

Komparasi Hasil Analisis Butir Soal Literasi Sains Budaya Bali menggunakan Program Berbeda

A. A. Istri Rai Sudiatmika

Pendidikan Fisika Universitas Pendidikan Ganesha, Singaraja Bali

Korespondensi: Jl. Pulau Adi I/18, Denpasar, Bali 80113. Email: r_sudiatmika@yahoo.co.id

N.Y. Rustaman

MIPA Pascasarjana Universitas Pendidikan Indonesia

Korespondensi: Jl. Curug Kencana 63-64 Kopo Bandung. Email: nuryani_rustaman@yahoo.com

A. Zainul

MIPA Pascasarjana Universitas Pendidikan Indonesia

Korespondensi: Jl. Gunung Batu 117 J-6 Bandung 40175. Email: perbo@indo.net.id

Abstract: A comparative study to contrast the results of test items analysis on scientific literacy in the context of Balinese Culture was carried out by using Anates program 4.0 version and Iteman program 3.00 version. A number of students of grade IX of Public Junior High Schools in Bali province was involved as research subjects ($n = 1648$). Using both Anates and Iteman, the study resulted in different results on degree of difficulty, discrimination index, item validity, and instrument reliability, but still in the same category range, except the index of discrimination of two test items. Result difference that is showed by both programe cause of data percentace analysed is different (Anates: 54%; Iteman: 100%). In Anates analysis program the quality of the distractor was detected from the number of distractor. Meanwhile in Iteman analysis program, it was detected from discrimination index for each distractor. Discrimination index can clearly show the quality of the distractors, as it can distinguish the higher and lower groups. A good distractor from the lower group is one with negative discrimination index. The result showed that 35 items that had been developed are good. The instrument reliability of test is high.

Kata kunci: komparasi, analisis butir soal, literasi sains, budaya Bali, anates, iteman.

Selama beberapa tahun ini telah dilaksanakan penelitian-penelitian tentang asesmen. Di Amerika Serikat diselenggarakan penelitian *National Assessment of Educational Progress (NAEP)* dan *Progress in International Reading Literacy Study (PIRLS)*. Penelitian NAEP dilaksanakan secara berkala di kelas 4, kelas 8 dan kelas 12. Materi yang diujikan antara lain: kesenian, kewarganegaraan, ekonomi, geografi, sejarah, membaca, menulis, matematika, dan sains (*National Assessment of Educational Progress, 2010*). Penilaiannya menekankan pada pencapaian materi subyek, pengalaman pengajaran dan lingkungan sekolah. Penelitian PIRLS dilaksanakan di kelas 4 yang menekankan pada prestasi membaca, membaca

prilaku dan sikap siswa (<http://nces.ed.gov/surveys/PIRLS/>). Pada level internasional penelitian tentang asesmen diselenggarakan oleh IEA (*International Association for the Evaluation of Educational Achievement*) dengan TIMSS (*Trend International Mathematics and Science Study*) dan OECD (*Organization for Economic Cooperation and Development*) dengan PISA (*Programme for International Student Assessment*). TIMSS meliputi bidang matematika dan sains untuk siswa kelas 4 dan kelas 8 (Martin, *et.al.*, 2003). PISA menguji tentang Literasi Membaca (*Reading Literacy*), Literasi Matematika (*Mathematical Literacy*) dan Literasi Sains (*Scientific Literacy*) siswa usia 15 tahun (*Organization for Economic Cooperation and Development, 2003*).



Selain penelitian tentang NAEP, PIRLS, TIMSS, dan PISA juga dilakukan penelitian untuk melihat komparasi hasil belajarnya. Penelitian-penelitian tersebut antara lain: *A Comparison of PISA and TIMSS 2003 achievement results in Mathematics and Science* (Wu, 2008). Penelitian ini membandingkan nilai hasil belajar matematika dan sains siswa kelas 8 tahun 2003. Berikutnya, *A Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments*, membandingkan nilai hasil belajar membaca siswa kelas 4 tahun ajaran 2002. Selanjutnya, *Comparing NAEP, TIMSS, and PISA in Mathematics and Science* membandingkan interpretasi rerata hasil matematika dan sains melalui butir soal TIMSS dan PISA. Terakhir, *Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science* yang menginterpretasikan hasil dari kedua penilaian (PIRLS dengan PISA), dan kemudian membandingkan hasilnya dengan penemuan NAEP.

