

METODE STANDARD SETTING UNTUK UJIAN NASIONAL DI SEKOLAH DASAR

¹⁾Sri Rejeki, ²⁾Djemari Mardapi, ³⁾Kumaidi

¹⁾Universitas PGRI Yogyakarta, ²⁾Universitas Negeri Yogyakarta,

³⁾Universitas Muhammadiyah Surakarta

¹⁾sri.rjk10@yahoo.com, ²⁾djemarimardapi@gmail.com, ³⁾kuma_426@yahoo.com

Abstrak

Penelitian ini bertujuan mengetahui karakteristik hasil *Cut Score* UN di SD tahun 2009 berdasarkan pengembangan implementasi metode Nedelsky, Ebel, *Bookmark* dan hasil modifikasi metode Ebel-*Bookmark*; keyakinan panelis terhadap implementasi metode *Standard Setting* untuk menetapkan *Cut Score*; dan akurasi implementasi berdasarkan standar deviasi masing-masing metode. Sampel meliputi 10 orang panelis pada putaran 1, 16 guru pada putaran 2, 8 pakar dan 1650 hasil pekerjaan siswa. Prosedur dan analisis data mengikuti langkah empat metode yang ditetapkan. Analisis *item* menggunakan program ITEMAN dan Bilog MG dengan 1 Parameter Logistik. Berdasarkan hasil analisis data disimpulkan (1) Rerata *Cut Score* berdasarkan metode Nedelsky untuk mata pelajaran Bahasa Indonesia 28,44, Matematika 23,47, IPA 27. Metode Ebel untuk Bahasa Indonesia 80,75, Matematika 60,21, IPA 77,54. Metode *Bookmark* untuk Bahasa Indonesia sebesar 51,63, Matematika 51,33, IPA 59,16. Modifikasi Ebel-*Bookmark* untuk Bahasa Indonesia 81,27, Matematika 79,06 dan IPA 77,44; (2) keyakinan panelis dalam implementasi metode *Standard Setting* untuk metode *Bookmark* 81,25%, modifikasi Ebel-*Bookmark* 62,50%, metode Ebel 43,75% dan metode Nedelsky 43,75%; (3) Metode *Bookmark* lebih akurat dalam menetapkan *cut score* mata pelajaran Bahasa Indonesia dan Matematika, modifikasi metode Ebel-*Bookmark* lebih akurat untuk menetapkan *cut score* mata pelajaran IPA.

Kata kunci: *standard setting, ujian nasional, sekolah dasar*

THE STANDARD SETTING METHOD FOR THE NATIONAL EXAMINATION IN THE ELEMENTARY SCHOOL

¹⁾Sri Rejeki, ²⁾Djemari Mardapi, ³⁾Kumaidi

¹⁾Universitas PGRI Yogyakarta, ²⁾Universitas Negeri Yogyakarta,

³⁾Universitas Muhammadiyah Surakarta

¹⁾sri.rjk10@yahoo.com, ²⁾djemarimardapi@gmail.com, ³⁾kuma_426@yahoo.com

Abstract

This study aims to investigate the characteristics results in Cut Score of UN elementary school in 2009 based on developing the implementation of Nedelsky, Ebel, Bookmarks methods, and modified Ebel method-Bookmark; panelists' confidence on the implementation of the methods for setting the Standard Setting Cut Score; and the implementation's accuracy based on the standard deviation of each method. The sample included 10 panelists in round 1, 16 teachers in round 2, 8 experts, and 1650 students' work. Procedures and analysis of data followed the steps of four methods specified. Item analysis used the program ITEMAN and Bilog MG with 1 Parameter Logistic. Based on the results of data analysis it is concluded 1). Cut Mean Score based on Nedelsky method for Indonesian is 28.44, Math is 23.47, and Science is 27. Ebel methods for Indonesian is 80.75, Mathematics is 60.21, and Science is 77.54. Bookmark method for Indonesian is 51.63, Mathematics is 51.33, Science is 59.16. Modified Ebel-Bookmark for Indonesian is 81.27, Mathematics is 79.06, and Science is 77.44; 2). panelists confidence in the implementation of the Standard Setting of Bookmark method is 81.25%, Ebel-modification Bookmark is 62.50%, Ebel methods is 43.75%, and methods Nedelsky is 43.75%; 3).The Bookmark method is more accurate in determining the cut score for Indonesian Language and Mathematics, modified Ebel-Bookmark method is more accurate to establish the cut score for the science subject.

