

Standard Setting Ujian Nasional Dengan Menggunakan Metode Angoff dan *Bookmark*

Arniati Prasedyawati Herkusumo
email: arniatiph@yahoo.com, Pusat Penilaian Pendidikan

Abstrak: Penelitian ini bertujuan untuk mengetahui batas skor kelulusan (*cut score*) yang tepat bagi peserta Ujian Nasional (UN). Seperti kita ketahui bahwa sampai saat ini penentuan kelulusan pada Ujian Nasional ditentukan melalui *judgement* dan merupakan kebijakan yang dikeluarkan oleh Pemerintah tanpa memperhatikan soal-soal yang ada pada paket-paket UN dan tanpa melibatkan tenaga-tenaga ahli (seperti guru bidang studi, ahli kurikulum, dan ahli penilaian). Penentuan kelulusan seperti ini kurang memenuhi kriteria sebagaimana yang diisyaratkan oleh ahli-ahli pengukuran pendidikan, seperti yang disebutkan dalam *Standards for Educational and Psychological Testing*: "When cut scores defining pass-fail, the judgmental process should be designed so that judges on bring their knowledge and experience to bear in a reasonable way" Metode ilmiah yang digunakan untuk menentukan standar atau *cut score* kelulusan pada studi ini adalah dengan menggunakan Metode Angoff dan *Bookmark*. Metode ini dapat digunakan untuk menentukan standar kelulusan UN menjadi lebih ilmiah.

Kata Kunci: *Ujian Nasional, standard setting, assessment, cut score, dan bookmark*

Abstract : The purpose of this research is to find out the cut score of graduation for the national exam participants, as we know so far the determined passing score through judgement and the policy of the Government without paying attention to the matter of matter that exists on the test devices, and without involving the experts (as curriculum subjects teachers, expert, assessment expert). Determining graduation like this does not meet the criteria as signaled by experts such as educational measurement experts called in *Standards for Educational and Psychological Testing* : "When cut scores defining pass-fail, the judgmental process should be designed so that judges on bring their knowledge and experience to bear in a reasonable way". Scientific method that can be used to determine the score a passing grade is the Angoff method and bookmark method. This method can be used to determine the standard graduation test national became more scientific.

Key words: *National Examination, standard setting, assessment, cut score, and bookmark*

Pendahuluan

Salah satu kegiatan yang penting dalam meningkatkan kualitas pendidikan adalah penilaian. Dalam banyak hal, hasil penilaian sering dipandang sebagai tolok ukur keberhasilan proses pembelajaran. Pada setiap sistem ujian, umumnya hasil penilaian dikategorikan menjadi dua yaitu lulus dan tidak lulus, terutama untuk kurikulum yang berbasis kompetensi. Lulus berarti peserta didik telah memiliki kompetensi minimum, sedang yang tidak lulus belum memiliki kompetensi minimum.

Dalam upaya meningkatkan kualitas pendidikan, pemerintah menerbitkan Peraturan Pemerintah (PP) No 19 tahun 2005 tentang Standar Nasional Pendidikan. Pada PP ini, peme-

rintah menetapkan 8 standar nasional pendidikan. Salah satu standar nasional pendidikan yang penting adalah Standar Kompetensi Lulusan (SKL), yaitu kemampuan minimum yang harus dimiliki peserta didik yang lulus dari suatu jenjang pendidikan.

SKL ini menjadi acuan dalam menyusun kisi-kisi ujian. Siapa saja yang menyusun soal apabila menggunakan kisi-kisi yang sama akan menghasilkan soal yang relatif sama. Soal yang relatif sama ini dapat digunakan untuk ujian, dan skor yang diperoleh dapat dibandingkan antar siswa atau antar sekolah.

Penilaian pada kurikulum berbasis kompetensi menggunakan acuan kriteria. Asumsi acuan ini adalah setiap peserta didik dapat belajar pelajaran

apa saja, hanya waktunya yang bervariasi. Hasil penilaian yang menggunakan acuan kriteria adalah lulus dan tidak lulus. Penetapan skor batas lulus atau dikenal dengan penetapan standar kelulusan dapat dilakukan melalui *judgment* (pertimbangan) dan kesepakatan pengambil keputusan atau dengan prosedur yang ilmiah. Selama ini prosedur yang digunakan adalah melalui *judgment*. Penetapan dengan cara ini memiliki kelemahan, yaitu tidak berdasarkan data empirik dan prosedur yang telah teruji. Untuk itu perlu dicari cara yang dapat dipertanggungjawabkan dari teori pengukuran.

Pada masa diberlakukannya Evaluasi Belajar Tahap Akhir Nasional (Ebtanas) pada tahun 1980-2002 penentuan standar kelulusan berdasarkan *judgment* (pertimbangan) yaitu pendapat sejumlah orang. Ketika dilaksanakan Ujian Akhir Nasional (UAN) tahun 2003-2004, penetapan standar kelulusan juga menggunakan *judgment*. Demikian pula saat mulai dilaksanakan Ujian Nasional (UN) sejak tahun 2005, batas kelulusan juga ditetapkan melalui *judgment* sejumlah orang. Penetapan standar kelulusan dengan cara *judgment* tentunya mengandung banyak kelemahan karena tidak saja merupakan pendapat sekelompok orang, namun kelompok tersebut juga tidak melibatkan tenaga ahli seperti guru bidang studi, ahli kurikulum, dan ahli penilaian.

Kriteria kelulusan atau batas kelulusan memegang peran penting dalam meningkatkan kualitas pendidikan. Peserta didik yang dinyatakan lulus harus disertai dengan penjelasan kemampuan apa saja yang telah dimiliki peserta didik tersebut untuk mata-mata pelajaran yang diujikan. Dengan demikian, kriteria kelulusan harus mencakup skor dan kemampuan yang dicapai. Oleh karena itu penetapan batas kelulusan harus menggunakan cara yang lebih baik dan lebih ilmiah, dan hasilnya dapat dipertanggungjawabkan.

Pada masa penggunaan Ebtanas, tingkat kelulusan siswa di semua jenjang pendidikan cenderung sangat tinggi. Standar nilai kelulusan sudah ditentukan, namun penentuan nilai akhir merupakan gabungan nilai hasil Ebtanas dan nilai pada rapor dengan bobot yang ditentukan oleh sekolah. Kebebasan sekolah dalam menentukan bobot nilai rapor dan bobot nilai Ebtanas, me-

nyebabkan besarnya bobot yang bervariasi. Besarnya bobot ini sering digunakan sebagai dasar untuk menentukan agar semua siswa lulus atau hampir semua lulus. Kelemahan Ebtanas ini diatasi dengan menggunakan sistem ujian akhir yang disebut dengan Ujian Akhir Nasional (UAN) sejak tahun 2002.

Pada ujian akhir nasional semua mata pelajaran yang ada di sekolah diujikan. Ada tiga mata pelajaran yang soalnya dikembangkan secara nasional. Sisanya dikembangkan di sekolah atau daerah. Semua mata pelajaran baik yang dikembangkan di pusat maupun di sekolah atau daerah menjadi bagian dari UAN. Kriteria kelulusan yang digunakan adalah peserta didik tidak boleh memiliki nilai 3,0 (tiga) ke bawah. Ada ujian praktik yang hasilnya digabung dengan hasil ujian teori yaitu bahwa mata ujian Indonesia dan bahasa Inggris. Siswa diperkenankan mengulang apabila pada nilai yang dicapai pada ujian pertama belum memenuhi batas kelulusan. Hasilnya hampir semua lulus, hal ini bisa disebabkan siswa bertambah giat belajar atau ada pengangkatan nilai siswa pada hasil UAN perbaikan agar lulus.

