

## KOMPARASI METODE *STANDARD SETTING* UNTUK PENENTUAN KKM MATA PELAJARAN MATEMATIKA KELAS VIII SMP

<sup>1)</sup>Susi Anto, <sup>2)</sup>Djemari Mardapi

<sup>1)</sup>LPMP Daerah Istimewa Yogyakarta, <sup>2)</sup>Universitas Negeri Yogyakarta  
<sup>1)</sup>susianto.lpmpjogja@gmail.com, <sup>2)</sup>djemarimardapi@yahoo.co.id

### Abstrak

Tujuan penelitian ini adalah menemukan skor batas Kriteria Ketuntasan Minimum (KKM) dengan memanfaatkan metode yang ada dalam *standard setting*. Metode yang digunakan adalah metode *Extended Angoff* dan metode Ebel. Penelitian ini merupakan penelitian deskriptif kuantitatif yang diperkuat dengan data kualitatif. Data kuantitatif yang digunakan dalam penelitian ini adalah pola respon peserta didik atas soal UKK SMP/MTs Kelas VIII Mapel Matematika Kabupaten Sleman 2011/2012. Selain itu, dalam penentuan *cut of score*, juga digunakan data kuantitatif yang diperoleh dari *expert judgement*. Sementara *expert judgement* yang bersifat kualitatif digunakan untuk menilai kualitas pelaksanaan pertemuan *standard setting*. Hasil penelitian ini menunjukkan bahwa *cutscore* yang diperoleh dengan menggunakan metode *Extended Angoff* maupun Ebel masing-masing **59** dan **50,98** pada skala **100**. *Cutscore* ini berbeda cukup signifikan dengan KKM sekolah yang ditentukan dengan menggunakan metode konvensional. Berdasarkan analisis validitas *standard setting*, metode *Extended Angoff* memberikan hasil *cutscore* yang relatif lebih valid dibanding metode Ebel. Validitas *standard setting* yang diukur dalam penelitian ini adalah validitas internal yang meliputi *method consistency*, *decision consistency*, *intra-judge consistency*, dan *inter-judge consistency*.

**Kata kunci:** *standard setting*, KKM, validitas *standard setting*

## COMPARISON OF STANDARD SETTING METHOD FOR DETERMINING MINIMUM MASTERY CRITERIA

<sup>1)</sup>Susi Anto, <sup>2)</sup>Djemari Mardapi

<sup>1)</sup>LPMP Daerah Istimewa Yogyakarta, <sup>2)</sup>Universitas Negeri Yogyakarta  
<sup>1)</sup>susianto.lpmpjogja@gmail.com, <sup>2)</sup>djemarimardapi@yahoo.co.id

### Abstract

The objective of the research is to find cutscore of Minimum Mastery Criteria (KKM) by utilizing methods existing in standard setting. The methods used are Extended Angoff and Ebel methods. This research is quantitative descriptive one supported by qualitative data. Quantitative data used in this research are the pattern of students' responses against the problems of Mathematics at the End of the Year Examination for SMP/MTs for eight graders in Kabupaten Sleman 2011/2012. In addition, quantitative data obtained from expert judgement are also used for determining cut of score. Meanwhile, qualitative expert judgement is used to assess the quality of standard setting meeting. The result of this research shows that cutscore gained using both Extended Angoff and Ebel methods is 59 and 50,98 respectively on a scale of 100. This cutscore is significantly different from school KKM defined using conventional method. Based on analysis of standard setting, Extended Angoff method would provide cutscore result that is relatively more valid compared to Ebel. The validity of standard setting measured in this research is the internal validity including method consistency, decision consistency, intra-judge consistency, and inter-judge consistency.

**Keywords:** *standard setting*, *minimum mastery criteria*, *standard setting validity*

## Pendahuluan

*Standard setting* adalah proses penetapan suatu titik dalam skala skor tes tertentu yang digunakan untuk menentukan level performa suatu kebijakan atau untuk membuat klasifikasi (Sireci dalam Wells, 2007). Dalam dunia pendidikan, *standard setting* banyak digunakan untuk menentukan skor batas kelulusan atau biasa disebut sebagai skor minimum kelulusan. Namun lebih dari itu, *standard setting* sesungguhnya dapat digunakan sebagai alat bantu untuk memetakan mutu pendidikan, sebab dengan *standard setting* kita bisa membuat suatu klasifikasi rendah-sedang-tinggi kompetensi seseorang atau prestasi suatu wilayah tertentu (Athanasou dan Lamprianou, 2009, p.272).