Keywords: *Standard Setting, National Examination, elementary school*

Pendahuluan

Undang-Undang Republik Indonesia Nomor 20 Tahun 2003, pada bab XVI pasal 57 sampai dengan 59 tentang Evaluasi menyatakan bahwa dalam rangka pengendalian mutu pendidikan secara nasional dilakukan evaluasi sebagai bentuk akuntabilitas penyelenggara pendidikan kepada pihak-pihak yang berkepentingan. Evaluasi dilakukan oleh lembaga yang mandiri secara berkala, menyeluruh, transparan, dan sistematis untuk menilai pencapaian standar nasional pendidikan dan proses pemantauan evaluasi tersebut harus dilakukan secara berkesinambungan, yang akhirnya akan dapat membenahi mutu pendidikan.

Evaluasi adalah bagian utama dalam proses belajar mengajar, evaluasi di sekolah tidak hanya mengevaluasi siswa tetapi juga kinerja guru dan sekolah, untuk menunjukkan efektivitas sekolah. Ini berarti asesmen dan evaluasi pada gilirannya merupakan bagian dari proses akuntabilitas. Secara tidak langsung perencanaan kegiatan asesmen dan evaluasi memerlukan aturan secara nasional untuk memenuhi tujuan pendidikan nasional dan lokal serta kebutuhan sekolah. Asesmen dan evaluasi diharapkan dapat memberikan informasi dan pertimbangan kelanjutan proses belajar mengajar baik untuk sekolah, daerah maupun pemerintah pusat.

Mengembangkan pembelajaran adalah tujuan utama sekolah. Asesmen merupakan kerangka kerja yang dapat menghasilkan dasar bagi perencanaan langkah-langkah pendidikan selanjutnya dalam merespon kebutuhan siswa. Penilaian pendidikan sebagai proses pengumpulan dan pengolahan informasi hasil evaluasi dalam rangka menentukan pencapaian hasil belajar siswa dan keberhasilan proses pembelajaran. Dalam penilaian pendidikan ada beberapa yang harus diperhatikan baik prinsip penilaian, standar penilaian, mekanisme, prosedur dan instrumen penilaian hasil peserta didik. Prinsip penilaian hasil peserta didik pada jenjang pendidikan dasar dan menengah berdasarkan penjelasan Peraturan

Menteri Pendidikan Nasional No. 20 Tahun 2007(2007, pp.7-8) adalah sah, objektif, adil, terpadu, terbuka, menyeluruh dan berkesinambungan, sistematis, beracuan kriteria dan akuntabel. Mekanisme dan prosedur penilaian hasil belajar dilaksanakan oleh pendidik, satuan pendidikan, dan pemerintah.

Ujian Nasional (UN)/Ujian Akhir Sekolah Berstandar Nasional (UASBN) di Indonesia (dahulu dengan istilah EBTANAS, UNAS, UAN) pada sejarah perkembangan dan implementasinya terdapat beberapa permasalahan. Permasalahan yang muncul adalah masalah administrasi, persepsi *stakeholders*, penetapan batas kelulusan, dan masalah geografis, pemahaman terhadap Kurikulum Tingkat Satuan Pendidikan dan adanya otonomi daerah yang berdampak pada otonomi pendidikan di Indonesia.

Administrasi Ujian Nasional selama ini memang berliku mulai penyusunan kisi-kisi UN, penyusunan soal baik tingkat nasional dan daerah, uji coba, penetapan soal untuk standar nasional dan daerah, cetak perangkat soal dan LJK, pengiriman, distribusi dan pendataan sesampainya di sekolah dan pengiriman kembali hasil pekerjaan peserta ujian. Permasalahan yang sering muncul dari administrasi UN adalah pada kerawanan cetak soal dan LJK, pengiriman dan distribusi soal. Permasalahan lain dari UN adalah persepsi *stakeholders* terhadap pelaksanaan dan hasil UN, seperti adanya pro-kontra Ujian Akhir Nasional dalam masyarakat, sosialisasi tentang batas kelulusan belum merata, masyarakat dalam memahami batas kelulusan belum seperti yang diharapkan dalam kebijakan pendidikan.

Penetapan batas kelulusan hasil UN di Indonesia banyak menuai kritikan karena batas kelulusan di Indonesia seperti di tingkat SMP dan SMA batas kelulusan masih sama, batas kelulusan di tingkat SD ditentukan oleh sekolah, belum dikembangkannya metode penetapan standar berdasarkan prosedur atau metode penetapan *standard setting* secara tunggal (nasional) ataupun tingkat daerah dan adanya bias

dalam penentuan batas kelulusan yang belum berbasis pada SKL (standar kompetensi lulusan) dan KKM (kriteria ketuntasan minimal) dan kurikulum yang berlaku. Di sisi lain, dalam penetapan standar kelulusan di Indonesia belum memanfaatkan metode *standard setting* yang berkembang dalam sistem pengujian dan penilaian.

