

PENGEMBANGAN ALAT UKUR PENILAIAN LITERASI SAINS PADA KONTEN STRUKTUR ATOM DAN IKATAN KIMIA MENGUNAKAN KONTEKS WAYANG KULIT

Aulia Rahim, Hernani, Ahmad Mudzakir

Universitas Pendidikan Indonesia

email: zakir66@hotmail.com

Abstract: The Development of Scientific Literacy Measurement on The Content of Atomic Structure and Chemical Bond Through the Means of Wayang Kulit. The research entitled “The Development of Scientific Literacy Measurement on The Content of Atomic Structure and Chemical Bond Through the Means of Wayang Kulit” aims to obtain the tools to measure the content of science literacy that contain validity, reliability, level of difficulty, differentiator means, distruction function, and correlation between each number of question with the total score which shows the empirical validity. The method that is used is descriptive quantitative. To analyze and test the reliability of questions by using Anates 4.0.2, Science’ second graders of one of the high school in Bandung were chosen as the research subject. The process of measuring the validity that was conducted by using CVR (Content Validity Ratio) was examined by five lecturers : two experts at science literacy, an expert at assessment evaluation, an expert at chemical bond, and an expert at organic chemical. The validity points that were examined include the constructional validity -suitability between indicators used to the basic and PISA’ scientific competency-, along with the content validity –suitability between indicator and questions as well as questions and the answers-. The result of CVR analysis to forty questions, it is suggested that most of the questions contain content and constructional validity, have high reliability level, and empirically valid (the correlation between the number of questions and total score). The questions which range from easy to difficult level are observed to need to advance certain points : the difficulty level, differentiator point, and the function of distraction.

Abstrak: Pengembangan Alat Ukur Penilaian Literasi Sains pada Konten Struktur Atom dan Ikatan Kimia Menggunakan Konteks Wayang Kulit. Penelitian yang berjudul “Pengembangan Alat Ukur Penilaian Literasi Sains pada Konten Struktur Atom dan Ikatan Kimia menggunakan Konteks Wayang Kulit” ini bertujuan untuk memperoleh alat ukur penilaian literasi sains yang teruji validitas, reliabilitas, tingkat kesukaran, daya pembeda, keberfungsian distraktor serta korelasi antara butir soal dengan skor total yang menunjukkan validitas empirisnya. Metode yang digunakan adalah metode deskriptif kuantitatif dengan subjek penelitian siswa kelas XI MIA di salah satu SMA Negeri di Bandung untuk menguji reliabilitas dan analisis butir soal dengan menggunakan *software* Anates Versi 4.0.2. Validasi diukur menggunakan CVR (*Content Validity Ratio*) yang dilakukan kepada 5 dosen ahli yang kompeten di bidangnya. Dua orang ahli di bidang literasi sains, satu orang ahli di bidang evaluasi pembelajaran (*assessment*), satu orang ahli di bidang ikatan kimia, dan satu orang ahli di bidang kimia organik. Validitas yang dinilai berupa validitas konstruk yaitu kesesuaian indikator dengan kompetensi dasar dan kompetensi ilmiah PISA serta validitas isi berupa kesesuaian indikator dengan butir soal dan butir soal dengan jawaban. Dari hasil analisis CVR 40 butir soal, pada umumnya soal memiliki validitas isi dan konstruk yang dinyatakan valid dengan kategori reliabilitas yang tinggi, validitas empiris (korelasi skor butir dengan skor total) dikatakan valid, tingkat kesukaran meliputi soal mudah, sedang dan sukar dengan perlu perbaikan proporsi tingkat kesukaran, daya pembeda yang baik, dan ada distraktor yang berfungsi dengan baik namun adapula distraktor yang tidak berfungsi dengan baik sehingga diperlukan perbaikan.

Kata-kata kunci: alat ukur penilaian literasi sains, ikatan kimia, kompetensi PISA, struktur atom, wayang kulit.

