸

Untuk mempermudah analisisn, maka proses pemilihan butir tes mengikuti prosedur sebagaimana digambarkan pada gambar 1. Analisis butir soal secara kuantitatif ini dilakukan dengan bantuan program MicroCat ITEMAN versi 3.00.

²⁷ Panduan Analisis Butir Soal, www.dikmenum.go.id, 13

²⁸ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*. (Jakarta: Bumi Aksara, 2001)

Sesuai dengan tujuan penelitian ini, produk yang dikembangkan adalah seperangkat instrument tes yang valid dan reliabel untuk mengukur keterampilan proses sains dasar kelas atas (KPSDKA).



Gambar 1. Proses Seleksi Butir Tes

Berdasarkan data hasil validasi ahli, kedua validator menyimpulkan bahwa butir soal dapat digunakan sebagai instrumen ketrampilan proses sains dasar kelas atas yang disingkat KPSDKA dengan syarat dilakukan beberapa perbaikan. Untuk keperluan ini maka kemudian dilakukan perbaikan pada beberapa butir. Rangkuman perbaikan butir soal oleh validator ahli adsa pada tabel 4.

Tabel 4. Perbaikan Butir oleh Validator Ahli

Nomor Butir	Awal	Perbaikan
2	Jawaban berupa angka tidak urut	Mengurutkan jawaban dari yang kecil ke besar
4	Pilihan jawaban disamping gambar	Pilihan jawaban dibawah gambar
7	“Ketika kamu memukul gendang akan terdengar bunyi dari karet tersebut”	“Ketika kamu memukul gendang akan terdengar bunyi dari gendang tersebut”
9	“Gendang telinga ditunjukkan pada gambar nomor berapa?”	“Berdasarkan gambar pada nomor soal diatas, gendang telinga ditunjukkan pada gambar nomor berapa?”
12	Kunci jawaban bias	Mengganti kunci jawaban
15	Petunjuk jawaban tidak ada	Memberi petunjuk jawaban pada gambar
19	Grafik berbentuk grafik pie tidak terlalu jelas	Grafik diubah menjadi grafik batang
20	“Berikut ini gambar	“Dibawah ini adalah gambar

	sehelai daun.”	sehelai daun.
20	Kata perintah hendaknya menggunakan tanda petik	Memberi tanda petik.
21, 23	“televise”	“televisi”
23	Tabel kurang besar, tidak terbaca	Tabel diperbesar dan tulisan diperjelas
23	“gelas millimeter”	“gelas milliliter”

Output ITEMAN 3.0, butir nomor 10 dan 21 mempunyai tingkat kesukaran lebih dari 0.75 dan daya pembeda kurang dari 0.25, menunjukkan bahwa butir tersebut terlalu mudah dan relatif mudah dan tidak terlalu kuat untuk membedakan siswa yang *high-achieving* dan *low-achieving*. Butir ini dikategorikan butir yang tidak baik. Butir nomor 10 dibuang sedangkan butir nomor 21 diperbaiki.

Untuk butir 10, butir dibuang karena daya pembeda kurang negatif yang menunjukkan bahwa siswa yang secara keseluruhan mempunyai skor rendah untuk instrumen ini menjawab benar untuk butir nomor 10 ini dengan jumlah siswa yang lebih banyak dibanding jumlah siswa yang rata-rata skor-nya baik yang menjawab benar untuk butir ini. Butir ini benar-benar tidak dapat membedakan mana siswa yang *high achiever* dan mana siswa yang *low achiever*. Butir nomor 10 ini dibuang. Secara lebih mudah, dapat dijelaskan bahwa suatu soal yang sulit ternyata siswa yang skor-nya rendah dapat menjawab soal tersebut dibandingkan siswa yang skor-nya tinggi. Berbeda dengan butir 10, butir nomor 21 diperbaiki karena meskipun daya pembeda dibawah 0.25, tetapi masih diatas 0.

Berbeda dengan analisis butir nomor 10, dengan daya pembeda yang sama-sama bernilai negatif, butir nomor 13 (daya beda = - 0.21) dapat diperbaiki, karena tingkat kesukaran berada pada range yang dapat diterima, yaitu $p = 0.643$ ($0.25 \leq p \leq 0.75$).