Dari semua penelitian tersebut di atas, tidak satupun mengkomparasikan hasilnya apakah memberikan hasil interpretasi yang sama jika dianalisis dengan program yang berbeda. Untuk itu, dalam penelitian ini dibandingkan hasil analisis butir soal literasi sains dalam konteks budaya Bali menggunakan program anates versi 4.0 (Karno & Wibisono, 2003) dan program iteman versi 3.00. Hal ini dimaksudkan untuk melihat apakah kedua program tersebut memberikan hasil interpretasi yang tepat dalam mengeliminasi butir soal sehingga diperoleh butir soal yang valid dan reliabel. Selain itu, komparasi kedua program juga dimaksudkan untuk memberikan gambaran apakah kedua program adalah tepat digunakan dalam menganalisis butir soal. Butir soal literasi sains dalam konteks budaya Bali yang akan dianalisis memiliki karakteristik lebih eksplisit terhadap ilmu pengetahuan, pemahaman dan keterampilan yang berfungsi efektif dalam kehidupan sehari-hari.

Karena siswa kelas IX SMP berada pada usia sekitar 14-15 tahun, maka alat ukur literasi sains untuk keperluan asesmen dibuat untuk siswa kelas IX SMP dengan dimensi literasi sains diadopsi dari PISA sebagai *benchmark*, tetapi dalam penyusunan butir soalnya mempertimbangkan kedekatan dengan siswa sesuai jenjang sekolahnya yaitu jenjang sekolah SMP. Ada tiga dimensi sains yang dikembangkan dalam literasi sains budaya Bali antara lain: isi sains (*science*

content), proses sains (*scientific processes*) dan konteks sains (*scientific context*).

Secara rinci ketiga dimensi sains menurut PISA (*Organization for Economic Cooperation and Development*, 1999: 62-63) dapat diuraikan sebagai berikut.

- 1) Isi Sains (*scientific content*): meliputi pemahaman konseptual yang dibutuhkan dalam penggunaan proses-proses.
- 2) Proses sains (*scientific processes*): meliputi proses mental yang terlibat untuk menyoroti:
 - a. Mengenali masalah-masalah yang dapat diselidiki secara ilmiah (*recognizing scientifically investigable questions*).
 - b. Mengidentifikasi bukti yang dibutuhkan dalam penyelidikan ilmiah (*identifying evidence needed in a scientific investigation*).
 - c. Menarik atau mengevaluasi kesimpulan-kesimpulan (*drawing or evaluating conclusions*).
 - d. Mengomunikasikan kesimpulan-kesimpulan valid (*communicating valid conclusions*).
 - e. Mendemonstrasikan pemahaman konsep-konsep sains (*demonstrating a understanding of scientific concepts*).

Dalam kaitannya dengan literasi sains budaya Bali, maka proses literasi sains budaya Bali yang dikembangkan melibatkan kemampuan (*ability*) literasi sains, baik sains asli Bali maupun yang telah berakulturasi dengan sains Veda dan sains modern. Literasi sains budaya Bali tersebut meliputi pengetahuan sains dan pemahaman konseptual yang dibutuhkan dalam penggunaan proses mental yang dihadirkan dalam kondisi atau lingkungan yang realistik bagi siswa yang disesuaikan dengan level perkembangan anak usia SMP.

- 3) Konteks sains (*scientific context*): meliputi aplikasi pengetahuan sains dan penggunaan proses-proses sains dalam situasi-situasi yang mencerminkan kondisi atau lingkungan yang realistik/nyata. Konteks sains yang dimaksud adalah sains lokal, sains nasional dan sains global. Sains lokal Bali meliputi sains Bali asli dan sains Weda yang menjunjung nilai-nilai agama. Sains nasional meliputi sains lokal yang juga umum digunakan secara nasional. Sains global meliputi sains modern yang diakulturasi dari budaya Barat.



METODE

Terdapat 35 butir soal literasi sains konteks budaya Bali yang diujikan dalam penelitian ini. Sebaran butir soal yang diujikan disajikan pada tabel 1.

Tingkat Kesukaran

Tingkat kesukaran (P) merupakan proporsi peserta tes yang menjawab benar (Crocker & Algina, 1986: 329).