Geografis Indonesia yang terdiri dari kepulauan juga menjadi permasalahan yang tak dapat dihindarkan dalam UN. Hal ini berkaitan erat dengan pengiriman dan distribusi soal, pemerataan fasilitas pendidikan, kualitas pendidikan yang belum merata dan jarak tempuh peserta ujian dengan sekolah.

Penilaian yang dilakukan seseorang atau suatu lembaga biasanya memiliki tujuan. Tujuan tersebut antara lain (a) menetapkan atau membuat keputusan seseorang peserta ujian lulus atau gagal, (b). menempatkan para peserta ujian pada level tertentu misalnya termasuk kategori dasar, menengah atau lanjut dalam suatu pencapaian kemampuan ketrampilan tertentu. Ranah/isi dan ketrampilan yang telah ditetapkan dalam kurikulum dan kategori kinerja untuk penafsiran skor ujian adalah konsep yang mendasar dalam sistem penilaian pendidikan yang bertujuan untuk menjelaskan apa yang diketahui dan bisa dilakukan oleh para siswa. Informasi hasil ujian juga bermanfaat untuk *stakeholders*.

Penetapan kriteria penilaian memerlukan standar tertentu yang disebut dengan *standard setting (setting performance standard)*. Penentuan standar pendidikan adalah penentuan nilai batas (*cut score*). Seseorang dikatakan sudah lulus/kompeten bila telah melewati nilai batas tersebut berupa nilai batas antara peserta didik yang sudah menguasai kompetensi tertentu dengan peserta didik yang belum menguasai kompetensi tertentu. Bila hal itu terjadi pada ujian nasional atau sekolah maka nilai batas berfungsi untuk memisahkan antara peserta didik yang lulus dan tidak lulus disebut batas kelulusan. Kegiatan penentuan batas kelulusan disebut *standard setting*.

Kata “standar” dapat digunakan dalam kaitannya dengan (1) muatan dan keterampilan yang dipandang perlu untuk dikuasai peserta ujian, dan (2) skor yang perlu mereka dapatkan untuk menunjukkan pengetahuan dan keterampilan yang relevan. Dalam konteks penilaian yang digunakan di lingkungan pendidikan, perbedaan harus dilakukan antara standar isi dan standar prestasi, karena ketidakjelasan mengenai perbedaan antara kedua konsep ini sering kali muncul di kalangan pengambil kebijakan, pendidik, dan masyarakat. Standar isi tercerminkan dalam kurikulum dan menetapkan apa yang semestinya diketahui dan bisa dikerjakan oleh peserta ujian. Standar isi memberikan arahan bagi pengajar mengenai apa yang perlu diajarkan. Standar prestasi dengan demikian dapat dipandang sebagai operasionalisasi standar isi bagi tes atau penilaian yang telah disusun untuk mengukur standar isi.

Secara umum metode *standard setting* diklasifikasikan dalam 3 kelompok (Cizek & Bunch, 2007, p.9) yakni *examine centered versus test centered; holistic models versus analytic models; norm referenced versus criterion referenced*. Menurut Jaeger (Cizek & Bunch, 2007, p.9) metode penentuan standar bisa diklasifikasikan menjadi dua kategori yaitu metode yang terpusat pada peserta ujian dan metode yang terpusat pada ujian. Kane (Cizek & Bunch, 2007, p.9) telah mengemukakan cara untuk mengklasifikasikan metode penentuan standar. Ia mengemukakan bahwa metode bisa diklasifikasikan sebagai model holistik yang berasumsi bahwa prestasi dan keterampilan sangat terintegrasi atau model analitis yang berasumsi bahwa prestasi bisa diukur menggunakan sampel kinerja. Ada tiga hal yang berpengaruh dalam implementasi metode *standard setting* dalam menentukan *cut score*: (1) terminologi mengenai standar, (2) peran panelis dalam menentukan standar prestasi, dan (3) perbedaan dalam metode penetapan standar antarjenis butir (Hambleton, 2006, p.436).