Butir nomor 3 dan 5 mempunyai tingkat kesukaran $p = 0.893$, yang berarti 89% siswa dapat menjawab dengan benar soal ini. Dengan tingkat kesukaran $p \geq 0.75$, maka soal ini dianggap terlalu mudah untuk dikerjakan. Daya pembeda butir ini juga sama yaitu 0.259 (≥ 0.25), meskipun masih dibawah 0.4 tetapi sudah agak dapat membedakan mana siswa yang *low* dan yang *high*. Keputusan untuk butir 3 dan 5 adalah diperbaiki.

Butir nomor 4, 6, 11, 12, 14, 16 dan 20 adalah butir dengan tingkat kesukaran yang dapat diterima yaitu lebih dari 0.25 dan kurang dari 0.75. range ini menunjukkan bahwa soal tidak terlalu mudah juga tidak terlalu sulit. Daya pembeda untuk butir-butir tersebut dibawah 0.25, yang mengindikasikan bahwa soal kurang dapat digunakan untuk

membedakan siswa pandai dan tidak pandai (*high-achieving* dan *low-achieving*). Untuk butir ini diperlukan perbaikan.

Butir nomor 9, 15, 19, dan 23 adalah butir dengan tingkat kesukaran yang dapat diterima yaitu lebih dari 0.25 dan kurang dari 0.75. Range ini menunjukkan bahwa soal tidak terlalu mudah juga tidak terlalu sulit. Daya pembeda untuk butir-butir tersebut antara 0.25 sampai 0.39, yang mengindikasikan bahwa soal sudah agak dapat digunakan untuk membedakan siswa pandai/*high* dan tidak pandai/*low* tapi masih perlu untuk diperbaiki menjadi lebih baik. Daya pembeda ditunjukkan dengan hasil *Point Biserial*.

Butir nomor 18, mempunyai tingkat kesukaran dengan $p = 0.893$, yang berarti 89% siswa dapat menjawab dengan benar soal ini. Dengan tingkat kesukaran $p \geq 0.75$, maka soal ini dianggap terlalu mudah untuk dikerjakan. Daya pembeda butir adalah 0.576, merupakan kriteria yang baik karena lebih dari 0.40, yang berarti sudah dapat digunakan untuk membedakan mana siswa yang *low* dan yang *high*. Keputusan untuk butir 18 adalah diperbaiki, agar tingkat kesukaran meningkat atau soal tidak terlalu mudah.

Butir nomor 24 mempunyai tingkat kesukaran $p = 0.071$, yang berarti hanya 7% siswa dapat menjawab dengan benar soal ini. Dengan tingkat kesukaran $p \leq 0.25$, maka soal ini dianggap terlalu sukar untuk dikerjakan. Daya pembeda butir yaitu 0.221, meskipun masih dibawah 0.4 tetapi masih bernilai positif. Keputusan untuk butir 24 adalah diperbaiki.

Butir soal yang memiliki kriteria “baik” karena memiliki tingkat kesukaran dalam range yang diterima yaitu $0.25 \leq p \leq 0.75$, dan daya pembeda ≥ 0.40 adalah butir nomor 1, 2, 7, 8, 17, dan 22.

Hasil analisis butir dengan kriteria seperti dijelaskan di atas, yang dibantu dengan program MicroCAT Iteman 3.0 tersebut dapat dirangkum pada tabel 5. Dari tabel terlihat dari 24 butir soal yang di uji cobakan dalam tes pilot, 10 butir soal diterima (nomor 1, 2, 7, 8, 9, 15, 17, 18, 22 dan 23), 13 butir soal diperbaiki (3, 4, 5, 6, 11, 12, 13, 14, 16, 19, 20, 21, 24), dan 1 soal dibuang / tidak digunakan (nomor 10).