Pada tahun 2004, batas nilai minimum untuk bisa lulus ditingkatkan menjadi 4,00. Kenaikan batas ini diharapkan meningkatkan kualitas pendidikan. Selain itu nilai praktek dipisahkan dengan nilai teori, sehingga tidak memungkinkan nilai pelajaran yang ada prakteknya seperti sains dan bahasa Inggris mengangkat nilai teori. Meningkatnya nilai batas lulus dari 3,01 menjadi 4,01 dan pemisahan nilai teori dan praktek diharapkan dapat mendorong peningkatan kualitas pembelajaran. Keputusan ini sebenarnya merupakan tantangan bagi kepala sekolah, guru, siswa, dan orang tua. Untuk mencapainya, guru dan siswa harus bekerja sama dalam meningkatkan proses pembelajaran.

Perkembangan terakhir, sejak tahun 2005 dilaksanakan ujian nasional, sebagai suatu kegiatan untuk mengetahui pencapaian standar nasional pendidikan, yaitu standar kompetensi lulusan. Batas kelulusan pada ujian nasional tahun pelajaran 2008/2009 adalah rata-rata minimum 5,50, boleh memiliki nilai 4,0 pada paling banyak 2 mata pelajaran, lainnya minimum 4,25.

Penentuan kelulusan berdasarkan *judgment* mengandung banyak kelemahan, diantaranya

tidak memperhatikan kondisi siswa, wilayah Indonesia yang sangat luas, dan kondisi sekolah yang bervariasi. Oleh karena itu, penetapan standar minimum yang harus dicapai peserta didik agar dapat lulus dari satuan pendidikan perlu menggunakan cara yang lebih baik. Terkait dengan kebijakan pemerintah dalam sistem ujian nasional tersebut perlu dilakukan penelitian untuk mengetahui berapa skor batas kelulusan peserta didik yang tepat untuk setiap mata pelajaran yang di UN kan.

Permasalahan yang mendorong perlunya dilakukan penelitian ini yaitu terkait dengan salah satu kebijakan yang diterapkan pemerintah dalam sistem Ujian Nasional yaitu penentuan batas kelulusan minimum yang harus dicapai peserta didik yang ditetapkan melalui *judgement* semata.

Tujuan Penelitian ini adalah dapat menghasilkan kriteria kelulusan peserta didik berupa batas skor kelulusan peserta didik untuk dapat lulus SMA/MA Program IPA dan IPS untuk mata pelajaran Bahasa Indonesia, Matematika IPS, Matematika IPA, Bahasa Inggris, Ekonomi, dan Fisika.

Kajian Literatur

Pengertian *standard setting*

Standard setting adalah proses yang digunakan untuk menentukan atau memilih suatu *passing score* (batas lulus) pada suatu ujian. Dengan kata lain, *standard setting* suatu proses menentukan *cut score* (batas lulus) terhadap instrumen pendidikan atau psikologi untuk menjawab pertanyaan "seberapa bagus yang disebut cukup bagus" (George Engelhard, Jr. Dan Stephen E. Cramer, 1995, dalam Wilson, dkk, 1997). Dari semua langkah-langkah dalam proses pengembangan tes, *standard setting* merupakan tahapan yang lebih dekat pada seni daripada sains (ilmu pengetahuan). Komponen esensial dalam *standard setting* melalui *judgment* seperti yang dikemukakan oleh Angoff (1971), Ebel (1972), Jaeger (1989), dan Nedelsky (1954) yaitu panelis atau penilai ahli (Plake, Melican, & Mills, 1991). Lebih lanjut, Jaeger (1989) mengidentifikasi delapan kualifikasi ahli bidang studi (*Subject Matter Expert, SME*) yakni: 1) terbaik dalam bidang spesialisasinya; 2) memiliki wawasan yang luas dalam bidang keahliannya; 3) memiliki kemampuan menyelesaikan masalah dengan cepat

sesuai bidangnya; 4) mampu mengkaji secara mendalam level konseptual dalam bidangnya dibandingkan orang baru; 5) menganalisis problem-problem dalam bidangnya secara kualitatif; 6) menilai problem secara lebih akurat dibandingkan orang baru; 7) mempunyai daya ingat semantik yang lebih kompleks; dan 8) memahami makna setiap skor yang diperoleh dalam suatu tes.

Hattie & Brown (2003) menyatakan bahwa *setting performance standard* merupakan suatu proses meminta pertimbangan rasional dari para ahli yang 1) memiliki pengetahuan tentang kebutuhan akan tes dan asesmen yang ingin ditetapkan standarnya; 2) memahami makna skor pada level yang bervariasi pada skala yang digunakan untuk menyimpulkan performansi peserta tes; dan 3) memahami sepenuhnya batasan tentang prestasi yang berhubungan dengan standar performansi yang dimintakan kepada mereka untuk ditetapkan.

Berbagai literatur pengukuran mengatakan bahwa terdapat lebih dari 30 metode *standard setting* yang berbeda (Glass, 1997; Hambleton, 1991; Jaeger, 1989), namun pada penelitian ini metode yang digunakan adalah metode Angoff dan metode *Bookmark*.

Metode Angoff

Pada metode Angoff klasik (1971), panelis diminta untuk berpikir tentang suatu kelompok peserta tes dan mengestimasi proporsi dari mereka yang akan dapat menjawab item dengan tepat. Kemungkinan ini kemudian dijumlahkan untuk semua item untuk memperoleh *passing score* minimum. Rerata *cut-off score* merupakan *cutting score* final untuk sebuah tes. Dengan kata lain bahwa, konsensus dari semua penilaian ahli menjadi *passing score* minimum. Keunggulan metode Angoff dibanding dengan prosedur yang lain (Impara, 2000:2 dan Goodwin, 1996: 253) yaitu prosedurnya sederhana dan mudah dilaksanakan, karenanya reliabilitas asesmen intra dan inter *judgement* menjadi tinggi. Metoda ini menggunakan statistik sederhana yang mudah dihitung dan dipahami. Adapun tahapan metode Angoff yang dilaksanakan yaitu: 1) panelis melihat butir soal pertama dan menilai tingkat kesulitannya; 2) setiap panelis secara individu meng-

estimasi persentase sekelompok peserta tes yang dapat menjawab butir soal dengan benar; 3) panelis mendiskusikan hasil estimasi mereka; 4) setiap hasil estimasi ditabulasikan dan dihitung rata-ratanya; 5) urutan di atas diulang untuk semua butir soal; dan 6) rata-rata hasil estimasi setiap butir dijumlah dan dirata-ratakan kembali untuk memperoleh *cut point*.

Penentuan batas kelulusan (*cut score*) dengan metode Angoff pada prinsipnya merupakan proses yang sangat sederhana. Dalam proses ini, sekelompok panelis masing-masing diminta untuk memikirkan kompetensi minimal yang dapat membedakan mereka yang lulus dan tidak lulus. Instruksi tipikal yang banyak digunakan pada metode ini adalah tiap panelis diminta mengestimasi proporsi 100 peserta ujian yang dapat menjawab butir-butir soal dengan benar.