Dalam masyarakat modern, kehidupan sehari-hari sangat dipengaruhi dan ditentukan oleh ilmu pengetahuan. Perkembangan ilmu pengetahuan pada masa kini menuntut manusia untuk memahami berbagai fenomena yang terjadi dalam kehidupan sehari-hari yang berorientasi sains dan teknologi. Munculnya kesadaran masyarakat tentang arti dan nilai penting sains dan teknologi dalam kehidupan sehari-hari dapat meningkatkan literasi sains masyarakat itu sendiri. Dengan meleknnya masyarakat terhadap literasi sains, dapat menjadi suatu langkah yang baik untuk dapat meningkatkan mutu pendidikan. Adapun hal yang dapat mencerminkan mutu pendidikan dapat dilihat dari evaluasi pendidikan yang dilakukan.

Salah satu bentuk penilaian pendidikan secara internasional dilakukan oleh lembaga OECD dengan salah satu programnya, yaitu PISA (*Programme for International Student Assessment*). Subjek penilaian dari PISA ini salah satunya adalah literasi sains. Literasi sains berkaitan dengan kapasitas siswa dalam memahami informasi proses terjadinya ilmu pengetahuan dan fakta yang ada dalam kehidupan sehari-hari dan kaitannya dengan masa yang akan datang, serta kemampuan menerapkan pengetahuan dalam kehidupan sehari-hari. PISA mengidentifikasi tiga dimensi besar literasi sains dalam pengukurannya, yakni konten sains, proses sains, dan konteks aplikasi sains (OECD, 2013, hlm.99).

Namun, berdasarkan hasil studi penilaian yang dilakukan oleh PISA (*Programme for International Student Assessment*) mengungkapkan bahwa pembelajaran sains di Indonesia kurang berhasil meningkatkan kemampuan literasi sains. Hal ini terungkap berdasarkan hasil studi PISA pada tahun 2000 Indonesia berada pada peringkat ke-38 dari 41 negara peserta PISA dengan nilai rerata tes 393; pada tahun 2003 Indonesia menempati peringkat ke-38 dari 40 negara peserta dengan nilai rerata tes 395; pada tahun 2006

Indonesia menempati peringkat ke-50 dari 57 negara peserta dengan nilai rerata tes 393; pada tahun 2009 Indonesia menempati peringkat ke-60 dari 65 negara peserta dengan nilai rerata tes 383; dan terakhir pada tahun 2012, Indonesia menempati peringkat ke-64 dari 65 negara peserta dengan nilai rerata tes 382 (Litbang, Kemdikbud, 2011).

Hal ini menunjukkan bahwa literasi sains anak Indonesia berada pada level rendah. Salah satu faktor penyebab rendahnya literasi sains siswa antara lain siswa Indonesia pada umumnya kurang terlatih dalam menyelesaikan soal-soal dengan karakteristik seperti soal-soal pada PISA. Hal ini dapat dilihat dari sebaran soal Ujian Nasional masih sangat didominasi oleh penghitungan, sehingga siswa hanya dituntut melakukan penghitungan dengan menerapkan rumus-rumus tanpa menekankan *problem solving* atau penalaran (Yunengsih, 2008, hlm. 23). Oleh karena itu, penelitian ini mencoba untuk mengembangkan alat ukur literasi sains menggunakan konteks wayang kulit pada konten struktur atom dan ikatan kimia.

Wayang kulit merupakan salah satu budaya Indonesia yang keberadaannya hampir jarang ditemukan. Sebagai salah satu bentuk pelestarian budaya Indonesia, wayang kulit dapat dijadikan sebagai konteks dalam kegiatan pembelajaran di sekolah. Seperti analisis yang dilakukan oleh Ahkamajaya (2014) terhadap pembuatan wayang kulit dan pagelaran wayang kulit, diperoleh tiga materi pokok yang dapat dihubungkan dengan wayang kulit yaitu struktur atom untuk pagelaran wayang kulit (*lighting* dalam pagelaran), ikatan kimia dan makromolekul untuk pembuatan wayang kulit.