Metode Penelitian

Populasi dalam penelitian ini adalah seluruh hasil UASBN siswa SD kelas VI tahun 2009 dan guru kelas VI SD di DIY. Sampel yang diambil meliputi guru SD sebagai panelis sejumlah 10 guru pada putaran 1, 16 guru pada putaran 2, 8 pakar dan 1650 hasil pekerjaan siswa. Penelitian ini dilakukan mengikuti prosedur dan analisis data dalam metode Nedelsky, Ebel, *Bookmark* dan modifikasi Ebel-*Bookmark*. Analisis *item* menggunakan program ITEMAN dan Bilog MG dengan 1 Parameter Logistik (1PL).

Hasil dan Pembahasan

Prosedur

Prosedur Metode Nedelsky

Pembahasan mengenai kelebihan dan kekurangan metode Nedelsky dalam implementasinya menunjukkan bahwa kelebihan dari metode ini pada saat digunakan, di satu sisi guru mengatakan metode ini lebih teliti karena per *item* ditelaah kunci jawaban yang benar dan kemungkinan jawaban yang salah, sehingga *cut score* yang dihasilkan akan lebih baik, tetapi guru mengalami kesulitan (dari hasil angket, guru/panelis ada yang mengatakan paling sulit) pada saat menetapkan soal yang dianggap benar dan diminta menetapkan distraktor yang ada, waktu yang digunakan untuk telaah soal dalam metode Nedelsky relatif lama. Permasalahan yang lain adalah guru merasa ditest saat mereka menelaah per *item* dan tidak diberi kunci jawaban untuk setiap mata pelajaran, sehingga metode ini di rasa sulit oleh panelis. Implementasi metode Nedelsky memperkenalkan kepada guru bagaimana kriteria batas kelulusan dapat ditetapkan berdasarkan telaah *item* soal ujian.

Prosedur Metode Ebel

Metode Ebel sebenarnya sudah merupakan perbaikan dari metode Nedelsky dan Angoff Ya/Tidak, yaitu panelisnya tidak hanya guru tapi juga ahli, namun panelis mengalami kesulitan saat harus menetapkan bobot kategori, karena harus memberikan

perkiraan berapa persen siswa dimungkinkan akan mampu mengerjakan soal tersebut dalam kategori relevansi dan tingkat kesulitan yang bervariasi. Panelis/guru merasa bahwa memberikan perkiraan dengan persentase akan menimbulkan bias, jika perkiraan tersebut terlalu tinggi atau terlalu rendah, sehingga akan berpengaruh pada *cut score*.

Prosedur Metode Bookmark

Prosedur *Bookmark* dirasa paling mudah bagi guru/panelis dalam menetapkan *cut score*, karena data sudah disiapkan terlebih dahulu, soal sudah tertata dari yang paling mudah ke paling sulit, panelis merasa lebih mudah menentukan dan memasukkan sampai halaman *Bookmark* ke berapa siswa dianggap mampu menjawab pertanyaan dengan benar sebagai penetapan *cut score*. Kelemahan implementasi prosedur *Bookmark* menurut para panelis adalah komposisi soal dari yang termudah ke yang paling sulit memungkinkan ada beberapa indikator dari kemampuan yang diujikan tidak termasuk dalam halaman *Bookmark* yang dipilih oleh panelis sebagai batas kemampuan siswa dalam menjawab soal.

Prosedur Modifikasi Metode Ebel-Bookmark

Kelebihan hasil modifikasi metode Ebel-*Bookmark* diartikan bahwa modifikasi metode ini memberikan kemudahan kepada guru dan ahli dalam menetapkan batas kelulusan untuk suatu kompetensi yang diharapkan, dan dapat menentukan level kinerja peserta ujian, dengan tiga cara. Pertama melihat pada relevansi dan tingkat kesulitan soal yang sudah terurut; kedua batas kelulusan berdasarkan pada tingkat kesulitan butir dan tingkat kemampuan siswa; ketiga penetapan batas kelulusan melibatkan ahli dan guru secara bersama.

Kelemahan dari metode ini adalah dalam mendeteksi kemampuan *examine* hanya dilihat dari jumlah butir yang mampu dijawab benar oleh *examine* dan belum menjelaskan kemampuan di standar kompetensi yang mana secara detail. Selama ini metode *standard setting* dan level kinerja yang ber-

kembang juga belum memberikan informasi yang demikian. Berdasarkan implementasi di lapangan dari prosedur modifikasi Ebel-Bookmark diperoleh informasi soal yang terurut dalam booklet (*OIB/Odered Item Booklet*) dari soal yang paling mudah ke soal yang paling sulit ditelaah oleh guru ada beberapa soal yang masuk kategori tingkat kesulitan “mudah” tetapi ternyata masuk kategori “sulit”, kategori “sedang” tetapi oleh panelis termasuk kategori “sulit”.