Kriteria tingkat kesukaran (Zainul, 1997: 160):

- 0 % - 25 % Sukar
- 26 % - 75 % Sedang
- 76 % - 100 % Mudah

Tabel 1 Sebaran Butir Soal Literasi Sains yang Diujikan

Konteks Proses	Sains Lokal/ SL		Sains Nasional/SN		Sains Global/SG		Total
	No. butir	Σ	No. butir	Σ	No. butir	Σ	
M	2, 36, 38	3	26, 34	2	24	1	8
I	4, 20, 37	3	18, 33	2	13, 14	2	7
S	8, 11, 17, 21	4	3, 19, 32	3	12	1	8
K	6, 22, 23, 27, 28	5	-	-	9, 16, 39	3	8
D	5, 10, 15, 29	4	35	1	1	1	6
Total		19		8		8	35

Keterangan Proses Sains:

- M = Mengenal masalah-masalah yang dapat diselidiki secara ilmiah
- I = Mengidentifikasi bukti yang dibutuhkan dalam penyelidikan ilmiah
- S = Menarik atau mengevaluasi kesimpulan-kesimpulan
- K = Mengomunikasikan kesimpulan-kesimpulan valid
- D = Mendemonstrasikan pemahaman konsep-konsep sains

Peserta yang terlibat dalam penelitian ini adalah siswa kelas IX SMP Negeri yang ada di Kabupaten Klungkung, Kabupaten Buleleng, dan Kota Denpasar ($n = 1648$). Laporan ini terpusat pada komparasi analisis butir soal ditinjau dari tingkat kesukaran, daya pembeda, validitas butir soal, reliabilitas instrumen tes (naskah soal), serta kualitas pengecohnya menggunakan program Anates versi 4.0 dan program Iteman versi 3.0. Selain itu untuk menentukan dieliminasi tidaknya butir soal tersebut seharusnya memenuhi lima kriteria sebagai berikut.

Daya Pembeda/Indeks Diskriminasi

Daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai/berkemampuan tinggi dengan siswa yang bodoh/berkemampuan rendah (Arifin, 2008). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi (D). Indeks diskriminasi (D) merupakan selisih antara proporsi kelompok atas (peserta tes berkemampuan tinggi) yang menjawab tes dengan benar (p_u) dengan proporsi kelompok bawah (peserta tes berkemampuan rendah) yang menjawab tes dengan benar (Crocker & Algina, 1986: 314).

Tabel 2. Kriteria dan Interpretasi Daya Pembeda

No.	Daya Pembeda (D)	Kategori	Interpretasi
1.	$D \geq 0,40$	Sangat baik	Butir soal berfungsi dengan baik
2.	$0,30 \leq D \leq 0,39$	Baik	Butir soal sedikit atau tidak perlu ada revisi
3.	$0,20 \leq D \leq 0,29$	Sedang	Butir soal direvisi
4.	$D \leq 0,19$	Lemah/jelek	Butir soal dibuang
5.	Negatif	Jelek sekali	Butir soal dibuang



Validitas Butir Soal

Validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu instrumen pengukur (butir soal) dalam melakukan fungsi ukurnya. Berdasarkan definisi tersebut dapat dikemukakan bahwa uji validitas butir dimaksudkan untuk melihat apakah item pertanyaan (butir soal) yang dipergunakan mampu mengases tingkat literasi sains siswa kelas IX SMP Negeri di Propinsi Bali.

Validitas (kesahihan) butir soal literasi sains diuji dengan menggunakan rumus korelasi *point bi serial* r_{pbi} (*point biserial correlation*) yaitu korelasi antar skor butir item dengan skor total. Butir soal dikatakan valid jika hasil perhitungan $>$ tabel. Berdasarkan daftar *Pearson Product Moment correlation coefficient* pada taraf signifikansi 5%, maka $=$ dengan $df = (n-2)$ untuk $n = 1648$ diperoleh $= 0,062$ (Guilford & Fruchter, 1978).

Reliabilitas instrumen tes

Naskah soal dikatakan reliabel (handal), jika tes memberikan hasil yang sama untuk kelompok yang sama yang diberikan tes pada waktu dan situasi yang berbeda (Slameto, 2001). Analisis reliabilitas tes menggunakan teknik *Alpha Cronbach*. Menurut Yu (2010), *Alpha Cronbach* dapat digunakan untuk tipe data biner maupun data bersekala besar.