Struktur atom merupakan salah satu topik kimia yang akan dipelajari oleh siswa SMA/MA kelas X pada semester 1. Topik ini termasuk topik yang sulit. Zoller, 1990; Harrison dan Treagust, 1996 dalam Sirhan (2007, hlm 3) melaporkan bahwa struktur atom merupakan salah satu topik kimia yang

sulit bagi siswa. Hal ini karena kebanyakan konsep dalam struktur atom berhubungan dengan konsep yang abstrak yang membuat siswa merasa kesulitan dalam mempelajarinya. Selain itu, menurut Kurniawan, *et al* (2006, hlm 362) siswa SMA/MA mengalami kesulitan belajar tentang topik struktur atom karena metode penyampaian materi struktur atom oleh pendidik (guru) yang monoton dan kurang menarik serta tidak dikaitkan dengan kemajuan teknologi yang ada atau tidak dikaitkan dengan kehidupan sehari-hari.

Ikatan kimia juga merupakan salah satu topik abstrak pada materi kimia. Sesuatu yang jauh dari pengalaman keseharian siswa sekolah menengah, yang mana siswa tidak bisa melihat atom, struktur atom, dan bagaimana atom-atom tersebut membentuk suatu ikatan, sehingga banyak siswa yang mengalami kesulitan dalam memahami konsep-konsep ikatan kimia.

Pembuatan alat ukur penilaian literasi sains ini mengacu kepada kompetensi dasar dan kompetensi ilmiah PISA yang disesuaikan dengan soal-soal karakteristik PISA. Soal-soal literasi sains memiliki ciri khas yaitu selalu diawali dengan teks pengantar berisi konteks yang dikenal siswa yang kemudian diikuti dengan pertanyaan-pertanyaan yang terkait dengan konten. Adapun empat aspek yang dinilai dalam alat ukur penilaian literasi sains

mencakup konteks sains, konten sains, kompetensi sains, dan sikap sains.

Suatu alat penilaian dikatakan mempunyai kualitas yang baik apabila alat tersebut memiliki atau memenuhi dua hal, yakni ketepatannya atau validitasnya dan ketetapan atau keajegannya atau reliabilitasnya (Sudjana, 2011, hlm. 12). Setelah diperoleh suatu alat ukur, untuk mengidentifikasi soal-soal yang baik, kurang baik, ataupun soal yang jelek dilakukan analisis butir soal. Ada sejumlah karakteristik butir yang diuji yaitu tingkat kesukaran, daya beda dan efektivitas pengecoh. Sehingga dalam penelitian ini dilakukan pengembangan alat ukur penilaian literasi sains beserta uji validitas dan reliabilitas serta analisis butir soal dari soal-soal yang telah dikembangkan pada konten struktur atom dan ikatan kimia menggunakan konteks wayang kulit.

Validitas

Validitas merupakan syarat yang terpenting dalam suatu alat evaluasi. Suatu teknik evaluasi dikatakan mempunyai validitas yang tinggi (disebut valid) jika teknik evaluasi atau tes itu dapat mengukur apa yang sebenarnya akan diukur. Salah satu pendekatan kuantitatif yang dapat digunakan untuk menguji validitas konten (isi) adalah dengan menggunakan CVR (*Content Validity Ratio*). Berikut tabel yang menunjukkan nilai minimum CVR yang diterima untuk sejumlah ahli (validator) yang berbeda.

