



PENGEMBANGAN COMPUTERIZED TESTLET UNTUK MENGUKUR KETERAMPILAN PROSES SAINS PADA MATERI STOIKIOMETRI

Development of Computerized Testlet to Measure Science Process Skill on Stoichiometry

Stefanus Kristiyanto, Ashadi, Sri Yamtinah*, dan Sri Mulyani*

*Program Studi Magister Pendidikan Kimia, FKIP, Universitas Sebelas Maret
Jl. Ir. Sutami No. 36A, Surakarta, Jawa Tengah 57126, Indonesia*

* Untuk korespondensi: email: jengtina@staff.uns.ac.id; srimulyaniuns@staff.uns.ac.id

Received: October 28, 2019

Accepted: December 20, 2019

Online Published: December 31, 2019

DOI : 10.20961/jkpk.v4i3.35198

ABSTRAK

Penelitian ini bertujuan mengembangkan instrumen penilaian computerized testlet untuk mengukur keterampilan proses sains pada materi stoikiometri di tinjau dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh yang memenuhi kriteria sebagai instrumen penilaian yang baik, serta dapat menampilkan profil individu keterampilan proses sains peserta didik. Subjek uji coba penelitian adalah siswa kelas X SMA Negeri di Boyolali. Uji validitas dilakukan dengan validitas isi, kriteria, dan validasi ahli. Uji reliabilitas dilakukan dengan mencari harga koefisien reliabilitas r . Uji karakteristik dilakukan dengan menentukan taraf kesukaran, daya pembeda dan indeks pengecoh. Profil keterampilan proses sains ditentukan dengan mengukur penguasaan proses sains peserta didik. Hasil penelitian dan pengembangan ini adalah computerized testlet yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan validitas soal lebih dari 0,79 (valid), memiliki reliabilitas tes pada uji coba lapangan utama 0,643 dan 0,610 pada uji coba pelaksanaan lapangan yang tergolong tinggi, memiliki daya pembeda dengan persentase 10% jelek, 66,7% cukup, dan 23,3% baik, memiliki tingkat kesukaran dengan persentase 20% sukar, 53,3% sedang dan 26,7% mudah.

Kata Kunci: *computerized testlet, keterampilan proses sains, stoikiometri*

ABSTRACT

This research aims to develop a computerized testlet assessment to measure science process skills on stoichiometry material in terms of validity, reliability, difficulty level, distinguishing features and deception indexes that meet the criteria as good assessment instruments, and can display individual profiles of science process skills students. The subject of the research trial was a grade X student of Senior High School in Boyolali. Validity test is done by content validity, criteria, and expert validation. Reliability test is done by finding the price of the reliability coefficient r . Characteristics test is done by determining the level of difficulty, distinguishing features and deception index. The profile of science process skills is determined by measuring students' mastery of the science process. The results of the research and development was declared feasible and met the criteria as a good question with a validity of questions more than 0.79 (valid), has a test reliability on the main field trials of 0.643 and 0.610 on the trial the implementation of the field is relatively high, has a distinguishing power with a percentage of 10% bad, 66.7% is sufficient, and 23.3% is good, has a difficulty level with a percentage of 20% difficult, 53.3% moderate and 26.7% easy.

Keywords: *computerized testlet, science process skills, stoichiometry*

PENDAHULUAN

Proses penilaian pembelajaran sains sementara ini hanya difokuskan pada ranah kognitif saja [1]. Pernyataan tersebut didukung oleh hasil penelitian yang menunjukkan bahwa penilaian yang digunakan untuk menilai keterampilan proses sains siswa 100% hanya mengukur aspek kognitif yakni menggunakan kuis, ulangan akhir pokok bahasan, ulangan umum, dan tugas rumah tanpa menilai unjuk kerja siswa [2]. Hal ini menyebabkan evaluasi pada aspek keterampilan dan sikap yang juga menjadi tuntutan kurikulum dalam penilaian proses pembelajaran di kelas belum dilakukan secara optimal. Sementara penilaian terhadap kinerja siswa dalam bentuk penugasan jarang dilakukan sebagai suatu model penilaian alternatif yang lebih bermakna. Kinerja siswa perlu dinilai pada saat kegiatannya sedang berlangsung [3].