Kriteria reliabilitas Naskah Soal (Russeffendi, 2001: 144)

No.	Nilai $\hat{\alpha}$	Kualifikasi Reliabilitas Naskah Soal
1.	0,00-0,39	rendah
2.	0,40-0,59	sedang
3.	0,60-1,00	tinggi

Kualitas Pengecoh

Ketika membicarakan soal pilihan ganda, setiap butir soal yang dibuat dilengkapi dengan beberapa

pilihan jawaban. Salah satu pilihan tersebut merupakan pilihan jawaban benar/terbaik dan yang lainnya merupakan pengecoh sebagai pilihan jawaban yang salah. Soal pilihan ganda yang dibuat untuk mengukur tingkat literasi sains siswa SMP terdiri dari 4 pilihan jawaban (*option*), sehingga satu *option* merupakan kunci jawaban dan tiga *option* lainnya berfungsi sebagai pengecoh. Pengecoh berfungsi dengan baik jika peserta tes dari kelompok bawah (peserta tes berkemampuan rendah) memilih pilihan jawaban yang salah lebih banyak daripada siswa kelompok atas (peserta tes yang berkemampuan tinggi). Jadi pengecoh yang baik jika yang terkecoh adalah kelompok bawah.

Berdasarkan uraian di atas, selain uji reliabilitas naskah soal dan kualitas pengecoh, validasi butir soal didasarkan pada analisis daya pembeda dan validitas butir soal. Pengambilan keputusan validasi butir soal disajikan pada tabel 3.

Kelima kriteria dalam mengeliminasi butir soal seperti di uraikan di atas, dua diantaranya yaitu: daya pembeda dan validitas butir soal memegang peranan penting di dalam pengambilan keputusan apakah butir soal itu diterima tanpa revisi, diterima dengan revisi, atau harus dibuang (disajikan pada tabel 3).

Alasan kriteria tingkat kesukaran, reliabilitas instrumen tes dan kualitas pengecoh tidak dijadikan patokan dalam pengambilan keputusan mengeliminasi butir soal karena: (1) Tingkat kesukaran butir soal merupakan seleksi lebih lanjut implementasi tes untuk keperluan apa tes tersebut diujikan kepada siswa. Misalnya apakah tes diujikan untuk penerimaan siswa baru atau untuk tingkat penguasaan kemampuan literasi sains siswa. Jika untuk mengukur tingkat kemampuan literasi sains siswa, apakah acuan penilaian menggunakan PAP (Penilaian Acuan Patokan) atau PAN (Penilaian Acuan Normal). Setelah itu baru ditentukan persentase jumlah butir soal mudah, sedang, dan sukar yang diperlukan untuk implementasi pengujian. (2) Instrumen tes yang reliabel belum tentu valid, sedangkan seluruh butir

Tabel 3 Kriteria Pengambilan Keputusan Butir Soal

Daya Pembeda (D)	Validitas butir soal	Kriteria Pengambilan Keputusan
$D \geq 0,40$	Signifikan/tidak	Butir soal bersangkutan tetap dipakai
$0,30 \leq D \leq 0,39$	Signifikan	Butir soal bersangkutan tetap dipakai
$0,30 \leq D \leq 0,39$	Tidak signifikan	butir soal bersangkutan direvisi
$0,20 \leq D \leq 0,29$	Signifikan/tidak	butir soal bersangkutan direvisi
$D \leq 0,19$	Signifikan/tidak	Soal sebaiknya dibuang



soal yang valid sudah tentu reliabel. Jadi kriteria validitas butir soal sudah mewakili kriteria reliabilitas instrumen tes. (3) Walaupun daya pembeda dan validitas butir soal digunakan sebagai pengambilan keputusan butir soal dapat diterima dengan atau tanpa revisi, tetap harus dicek hasil statistik kualitas pengecoh yang direvisi. Setiap pengecoh yang seharusnya direvisi tidak harus direvisi jika dalam menarasikan pengecoh dilakukan dengan mengubah narasi kunci jawaban yang mengikuti pola struktur konseptual yang benar.