Probabilitas yang diperoleh dari tiap butir soal kemudian dijumlah dan hasilnya dinamakan sebagai Kriteria Kelulusan Minimal (KKM) atau *Minimum Passing Level* (MPL) dari panelis. Rata-rata KKM dari sekelompok panelis merupakan batas kelulusan final. Kesalahan baku KKM juga dapat dihitung. Kesalahan baku yang kecil menunjukkan bahwa tingkat kesepakatan estimasi proporsi dari para panelis tinggi (Ricker, 2009). Kesalahan baku yang besar berarti tingkat kesepakatan estimasi proporsi dari para panelis

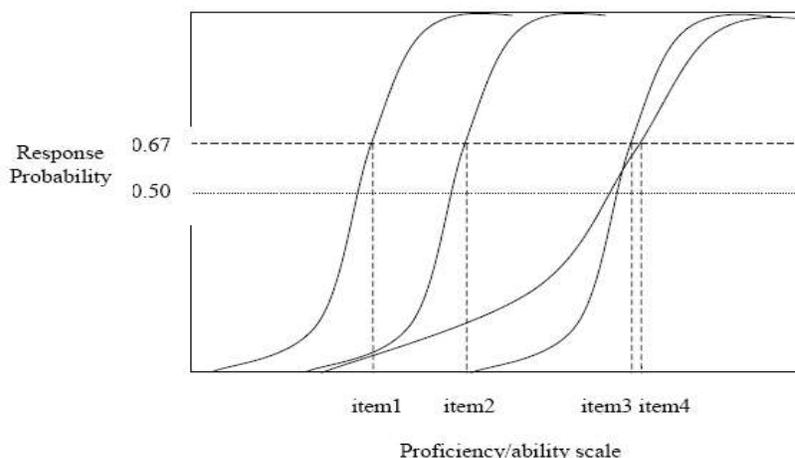
rendah.

Metode Bookmark

Metode *Bookmark* didasarkan pada *Item Response Theory/IRT* (Teori Respon Butir) (Lord, 1980) yang menggabungkan secara simultan antara karakteristik kemampuan peserta dan tingkat kesulitan butir soal. Setiap butir terskalakan dalam IRT dapat dinyatakan dengan kurva karakteristik yang menyatakan hubungan antara kemampuan peserta terhadap suatu butir (Gambar 1). Teori respon butir menyebabkan hal ini memungkinkan untuk mengurutkan berdasarkan kemampuan atau skor skala yang diperlukan suatu probabilitas khusus dari kesuksesan. Butir yang dipetakan pada suatu lokasi dalam skala IRT sedemikian rupa sehingga siswa dengan skor skala dekat pada butir spesifik, dapat disimpulkan memiliki pengetahuan ketrampilan dan kemampuan yang diperlukan untuk merespon dengan sukses pada butir dengan probabilitas tertentu. Pada prosedur *bookmark*, probabilitas kesuksesan diset 0,67. Angka 0,67 sebagai probabilitas respon (*response probability*, RP) didukung oleh penelitian Huynh tahun 1998, seperti yang digambarkan pada kurva karakteristik soal berikut ini.

Huynh menyatakan bahwa pada model 3 PL ini, fungsi informasi butir termaksimumkan ketika q berada pada $P(q) = (c + 2)/3$. Sebagai akibatnya

Item Characteristic Curves (ICCs) for SR Items Mapped at RP = 0.67



Gambar 1. Kurva karakteristik butir yang dipetakan pada RP 0,67 (Diadaptasi dari Mitzel, Lewis, Patz, & Green (2001), p. 261)

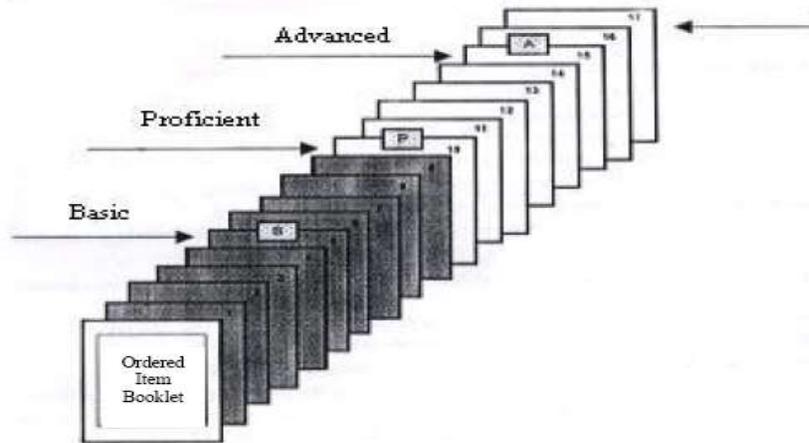
pada model 2 PL, ketika guessing (c) tidak ada, nilai $RP = 2/3$. Material utama yang sering digunakan pada penentuan standar seting dengan *bookmark* adalah buku tes (*booklet*). Menggunakan parameter b, butir soal diurutkan dari yang mudah ke yang sulit dalam buku tes. Seperti diilustrasikan pada Gambar 2 berikut ini, buku tes dengan butir terurut memiliki satu butir soal tiap halaman dengan halaman pertama berisi butir soal termudah dan yang terakhir butir soal tersulit. Pengurutan tingkat kesukaran tersebut menurut Lewis, dkk (1998), untuk membantu panelis menyusun suatu konsep terintegrasi dalam membuat *cut score*. Selain buku tes, juga akan sangat membantu jika disiapkan petunjuk penggunaan, tabel rangkuman *bookmark* untuk menghitung rata-rata *cut score* atau Kriteria Kelulusan Minimal (KKM) yang diperoleh dari para

panelis, dan format identifikasi deskriptor KKM.

Metodologi Penelitian

Responden dalam studi ini adalah guru SMA/MA program IPA dan IPS yang mengajar Bahasa Indonesia, Matematika, Bahasa Inggris, Ekonomi, dan Fisika yang berpengalaman mengajar minimum 10 tahun dan mengajar kelas 12 pada mata pelajaran tersebut minimal selama 5 tahun. Jumlah guru setiap mata pelajaran adalah 11 orang yang diberasal dari propinsi Bengkulu, Jawa Tengah, D.I. Yogyakarta, dan Nusa Tenggara Timur yang mewakili berbagai kualitas sekolah. Keempat propinsi tersebut dipilih karena pelaksanaan UN-nya dianggap bersih dari kecurangan. Proporsi guru tiap mata pelajaran dilihat dari keempat propinsi tersebut ditunjukkan pada Tabel 1.