Kegiatan *assessment* pembelajaran yang hanya mengukur pengetahuan peserta didik saja tanpa melatih dan mengembangkan keterampilan proses sains peserta didik menyebabkan keterampilan proses sains pada peserta didik tergolong rendah. Keterampilan proses sains peserta didik yang rendah disebabkan oleh beberapa faktor meliputi: rendahnya latar belakang sains, minimnya prasarana laboratorium [4], buku satu-satunya pedoman dalam pembelajaran [5], administrasi sekolah belum menginisiasi pembelajaran kontekstual [6], hanya menekankan penguasaan konsep, serta kegiatan pembelajaran yang belum mengeksplorasi keterampilan proses sains peserta didik [7]. Keterampilan proses sains yang tidak mengikutsertakan penilaiannya akan mengakibat-

kan pembelajaran yang dilakukan sia-sia [8], dan kegiatan asesmen yang tidak menyeluruh atau hanya dilakukan di akhir pembelajaran akan tidak bermakna [9].

Hasil survei PISA pada tahun 2006, rata-rata skor prestasi sains peserta didik menurun, yaitu 393 dan urutannya 50 dari 57 negara dan pada tahun 2009, rata-rata skor prestasi sains peserta didik kembali menurun, yaitu 383 dengan menempati urutan 60 dari 65 negara. Hasil Studi PISA tahun 2009 menunjukkan tingkat penguasaan sains siswa Indonesia yang tidak jauh berbeda dengan hasil studi tahun 2006. Tingkat penguasaan sains siswa Indonesia berada pada peringkat ke 57 dari 65 negara peserta dengan skor yang diperoleh 383 dan skor ini berada di bawah rata-rata standar dari PISA [10]. Tercatat skor sains Indonesia dalam PISA 2015 adalah 403, menempati urutan 62 dari 72 negara. Rata-rata skor prestasi sains peserta didik Indonesia hanya mencapai di bawah skor rata-rata yaitu dibawah 50 [11].

Desain Testlet adalah suatu set item yang memberikan stimulus. Hal ini telah banyak di gunakan dalam dunia pendidikan dan tes psikologi [12-13]. Hal ini telah banyak di gunakan dalam dunia pendidikan dan tes psikologi. Banyak pengembang test yang menemukan desain Testlet ini menarik karena efisien dalam penulisan itemnya. Dalam sains beberapa topik adalah hierarkis, sebagai contoh struktur hierarki kemampuan mental. Beberapa peneliti telah mengembangkan jenis "latent trait" untuk mengukur berbagai macam keterampilan, seperti diagnosis kognitif [14]. ,model multidimensi dengan struktur hierarkis [15-16].

Dua konsep umum yang biasa digunakan pada faktor model Testlet adalah independensi tiap item dan multi dimensional. Testlet yang digunakan biasanya tidak terlepas dari 2 bentuk tersebut. Testlet dengan independensi artinya setiap item tes yang dikembangkan tidak berhubungan dengan item lainnya, sebaliknya dengan menggunakan konsep multidimensi setiap item yang dikembangkan mungkin berhubungan dengan item lainnya sebagai contoh ketika disajikan suatu data grafik pertanyaan pertamanya adalah apa yang didapatkan dari grafik, sedangkan pertanyaan selanjutnya adalah mengapa itu terjadi. Multidimensi ini mungkin terlihat lebih beralasan ketika diterapkan pada multi item yang berhubungan dengan konteks tetapi tidak dibuat secara langsung dari satu sama lain. Dalam jurnal ini dijelaskan pula faktor yang mempengaruhi Testlet model jika menggunakan model independen dan multidimensi [17].