Tabel 5. Hasil Analisis Butir untuk KPSDKA pada Tes Pilot

No. Butir	Tingkat Kesukaran	Daya Pembeda	Keputusan
1	0.571	0.4	Diterima
2	0.750	0.431	Diterima
3	0.893	0.259	Diperbaiki
4	0.786	0.181	Diperbaiki

5	0.893	0.259	Diperbaiki
6	0.500	0.184	Diperbaiki
7	0.785	0.569	Diterima
8	0.393	0.553	Diterima
9	0.250	0.332	Diterima
10	0.964	-0.120	Dibuang
11	0.607	0.150	Diperbaiki
12	0.607	0.225	Diperbaiki
13	0.643	-0.210	Diperbaiki
14	0.536	0.204	Diperbaiki
15	0.393	0.302	Diterima
16	0.321	0.220	Diperbaiki
17	0.429	0.442	Diterima
18	0.893	0.576	Diterima
19	0.571	0.276	Diperbaiki
20	0.679	0.042	Diperbaiki
21	0.821	0.220	Diperbaiki
22	0.536	0.573	Diterima
23	0.286	0.368	Diterima
24	0.071	0.221	Diperbaiki

Reliabilitas berhubungan dengan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran yang sama. Untuk mengetahui konsistensi dari perangkat tes yang dikembangkan

Reliabilitas berhubungan dengan tingkat konsistensi atau kemantapan hasil terhadap hasil dua pengukuran yang sama. Untuk mengetahui konsistensi dari perangkat tes yang dikembangkan dilakukan dengan metode konsistensi internal, dengan melihat konsistensi antar item atau antar bagian dalam tes itu sendiri.

Hasil tes reliabilitas dengan bantuan SPSS for windows diperoleh nilai alpha 0.524. Menurut kriteria tinggi rendahnya reliabilitas sebuah perangkat tes, nilai tersebut berada pada kategori cukup yaitu antara 0,400 – 0.599.

Hasil reliabilitas tersebut dilakukan pada hasil dari tes pilot. Untuk memastikan tingkat reliabilitas yang lebih baik, diperlukan lagi tes pilot ke-2 engan perangkat yang telah melalui proses revisi setelah tes pilot pertama. Pada penelitian ini tidak dilakukan tes pilot lagi.

PENUTUP

Tujuan penelitian ini adalah mengembangkan tes keterampilan proses sains dasar kelas atas SD/MI yang valid dan reliabel. Sesuai dengan tahapan pengembangan tes dapat disimpulkan sebagai berikut:

1. Dilakukan pengembangan perangkat tes tahap pertama yang meliputi: 1) mengidentifikasi tes objektif; 2) melakukan spesifikasi isi/konten dari tes atau mendeskripsikan keterampilan proses sains yang akan diteskan; 3) menyusun kisi-kisi instrumen termasuk indikator dari masing-masing keterampilan dan butir soal per indikator; 4) menulis item/butir soal sesuai dengan indikator; dan 5) melakukan *expert judgement* (untuk meyakinkan *face dan content validity*). Hasil dari tahap ini adalah 24 butir soal pilihan ganda yang telah valid setelah dilakukan berbagai revisi sesuai saran ahli.
2. Dilakukan pengembangan perangkat tes tahap kedua yang merupakan analisis psikometrik yang dilakukan dengan melakukan tes pilot untuk menentukan reliabilitas internal sekaligus mengetahui tingkat kesulitan dan diskriminansi. Butir yang diterima adalah butir yang mempunyai tingkat kesulitan, daya beda, serta distribusi jawaban sesuai criteria salam penelitian ini. Hasil dari tahap ini adalah terdapat 10 butir soal diterima (butir nomor 1, 2, 7, 8, 9, 15, 17, 18, 22 dan 23), 13 butir soal diperbaiki/direvisi (butir nomor 3, 4, 5, 6, 11, 12, 13, 14, 16, 19, 20, 21, 24), dan 1 soal dibuang / tidak digunakan (nomor 10). Tingkat reliabilitas butir soal tes adalah 0.524, merupakan tingkat reliabilitas dengan klasifikasi sedang.