Metode yang digunakan untuk mengumpulkan data dalam penelitian ini adalah dokumentasi



Gambar 2. Ilustrasi Buku Tes dengan Butir Terurut pada Prosedur *Bookmark* (Diadaptasi dari Mitzel, Lewis, Patz, & Green (2001: 263))

Tabel 1. Proporsi Sampel Tiap Mata Pelajaran

Provinsi	Banyaknya Guru	Sekolah		Jenis Kelamin Guru	Jumlah
		Prestasi	Asal		
Bengkulu	1	Sedang	Kota	Laki-laki	2
	1	Rendah	Pinggiran/Desa	Perempuan	
NTT	1	Sedang	Pinggiran/Desa	Perempuan	2
	1	Rendah	Kota	Laki-laki	
Jawa Tengah	1	Sedang	Kota	Laki-laki	2
	1	Rendah	Pinggiran/Desa	Perempuan	
DI Yogya-karta	1	Sedang	Kota	Perempuan	5
	1	Rendah	Pinggiran/Desa	Laki-laki	
	1	Sedang	Pinggiran/Desa	Laki-laki	
	1	Sedang	Kota	Perempuan	
Total Guru Tiap Mata Pelajaran					11

Sumber: Puspendik 2010

dan wawancara. Instrumen yang digunakan adalah daftar dokumentasi, pedoman wawancara, dan kuesioner. Teknik analisis data menggunakan Metode Angoff dan *Bookmark*. Adapun data yang digunakan dalam penelitian ini adalah data Ujian Nasional 2009 untuk mata pelajaran Bahasa Indonesia, Matematika, Bahasa Inggris, Ekonomi, dan Fisika.

Hasil Penelitian dan Pembahasan

Peserta penetapan (panelis) *standard setting* untuk seluruh mata pelajaran yang diteliti terdiri atas 11 orang guru. Semua panelis berlatar belakang sarjana pendidikan sesuai dengan mata pelajaran yang diteliti. Kesebelas panelis ini berasal dari sekolah dengan prestasi sedang dan rendah (berdasarkan hasil UN), memiliki pengalaman mengajar lebih dari 10 tahun, dan telah mengampu kelas XII selama minimal 5 tahun. Untuk setiap mata pelajaran, panelis berasal dari Bengkulu (2 orang), Nusa Tenggara Timur (3 orang), Jawa Tengah (2 orang), dan dari DI Yogyakarta (5 orang).

Penentuan *standard setting* dengan metode Angoff ini dilakukan melalui 3 tahapan, yaitu tahap latihan, putaran 1, dan putaran 2. Pada tahap latihan, panelis diminta mencermati terlebih dahulu butir-butir soal UN mata pelajaran sesuai dengan latar belakang pendidikan dan mata pelajaran yang diampu, yang terdiri dari 9 butir. Para panelis diminta untuk menuliskan deskriptor dari setiap butir soal tersebut. Selanjutnya, para panelis diminta menuliskan angka "1" untuk butir-butir soal yang diperkirakan mampu dijawab benar oleh para siswanya. Rerata yang diperoleh berdasarkan banyaknya "1" pada butir-butir tes tiap panelis merupakan Kriteria Kelulusan Minimal. Putaran I dilakukan dengan cara yang sama dengan tahap latihan, namun dengan menggunakan keseluruhan butir tes UN, pada putaran II dilakukan hal yang sama dengan putaran I. Rerata yang dihasilkan dari putaran I dan II tersebut merupakan

batas lulus dari mata pelajaran yang diujikan.

Penentuan *standard setting* dengan menggunakan metode *Bookmark* juga dilakukan melalui 3 tahapan yaitu tahap latihan, putaran 1, dan putaran 2. Pada setiap tahapan dilakukan 1) review buku tes secara individual; 2) latihan memberi tanda (*bookmark*) pada buku tes secara individual; 3) mengidentifikasi deskriptor KKM dengan Lembar Identifikasi Deskriptor KKM secara individual; 4) merangkum *bookmark* panelis secara kelompok; 5) mendiskusikan KKM yang diperoleh berdasarkan rangkuman *Bookmark* panelis secara kelompok; dan 6) mendiskusikan deskriptor KKM secara kelompok.

Setelah mereview butir-butir tersebut, panelis diminta menuliskan batas tanda (*bookmark*), pada halaman berapa peserta sudah tidak dapat menjawab dengan peluang benar sebesar 0,67. Rerata halaman batas tanda dari butir-butir tes yang telah diurutkan dalam Booklet Butir Terurut dari tiap panelis merupakan Kriteria Ketuntasan Minimal dari setiap mata pelajaran.

Mata Pelajaran Bahasa Indonesia

Metode Angoff

Berdasarkan hasil rerata dari panelis, *cut score* dari UN mata pelajaran Bahasa Indonesia pada putaran I adalah sebanyak 36,250 butir benar dari 47 butir yang terdapat pada tes (77,1%). Dengan menggunakan *cut score* ini, ternyata hanya sebanyak 28 % dari 37.502 peserta UN tahun 2009 dari mata pelajaran ini yang dapat dikategorikan lulus. Pada putaran kedua, diperoleh hasil *cut score* yang merupakan banyaknya butir benar sebanyak 33,50 (71,3%) dari 47 butir yang terdapat pada tes, dengan persentase kelulusan meningkat menjadi 50,0%. Rerata banyaknya butir benar putaran I dan II sebesar 34,875 (74,202%), dan persentase kelulusannya sebanyak 75,6%. (Pada putaran I dan II menggunakan sebanyak 47 butir soal UN dari jumlah butir yang seharusnya sebanyak 50 butir.

Tabel 2. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Bahasa Indonesia

Putaran	Banyaknya butir benar	Persentase Butir Benar	Persentase Kelulusan
I	36,250	77,1%	28,0%
II	33,50	71,3%	50,0%
Rerata	34,875	74,2%	39,0%

Berdasarkan analisis butir soal, 3 butir dinyatakan tidak baik karena memiliki korelasi yang negatif dengan skor total) Hasil selengkapnya disajikan pada Tabel 2.

Metode Bookmark

Dengan menggunakan metode *Bookmark*, hasil yang diperoleh pada putaran I *cut score* pada skala kemampuan - 0,313 (46,089 pada skala 100), dengan persentase kelulusan sebesar 70,1%. Hasil ini diperoleh dengan *bookmark* pada halaman 33. Sementara itu, pada putaran II *cut score* pada skala kemampuan 0,116 (51,447 pada skala 100), dengan persentase kelulusan sebesar 50,4%. Hasil ini diperoleh dengan *bookmark* pada halaman 35. Hasil selengkapnya disajikan pada

peserta UN tahun 2009 pada mata pelajaran ini dikategorikan lulus, dan hanya 4% yang tidak lulus. Hasil yang hampir sama diperoleh pada putaran kedua. Banyaknya butir benar sebesar 21,64 (54,09%), namun persentase kelulusan tetap 96%. Rerata banyaknya butir benar putaran I dan II sebesar 21,37 (53,43%), dan persentase kelulusan juga 96%. Hasil selengkapnya disajikan pada Tabel 4.

Metode Bookmark

Pada metode *Bookmark*, hasil yang diperoleh pada putaran II lebih tinggi dibandingkan putaran I. Putaran I *cut score* pada skala kemampuan - 1.04 (36.96 pada skala 100), dengan persentase kelulusan 84.92%. Putaran II *cut score* pada skala

Tabel 3. Hasil Standard Setting Metode Bookmark pada Bahasa Indonesia

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	- 0,313	0,525	46,089	6,567	70,1%
II	0,116	0,472	51,447	5,901	50,4%
Rerata	- 0,098		48,768		60,25%

Tabel 3.