Untuk mengukur dan mengevaluasi keterampilan proses sains dan keterampilan berpikir kritis tidak hanya dapat dilakukan dengan instrumen pilihan ganda, tetapi dapat menggunakan wawancara juga observasi [18]. Melakukan penilaian keterampilan proses dengan tidak menyertakan pemahaman konsep yang tidak dimiliki oleh siswa di dalamnya adalah sesuatu yang tidak valid. Oleh karena ini, penting untuk melakukan penilaian keterampilan proses hanya yang berhubungan dengan konten dimana pengetahuan konseptual yang tidak akan menjadi permasalahan untuk menggunakan keterampilan proses [19].

Berdasarkan hal tersebut maka perlu pengembangan instrumen *assessment* yang

bersifat autentik, dan bervariasi sehingga tidak hanya menilai kemampuan kognitif. Instrumen *assessment* yang dikembangkan berbasis komputer (*Computer Based Test*) untuk mengukur keterampilan proses sains peserta didik pada materi stoikiometri. Materi dipilih berdasarkan kaitannya dengan tema dalam kehidupan dan mengacu pada soal PISA. Instrumen penilaian testlet untuk mengukur keterampilan proses sains ini sebenarnya sudah ada namun masih terbatas pada paper test [20]. Melalui pengembangan instrumen berbasis komputer ini diharapkan para pendidik dapat dengan mudah mengukur kemampuan keterampilan proses sains peserta didik dan membiasakan peserta didik dengan soal-soal berstandar internasional dalam pembelajaran di Indonesia sehingga mampu mendukung peningkatan kualitas pendidikan di tingkat dunia, khususnya pada keterampilan proses sains.

METODE PENELITIAN

Penelitian pengembangan instrumen penilaian testlet berbasis komputer untuk mengukur keterampilan proses sains ini menggunakan desain penelitian dan pengembangan (R&D) model Borg dan Gall [21] yang terdiri dari 10 tahap yaitu: 1) penelitian pendahuluan dan pengumpulan informasi, 2) perencanaan, 3) pengembangan bentuk produk awal, 4) uji coba lapangan awal, 5) revisi produk awal, 6) uji coba lapangan utama, 7) revisi produk utama, 8) uji coba pelaksanaan lapangan, 9) revisi produk akhir, 10) diseminasi.

Penelitian pendahuluan dan pengumpulan informasi meliputi studi literatur, studi lapangan, serta identifikasi dan analisis

kebutuhan. Penelitian dilaksanakan di tiga SMA di Kabupaten Boyolali, Uji karakteristik dilakukan dengan menentukan taraf kesukaran, daya pembeda, dan proporsi atau perbandingan kategori keterampilan proses sains dalam instrumen. Instrumen yang dikembangkan digunakan untuk mengukur kemampuan keterampilan proses sains siswa.

Uji validitas meliputi validitas isi, kriteria, dan validasi oleh ahli [22]. Uji validitas isi dilakukan melalui telaah instrumen oleh 11 ahli, sedangkan uji validitas kriteria digunakan untuk menentukan validasi menggunakan teknik Aiken. Uji validasi oleh ahli penting untuk mengetahui tingkat validitas instrumen penilaian berbasis komputer yang dikembangkan. Pengujian dilakukan dengan menggunakan angket validasi yang diisi oleh validator.

Uji reliabilitas dilakukan dengan menghitung harga koefisien reliabilitas model Iteman. Uji kemampuan keterampilan proses sains dilakukan dengan menghitung persentase penguasaan keterampilan proses sains. Daya pembeda dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi item, yakni kelompok yang tergolong pandai dan kelompok bawah (the lower group) yaitu kelompok peserta yang mendapat nilai rendah [23].