Tabel 1. Nilai Kritis untuk Lawshe (1975) Content Validity Ratio ($CVR_{critical}$)

N (Jumlah Validator)	Level signifikansi untuk tes <i>One-Tailed</i>					
	0, 1	0, 05	0, 025	0, 01	0, 005	0, 001
	Level signifikansi untuk tes <i>Two-Tailed</i>					
	0, 2	0,1	0, 05	0, 02	0, 01	0, 002
5	0, 573	0, 736	0, 877	0, 99	0, 99	0, 99
6	0, 523	0, 672	0, 800	0, 950	0, 99	0, 99
7	0, 485	0, 622	0, 741	0, 879	0, 974	0, 99
8	0, 453	0, 582	0, 693	0, 822	0, 911	0, 99
9	0, 427	0, 548	0, 653	0, 775	0, 859	0, 99
10	0, 405	0, 520	0, 620	0, 736	0, 815	0, 977

(Wilson, R, *et al*, 2012, hlm. 206).

Reliabilitas

Reliabilitas dapat diartikan sama dengan konsistensi atau keajegan. Suatu instrumen evaluasi, dikatakan mempunyai nilai reliabilitas tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur (Sukardi, 2008, hlm. 43). Berikut tabel yang menunjukkan kriteria nilai reliabilitas suatu soal:

Tabel 2. Kriteria Reliabilitas Soal

Nilai	Kriteria
0 – 0,2	Tidak reliabel
0,21 – 0,4	Rendah
0,41 – 0,70	Sedang
0,71 – 0,90	Tinggi
0,91 – 1	Sangat tinggi

(Subana, dkk, 2005, hlm 132).

Tingkat kesukaran

Analisis tingkat kesukaran dimaksudkan untuk mengkaji soal yang mudah, sedang dan

sukar sehingga bisa menyeimbangkan proporsi soal yang mudah, sedang, dan sukar dalam tes (Hamzah, 2012, hlm. 156).

Indeks kesukaran sering diklasifikasikan sebagai berikut:

Tabel 3. Kriteria Indeks Kesukaran

Nilai P (Tingkat kesukaran)	Interpretasi
0,00 – 0,30	Soal Sukar
0,31 – 0,70	Soal Sedang
0,71 – 1,00	Soal Mudah

(Arikunto, 2012, hlm. 225)

Daya Pembeda

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang berkemampuan rendah. Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi (D). Berikut tabel yang menunjukkan kriteria daya pembeda:

Tabel 4. Kriteria Daya Pembeda

Nilai Daya Pembeda (D)	Interpretasi
$0,70 \leq D \leq 1,0$	Baik sekali
$0,40 \leq D \leq 0,69$	Cukup baik
$0,30 \leq D \leq 0,39$	Memerlukan revisi sedikit atau tidak
$0,20 \leq D \leq 0,29$	Memerlukan revisi atau disisihkan
$D \leq 0,19$	Disisihkan atau direvisi total

(Firman, 2013, hlm. 60)

Efektivitas Pengecoh

Pengecoh (*distractor*) yang juga dikenal dengan istilah penyesat atau penggoda adalah pilihan jawaban yang bukan merupakan kunci jawaban. Pengecoh diadakan untuk menyesatkan siswa agar tidak memilih kunci jawaban. Kriteria pengecoh yang baik adalah apabila pengecoh tersebut dipilih oleh paling sedikit 5% dari peserta tes (Hamzah, 2012, hlm. 180).

METODE

Metode penelitian yang digunakan dalam penelitian ini adalah deskriptif kuantitatif.

Penelitian deskriptif adalah penelitian yang hanya melibatkan satu variabel pada satu kelompok, tanpa menghubungkan dengan variabel lain atau membandingkan dengan kelompok lain (Purwanto, 2010, hlm. 177). Penelitian deskriptif memaparkan suatu fenomena dalam pembelajaran dengan ukuran-ukuran statistik, seperti frekuensi, persentase, rata-rata, variabilitas (rentang dan simpangan baku), serta citra visual dari data misalnya dalam bentuk grafik (Firman, 2013, hlm. 10).