HASIL DAN PEMBAHASAN

Dari hasil analisis tingkat kesukaran, daya pembeda, dan validitas butir soal yang ditunjukkan pada tabel 4 terlihat bahwa: a) Komparasi tingkat kesukaran butir soal hasilnya tidak berbeda jauh (berada pada kategori yang sama); b) Komparasi daya pembeda butir soal hasilnya tidak berbeda jauh (berada pada kategori yang sama) kecuali dua butir soal nomor 9 dan 13; c) Komparasi validitas butir soal hasilnya tidak ada perbedaan yang berarti (semua butir soal kategorinya signifikan); d) Hasil komparasi reliabilitas keseluruhan naskah butir soal tidak berbeda jauh, sama-sama berada pada rentang 0,60-1,00 (kualifikasi reliabilitas yang tinggi).

Perbedaan hasil tingkat kesukaran, daya pembeda, validitas butir soal dan reliabilitas naskah

soal disebabkan karena kedua program dalam analisisnya menggunakan bobot persentase data kelompok atas dan kelompok bawah yang berbeda (Anates: 54%; Iteman: 100%). Penggunaan bobot persentase data yang berbeda (Anates: 54%; Iteman: 100%) menyebabkan proporsi jumlah jawaban yang benar hasilnya berbeda untuk nomor butir yang sama baik dari kelompok atas maupun kelompok bawah, sehingga hasil daya pembeda butir soal berbeda. Makin besar perbedaan proporsi jumlah jawaban yang benar untuk nomor butir yang sama, makin besar pula perbedaan hasil daya pembeda yang ditunjukkan oleh kedua program tersebut termasuk kategorinya juga akan berbeda seperti yang ditunjukkan butir soal nomor 9 dan 13. Selain itu, hasil analisis butir soal khususnya pada pengkategorian validitas butir soal menggunakan program Anates perlu harus di cek ulang mengingat hasil pengkategorian setiap butir soal masih ada yang salah. Meskipun masih ada perbedaan hasil seperti diuraikan di atas, seluruh naskah butir soal memiliki reliabilitas yang tinggi dan semua butir soal dapat dipakai (tidak dieliminasi).

Selanjutnya dilakukan analisis kualitas pengecoh. Analisis kualitas pengecoh menggunakan program Anates hanya ditunjukkan berdasarkan tingkat baik buruknya pengecoh yang didasarkan pada banyaknya siswa yang memilih masing-masing pilihan jawaban/option. Hasil tersebut tidak memperlihatkan kelemahan dari masing-masing pengecoh karena

Tabel 4. Komparasi hasil analisis tingkat kesukaran, daya pembeda, dan validitas butir soal menggunakan program Anates versi 4.0 dan Iteman versi 3.0.

Konteks dan Proses Sains	Jml. Betul	Tingkat Kesukaran / Prop. Correct (%)			Daya Pembeda			Validitas Butir Soal [$r_{pbi} = (\alpha) 5\% = 0,062$]			Keputusan		
		Anates	Iteman	Kategori	Anates	Iteman	Kategori	Anates	Iteman	Kategori	1	2	3
SGD-01.	1018	62,56	62,5	Sedang	0,5281	0,485	SB	0,383	0,380	Valid	√		
SLM-02	692	49,15	49,3	Sedang	0,5618	0,600	SB	0,483	0,479	Valid	√		
SNS-03.	759	55,64	55,6	Sedang	0,6292	0,599	SB	0,479	0,476	Valid	√		
SLI-04.	712	42,66	42,7	Sedang	0,4045	0,403	SB	0,321	0,319	Valid	√		
SLK-05.	690	47,51	47,5	Sedang	0,5798	0,558	SB	0,449	0,445	Valid	√		
SLD-06.	380	26,82	26,7	Sedang	0,5079	0,647	SB	0,484	0,481	Valid	√		
SLS-07	793	53,40	53,4	Sedang	0,6899	0,674	SB	0,539	0,537	Valid	√		
SGK-08.	447	28,22	28,3	Sedang	0,4494	0,625	SB	0,470	0,469	Valid	√		
SLD-9.	1406	84,95	84,8	Mudah	0,3764	0,486	B/SB	0,322	0,319	Valid	√		
SGS-10.	1104	67,72	67,5	Sedang	0,4787	0,506	SB	0,387	0,389	Valid	√		
SGS-11.	1061	67,72	68,0	Sedang	0,4067	0,430	SB	0,328	0,329	Valid	√		
SGI-12.	647	42,23	42,2	Sedang	0,4022	0,443	SB	0,352	0,351	Valid	√		
SGI-13.	1319	83,50	82,8	Mudah	0,3708	0,524	B/SB	0,383	0,354	Valid	√		