Tingkat kesulitan atau tingkat kesukaran item diambil dari program analisis butir soal ITEMAN yang digunakan, tingkat kesukaran soal dapat diperoleh dari harga *prop. Correct*. Indeks pengecoh atau distraktor dapat diperoleh dari program ITEMAN, indeks pengecoh diperoleh dari harga *prop. endorsing* dan *biser*. Dimana pengecoh dikatakan telah berfungsi dengan baik apabila harga *prop. endorsing* minimal 0,020

(2%) dan harga dari *biser* adalah negatif. Harga negatif dari *biser* memiliki arti bahwa pengecoh lebih banyak dipilih oleh peserta didik dari kelompok bawah.

HASIL DAN PEMBAHASAN

1. Analisis Kualitas Tes

a. Validitas

Validasi instrumen pada penelitian ini menggunakan formula Aiken. Formula ini memiliki standar validitas mengikuti variabel penentunya. Salah satu variabel penentu validitas Aiken ini adalah jumlah validator dan jumlah kriteria penentu validitas yang digunakan. Semakin banyak jumlah validator, dan semakin banyak jumlah kriteria yang digunakan, maka nilai standar untuk valid akan semakin mengecil. Hasil dari perhitungan validasi menggunakan formula Aiken menunjukkan semua item soal dinyatakan valid karena nilai V lebih tinggi dari V_{tabel} (0,79).

b. Reliabilitas

Pengujian reliabilitas hanya dilakukan pada tahap Main Field Test dan Operational Field Test. Hasil dari Tabel 1, reliabilitas berada pada rentang 0,61-0,80 yang dapat dikategorikan tinggi. Hal ini dapat diartikan bahwa Instrumen Penilaian Computerized Testlet memiliki tingkat konsistensi atau keajegan dalam mengukur yang tinggi. Perhitungan reliabilitas menggunakan formula Iteman disajikan pada Tabel 1.

Tabel 1. Hasil Pengujian Reliabilitas

Tahap Uji	Nilai Reliabilitas	Keterangan
Main Field Test	0,643	Tinggi
Operational Field Test	0,610	Tinggi

c. Daya Pembeda

Hasil Analisis Daya Pembeda disajikan pada Gambar 1 . Pada Gambar 1 ditunjukkan adanya perubahan kualitas daya pembeda dari *Main Field Test* dan *Operational Field Test*.

d. Tingkat Kesukaran

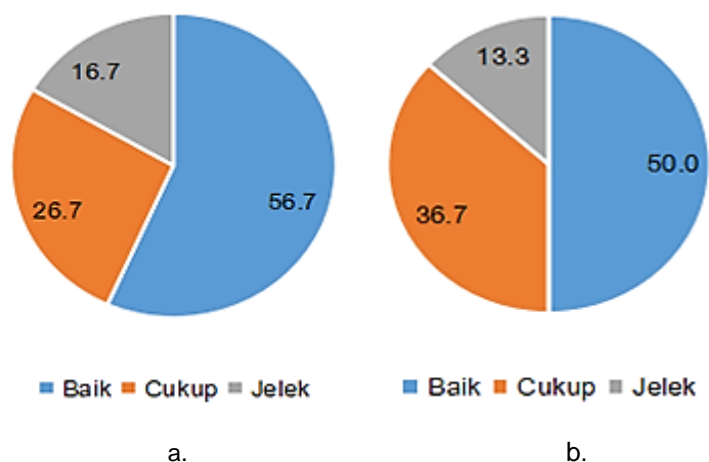
Hasil Analisis Tingkat Kesukaran disajikan pada Gambar 2 yang menunjukkan perbedaan hasil dari analisis tingkat kesukaran pada uji coba lapangan utama (*Main Field Test*) dan uji coba pelaksanaan lapangan (*Operational Field Test*).