Penelitian ini difokuskan pada pengembangan alat ukur literasi sains dan

pengujian kualitas alat ukur menggunakan tiga parameter uji yaitu validitas, reliabilitas, dan analisis butir soal. Dalam penelitian deskriptif ini, digunakan pendekatan kuantitatif non-eksperimental yaitu pengumpulan dan pengukuran data yang berbentuk angka-angka tanpa adanya suatu perlakuan terlebih dahulu. Subjek dalam penelitian ini adalah siswa kelas XI MIA 3 di salah satu SMA Negeri di Bandung. Adapun instrumen penelitian yang digunakan untuk pengumpulan data berupa lembar penyusunan teks pengantar soal, lembar validasi, dan alat ukur penilaian literasi sains yang dikembangkan dalam penelitian ini.

Data yang diperoleh dalam penelitian ini berupa hasil validitas konstruk untuk kesesuaian indikator dengan kompetensi dasar dan kompetensi PISA serta validitas isi untuk kesesuaian indikator dengan butir soal dan butir soal dengan jawaban yang diolah dengan menggunakan rumus CVR dan CVI, sedangkan data berupa hasil uji reliabilitas, uji validitas berdasarkan korelasi skor butir dan skor total, uji tingkat kesukaran, uji daya pembeda, dan uji efektifitas pengecoh diolah dengan menggunakan software Anates Versi 4.0.2 dengan kriteria reliabilitas menurut Subana (2005), kriteria validitas dan pengecoh menurut Hamzah (2012), kriteria tingkat kesukaran menurut Arikunto (2012) dan kriteria daya pembeda menurut Firman (2013).

HASIL DAN PEMBAHASAN

Hasil Pengembangan

Pengembangan alat ukur penilaian literasi sains disesuaikan dengan kompetensi dasar kurikulum 2013 dan kompetensi PISA 2012 dengan mengacu pada bahan ajar kimia dalam wayang kulit yang telah dikembangkan pada penelitian sebelumnya. Alat ukur yang dikembangkan berupa soal pilihan berganda sebanyak 40 butir soal yang mencakup konten struktur atom dan ikatan kimia dengan konteks wayang kulit.

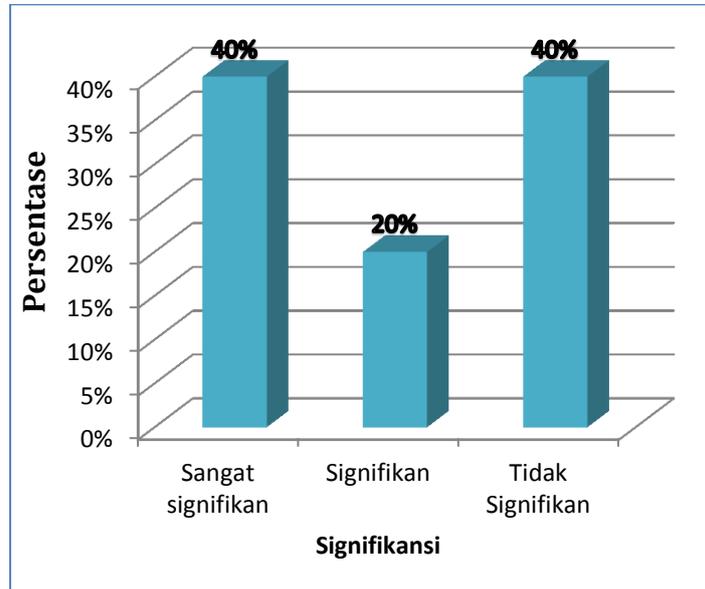
Berdasarkan hasil validasi judgement dari 5 dosen ahli mengenai validitas konstruk (kesesuaian indikator dengan kompetensi dasar dan kompetensi PISA) serta validitas isi (kesesuaian indikator dengan butir soal dan butir soal dengan jawaban) diperoleh nilai CVI untuk validitas konstruk sebesar 0,99 karena terdapat satu validator yang menyatakan tidak sesuai antara indikator dengan kompetensi dasar, dan antara indikator dengan kompetensi PISA yaitu pada soal nomor 4. Nilai CVI sebesar 0,99 menunjukkan bahwa indikator yang telah dikembangkan dapat mengukur aspek berpikir seseorang dan mengukur secara tepat aspek pengetahuan, keterampilan, dan sikap dengan tuntutan penilaian literasi sains PISA serta kurikulum 2013. Hal ini sesuai dengan aturan pada standar proses BSNP (2007) yang menyatakan bahwa indikator adalah perilaku yang dapat diukur dan/atau diobservasi untuk menunjukkan ketercapaian kompetensi dasar tertentu yang menjadi acuan penilaian mata pelajaran yang dirumuskan dengan menggunakan kata kerja operasional yang dapat diamati, diukur, yang mencakup pengetahuan, sikap, dan keterampilan.