Secara garis besar, *standard setting* dibedakan menjadi dua golongan. Pertama yang menggunakan acuan norma sementara kedua yang menggunakan acuan kriteria. Dalam interpretasi nilai beracuan norma, kemampuan peserta tes diperbandingkan dengan kemampuan orang lain dalam kelompok acuan. Sementara dalam interpretasi nilai beracuan kriteria, kemampuan peserta tes diperbandingkan dengan level kemampuan tertentu (Reynolds, Livingston, dan Willson, 2010, pp.62-64).

Penentuan skor minimal kelulusan di Indonesia belum menerapkan kedua pendekatan di atas, tapi ditetapkan berdasarkan kesepakatan dan masukan dari lapangan. Begitu pula dalam menilai ketuntasan suatu mata pelajaran. Penentuan Kriteria Ketuntasan Minimum (KKM) selama ini dilakukan oleh masing-masing sekolah dengan menggunakan metode di luar *standard setting* atau ditetapkan oleh Dinas Pendidikan di tingkat kabupaten/kota dengan tujuan tertentu yang terkadang justru kontraproduktif dengan tujuan pendidikan secara luas.

Dalam hal penentuan KKM, metode yang selama ini dikenal oleh guru adalah dengan menentukan subskornya berdasarkan *intake* siswa, daya dukung, dan kompleksitas. Dari ketiga subskor tersebut selanjutnya diambil rata-rata yang kemudian digunakan sebagai KKM. Penggunaan teknik ini tentu

menimbulkan masalah yang cukup serius terutama berkenaan dengan masalah reliabilitas, sebab menentukan nilai *intake* siswa, daya dukung, dan kompleksitas membawa konsekuensi tingginya variabilitas nilai yang mungkin muncul dari para penilai. Permasalahan reliabilitas ini bisa dieliminasi, bahkan dianulir jika KKM ditentukan dengan menggunakan *standard setting*.

Penelitian ini memilih metode *Extended Angoff* dan Ebel dengan alasan sebagai berikut. Pertama, metode ini relatif mudah jika diterapkan pada guru. Kedua, metode ini tidak memerlukan persyaratan analisis butir yang rumit, tetapi cukup dengan menggunakan teori tes klasik yang sudah cukup dikenal di kalangan para guru.

Penelitian ini bertujuan untuk: (1) mengetahui karakteristik perangkat soal UKK jenjang SMP/MTs kelas VIII mata pelajaran Matematika Kabupaten Sleman tahun pelajaran 2011/2012, (2) mengetahui skor KKM bidang studi Matematika untuk siswa SMP/MTs kelas VIII di Kabupaten Sleman jika ditentukan dengan menggunakan metode *standard setting* berbasis tes (*test-centered*) yaitu metode *Extended Angoff* dan metode Ebel, (3) mengetahui persentase ketuntasan siswa kelas VIII jenjang SMP/MTs untuk mata pelajaran Matematika di Kabupaten Sleman jika menggunakan KKM yang ditentukan dengan metode *Extended Angoff* dan metode Ebel, (4) mengetahui deskripsi kemampuan siswa kelas VIII jenjang SMP/MTs di Kabupaten Sleman untuk mata pelajaran Matematika yang telah dikatakan tuntas jika menggunakan skor KKM berdasarkan metode *Extended Angoff*, (5) membuat interpretasi terhadap perbedaan skor KKM yang dihasilkan dengan menggunakan metode *standard setting* yang berbeda, (6) mengetahui metode yang mampu menghasilkan *cutscore* dengan validitas terbaik, (7) mengetahui kompatibilitas metode *Extended Angoff* dan metode Ebel bagi para guru untuk menentukan KKM, dan (8) mengetahui metode yang lebih mudah dalam penentuan KKM diantara 2 metode yaitu metode konvensional dan metode *standard setting* (*Extended Angoff* dan Ebel).