Batas Kelulusan

Metode penetapan batas kelulusan (*cut score*) (*cut score*) diklasifikasikan secara umum dalam 3 kelompok (Cizek & Bunch, 2007, p.9) yakni *examine centered versus test centered*; *holistic models versus analytic models*; *norm referenced versus criterion referenced*.

Mata Pelajaran Bahasa Indonesia

Berdasarkan implementasi 4 metode *standard setting* yang berdasar *test centered*, *cut score* untuk mata pelajaran Bahasa Indonesia disajikan pada Tabel 1.

Tabel 1. Perbandingan *Cut Score* Mata pelajaran Bahasa Indonesia ditinjau dari 4 Metode *Standard setting*

	Nedelsky	Ebel	Bookmark	Ebel-Bookmark
Putaran 1	30	80,227	47,553	
Putaran 2	26,875	81,226	55,714	82,135
Rerata	28,4375	80,7465	51,6335	

Mata Pelajaran Ilmu Pengetahuan Alam

Berdasarkan implementasi 4 metode *standard setting* yang berdasar *examine centered* dan *test centered* (Metode *Bookmark*) dan berdasarkan *test centered* (Metode Nedelsky, dan Ebel), *cut score* untuk mata pelajaran IPA dijelaskan pada Tabel 2.

Tabel 2. Perbandingan *Cut Score* Mata Pelajaran IPA ditinjau dari implementasi 4 Metode *Standard setting*

	Nedelsky	Ebel	Bookmark	Ebel-Bookmark
Putaran 1	29	70,854	54,42	
Putaran 2	25	74,54	63,84	77,4359
Rerata	27	77,54	59,155	

Perbandingan *cut score* antarmata pelajaran

Berdasarkan hasil analisis implementasi 5 metode *standard setting* yang dikembangkan diperoleh beberapa data yang dapat dijelaskan dalam beberapa aspek. Salah satunya adalah dengan membandingkan *cut score* 2 mata pelajaran yang diujikan (Mata Pelajaran Bahasa Indonesia dan Ilmu Pengetahuan Alam).

Tabel 3. Perbandingan *cut score* mata pelajaran USBN 2009

Putaran	Metode <i>Standard setting</i>	Mata Pelajaran	
		Bahasa Indonesia	IPA
Putaran 1	Nedelsky	30	29
	Ebel	80,227	70,854
	<i>Bookmark</i>	47,553	54,42
Putaran 2	Nedelsky	26,4375	25
	Ebel	81,266	74,54
	<i>Bookmark</i>	55,714	63,84
Rerata	Ebel- <i>Bookmark</i>	82,135	77,4359
	Nedelsky	28,4378	27
	Ebel	80,7465	77,54
	<i>Bookmark</i>	51,6335	59,155

Kemampuan lulusan

Berdasarkan hasil penelitian, kemampuan lulusan dapat dilihat dari beberapa aspek. Aspek itu antara lain kemampuan lulusan dilihat dari deskriptor implementasi tiap-tiap metode setiap putaran atau reratanya, perbedaan nilai yang diperoleh dalam

setiap mata pelajaran, persentase menjawab benar setiap mata pelajaran, jumlah kelulusan dari setiap mata pelajaran berdasarkan sampel penelitian, pendapat panelis dalam telaah soal.

Kemampuan lulusan ditinjau dari deskriptor.

Berdasarkan implementasi 4 metode *standard setting* dan modifikasi metode *Ebel-Bookmark* kemampuan lulusan dapat dilihat dari deskriptor masing-masing mata pelajaran. Deskriptor berdasarkan implementasi metode *standard setting* dapat dijelaskan sebagai berikut.

1) Metode Nedelsky

Dalam menetapkan deskriptor dilakukan analisis panelis terhadap soal UASBN berkaitan dengan SKL dan indikatornya. Deskriptor diambil berdasarkan jumlah panelis yang menyatakan bahwa suatu soal diperkirakan mampu dijawab benar oleh siswa minimal lebih dari separoh ($1/2$) dari jumlah panelis ($p > 5$) untuk putaran 1 dan ($p > 8$) untuk putaran 2. Berikut ini Tabel deskriptor untuk putaran 1 dan 2 dengan metode Nedelsky.