e. Interpretasi Kelayakan Butir Soal

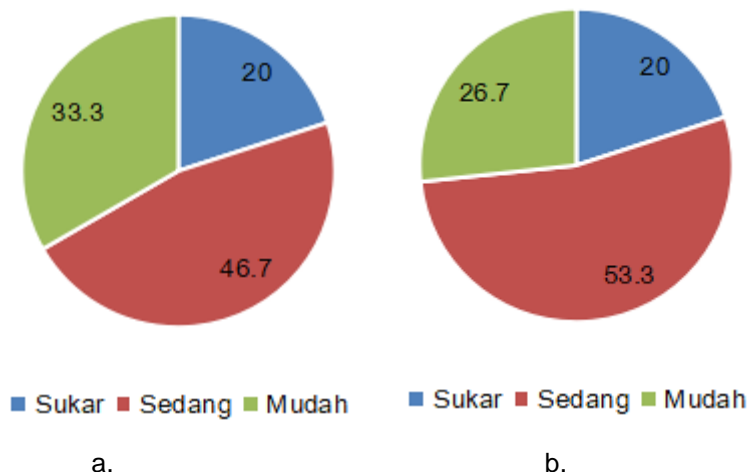
Interpretasi kelayakan butir soal dilihat dari tingkat kesukaran dan daya pembeda diperlihatkan pada Tabel 2.

Tabel 2. Interpretasi kelayakan butir soal

Interpretasi Butir Soal	%	Nomor Soal
Diterima	90	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,16,17, 18,20,21,22,23,24, 25,26,27,29,30
Ditolak	10	15,19,28



Gambar 1. Persentase Analisis Daya Pembeda (a) *Main Field Test*, (b) *Operational Field Test*



Gambar 2. Persentase Analisis Tingkat Kesukaran (a) *Main Field Test*, (b) *Operational Field Test*

f. Ketuntasan Indikator KPS

Pada tahap uji coba pelaksanaan lapangan instrumen diujikan kepada tiga kelas peserta didik di masing-masing satuan

pendidikan. Hasil uji coba ini selain dapat digunakan untuk menganalisis butir soal, tes ini juga dapat berfungsi untuk mengetahui KPS yang dimiliki peserta didik. Hasil Analisis KPS pada tahap ini disajikan pada Tabel 3.

Tabel 3. Hasil Analisis KPS pada tahap uji coba pelaksanaan lapangan

Indikator	Persen Ketuntasan			
	SMA A	SMA B	SMA C	Total
Mengobservasi peristiwa ilmiah	71,3	62,5	50,1	61,3
Menginterpretasi data	63,5	54,6	44,9	54,3
Memprediksi perubahan	55,6	40,7	35,4	43,9
Menerapkan konsep	42,1	32,7	21,4	32,1
Membuat kesimpulan	38,6	29,5	19,3	29,1

Indikator yang paling dikuasai oleh peserta didik adalah indikator mengobservasi peristiwa ilmiah, sedangkan untuk indikator lainnya di setiap satuan pendidikan memberikan proporsi yang hampir sama. Hasil yang hampir sama dari pengujian sebelumnya ini dikarenakan kualitas soal yang tidak banyak berubah. Sebaran tingkat kesulitan soal pun tidak banyak berubah.

Soal-soal pada indikator keterampilan proses sains mengobservasi peristiwa ilmiah dan menginterpretasi data cenderung lebih mudah dikerjakan dari pada indikator keterampilan proses sains yang lain, terlebih kebanyakan soal pada indikator keterampilan proses sains mengobservasi peristiwa ilmiah dan menginterpretasi data terdapat pada soal tingkat pertama setiap stem, artinya soal ini memang soal yang lebih mudah dibandingkan dengan soal lainnya. Hal ini menjadikan indikator mengobservasi peristiwa ilmiah dan menginterpretasi data memiliki persentase jawaban yang lebih tinggi dibandingkan dengan indikator yang lain. Selain dilakukan pengujian secara empiris, untuk lebih meyakinkan dan mendapatkan justifikasi dan

dukungan data tentang instrumen yang dikembangkan maka dilakukan diseminasi di akhir tahap penelitian dan pengembangan yang dilakukan. Diseminasi ini dimaksudkan untuk mensosialisasikan hasil penelitian dan pengembangan yang telah dilakukan kepada para pendidik satuan pendidikan sebagai pengguna instrumen yang dikembangkan. Diseminasi dilakukan terhadap 33 orang pendidik kimia SMA di Kabupaten Boyolali.