Sedangkan untuk validitas isi diperoleh nilai CVI sebesar 1 yang menunjukkan bahwa alat ukur yang dikembangkan telah mempunyai validitas isi karena alat ukur tersebut dapat mengukur hal-hal yang mewakili keseluruhan isi bahan pelajaran yang akan di ukurnya. Nilai CVI sebesar 1 menunjukkan bahwa validitas isi yang dicapai tinggi sehingga pada alat ukur tersebut, materi tesnya representatif (mewakili) semua pengetahuan yang di ajarkan, dari segi lingkup maupun proses penalaran. Hal ini sesuai dengan penelaahan kajian materi terhadap bahan ajar yang telah dikembangkan pada penelitian sebelumnya yang sudah melalui proses validasi dan dinyatakan valid.

Validitas Empiris

Hasil validitas empiris diperoleh korelasi skor butir dengan skor total sebanyak 40 %

soal sangat signifikan, 20% soal signifikan, dan 40% soal tidak signifikan. Hasil ini dapat dilihat pada gambar di bawah ini:



Gambar 1. Grafik Korelasi Butir Soal dengan Skor Total pada Alat Ukur Penilaian Literasi Sains yang Dikembangkan

Soal yang signifikan adalah soal yang telah valid secara empiris. Hal ini dapat diketahui dari nilai korelasi antara skor butir dengan skor total yang dibandingkan dengan tabel harga kritik $r_{\text{product moment}}$. Untuk jumlah butir soal sebanyak 40 soal, nilai r_{tabel} adalah 0,304.

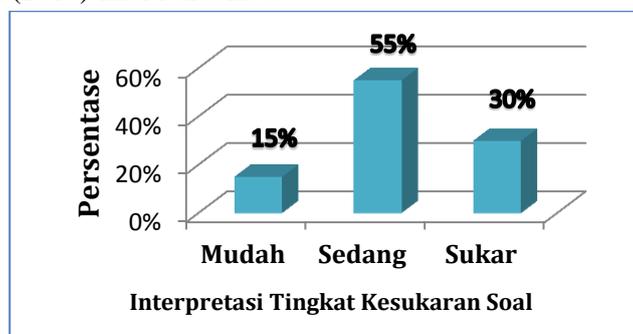
Reliabilitas

Hasil pengolahan data menggunakan software Anates Versi 4.0.2 diperoleh nilai reliabilitas sebesar 0,73. Berdasarkan kriteria reliabilitas menurut Subana (2005) angka 0,73 menunjukkan bahwa reliabilitas soal tinggi. Menurut Firman (2013) nilai reliabilitas

yang tinggi ini menunjukkan bahwa pengukuran yang dilakukan berulang-ulang dengan alat ukur yang telah dikembangkan terhadap subjek yang sama dalam kondisi yang sama akan menghasilkan informasi yang sama atau mendekati sama.