Konteks dan Proses Sains	Jml. Betul	Tingkat Kesukaran / Prop. Correct (%)			Daya Pembeda			Validitas Butir Soal [$r_{pbi} = (\alpha) 5\% = 0,062$]			Keputusan		
		Anates	Iteman	Kategori	Anates	Iteman	Kategori	Anates	Iteman	Kategori	1	2	3
SLD-14.	972	61,59	62,0	Sedang	0,4067	0,503	SB	0,391	0,395	Valid	√		
SGK-15.	981	62,14	62,7	Sedang	0,7213	0,767	SB	0,600	0,601	Valid	√		
SLS-16	1162	74,51	75,0	Sedang	0,4652	0,592	SB	0,426	0,434	Valid	√		
SNI-17.	618	37,86	38,0	Sedang	0,4719	0,492	SB	0,390	0,386	Valid	√		
SNS-18.	509	36,10	36,2	Sedang	0,6000	0,601	SB	0,470	0,468	Valid	√		
SLI-19.	453	31,67	31,7	Sedang	0,4719	0,548	SB	0,421	0,420	Valid	√		
SLS-20.	345	24,64	24,6	Sukar	0,5730	0,742	SB	0,543	0,543	Valid	√		
SLK-21.	939	61,71	61,6	Sedang	0,5618	0,637	SB	0,500	0,500	Valid	√		
SLK-22.	924	61,35	61,3	Sedang	0,5910	0,649	SB	0,513	0,510	Valid	√		
SGM-23.	839	57,89	57,9	Sedang	0,6427	0,613	SB	0,487	0,485	Valid	√		
SLK-24.	527	33,43	33,7	Sedang	0,4045	0,522	SB	0,400	0,403	Valid	√		
SLK-25.	1266	78,09	77,9	Mudah	0,4022	0,548	SB	0,393	0,392	Valid	√		
SLD-26.	841	56,55	56,7	Sedang	0,5348	0,590	SB	0,468	0,468	Valid	√		
SNS-27.	1061	69,78	70,0	Sedang	0,6449	0,732	SB	0,554	0,555	valid	√		
SNI-28.	577	41,26	41,1	Sedang	0,6629	0,613	SB	0,485	0,484	Valid	√		
SNM-29.	745	51,70	51,9	Sedang	0,4652	0,515	SB	0,416	0,411	Valid	√		
SND-30.	746	45,75	45,8	Sedang	0,4090	0,467	SB	0,373	0,372	Valid	√		
SLM-31.	1252	78,40	78,5	Mudah	0,5955	0,760	SB	0,541	0,541	Valid	√		
SLI-32.	419	30,04	30,1	Sedang	0,5438	0,611	SB	0,467	0,464	Valid	√		
SLM-33.	721	44,42	44,4	Sedang	0,4022	0,405	SB	0,318	0,314	Valid	√		
SGK-34.	823	53,46	53,3	Sedang	0,5955	0,613	SB	0,488	0,488	Valid	√		
SNM-35.	673	42,42	42,4	Sedang	0,5506	0,502	SB	0,396	0,398	Valid	√		

Keterangan:

- Daya pembeda: SB = sangat baik; B= Baik (butir soal nomor 9 dan 13, hasil daya pembeda ditulis B/SB artinya dianalisis dengan program anates termasuk kategori baik, tetapi dianalisis dengan program iteman termasuk kategori sangat baik.
- Keputusan: 1 = diterima tanpa revisi; 2 = diterima dengan revisi; 3 = dibuang
- Reliabilitas naskah soal ($\hat{\alpha}$) = 0,83 (program Anates) dan ($\hat{\alpha}$) = 0,876 (program Iteman).