**Metode Penelitian**

Penelitian ini merupakan penelitian deskriptif kuantitatif yang diperkuat dengan data kualitatif. Data kuantitatif yang digunakan dalam penelitian ini adalah pola respon peserta didik atas soal UKK SMP/MTs Kelas VIII Mapel Matematika Kabupaten Sleman 2011/2012. Selain itu, dalam penentuan *cut of score*, juga digunakan data kuantitatif yang diperoleh dari *expert judgement*. Sementara *expert judgement* yang bersifat kualitatif digunakan untuk menilai kualitas pelaksanaan pertemuan *standard setting*.

Penelitian ini mengambil tempat di Kabupaten Sleman. Alasan penentuan lokasi penelitian ini adalah: (1) Sleman memiliki wilayah yang cukup luas, (2) Sleman memiliki kondisi sosial ekonomi yang cukup lengkap yang mewakili tiga kondisi yang berbeda, yaitu urban, suburban, dan rural.

Penelitian dilaksanakan pada bulan April hingga Agustus 2012. Serangkaian kegiatan persiapan dilakukan pada bulan April 2012, sementara pengambilan data respon siswa dilakukan pada bulan Juni 2012. Pada bulan Juli 2012 dilakukan analisis butir soal berdasarkan data respon siswa untuk mengetahui karakteristik perangkat tes. Kegiatan selanjutnya adalah pengambilan data kuantitatif dan kualitatif melalui panel bersama *expert* dalam bentuk *focused group discussion* (FGD) yang dilaksanakan pada bulan Agustus 2012.

Data set dalam penelitian ini berupa respon peserta didik SMP/MTs kelas VIII Kabupaten Sleman tahun pelajaran 2011/2012 sebanyak 3504 siswa. Siswa tersebut terdistribusi dalam 24 SMP/MTs yang dipilih mengikuti distribusi perolehan nilai UKK jenjang SMP/MTs kelas VIII Kabupaten Sleman tahun pelajaran 2011/2012 bidang studi Matematika. Dengan demikian, sekolah yang dipilih sebagai sampel mewakili kelompok-kelompok perolehan nilai baik tinggi, sedang, maupun rendah.

Tabel 1. Sumber Data Penentuan KKM Menggunakan Metode *Extended Angoff*

No	Sekolah	Jumlah siswa	Nilai Rata-Rata
1	SMPN 1 Godean	129	8,35
2	SMPN 1 Minggir	201	6,22
3	SMPN 1 Mlati	143	5,08
4	SMPN 2 Pakem	104	4,51
5	SMPN 1 Berbah	108	4,41
6	SMPN 2 Ngemplak	207	4,16
7	SMPN 1 Ngaglik	203	3,80
8	MTsN Godean	146	3,67
9	SMPN 2 Moyudan	188	3,41
10	MTsN Tempel	160	3,25
11	SMPN 3 Turi	89	3,02
12	SMP Muh. 1 Gamping	85	2,76
	Jumlah	1763	52,64
	Nilai rata-rata sampel		4,39
	Nilai rata-rata populasi		4,36

Sumber data kuantitatif dalam penelitian ini dapat dilihat pada Tabel 1 dan Tabel 2. Selain pertimbangan pola distribusi skor siswa, pemilihan sekolah sampel juga memperhatikan letak sekolah secara geografis. Dengan pertimbangan tersebut diharapkan sampel merepresentasikan seluruh kondisi daerah yang ada, baik secara geografi maupun secara sosial ekonomi. Secara geografi, sampel diharapkan mampu mewakili wilayah Sleman Barat, Sleman Tengah, dan Sleman Timur. Sementara secara sosial ekonomi, sampel diharapkan mampu mewakili daerah urban, suburban, maupun rural.