DAFTAR PUSTAKA

- Azwar, Saifuddin. (2007). *Relibilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Bundu, Patta. (2006) *Penilaian Keterampilan Proses dan Sikap Ilmiah dalam Pembelajaran Sains SD*. Dirjen Dikti Depdiknas.
- Cohen, Manion, Morrison. (2007). *Research Method in Education*. UK: Routledge.
- Collette, A.T., & Chiappetta, E.L. (1994). *Science instruction in the middle and secondary schools*. 2nd Edition. New York: Macmillan Pub. Co.
- Friedl, Alfred E., *Teaching Science to Children An Integrated Approach*, Second Edition, New York: McGraw-Hill, Inc., 1991.
- Ghozali, Imam. *Aplikasi Analisis Multivariat dengan Program SPSS*. (Semarang: Badan Penerbit Universitas Diponegoro), 2006.
- Glover, John A., Bruning, Roger H. (1990). *Educational Psychology: Principles and Application*. New York: Harper Collins Publisher.
- Harlen, Wynne, (1996) *The Teaching of Science in Primary Schools* (London: David Fulton Publisher, Ltd), 1996, 27-34
- Hidayati, Titik. Pengembangan Tes Diagnostik untuk mengidentifikasi Keterampilan Proses Sains dengan Tema Energi pada Pembelajaran IPA Terpadu, *Unnes Science Education Journal* 2 (2) 2013, 311-319.

- Kamalia, Poppy. (2010). Keterampilan Proses dalam Pembelajaran IPA, Pusat Pengembangan dan Pemberdayaan Pendidikan dan Tenaga Kependidikan Ilmu Pengetahuan Alam.
- Kartowagiran, Badrun. (2009). Penyusunan Instrumen Kinerja SMK-SBI. *Makalah dalam Workshop Evaluasi Kinerja SMK-SBI P4TK Matematika* Yogyakarta.
- Kazeni Mungandi Monde Monika, (2005). *Development and Validation of a test of Integrated Science Process Skill for the Further Education and Training Learners*, (Disertasi pada University of Pretoria South Africa).
- Mardapi, Djemari. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta, Mitra Cendikia.
- Mollitor, L.L. & George, K.D. Development of a test of science process skill. *Journal of Research in Science Teaching*, 13, 405-412.
- McMillan, J.H. & Schumacher., S. Research in Education: A conceptual introduction, Ed. Ke-3. Glenview, IL: Scott, Foresman and Co, 1989.
- McDermott, L. C. (2001). Physics education research- the key to student learning. *American Journal of Physics*, 69(11), 1127-1137.
- Mertler, Craig A. and Charles, C. M., *Introduction to Educational Research*. Boston: Pearson, 2005.
- Neuman, Donald B. (1993). *Experiencing Elementary Science*. California: Wadsworth.
- Ong Eng Tek et al, (2011). The Development and validation of an All Encompassing Malaysian-Based Science Process Skills Test for Secondary School. *Journal of Science and Mathematics Education in South East Asia*, 2011 Vol. 34 No. 2, 203-236.
- Popham, W. James. Classroom Assessment what Teachers need to Know. Allyn and Bacon, 1994.
- Panduan Analisis Butir Soal, www.dikmenum.go.id, 11
- Permendikbud No. 64 Tahun 2013 tentang Standar Isi
- Puskur Depdiknas, Standar Isi 2006
- Prasetyo, Zuhdan K.. (1998) Taksonomi untuk pendidikan fisika (sains) Yogyakarta: *Cakrawala Pendidikan Majalah Ilmiah Kependidikan*. Edisi Khusus Dies, Mei, 146-151.
- Sugiyono. (2008). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Rezba, Richard J., et.al. *Learning and Assessing Science Process Skill*. (Iowa: Kendall/Hunt Publishing Co, 1995).
- Shahali, Edy Hafizan Mohd, Halim, Lilia. (2010). Development and Validation of a Test of Integrated Science Process Skill. *Procedia Social and Behavioral Sciences*, 142-146.
- Santi, Kartika, et al. Pengembangan Instrumen Penilaian Keterampilan Proses Sains pada Mata Pelajaran Fisika.
- Suharsimi Arikunto. (2001) *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Winkel, W.S. (1996). *Psikologi Pengajaran*. Jakarta: Grasindo.