Tingkat Kesukaran

Berdasarkan hasil pengolahan data menggunakan software Anates Versi 4.0.2 diperoleh tingkat kesukaran soal sebanyak 15% soal mudah, 55% soal sedang, dan 30% soal susah. Hasil ini dapat dilihat pada gambar di bawah ini:

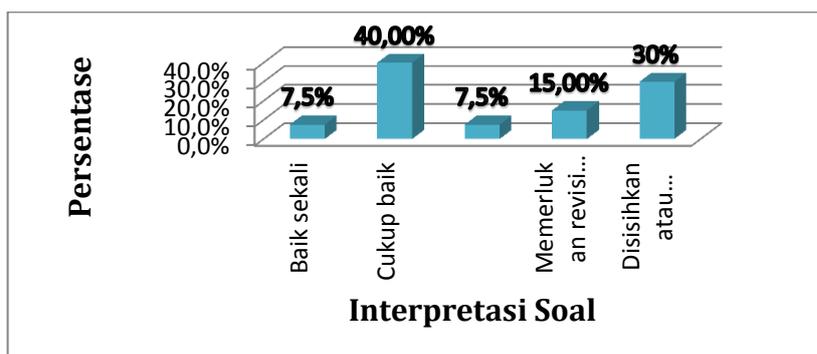


Gambar 2. Grafik Tingkat Kesukaran Butir Soal Alat Ukur Penilaian Literasi Sains yang Dikembangkan

Dari grafik di atas, dapat diketahui bahwa proporsi tingkat kesukaran soal yang dikembangkan memiliki proporsi yang belum memenuhi kurva normal. Menurut Ekawatiningsih, untuk menyusun suatu naskah ujian sebaiknya digunakan butir soal yang mempunyai tingkat kesukaran berimbang, yaitu soal berkategori sukar sebanyak 25%, kategori sedang 50%, dan kategori mudah 25%. Untuk data di atas, perlu dilakukan penambahan soal dengan kategori mudah dan pengurangan soal kategori sedang dan sukar, sehingga diperoleh proporsi yang berimbang.

Daya Pembeda

Pada umumnya daya pembeda butir soal pada alat ukur literasi sains yang dikembangkan berada pada kategori cukup baik dengan persentase sebesar 40%. Namun, ada juga soal yang sebaiknya disisihkan atau direvisi total karena memiliki nilai daya pembeda yang kecil yaitu sebesar 30% atau terdapat 12 soal, yaitu soal nomor 16, 17, 18, 20, 22, 23, 26, 27, 28, 31, 32, dan 37. Hal ini karena keduabelas soal tersebut tidak dapat membedakan siswa yang memiliki kemampuan tinggi ataupun kemampuan rendah. Adapun daya pembeda butir soal dari alat ukur yang dikembangkan secara keseluruhan dapat dilihat pada Gambar 3.

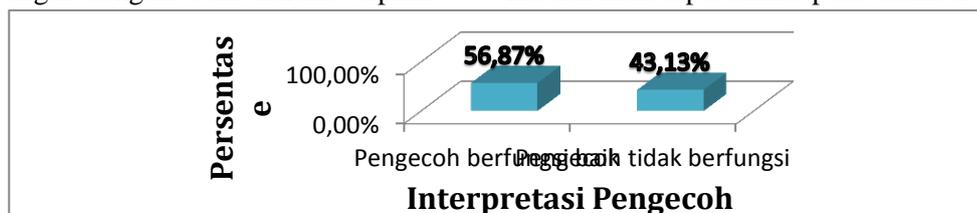


Gambar 3. Grafik Daya Pembeda Butir Soal Alat Ukur Penilaian Literasi Sains yang Dikembangkan

Efektifitas Pengecoh

Berdasarkan data hasil analisis keefektifan pengecoh, dari 160 butir distraktor atau pengecoh option dari alat ukur literasi sains yang dikembangkan terdapat sebanyak 91 butir distraktor yang berfungsi baik, dengan kata lain terdapat 56,9 % distraktor yang berfungsi dengan baik karena dipilih

oleh paling sedikit 5% dari siswa yang mengikuti tes dan terdapat 69 butir distraktor atau 43,2 % distraktor yang tidak berfungsi dengan baik karena dipilih kurang dari 5% siswa yang mengikuti tes. Sebaran pengecoh atau distraktor option dari butir soal alat ukur literasi sains yang dikembangkan secara keseluruhan dapat dilihat pada Gambar 4.