Tabel 2. Sumber Data Penentuan KKM Menggunakan Metode Ebel

No	Sekolah	Jumlah Siswa	Nilai Rata-Rata
1	SMPN 4 Pakem	119	7,86
2	SMPN 1 Turi	108	6,24
3	SMPN 2 Kalasan	211	5,21
4	SMPN 4 Ngaglik	140	5,02
5	SMPN 1 Cangkringan	143	4,74
6	SMP Muh. 3 Depok	136	4,03
7	SMPN 3 Mlati	140	3,64
8	SMPN 2 Gamping	206	3,64
9	SMP Muh. 1 Berbah	100	3,50
10	SMP Muh. 1 Minggir	169	3,29
11	MTsN Seyegan	202	3,25
12	SMP Muh. 2 Gamping	69	2,24
	Jumlah	1743	52,65
	Nilai rata-rata sampel		4,39
	Nilai rata-rata populasi		4,36

Sumber data kualitatif dan kuantitatif lainnya adalah guru-guru pengampu bidang studi Matematika kelas VIII yang bertugas di 24 SMP/MTs yang digunakan sebagai sampel data set yang memenuhi kualifikasi tertentu. Kualifikasi dimaksud berkenaan dengan latar belakang pendidikan dan pengalaman mengajar guru yang bersangkutan. Guru-guru yang ditunjuk sebagai panelis harus memiliki latar belakang pendidikan yang sesuai dan memiliki pengalaman mengajar bidang studi Matematika jenjang SMP/MTs sekurang-kurangnya 10 tahun dan sekurang-kurangnya 5 tahun terakhir

mengampu pelajaran Matematika di kelas VIII.

Banyaknya panelis yang digunakan dalam penelitian ini untuk masing-masing metode adalah 12 untuk menjaga agar validitas *intra judge* dan *inter judge* tetap tinggi (Fowell, S.L., Fewtrell, R., & McLaughlin, P.J., 2008, p.11). Selain itu, panelis untuk metode Angoff dibedakan dengan panelis untuk metode Ebel untuk menghindarkan bias dan interferensi dalam estimasi mereka.

Data yang diperlukan dalam penelitian ini dikumpulkan dengan menggunakan teknik dokumentasi dan melalui *focused group discussion* (FGD). Instrumen pengumpul data yang digunakan adalah lembar jawab siswa, lembar kerja panelis, dan angket panelis.

Karakteristik perangkat tes perlu dianalisis terlebih dahulu untuk memutuskan apakah butir-butir yang ada memenuhi syarat untuk digunakan sebagai basis penentuan *cutscore*. Untuk mengetahui karakteristik butir, respon jawaban peserta akan dikaji secara kuantitatif melalui analisis empirik dengan mengacu pada teori tes klasik. Perangkat bantu yang digunakan dalam analisis tersebut adalah program IteMan versi 3.0.

Selanjutnya untuk menentukan *cutscore*, penelitian ini menggunakan dua metode, yaitu metode *Extended Angoff* dan metode Ebel. Impara dan Plake dalam Stahl (2008, p.7) mengusulkan suatu modifikasi untuk lebih menyederhanakan proses pada metode *Original Angoff*. Modifikasi tersebut mengusulkan agar panelis memutuskan apakah peserta ujian secara individual mampu atau tidak mampu menjawab butir soal dengan benar. Cara ini dinamakan prosedur Ya/Tidak. Jawaban diskor 1 bila Ya dan diskor 0 jika Tidak. Skor yang diperoleh dari tiap butir soal pada masing-masing panelis kemudian dijumlahkan. Hasil penjumlahan itu merupakan kriteria kelulusan minimum atau *minimum passing level* dari panelis. Nilai rata-rata *minimum passing level* dari masing-masing panelis inilah yang merupakan *final cutting score* perangkat tes.

Metode kedua yang digunakan dalam penelitian ini adalah metode Ebel. Metode yang dikemukakan Ebel mengusulkan suatu prosedur yang mirip dengan Angoff, namun dengan pertimbangan keterkaitan menyangkut relevansi isi tiap butir dan tingkat kesukarannya dapat mempengaruhi pertimbangan tentang bagaimana kualifikasi minimal peserta tes yang diharapkan untuk mengerjakan suatu butir (Ebel, 1979, p.340). Teknik ini menggunakan suatu jaringan dua dimensi untuk membuat kategori setiap butir. Dimensi satu menyangkut tingkat relevansi isi butir, sementara dimensi dua menyangkut tingkat kesukaran butir. Tingkat relevansi butir dibedakan menjadi empat kategori yang meliputi *essential* (sangat penting), *important* (penting), *acceptable* (dapat diterima), dan *questionable* (dipertanyakan). Tingkat kesukaran dibedakan menjadi tiga kategori yang meliputi mudah, sedang, dan sulit.