Mata Pelajaran Matematika Program IPS Metode Angoff

Berdasarkan rerata dari panelis, *cut score* dari UN mata pelajaran Matematika IPS pada putaran I sebesar 21,11 butir benar dari 40 butir yang terdapat pada tes (52,78%). Dengan menggunakan *cut score* ini, sebanyak 96% dari 65.365

kemampuan -0.36 (45.48 pada skala 100), dengan persentase kelulusan 61.70%. Hasil akhir dari putaran I dan II ini diperoleh *cut score* pada skala kemampuan -0.70 (41.22 pada skala 100), atau dengan *bookmark* pada halaman 31. Persentase kelulusan dengan *cut score* ini sebesar 85,61%. Hasil selengkapnya disajikan pada Tabel

Tabel 4. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Matematika IPS

Putaran	Banyaknya butir benar	Persentase Butir Benar	Persentase Kelulusan
I	21,11	52,78	96,00%
II	21,64	54,09	96,00%
Rerata	21,37	53,43	96,00%

Tabel 5. Hasil Standard Setting Metode *Bookmark* pada Mata Pelajaran Matematika IPS

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	-1,04	0,33	36,96	4,16	84,92%
II	-0,36	0,39	45,48	4,81	61,70%
Rerata	-0,70		41,22		73%

5.

Mata Pelajaran Matematika Program IPA Metode Angoff

Berdasarkan rerata dari panelis, *cut score* dari UN mata pelajaran Matematika IPA pada putaran I sebesar 23,00 butir benar dari 40 butir yang terdapat pada tes (57,5%). Dengan menggunakan *cut score* ini, sebanyak 85,61% dari 14.658 peserta UN tahun 2009 mata pelajaran ini dikategorikan lulus, dan hanya 14,39% yang tidak lulus. Pada putaran kedua, diperoleh hasil *cut score* berupa banyaknya butir benar sebesar 23,75 (59,38%), dan persentase kelulusan menjadi 82,22%. Rerata banyaknya butir benar putaran I dan II sebesar 23,37 (83,92%), dan persentase kelulusan juga 83,92%. Hasil selengkapnya

dari putaran I dan II ini diperoleh *cut score* pada skala kemampuan -1,034 (37,075 pada skala 100), atau dengan bookmark pada halaman 16. Persentase kelulusan dengan *cut score* ini sebesar 85,61%. Hasil selengkapnya disajikan pada Tabel 7.

Mata Pelajaran Bahasa Inggris Program IPA/ IPS

Metode Angoff

Berdasarkan rerata dari panelis, *cut score* dari UN mata pelajaran Bahasa Inggris pada putaran I sebesar 26,364 butir benar dari 49 butir yang terdapat pada tes (53,8%). Dengan menggunakan *cut score* ini, sebanyak 69,7% dari 32.985 peserta UN tahun 2009 pada mata pelajaran ini

Tabel 6. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Matematika IPA

Putaran	Banyaknya butir benar	Persentase Butir Benar	Persentase Kelulusan
I	23,00	57,50%	85,61%
II	23,75	59,38%	82,22%
Rerata	23,375	58,44%	83,92%

disajikan pada Tabel 6.

Metode Bookmark

Pada metode Bookmark, hasil yang diperoleh pada putaran I dan II cukup konsisten. Putaran I *cut score* pada skala kemampuan -1,154 (35,58 pada skala 100), dengan persentase kelulusan 88,57%. Putaran II *cut score* pada skala kemampuan -0,914 (38,57 pada skala 100), dengan persentase kelulusan 85,61%. Hasil akhir

dikategorikan lulus. Pada putaran kedua, diperoleh hasil *cut score* berupa banyaknya butir benar sebesar 24,64 (50,0%), dan persentase kelulusan menjadi menjadi 75,6%. Rerata banyaknya butir benar putaran I dan II sebesar 25,5 (65,4%), dan persentase kelulusan juga 75,6%. (Pada mata pelajaran Bahasa Inggris, dari 40 butir soal, hanya digunakan 39 butir, dikarenakan 1 butir korelasi biserialnya negatif yaitu nomor 23). Hasil seleng-

Tabel 7. Hasil Standard Setting Metode Bookmark pada Mata Pelajaran Matematika IPA

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	-1,154	0,368	35,580	4,600	88,57%
II	-0,914	0,182	38,570	2,274	85,61%
Rerata	-1,034		37,075		85,61%

Tabel 8. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Bahasa Inggris

Putaran	Banyaknya butir benar	Persentase Butir Benar	Persentase Kelulusan
I	26,364	53,8%	69,7%
II	24,64	50,0%	75,6%
Rerata	25,5	65,4%	75,6%

kapnya disajikan pada Tabel 8 berikut.

Metode Bookmark

Dengan menggunakan metode *Bookmark*, hasil yang diperoleh pada putaran pertama dan kedua adalah sama/konsisten. Baik pada putaran pertama maupun putaran kedua, *cut score* pada skala kemampuan -1,66 (29,25 pada skala 100), dengan persentase kelulusan 95,04%. Hasil ini diperoleh dengan *bookmark* pada halaman 20.

tersebut ditunjukkan pada Tabel 10.

Metode Bookmark

Hasil putaran pertama dengan metode *Bookmark* kelompok mata pelajaran Ekonomi memperoleh rata-rata KKM sebesar -1,352 dengan standar deviasi 0,277 pada skor baru atau 33,102 dengan standar deviasi 3,458 pada skala 100. Dari 11 panelis yang ada, *Bookmark* tertinggi dicapai pada halaman 21, sedang yang terendah adalah pada

Tabel 9. Hasil Standard Setting Metode Bookmark pada Mata Pelajaran Bahasa Inggris

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	-1,66	0,45	29,25	5,61	95,04%
II	-1,66	0,45	29,26	5,61	95,04%
Rerata	-1,66		29,26		95,04%

Hasil selengkapnya disajikan pada Tabel 9.

Mata Pelajaran Ekonomi Metode Angoff

Hasil putaran pertama dengan metode Angoff diperoleh rata-rata KKM sebesar 23,818 dari 39 soal UN yang korelasinya terhadap skor total positif. Rata-rata KKM tersebut sama dengan 61,1% soal benar. Rata-rata KKM ini menyebabkan 85,6% dari 16.701 peserta UN tahun 2009 dari keempat propinsi yang digunakan dalam penelitian ini lulus.

Hasil putaran kedua metode ini memperoleh rata-rata KKM sebesar 24,182 soal benar atau 62,0% soal benar. Rata-rata KKM ini menyebabkan 82,4% sampel peserta UN tahun 2008 dari keempat propinsi yang digunakan dalam penelitian ini lulus. Perbandingan rata-rata KKM dan persen kelulusan kedua putaran metode Angoff

halaman 8 dari booklet butir terurut. Rata-rata KKM ini menyebabkan 92,4% dari 16.701 peserta UN tahun 2009 dari keempat propinsi yang digunakan dalam penelitian ini lulus.

Hasil putaran kedua metode ini memperoleh rata-rata KKM sebesar -1,030 dengan standar deviasi 0,215 pada skor baku atau 37,128 dengan standar deviasi 2,687 pada skala 100. Dari 11 panelis yang ada, *Bookmark* tertinggi dicapai pada halaman 21, sedang yang terendah adalah pada halaman 16 dari booklet butir terurut. Rata-rata KKM ini menyebabkan 87,8% sampel peserta UN tahun 2008 dari keempat propinsi yang digunakan dalam penelitian ini lulus. Perbandingan rata-rata KKM dan persentase kelulusan kedua putaran metode *Bookmark* tersebut ditunjukkan pada Tabel 11.