a) Mata Pelajaran Bahasa Indonesia

Berdasarkan Tabel analisis panelis pada putaran pertama dan kedua kemampuan lulusan dalam ujian bahasa Indonesia berdasar SKL meliputi membaca dan menulis. Kemampuan lulusan dalam membaca meliputi menentukan isi bacaan (kecuali indikator menjawab pertanyaan mengapa), menentukan unsur intrinsik dongeng yang dapat dikerjakan siswa berdasar analisis hanya amanat dalam cerita dan membedakan watak tokoh dalam dongeng, menentukan isi laporan, menentukan isi tersurat dan tersirat dari rubrik yang dibaca, menentukan unsur intrinsik drama anak-anak, menentukan unsur intrinsik dan makna kalimat cerita anak-anak dan menentukan denah. Kemampuan lulusan di dalam menulis adalah kemampuan menulis di-

alog, mengisi formulir, menyusun kata majemuk, menyusun paragraf, menulis iklan menulis petunjuk pemakaian, menggunakan sinonim dan antonim, mendeskripsikan Gambar dengan kalimat yang tepat, dan mengurutkan Gambar seri.

b) Mata pelajaran IPA

Berdasarkan hasil diskusi antar-panelis diperoleh hampir semua indikator dalam SKL mata pelajaran IPA mampu dijawab oleh siswa, indikator kompetensi yang dimiliki oleh siswa berdasar Standar Kompetensi Lulusan untuk mata pelajaran IPA yang **tidak dapat dijawab** oleh peserta ujian menurut panelis adalah menjelaskan akibat dari penggunaan bahan tertentu, menjelaskan fungsi organ pada tumbuhan, mengidentifikasi bentuk energi, menjelaskan cara penghematan energi, menjelaskan dampak pemanasan global, terjadinya gerhana bulan/matahari.

2) Metode Ebel

Kemampuan lulusan berdasarkan hasil analisis panelis dalam Metode Ebel terdapat pada tingkat relevansi dan tingkat kesulitan soal yang dimungkinkan mampu dijawab benar oleh peserta ujian. Rerata *cut score* mata pelajaran Matematika sebesar 60,21 (menjawab benar 23,48) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran Matematika sebesar 6.02. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650 peserta) maka yang dinyatakan lulus sebanyak 67,33% (1111 peserta) dan yang tidak lulus berdasarkan rerata *cut score* Matematika sebanyak 539 siswa (32,67 %). Rerata *cut score* mata pelajaran Bahasa Indonesia sebesar 80,75% (menjawab benar 37,14) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran Bahasa Indonesia sebesar 8,127. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650

peserta) maka yang dinyatakan lulus sebanyak 60,36% (996 peserta) dan yang tidak lulus berdasarkan rerata *cut score* Bahasa Indonesia sebanyak 654 siswa (39,64 %).

Rerata *cut score* mata pelajaran IPA sebesar 72,695% (menjawab benar 28 dari 39 soal) artinya siswa dinyatakan lulus jika nilai minimal yang dicapai dalam UASBN dalam mata pelajaran IPA sebesar 7,18. Apabila diimplementasikan pada peserta ujian yang menjadi sampel penelitian (1650 peserta) maka yang dinyatakan lulus sebanyak 61,58% (1016 peserta) dan yang tidak lulus berdasarkan rerata *cut score* IPA sebanyak 1016 siswa (38,42 %).

3) Metode Bookmark

Berdasarkan hasil analisis panelis dalam metode *Bookmark* dapat diketahui kemampuan lulusan yang diasumsikan berdasarkan rerata *Bookmark* yang ditetapkan oleh masing-masing panelis dimana peserta dianggap mampu menyelesaikan soal. Tabel-Tabel di bawah ini menunjukkan kemampuan lulusan untuk implementasi metode *Bookmark*.

4) Modifikasi Ebel-Bookmark

Kemampuan lulusan dalam analisis modifikasi Ebel *Bookmark* dapat dilihat dari asumsi panelis tentang tingkat kesulitan yang dimungkinkan dialami oleh siswa dalam menjawab soal ujian, meskipun dalam booklet sudah terurut menurut tingkat kesulitan. Berdasarkan modifikasi Ebel-*Bookmark* maka untuk mata pelajaran Bahasa Indonesia peserta ujian diasumsikan mampu menjawab *item* soal sebesar 81,266%, Matematika 79,06% dan IPA 77,44 %.

Kemampuan lulusan ditinjau dari perbedaan nilai UASBN tiap mata pelajaran.