Setelah *cutscore* ditentukan, langkah selanjutnya adalah menganalisis persentase dan deskripsi ketuntasan siswa kelas VIII jenjang SMP/MTs di Kabupaten Sleman untuk mata pelajaran Matematika. Analisis ini penting karena beberapa alasan. Pertama, untuk mengetahui peta pencapaian siswa kelas VIII jenjang SMP/MTs di Kabupaten Sleman dalam pembelajaran Matematika. Kedua, untuk membantu membuat proyeksi persentase kelulusan siswa kelas VIII jenjang SMP/MTs di Kabupaten Sleman setahun yang akan datang pada mata pelajaran Matematika. Ketiga, untuk membantu membuat proyeksi terhadap deskripsi kemampuan yang betul-betul sudah dikuasai siswa atau masih perlu upaya keras baik oleh guru maupun siswa untuk menguasainya. Keempat, untuk membuat penilaian terhadap estimasi panelis tentang kemampuan siswa-siswinya, apakah estimasi tersebut sudah tepat atautkah *under* atau bahkan *over-estimate*.

Validitas *cutscore* yang dihasilkan juga akan dihitung untuk menentukan *cutscore* manakah dari kedua metode yang digunakan (*Extended Angoff* dan Ebel) yang memiliki tingkat validitas terbaik. Validitas *cutscore*

akan diukur melalui validitas internal yang meliputi: *method consistency*, *decision consistency*, *intra-judge consistency*, dan *inter-judge consistency*.

Konsistensi metode akan diukur melalui *standard error of judgement* (SEj) atau kesalahan baku penjurian (Norcini et al., 1987, p.60). Menurut Cohen et al., (1999, p.362), konsistensi metode dapat dikatakan memenuhi jika nilai SEj sama dengan atau lebih kecil dari setengah nilai *standar error of measurement* (SEM) tes.

Konsistensi keputusan akan diukur dengan menggunakan persamaan:

$$|Z| = (C_x - M - 0,5) / S_x$$

Artinya  $C_x$  adalah *cutscore* tes,  $M$  adalah rata-rata skor peserta tes, dan  $S_x$  adalah simpangan baku skor peserta tes (Cizek & Bunch, 2007, p.307). Harga mutlak  $Z$  kemudian digunakan untuk menentukan estimasi koefisien kesepakatan ( $p_0$ ) dan nilai kappa ( $k$ ) dari dua tabel Subkoviak yang tercantum pada Cizek & Bunch (2007, pp.310-311). Semakin tinggi koefisien kesepakatan ( $p_0$ ) maka semakin tinggi pula proporsi keputusan konsisten yang diharapkan jika prosedur pengujian ulang dilakukan. Begitu pula semakin besar nilai kappa ( $k$ ) mengindikasikan makin besarnya pengaruh prosedur pengujian terhadap konsistensi keputusan. Rendahnya nilai kappa ( $k$ ) menunjukkan bahwa *cutscore* yang dihasilkan terletak jauh dari area distribusi skor dengan tingkat densitas tertinggi dan estimasi reliabilitas tes rendah (Cizek & Bunch, 2007, p.312).

*Intra-judge consistency* yang akan diukur adalah *intra-judge consistency* baik di dalam putaran maupun antar putaran. *Intra-judge consistency across rounds* akan dilihat dari magnitudo perubahan rating butir dari putaran ke putaran dan banyaknya rating butir yang berubah dari putaran ke putaran (Loomis et al., 2000, p.13). *Intra-judge consistency within rounds* akan ditentukan berdasarkan pemetaan keputusan rating setiap butir pada tabel tingkat kesulitan *item* (Loomis et al., 2000, p.15). Pemetaan itu akan diuji melalui nilai korelasinya. Nilai korelasi yang tinggi mengindikasikan tingkat konsistensi

yang memuaskan. Jika diestimasi bahwa tingkat kesulitan empiriknya tinggi -bahkan untuk panelis yang telah *expert* sekalipun, maka nilai korelasi di atas 0,30 sudah dianggap memuaskan (Alderson, 1993, p.55).