Tabel 10. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Ekonomi

Putaran	Banyak Butir Benar	Persentase Butir Benar	Persentase Kelulusan
I	23,818	61,1%	85,6%
II	24,182	62,0%	82,4%
Rerata	24	61,6%	84,0%

Tabel 11. Hasil Standard Setting Metode *Bookmark* pada Mata Pelajaran Ekonomi

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	-1,352	0,277	33,102	3,458	92,4%
II	-1,030	0,215	37,128	2,687	87,8%
Rerata	-1,191		35,115		90,10%

Mata Pelajaran Fisika

Metode Angoff

Hasil putaran pertama dengan metode Angoff diperoleh rata-rata KKM (*final cut score*) sebesar 22,667 dari 39 soal UN yang korelasinya terhadap skor total positif. Rata-rata KKM tersebut sama dengan 58,1% soal benar. Rata-rata KKM ini menyebabkan 92,8% dari 59.681 peserta UN tahun 2009 dari keempat propinsi yang digunakan dalam penelitian ini lulus.

Hasil putaran kedua metode ini memperoleh rata-rata KKM sebesar 21,833 soal benar atau 56,0% soal benar. Rata-rata KKM ini menyebabkan 94,2% sampel peserta UN tahun 2008 dari keempat propinsi yang digunakan dalam penelitian ini lulus. Perbandingan rata-rata KKM dan persen kelulusan kedua putaran metode Angoff tersebut ditunjukkan pada Tabel 12.

Metode Bookmark

Hasil putaran pertama metode *Bookmark* memperoleh rata-rata KKM (*final cut score*) sebesar -1,846 dengan standar deviasi 0,393 pada skor baru atau 26,927 dengan standar deviasi 4,914 pada skala 100. Dari 11 panelis yang ada, *Bookmark* tertinggi dicapai pada halaman 20, sedang yang terendah adalah pada halaman 6 dari booklet butir terurut. Rata-rata KKM ini menyebabkan 96,9% dari 59.681 peserta UN tahun 2009 dari keempat propinsi yang digunakan dalam penelitian ini lulus.

Hasil putaran kedua metode ini memperoleh rata-rata KKM sebesar -1,424 dengan standar

deviasi 0,208 pada skor baku atau 32,205 dengan standar deviasi 2,606 pada skala 100. Dari 11 panelis yang ada, *Bookmark* tertinggi dicapai pada halaman 23, sedang yang terendah adalah pada halaman 12 dari booklet butir terurut. Rata-rata KKM ini menyebabkan 94,2% sampel peserta UN tahun 2008 dari keempat propinsi yang digunakan dalam penelitian ini lulus. Perbandingan rata-rata KKM dan persen kelulusan kedua putaran metode *Bookmark* tersebut ditunjukkan pada Tabel 13.

Pembahasan

Metode Angoff

Rangkuman Kriteria Ketuntasan Minimum (KKM) dari keenam mata pelajaran yang dihitung berdasarkan persentase butir benar dapat dilihat pada Gambar 3. Tampak pada gambar pada putaran pertama KKM Bahasa Indonesia menduduki posisi yang paling tinggi disusul Ekonomi, Fisika, Matematika IPA, Matematika IPS, dan terakhir Bahasa Inggris.

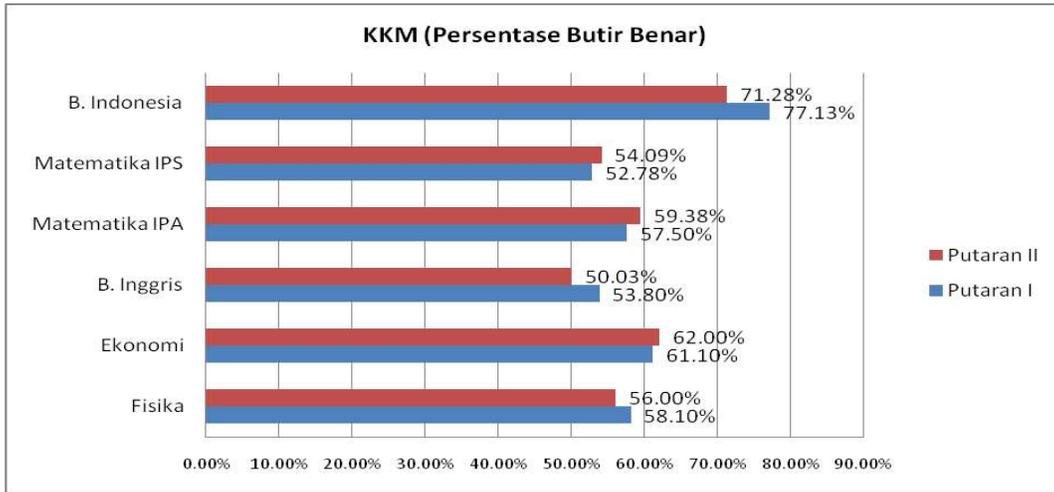
Pada putaran kedua KKM tertinggi adalah Bahasa Indonesia, disusul Ekonomi, Matematika IPA, Fisika, Matematika IPS, dan terakhir Bahasa Inggris. Perubahan urutan ini disebabkan para panelis telah mengetahui dampak penilaian butir soal yang telah dilakukan terhadap persentase kelulusan peserta ujian, seperti yang ditunjukkan pada Gambar 3. Panelis yang menginginkan banyak siswa yang lulus akan mengurangi jumlah butir soal yang harus dapat dikerjakan oleh peserta ujian. Namun tentunya tidak semua panelis melakukan hal tersebut, karena pengurangan butir soal yang benar berarti juga

Tabel 12. Hasil Standard Setting Metode Angoff pada Mata Pelajaran Fisika

Putaran	Banyak Butir Benar	Persentase Butir Benar	Persentase Kelulusan
I	22,667	58,1%	92,8%
II	21,833	56,0%	94,2%
Rerata	22,25	57,1%	93,5%

Tabel 13. Hasil Standard Setting Metode *Bookmark* pada Mata Pelajaran Fisika

Putaran	Skala Kemampuan		Skala (100)		Persentase Kelulusan
	Rerata	Stdev	Rerata	Stdev	
I	-1,846	0,393	26,927	4,914	96,9%
II	-1,424	0,208	32,205	2,606	94,2%
Rerata	-1,635		29,566		95,6%



Gambar 3. KKM (Persentase Butir Benar) Hasil Metode Angoff

pengurangan deskriptor. Selain itu, proses penilaian soal dilakukan secara individual, sehingga dampak tindakan seorang panelis hanya dapat diketahui setelah diskusi kelompok yang memaparkan KKM berdasar soal yang benar, KKM berdasar persentase soal benar, dan persentase kelulusan. Bukti tentang hal ini ditunjukkan pada Gambar 4, yaitu pada putaran kedua ternyata 3 dari 6 mata pelajaran pada putaran kedua persentase soal benarnya lebih tinggi dari putaran pertama. Gambar 4 menunjukkan persentase kelulusan peserta didik menurut metode Angoff.