Berdasarkan implementasi 4 metode *standard setting* dan *cut score* dihasilkan, batas minimal nilai kelulusan untuk setiap mata pelajaran. Nilai minimal yang dapat digunakan untuk menetapkan kelulusan *examine* dalam mata pelajaran bahasa Indonesia ber-

dasarkan hasil rerata dalam metode Nedelsky sebesar 6,183, metode Ebel 8,075, Metode *Bookmark* 5,164 dan modifikasi Ebel-*Bookmark* sebesar 8,127. Mata pelajaran Matematika metode Nedelsky sebesar 6,02, metode Ebel 6,021, metode *Bookmark* 5,133 dan Modifikasi Ebel *Bookmark* 7,906. Mata Pelajaran IPA untuk metode Nedelsky 6,923, metode Ebel 7,754, metode *Bookmark* 5,916 dan modifikasi Ebel-*Bookmark* 7,744. Dengan demikian, dapat ditemukan bahwa metode *Bookmark* memberikan peluang peserta ujian yang lulus akan lebih banyak, karena nilai minimal kelulusan lebih rendah dari metode yang lain untuk semua mata pelajaran yang di ujikan.

Tingkat Keyakinan/Kepercayaan Guru terhadap metode standard setting yang diimplementasikan

Berdasarkan hasil angket kepada para panelis dan analisis berdasarkan standar deviasi masing-masing setelah selesai pelaksanaan implementasi metode *standard setting* diperoleh hasil sebagai berikut.

- 1) Keyakinan/kepercayaan panelis dalam implementasi *standard setting* metode *Bookmark* (81,25%), metode Angoff Ya/Tidak (68,75%), modifikasi Ebel-*Bookmark* (62,50%), metode Ebel (43,75%) dan metode Nedelsky (43,75%).
- 2) Penetapan batas kelulusan (*cut score*) akan lebih baik dilaksanakan dengan metode *standard setting* karena dengan anggapan bahwa dengan metode *standard setting* akan menemukan *cut score* yang baik dan kredibel, meskipun pada akhirnya akan ditetapkan dengan kebijakan namun hendaknya berdasarkan pada hasil *cut score* dengan metode *standard setting*;
- 3) Para panelis percaya bahwa dengan metode *standard setting* penetapan standard kelulusan tidak akan mengurangi pro kontra masalah UN;
- 4) Berdasarkan putaran pertama dan kedua serta uji coba terbatas lima metode *standard setting* para panelis menyatakan bahwa metode Angoff Ya/Tidak paling mudah di implementasikan, namun peluang kelulusan peserta ujian rendah, metode *Bookmark* dikatakan mudah diimplementasikan

dan memberikan peluang paling besar kepada peserta ujian untuk lulus UN, modifikasi Ebel-Bookmark mudah dilaksanakan namun membutuhkan waktu dalam menganalisis kembali Booklet yang sudah terurut dalam memasukkan ke dalam kriteria relevansi dan tingkat kesulitan maupun dalam memberikan skor harapan, metode Nedelsky dan Ebel dirasa sulit oleh panelis karena harus menganalisis satu-pesatu *item* sehingga mereka para panelis merasa seperti diuji;

- 5) Perlu adanya pelatihan tentang metode *standard setting* kepada guru-guru khususnya guru SD, karena batas kelulusan (*cut score*) di SD ditetapkan oleh sekolah.

Rekomendasi Hasil

Maksud rekomendasi dalam penelitian ini adalah memberikan masukan dan sebagai bahan pertimbangan berdasarkan hasil penelitian tentang *standard setting*. Dalam berbagai hasil penelitian tentang *standard setting* yang menggunakan beberapa metode *standard setting* akan berusaha untuk memberikan rekomendasi, metode *standard setting* yang mudah dan dapat dimanfaatkan untuk menetapkan *cut score* dengan akurat serta diimplementasikan oleh para guru ataupun pemangku kepentingan.

Dalam penelitian ini ada lima metode *standard setting* yang digunakan untuk menetapkan *cut score*. Dasar penetapan rekomendasi adalah besarnya standar deviasi dari masing-masing metode *standard setting*, dengan asumsi semakin kecil standar deviasi maka metode *standard setting* tersebut semakin kecil tingkat kesalahan dalam menetapkan *cut score*.

Penetapan standar deviasi dengan mencari rerata standar deviasi antarpanelis pada putaran kedua. Dipilihnya putaran kedua karena panelis jumlahnya lebih banyak, 68,75% panelis telah mengikuti putaran pertama. Hasil Analisis standar deviasi diperoleh bahwa baik untuk mata pelajaran Bahasa Indonesia, IPA dan Matematika metode *Bookmark* lebih akurat dalam menentukan *cut score* selengkapnya disajikan pada Tabel 4.