tidak secara langsung menunjukkan yang terkecoh itu dari siswa kelompok atas atau kelompok bawah. Pada hasil analisis butir soal nomor 1 juga tidak tertera kode kunci jawaban. Berarti analisis butir soal dengan program Anates belum sempurna. Pengecoh berfungsi dengan baik jika yang terkecoh adalah siswa dari kelompok bawah. Berarti pengecoh berfungsi dengan baik jika daya pembeda setiap pengecoh (selisih antara proporsi kelompok atas dengan proporsi kelompok bawah yang memilih setiap pengecoh) adalah negatif. Hasil analisis kualitas pengecoh menggunakan program Iteman seperti dipaparkan pada lampiran 2 (kolom *alternative statistics*, sub kolom *Biser*), secara jelas memperlihatkan kualitas masing-masing pengecoh karena mampu membedakan siswa yang terkecoh apakah siswa dari kelompok atas (peserta tes

berkemampuan tinggi) atau dari kelompok bawah (peserta tes berkemampuan rendah). Hasil analisis kualitas pengecoh menggunakan program Iteman ternyata semua pengecoh berfungsi dengan baik kecuali pengecoh butir soal nomor 1 pengecoh C dan nomor 8 pengecoh B. Selain itu, dengan program Iteman juga mempertanyakan jika daya pembeda kunci jawaban < pengecoh, apakah kunci jawaban atau *option* pengecoh yang salah.

SIMPULAN DAN SARAN

Berdasarkan uraian di atas dapat disimpulkan beberapa hal. Temuan analisis butir soal literasi sains Budaya Bali menggunakan program Iteman memberikan hasil lebih tepat dibandingkan dengan menggunakan program Anates. Semua butir soal



yang dikembangkan (35 butir soal) dapat diterima tanpa revisi. Instrumen tes memiliki reliabilitas yang tinggi. Bertolak dari kenyataan tersebut, dalam analisis butir soal disarankan menggunakan program IteMan daripada menggunakan program Anates.

DAFTARRUJUKAN

- Anonim, 2002. A Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments. [Online]. Tersedia <http://nces.ed.gov/PUB-SEARCH/pubsinfo.asp?pubid=200310>.
- Arifin, M., 2008. Menganalisis Hasil Test. [Online]. Tersedia: <http://exalute.wordpress.com/2008/06/20/menganalisis-hasil-test/>
- Assessment Systems Corporation, 1988. *Item and Test Analysis Program ITEMAN (tm) Version 3.00*.
- Crocker, L. & Algina, J., 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston, Inc.
- Guilford, J. P. & Fruchter, B., 1978. *Fundamental Statistics in Psychology and Education*. Singapore: McGraw-Hill. [Online]. Tersedia: <http://www.watpon.com/table/pearson.pdf>.
- KarnoTo & Wibisono, Y., 2004. *Program ANATES Pilihan Ganda Versi 4.1.0*.
- Martin, O. et al., 2003. *International Mathematics Report: Finding from IEA's Repeat of the Third International Mathematics and Science Study at the Eight Grade (TIMSS)*. Boston: ISC.
- National Assessment of Educational Progress (NAEP)*, 2010. Overview. <http://nces.ed.gov/nation-sreportcard/about/>.
- Organization for Economic Cooperation and Development (OECD), 2003. *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. [Online]. Tersedia <http://www.pisa.oecd.org/dataoecd/43/9/33690591.pdf>
- Progress in International Reading Literacy Study (PIRLS)*, 2010. <http://nces.ed.gov/surveys/PIRLS>.
- Ruseffendi, E.T., 2001. *Dasar-Dasar Penelitian Pendidikan Dan Bidang Non-Eksata lainnya*. Semarang: IKIP Semarang Press.
- Slameto, 2001. *Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Turner, R. & Adams, R. J., 2007. The programme for international student assessment: An overview. *Journal of Applied Measurement*, 8(3), 237-248.
- Wu, M., 2008. *A Comparison of PISA and TIMSS 2003 achievement results in Mathematics and Science*. [Online]. Tersedia http://www.iea.nl/fileadmin/user_upload/IRC2008/Papers/TIMSS_Mathematics/Wu.pdf
- Yu, A. (t.t), 2010. *Using SAS for Item Analysis and Test Construction*. [Online]. Tersedia: <http://www.creative-wisdom.com/teaching/assessment/alpha.html>.
- Zainul, A., 1997. *Penilaian Hasil Belajar*. Jakarta: Pusat antar Universitas untuk Peningkatan dan Pengembangan Aktivitas Instruksional, Direktur Jenderal Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan.