*Inter-judge consistency* merupakan kriteria yang penting untuk menilai apakah suatu proses *standard setting* memiliki alasan yang cukup memadai atau tidak (Loomis et al., 2000, p.13). *Inter-judge consistency* akan dilihat dari simpangan baku *cutscore*. Secara umum simpangan baku antar putaran akan semakin mengecil untuk *cutscore* pada level cakap dan ahli, dan akan membesar untuk *cutscore* pada level dasar (Loomis et al., 2000, p.13).

Untuk menjawab pertanyaan apakah metode *Extended Angoff* dan metode Ebel cukup kompatibel digunakan oleh guru untuk menentukan KKM serta pertanyaan apakah metode *Extended Angoff* dan metode Ebel lebih mudah digunakan untuk menentukan KKM jika dibandingkan dengan metode konvensional akan digunakan analisis statistik deskriptif. Data yang dianalisis merupakan *feedback* panelis atas penyelenggaraan pertemuan *standard setting* yang telah dilakukan. Data tersebut diperoleh dari respon panelis atas instrumen angket yang disebarkan setelah putaran penentuan *standard setting* dilakukan.

## Hasil Penelitian dan Pembahasan

Sebagaimana telah diungkap pada bagian sebelumnya, karakteristik butir soal perangkat tes perlu diketahui terlebih dahulu sebelum digunakan dalam penentuan *cutscore*. Karakteristik butir tersebut digunakan sebagai dasar keputusan apakah instrumen tes dimaksud memenuhi syarat atau tidak untuk digunakan sebagai basis penentuan *cutscore*, sebab kelayakan instrumen menjadi prasyarat utama yang harus dipenuhi dalam penentuan *cutscore* yang berbasis alat tes (*test-centered models*).

Penentuan *cutscore* yang berbasis alat tes (*test-centered models*) mengasumsikan

bahwa perangkat tes yang digunakan adalah perangkat tes yang terstandar. Definisi terstandar ini mencakup terstandar tahapan penyusunannya, terstandar cakupan dan kedalaman materinya, maupun terstandar karakteristik tiap butir penyusunannya.

Dilihat dari sisi tahapan penyusunannya, perangkat tes dimaksud sudah bisa dikatakan terstandar karena disusun melalui prosedur yang telah terstandar oleh tim di bawah Musyawarah Kerja Kepala Sekolah (MKKS) jenjang SMP/MTs di Kabupaten Sleman. Begitu pula dari sisi cakupan dan kedalaman materinya, perangkat tes dimaksud juga sudah bisa dikatakan terstandar karena disusun berdasarkan SK dan KD yang sesuai dengan kelas dan jenjangnya sebagaimana termaktub dalam Standar Isi. Sementara dari karakteristik tiap butir penyusunannya, perlu dilakukan analisis untuk menyimpulkan apakah perangkat tes tersebut sudah bisa dikatakan terstandar atau belum.

Berdasarkan analisis butir menggunakan program Iteman tersebut, diperoleh kesimpulan bahwa seluruh butir yang terdapat dalam instrumen tes dapat digunakan dalam penentuan *cutscore*, karena semua indikator memberikan informasi yang bagus. Indikator-indikator itu adalah tingkat kesulitan butir, daya beda butir, efektivitas distraktor, dan keberfungsian kunci jawaban.

### Metode *Extended Angoff*

Putaran pertama metode *Extended Angoff* menghasilkan *cutscore* sebesar 23 butir benar dari 40 butir soal UKK atau sebesar 58%. Dengan menggunakan *final cutscore* 1 ini sebagai KKM, ternyata ada 22% siswa di sekolah sampel yang bisa dikatakan telah tuntas dalam mengikuti materi pembelajaran matematika SMP/MTs kelas VIII. Sementara pada putaran kedua, menghasilkan *cutscore* sebesar 24 butir benar dari 40 butir soal UKK atau sebesar 60%. Dengan menggunakan *final cutscore* 2 ini sebagai KKM, tingkat ketuntasan turun menjadi 19%.