Tabel 4. Rangkuman Standard Deviasi 4 Metode *Standard Setting*

	Nedelsky	Ebel	<i>Bookmark</i>	Modifikasi Ebel- <i>Bookmark</i>
Bahasa Indonesia	12,45	6,097	1,774	6,779
Matematika	9,136	5,471	1,736	4,725
IPA	18,97	5,283	4,288	4,247

Berdasarkan pendapat para panelis dan hasil perhitungan standar deviasi antarpanelis di setiap metode yang diimplementasikan dan dikembangkan dan *cut score* yang diperoleh menunjukkan bahwa metode *Bookmark* merupakan metode yang mudah dilaksanakan dan memberikan peluang lulus yang paling besar untuk mata pelajaran Bahasa Indonesia dan Matematika diantara metode yang lain. Khusus untuk mata pelajaran IPA dengan modifikasi metode Ebel-*Bookmark*.

Adanya perbedaan dalam mengidentifikasi metode yang dapat diimplementasikan dalam penetapan *cut score* yaitu panelis mengidentifikasi dengan mengurutkan metode mana yang mudah diimplementasikan tanpa melihat mata pelajaran yang diujikan, sedangkan dalam penetapan standar deviasi didasarkan pada hasil rerata standar deviasi antarpanelis disetiap implementasi metode *standard setting* dan berdasarkan identifikasi untuk setiap mata pelajaran .

Simpulan

Berdasarkan hasil analisis data dapat disimpulkan bahwa: (1) rerata *Cut Score* berdasarkan metode Nedelsky untuk mata pelajaran Bahasa Indonesia sebesar 28,44, Matematika 23,47, IPA 27. Rerata *Cut Score* metode Ebel untuk mata pelajaran Bahasa Indonesia sebesar 80,75, Matematika 60,21, IPA 77,54. Rerata *Cut Score* metode *Bookmark* untuk mata pelajaran Bahasa Indonesia sebesar 51,63, Matematika 51,33, IPA 59,16. *Cut Score* modifikasi Ebel-*Bookmark* untuk mata pelajaran Bahasa Indonesia sebesar 81,27, Matematika 79,06 dan IPA 77,44; (2) keyakinan/kepercayaan panelis dalam implementasi metode *Standard Setting*

untuk metode *Bookmark* sebesar 81,25%, modifikasi Ebel-*Bookmark* 62,50%, metode Ebel 43,75% dan metode Nedelsky 43,75%; (3) metode *Bookmark* lebih akurat dibandingkan dengan metode yang lain untuk menetapkan *cut score* mata pelajaran Bahasa Indonesia dan Matematika dalam UN SD 2009, modifikasi metode Ebel-*Bookmark* lebih akurat dibandingkan dengan metode lain untuk menetapkan *cut score* mata pelajaran IPA dalam UN SD 2009.

Keterbatasan Penelitian

Penelitian ini dilaksanakan terbatas di wi-layah Propinsi Daerah Istimewa Yogyakarta dengan jumlah panelis yang terbatas mewakili 5 daerah kabupaten dan kota di DIY. Pendekatan penelitian dengan model pe-ngembangan ini terbatas pada empat implementasi metode yang sudah ada dan satu modifikasi metode baik dalam rubrik maupun analisisnya. Kemampuan peserta ujian belum dapat terdeteksi secara lengkap, baru berdasarkan pada analisis *item* dan deskriptor.

Saran

Pendidikan dasar meliputi SD dan SMP, di tingkat sekolah dasar apabila semua siswa yang menempuh UN akan dinyatakan lulus maka untuk melanjutkan ke jenjang SMP maka diperlukan kategori kinerja/level

kinerja tertentu seperti *basic*, *advance* dan *provincient* sehingga dapat menempatkan siswa pada sekolah lanjutan pertama dengan lebih tepat sesuai kompetensi yang dimilikinya.

Daftar Pustaka

- Angoff Ya/Tidak, W.H. (1971). *Scales, norms, and equivalent scores*. Thorndike (Ed), *Educational Measurment* (2nd ed). Washington, DC: American Council on Education
- Badan Standar Nasional Pendidikan. (2007). *Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian Pendidikan*. Jakarta: BSNP
- Cizek, J Gregory, Bunch Michael B. (2007). *Standard Setting: A Guide to Estlabishing and Evaluating Performance standards on Tests*. London: Sage Publications.
- Hambleton, Ronald K & Pitoniak, Mary J. (2006). *Setting Performance Standards dalam Brennan, L. Robert. Educational Measurement (Fourth Edition)*, American Council on Education .p. 433-470.
- Depdiknas. (2003). *Peraturan Pemerintah No. 20 Tahun 2003, tentang Undang-Undang Sistem Pendidikan Nasional*.