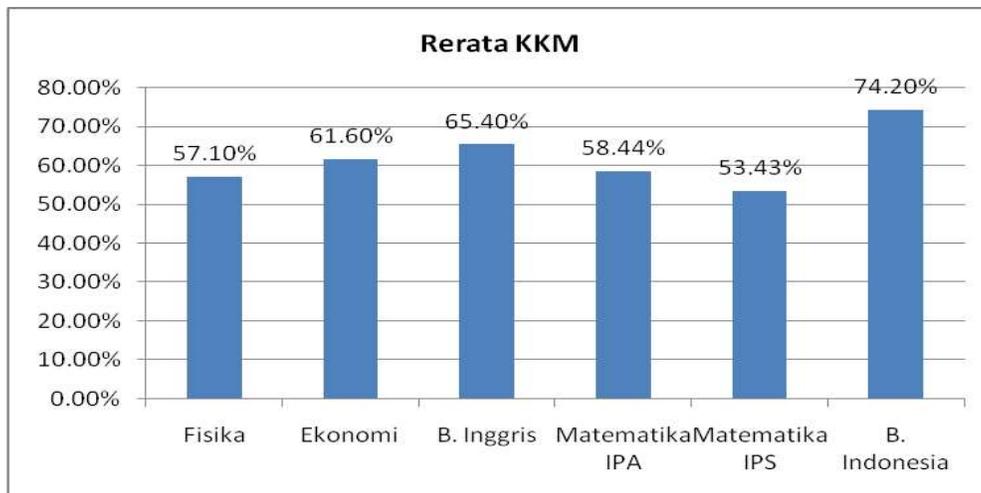
Rerata KKM *standard setting* 2 putaran metode Angoff menunjukkan rerata KKM-nya semua di atas 50% soal benar. Rerata KKM 6 mata pelajaran tersebut semua berada di atas batas kelulusan 42,5 skala 100 atau 4,25 skala 10 (Keputusan

BSNP No 1512/BSNP/XII/2008).

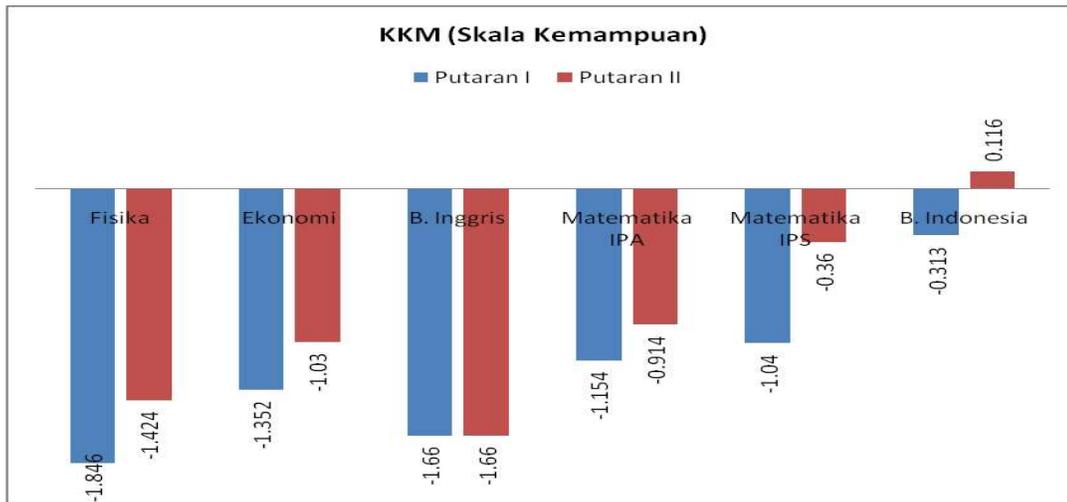
Metode Bookmark

Rangkuman KKM berdasarkan kemampuan peserta ujian hasil *standard setting* dengan metode *Bookmark* dari 6 mata pelajaran dalam skor baku ditunjukkan Gambar 5.

Gambar di atas menunjukkan bahwa pada putaran pertama Bahasa Indonesia memiliki KKM yang paling tinggi, disusul dengan Matematika IPS, Matematika IPA, Ekonomi, Bahasa Inggris, dan Fisika. Pada putaran kedua Bahasa Indonesia tetap memiliki KKM yang paling tinggi, disusul dengan Matematika IPS, Matematika IPA, Ekonomi, Fisika, dan Bahasa Inggris. Pada Gambar 6 menunjukkan KKM hasil metode *bookmark* dengan menggunakan skala 100. Pada putaran kedua Bahasa Indonesia tetap memiliki KKM yang



Gambar 4. Rerata KKM Hasil Metode Angoff Dua Putaran



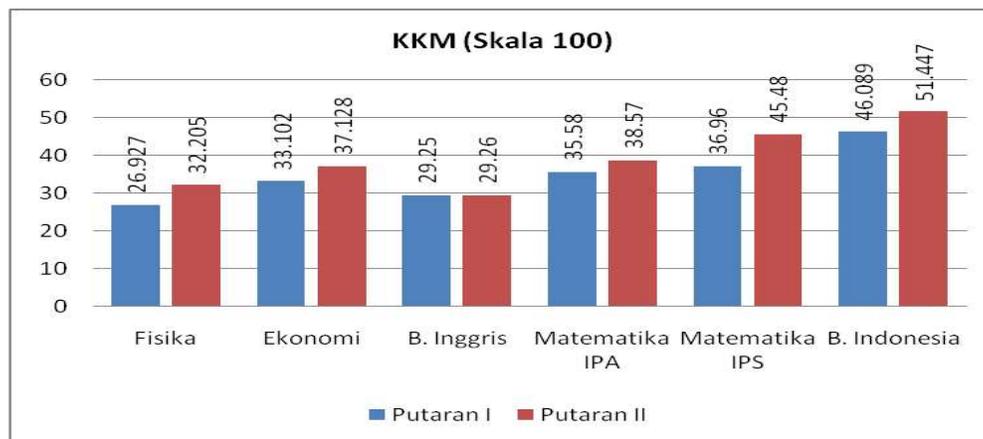
Gambar 5. KKM (Skala Kemampuan) Hasil Metode Bookmark

paling tinggi, disusul dengan Matematika IPS, Matematika IPA, Ekonomi, Fisika, dan Bahasa Inggris.

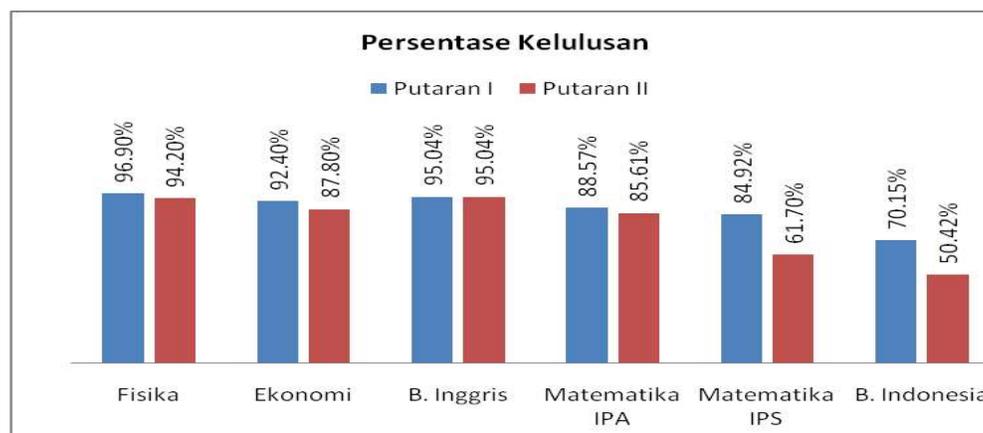
Semakin tinggi KKM menyebabkan semakin rendahnya tingkat kelulusan. Hal ini dapat dilihat

pada Gambar 7.