Gambar 4. Grafik Efektifitas Pengecoh pada Butir Soal Alat Ukur Penilaian Literasi Sains yang Dikembangkan

SIMPULAN

Soal literasi sains yang dikembangkan pada konten struktur atom dan ikatan kimia menggunakan konteks wayang kulit dapat dikatakan valid ditinjau dari validitas isi dan validitas konstruk Berdasarkan hasil uji coba yang dilakukan di salah satu SMA Negeri di Kota Bandung, soal literasi sains yang dikembangkan pada konten struktur atom dan ikatan kimia menggunakan konteks wayang kulit secara umum memiliki nilai reliabilitas dengan kategori tinggi, validitas empiris (korelasi skor butir dengan skor total) dikatakan valid, tingkat kesukaran meliputi soal mudah, sedang dan sukar dengan perlu perbaikan proporsi tingkat kesukaran, daya pembeda cukup baik sehingga dapat membedakan siswa kemampuan tinggi dan rendah, dan ada pengecoh yang berfungsi dengan baik namun adapula pengecoh yang tidak berfungsi dengan baik sehingga diperlukan perbaikan atau revisi.

DAFTAR RUJUKAN

- Ahkamajaya, H. (2014). *Pengembangan buku ajar materi struktur atom dan ikatan kimia menggunakan konteks wayang kulit untuk meningkatkan literasi sains dan teknologi Siswa*. (Tesis). Sekolah Pascasarjana, Universitas Pendidikan Indonesia, Bandung
- Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan edisi 2*. Jakarta: PT. Bumi Aksara.
- Ekawatiningsih, P. (t.t). Materi Kuliah Evaluasi Pembelajaran: Analisis Butir Soal. [online]. diakses dari: <http://staff.uny.ac.id/sites/default/files/pendidikan/Prihastuti%20Ekawatiningsih,%20S.Pd.,M.Pd./22.%20Materi%20Kuliah%20Evaluasi%20Pembelajaran.pdf>
- Firman, H. (2013). *Evaluasi pembelajaran kimia*. Bandung: Jurusan Pendidikan Kimia FPMIPA UPI.
- Hamzah. (2012). *Assesment Pembelajaran*. Jakarta: Bumi Aksara.
- Kurniawan, K, *et al.* (2006). Prototipe percobaan rutherford sebagai alat peraga pembelajaran model atom rutherford untuk meningkatkan hasil belajar siswa di SMA 2 kendal. *Kumpulan Makalah PKMI* (hlm. 362-369). Malang: Student Center Kampus III UMM.
- Litbang, Kemdikbud. (2011). *Survei Internasional PISA*. [Online] . diakses dari <http://litbang.kemdikbud.go.id/index.php/survei-internasional-pisa>.
- OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.
- Purwanto. (2010). *Metodologi Penelitian Kuantitatif untuk Psikologi dan Pendidikan*. Yogyakarta: Pustaka Pelajar.
- Sirhan, G. (2007). Learning Difficulties in Chemistry: An Overview. *Journal of Turkish Science Education*, 4(2), 2-20.
- Subana, dkk. (2005). *Dasar-Dasar Penelitian Ilmiah*. Bandung: Pustaka Setia.
- Sudjana. (2011). *Penilaian hasil proses belajar mengajar*. Bandung: PT. Remaja Rosdakarya.

Sukardi. (2008). *Evaluasi pendidikan; prinsip dan operasionalnya*. Jakarta: Bumi Aksara.

Wilson, R., Pan, W., & Schumsky, D. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197-210.

Yunengsih, Y. (2008) . *Ujian Nasional: Dapatkah Menjadi Tolak Ukur Standar Nasional Pendidikan (Hasil Kajian Ujian Nasional Matematika pada Sekolah Menengah Pertama)*. Jakarta: Sampoerna Foundation