Menurut responden pendidik pada forum diseminasi sebanyak 54,55% pendidik menyatakan indikator keterampilan proses sains yang digunakan relevan dengan keterampilan proses sains yang dibutuhkan peserta didik, 30,30% menyatakan cukup relevan, 12,12% menyatakan kurang relevan dan 3,03% menyatakan tidak relevan. Sebanyak 63,64% pendidik menyatakan bahwa instrumen dalam bentuk Computerized Testlet untuk mengukur keterampilan proses sains yang dikembangkan telah efektif, 24,24% menyatakan cukup efektif, 9,09% menyatakan kurang efektif, dan 3,03% menyatakan tidak efektif. Profil peserta didik yang telah dikembangkan dinilai mudah dalam

penggunaannya oleh 45,45% pendidik, dinilai cukup mudah oleh 33,33% pendidik, dinilai cukup sulit oleh 15,15% pendidik, dan dinilai sulit oleh 6,06% pendidik.

Berdasarkan hasil diseminasi didapatkan bahwa karena program CBT yang dikembangkan ini menggunakan sistem online, maka program ini relatif mudah untuk dioperasikan. Program ini berisikan analisis butir, skor kognitif peserta didik, skor testlet peserta didik, peserta remedial, materi remedial, ketuntasan indikator pembelajaran, dan profil peserta didik. Hasil penilaian dan analisis hasil penilaian dapat langsung dilihat di akhir tes.

2. Tampilan Computerized Testlet

Hasil pengembangan dan penelitian ini seiring dengan fenomena kecanggihan teknologi di semua bidang yang tidak bisa dihindari termasuk dalam hal tes atau penilaian dalam kurikulum 2013. Menanggapi serta mengikuti kemajuan dan perkembangan teknologi sudah seharusnya sistem penilaian yang konvensional beralih ke sistem yang terkomputerisasi.

Perubahan model tes secara konvensional dengan kertas menjadi tes berbasis komputer yang biasa dikenal dengan Computer Assisted Testing, Computerized Assessment, Computer Based Testing (CBT), Computer Aided Assessment (CAA), Computer Based Assessment (CBA), Online Assessment, E Assessment dan Web-Based Assessment [24].

Computerized testlet yang telah dikembangkan merupakan tes berbasis komputer yang memanfaatkan komputer untuk menggantikan kertas atau paper-pencil

dalam pengadministrasian tes [25]. Kementerian Pendidikan dan Kebudayaan Republik Indonesia menjelaskan bahwa Computer Based Test (CBT) adalah sistem pelaksanaan ujian dengan menggunakan komputer sebagai media ujiannya [26]. Computerized testlet yang dikembangkan merupakan bagian dari CBT.

Penilaian dengan computerized testlet memiliki perbedaan dengan sistem PBT atau Paper Based Test dalam hal media pengerjaan. Peserta didik yang menggunakan sistem computerized testlet langsung dapat memilih jawaban yang benar pada layar komputer. Jika menggunakan sistem PBT peserta diwajibkan menghitamkan bulatan di kertas lembar jawaban. Pada sistem computerized testlet peserta didik hanya perlu menggeser dan klik kursor pada jawaban yang dianggap benar. Namun dengan sistem PBT, peserta ujian juga dituntut untuk teliti supaya jawaban bisa terbaca saat dikoreksi dengan mesin pemindai. Kertas juga rentan berlubang jika menggunakan penghapus yang kasar. Computerized testlet memiliki kelebihan, antara lain: dibandingkan tes tertulis, dalam tes ini peserta dapat segera tau hasil tes, jika dibandingkan dengan tes lisan, tes ini dapat dilaksanakan serentak dengan peserta banyak dengan waktu yang relatif singkat, siswa merasa lebih bebas dan percaya diri dalam mengerjakan soal, mengurangi terjadinya kecurangan dalam pelaksanaan tes, karena setiap peserta akan mendapat soal yang berbeda dengan tingkat kesulitan yang sama [17].