Pada putaran pertama persentase kelulusan Fisika paling tinggi, disusul Bahasa Inggris, Ekonomi, Matematika IPA, Matematika IPS, dan Bahasa Indonesia. Pada putaran kedua Bahasa



Gambar 6. KKM (Skala 100) Hasil Metode Bookmark



Gambar 7. Persentase Kelulusan Hasil Metode Bookmark

Inggris paling tinggi, disusul Fisika, Ekonomi, Matematika IPA, Matematika IPS, dan Bahasa Indonesia.

Rendahnya KKM yang ditentukan dengan *bookmark* disebabkan karena beberapa hal, yang merupakan kelemahan dari metode *bookmark*. Penentuan metode *bookmark* dilakukan terlebih dahulu dengan mengestimasi parameter butir, urutan tingkat kesulitan hasil estimasi berbeda dibandingkan dengan urutan tingkat kesulitan berdasarkan pengalaman panelis. Hal ini menyebabkan panelis berhenti dan membuat *bookmark* pada butir yang siswa tidak bisa mengerjakan, padahal butir selanjutnya ada yang bisa dikerjakan siswa. Dengan kata lain, tingkat kesulitan butir antardaerah berbeda-beda, sehingga hasil kalibrasi parameter kesulitan menggunakan keseluruhan siswa berbeda dengan tingkat kesulitan menurut panelis.

Rendahnya *cut score* dengan metode *bookmark* dapat disebabkan pula oleh rendahnya tingkat kesulitan butir. Berdasarkan hasil analisis dengan model 1 parameter, diperoleh bahwa rerata tingkat kesulitan mata pelajaran Bahasa Inggris sebesar -1,850, Fisika sebesar -1,807, Matematika IPA sebesar -1,371, dan Matematika IPS sebesar -2.039. Tingkat kesulitan ini mempengaruhi kemampuan yang diestimasi pada saat peluang menjawab benar siswa sebesar 0,67. Semakin rendah tingkat kesulitan, semakin rendah pula kemampuan pada saat proporsi kemampuan sebesar 0,67.

Simpulan dan Saran

Simpulan

Berdasarkan temuan penelitian ini dapat disimpul-

kan sebagai berikut. Pertama, hasil analisis dengan menggunakan metode Angoff maupun metode *Bookmark* menunjukkan bahwa batas kelulusan dari 6 mata pelajaran yang diteliti berbeda-beda. Kedua, berdasarkan metode Angoff, rerata KKM dari dua putaran diperoleh *cut score* dan persentase kelulusan untuk mata pelajaran Fisika 22,25 (57,10%); Ekonomi 24 (61,60%); Bahasa Inggris 25,5 (65,4%); Matematika IPA 23,375 (58,44%); Matematika IPS 21,37 (53,43%); dan Bahasa Indonesia 34,875 (74,25%). Ketiga, berdasarkan metode *Bookmark* rerata KKM dari dua putaran diperoleh *cut score* dan persentase kelulusan untuk mata pelajaran Fisika 29,566 (95,6%); Ekonomi 35,115 (90,10%); Bahasa Inggris 29,26 (95,04%), Matematika IPA 37,075 (85,61%), Matematika IPS 41,22 (73%); dan Bahasa Indonesia 48,768 (60,25%).

Saran

Atas dasar simpulan disarankan: 1) perlu dipertimbangan untuk menggunakan batas kelulusan yang berbeda untuk mata pelajaran yang berbeda dalam UN; 2) perlu melakukan berbagai upaya untuk mencapai kemampuan lulusan seperti yang tertuang dalam SKL. Upaya tersebut antara lain: menambah variasi metode pembelajaran dengan fokus pemahaman konsep mata pelajaran, meningkatkan minat membaca, menggunakan media *listening* dalam pelajaran Bahasa Inggris, dan meningkatkan kemampuan *listening* bagi gurunya; 3) soal-soal yang ada dalam UN perlu disusun dengan tingkat kesulitan yang memadai bagi siswa; dan 4) perlu dilakukan penelitian lebih lanjut tentang penentuan *standard setting* dalam skala yang lebih luas.

Pustaka Acuan

- Angoff, W. H. 1971. Scale, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Ebel, Robert L. 1972. *Essentials of educational measurement*. Englewood Cliffs: Prentice-Hall.
- Glass, C.A.W., van der Linden, W. J., Hambleton, R. K. 1997. A Step Model to Analyze Partial Credit. In Hambleton, K.R & vander Linden W.J. (eds, 1997), *Handbook of Modern Item Response Theory*. New York: Springer.
- Goodwin, L.D. 1996. Relation between observed item difficulty levels and Angoff minimum passing levels for a group of borderline examniness. *Applied Measurement in Education*, 12, 13-28.
- Hambleton, K.R., van der Linden W.J. 1991. *Fundamentals of Item Response Theory*. New York: Sage Publications.

- Hattie, J.A., & Brown. G.T.L. *Standard setting for asTTle Reading: A Comparison of Method*. asTTle Technical Report #21, (2003 August), University Aukland/Ministry of Education.
- Huynh, Huynh, 1998. A Clarification on The Response Probability Criterion RP67 for Standard Setting Based on Bookmark and Item Mapping, *Educational Measurement: Issue and Practice*, vol 25.
- Impara James C. & Plake Barbara S. 2000. *A Comparison of Cut Scores using Multiple Standard Setting Methods*. Paper presented at the Large Scale Assessment Conference. Snowbird, UT June, 2000. at [http://www.unl.edu/BIACO/coop/Isac/aeramillardsimpfinal .pdf](http://www.unl.edu/BIACO/coop/Isac/aeramillardsimpfinal.pdf) diambil tanggal 27 November 2005
- Jaeger, R. M. 1989. *Selection of judges for standard-setting*. *Educational Measurement: Issues and Practice*, 10(2), 3-6, 10.
- Lewis, D.M., Patz, R.J., & Green, D.R. 1998. *The Bookmark Procedure: Psychological Perspectives*. NJ: Lawrence Erlbaum Assoc
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. 2001. The Bookmark Procedure: Psychological Perspectives. In G.J. Cizek (Ed.), *Setting Performance Standards*. Mahwah, NJ.
- Nedelsky, L. 1954. Absolute grading standards for objective test. *Educational and Psychological Measurement*, 14, 3-19.
- Peraturan Pemerintah Nomor 19 Tahun 2005 Tentang Standar Nasional Pendidikan
- Plake B. S., Impara, J. C., & Irwin, P. 2000. Consistency of Angoff-based predictions of item performance: Evidence of technical quality of results from the Angoff standard setting method. *Journal of Educational Measurement*, 37(4), 347-355.
- Ricker, K. L. 2009. *Setting Cut Scores: Critical Review of Angoff and Modified-Angoff Methods*. Edmonton (Alberta, Canada): Centre for Research in Applied Measurement and Evaluation University of Alberta.
- Wilson, Mark, George Engelhard, Jr., Karen Draney. 1997. *Objective Measurement: Theory Into Practice*, Greenwich CT: Ablex Publishing Corporation. 1997.