Instrumen penilaian yang dikembangkan dalam penelitian ini berbentuk *Com-*

puterized Testlet untuk mengukur keterampilan proses sains. Tahapan ilmiah dari instrumen ini adalah bentuk penilaian yang menggunakan komputer sebagai bagian integral dari penyampaian pertanyaan, penyimpanan jawaban, dan pembuatan laporan hasil tes atau latihan. Selain itu juga untuk membantu meringankan tugas pendidik dalam melakukan penilaian dengan membuat suatu software atau program analisis yang dapat melaksanakan semua tugas evaluasi pendidik dalam satu waktu yang singkat, baik untuk melakukan penilaian kognitif peserta didik, melakukan analisis ketuntasan indikator pembelajaran, dan analisis butir soal.

Software atau program yang dibuat untuk melengkapi instrumen penilaian dalam bentuk *Computerized Testlet* ini berbasis *Computer Based Test* (CBT). Program CBT adalah sistem pelaksanaan penilaian atau ujian dengan menggunakan komputer sebagai media ujiannya. Dalam pelaksanaannya, CBT berbeda dengan sistem ujian berbasis kertas atau Paper Based Test (PBT) yang selama ini sudah berjalan.

Profil peserta didik pada program ini berisikan data skor total dari tes yang telah dilakukan, kategori keterampilan proses sains peserta didik serta indikator soal yang sudah dan belum tercapai. Diharapkan Program ini dapat membantu tugas pendidik dalam melakukan evaluasi pembelajaran dan melakukan penilaian. Profil peserta didik yang menjadi ciri pembeda dari program lainnya diharapkan dapat menjadi solusi permasalahan penilaian yang dirasakan oleh pendidik.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa instrumen penilaian *Computerized Testlet* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan hasil memiliki angka validitas lebih dari 0,79 yang dikatakan valid, memiliki reliabilitas tes pada uji coba lapangan utama 0,643 dan 0,610 pada uji coba pelaksanaan lapangan yang tergolong tinggi, memiliki daya pembeda dengan persentase 13,3% jelek, 36,7% cukup, dan 50% baik, memiliki tingkat kesukaran dengan persentase 20,0% sukar, 53,3% sedang dan 26,7 % mudah.

DAFTAR RUJUKAN

- [1] S. Maryam. "Peer group dan aktivitas harian (belajar) pengaruhnya terhadap prestasi belajar remaja studi kasus pada SMU bina bangsa sejahtera plus di kota Bogor tahun 2002," *Jurnal pendidikan dan kebudayaan*. vol.13, no. 58, pp. 114-131, 2006.
- [2] I W. Suastra, "Mengembangkan Kemampuan Berpikir Kreatif Melalui Pembelajaran Sains," *Jurnal IKA*, vol. 4, no. 2, pp. 23-34, 2006.
- [3] I. W. Suastra, *Belajar dan Pembelajaran Sains*. Buku Ajar. Universitas Pendidikan Ganesha Singaraja, 2007.
- [4] G. U. Jack, "The Influence of Identified Student and School Variables on Student Science Process Skill Acquisition," *Journal of Education and Practice*. vol. 4, no. 5, pp. 16-22, 2013.
- [5] E. Igboegwu, "Effects of CoOperative Learning Strategy and Demonstration Method on Acquisition of Science Process Skills by Chemistry Students of Different Levels of Scientific Literacy" *Journal of research and Development*. vol. 3, no. 1, pp. 204-212. 2011.

- [6] L. L. Chaguna & D. M. Yango, "Science Process Skills Proficiency of The Grade VI Pupils in The Elementary Diocesan Schools of Baguio and Benguet," *Research Journal*. vol. 16, no. 4, pp. 22-32, 2008.
- [7] Sukarno, A. Permanasari, & I. Hamidah, "The Profile of Science Process Skill (SPS) Student at Secondary High School (Case Study in Jambi)," *International Journal of Scientific Engineering and Research*. vol. 1, no. 1, pp. 79-83, 2013.
- [8] Arifin. *Evaluasi Pembelajaran*. Bandung: Rosda, 2009.
- [9] W. Harlen, "Inquiry-Based Learning in Science and Mathematics," *Review of Science Mathematics & ICT Education*, vol. 7, no. 2, pp. 9–33, 2013.
- [10] Organisation for Economic Cooperation and Development (OECD). *PISA 2009 Assessment Framework. Key Competencies in Reading, Mathematics and Science*. Paris: OECD, 2009.
- [11] Organisation for Economic Cooperation and Development (OECD). *PISA 2015 Assessment Framework. Key Competencies in Reading, Mathematics and Science*. Paris: OECD, 2015.
- [12] Komisi Nasional Indonesia untuk UNESCO. *Pendidikan untuk Pembangunan Berkelanjutan (Education for Sustainable Development/ESD) di Indonesia*. Jakarta: Kemdikbud, 2014.
- [13] H.-Y. Huang, W.-C. Wang & P.-H. Chen, "An item response model with hierarchical latent traits," *Paper presented at the annual meeting of the American Educational Research Association*, Denver, CO. 2012.
- [14] D. D. L. Torres, "Higher-order latent trait models for cognitive diagnosis," *Journal of Psychometrika*. Vol. 69, ppp. 333–353, 2004.
- [15] Y. Sheng & C. Wikle, "Bayesian Multidimensional IRT Models With A Hierarchical Structure," *Journal of Psychology Measurement*. vol. 68, no. 3, pp. 413-430, 2008.
- [16] D. L. Torre, *Bayesian Psychometric Modelling*. Information Age Publishing, Inc. USA. 2009.
- [17] D. Mars & E. Christine, "Confirming Testlet Effect," *Journal of Applied Psychological Measurement*. March, 27. 2012.
- [18] H. Aktamis & N. Yenice, "Determination Of The Science Process Skills And Critical Thinking Skill Levels," *Procedia-Journal of Social and Behavioral Sciences*, vol. 2, no. 2, pp. 3282-3288, 2010.
- [19] E. H. M. Shahali & L. Halim, "Development and Validation of a Test of Integrated Science Process Skills," *Journal of Procedia-Social and Behavioral Sciences*, vol. 9, no. 9, pp. 142-146, 2010.
- [20] A. S. Shidiq & S. Yamtinah, "Gender Differences In Students' Attitudes Toward Science: An Analysis Of Students' Science Process Skill Using Testlet Instrument," *AIP Conference Proceedings* 1868, 030003. 2017.
- [21] W. R. Borg, & M. D. Gall, *Educational Research: An Introduction*. New York: Longman. 1989.
- [22] D. Mardapi. *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Medika. 2012.
- [23] Sudijono. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada. 2011.
- [24] S. Hadi. *Pengembangan Computerized Adaptive Test Berbasis Web*. Yogyakarta: Aswaja Pressindo. 2013.
- [25] M. Jamil, R. H. Tariq, & R. H. Shami, "Computer-Based Vs Paper-Based Examinations: Perceptions of University Teachers," *The Turkish Online Journal of Educational Technology*, vol.11, no. 4, pp. 371-381, 2012.
- [26] D. Ratnasari, Sukarmin & Suparmi, "Analisis Implementasi Instrumen Two-Tier Multiple Choice untuk Mengukur Keterampilan Proses Sains," *Jurnal Pendidikan dan Kebudayaan*, vol. 2, no. 2, pp. 166-179, 2017.
- [27] R. Pakpahan. *Computer-Based National Exam Model: Its Benefits And Barriers*. Jakarta: Pusat Penilaian Pendidikan, Balitbang, Kemdikbud. 2016.