Tabel 3. Rekapitulasi *Cutscore* Metode *Extended Angoff*

No. Panelis	<i>Cutscore</i> Panelis				Perubahan Keputusan	
	Putaran 1		Putaran 2			
1	19	48%	20	50%	1	2,5%
2	14	35%	15	38%	1	2,5%
3	25	63%	26	65%	1	2,5%
4	24	60%	23	58%	1	2,5%
5	23	58%	23	58%	0	0,0%
6	22	55%	27	68%	5	12,5%
7	25	63%	28	70%	3	7,5%
8	23	58%	25	63%	2	5,0%
9	22	55%	20	50%	2	5,0%
10	24	60%	24	60%	0	0,0%
11	31	78%	31	78%	0	0,0%
12	24	60%	26	65%	2	5,0%
Rata-Rata	23	58%	24	60%	1,5	3,75%
KKM Final	23,5		59%			

Tabel 3 menampilkan rekapitulasi *cutscore* para panelis pada putaran pertama dan kedua metode *Extended Angoff*. Rata-rata estimasi para panelis pada putaran pertama dan kedua menunjukkan perbedaan angka yang sudah dapat diterima yaitu 1,5 butir dari 40 butir soal (3,75%). Oleh karena itu tidak diperlukan adanya putaran ketiga pada metode *Extended Angoff* ini.

Tabel 4. Hasil Penentuan KKM Bidang Studi Matematika Jenjang SMP/MTs Kelas VIII di Kabupaten Sleman Menggunakan Metode *Extended Angoff*

Putaran	Banyak		Persentase	
	Butir Benar	Persentase Benar	Banyak Siswa Tuntas	Persentase Siswa Tuntas
Pertama	23	58%	765	22%
Kedua	24	60%	682	19%
Rata-Rata	23,5	59%	682	19%

KKM sesungguhnya diperoleh dengan mengambil rata-rata *cutscore* dari putaran pertama dan kedua. Tabel 4 menunjukkan skor KKM akhir sebesar 59%. De-

ngan menggunakan KKM ini, ada 682 dari 3504 (sekitar 19%) siswa di sekolah sampel yang bisa dikatakan telah tuntas dalam mengikuti materi pembelajaran matematika SMP/MTs kelas VIII.

Setelah menentukan *cutscore* pada putaran pertama, para panelis mendiskusikan deskriptor dari *cutscore* yang ditetapkan dengan metode *Extended Angoff*. Deskriptor yang digunakan merupakan kompetensi yang diukur oleh tiap butir soal hasil diskusi dan disesuaikan dengan frekuensi panelis yang menjawab Ya. Hasil diskusi menunjukkan ada 38 deskriptor yang memperoleh minimal 1 jawaban Ya dari panelis. Itu berarti populasi siswa yang telah melampaui skor KKM atau siswa yang telah tuntas dapat dikatakan setidaknya memiliki 38 kompetensi sebagaimana diukur oleh ke-38 butir tersebut.

#### Metode Ebel

Putaran pertama metode Ebel menghasilkan *cutscore* sebesar 19,93 atau 49,82 pada skala 100. Dengan menggunakan *final cutscore* 1 ini sebagai KKM, ada 32% siswa di sekolah sampel yang bisa dikatakan telah tuntas dalam mengikuti materi pembelajar-

an Matematika SMP/MTs kelas VIII. Sementara pada putaran kedua, penentuan *cutscore* dengan metode Ebel menghasilkan *cutscore* sebesar 20,85 atau 52,13 pada skala 100. Dengan menggunakan *final cutscore* 2 ini sebagai KKM, tingkat ketuntasan menurun menjadi 28%.

Tabel 5 menampilkan rekapitulasi *cutscore* para panelis pada putaran pertama dan

kedua metode Ebel. Rata-rata estimasi para panelis pada putaran pertama dan kedua menunjukkan perbedaan angka yang sudah dapat diterima yaitu 1,82 butir dari 40 butir soal (4,55%). Oleh karena itu tidak diperlukan adanya putaran ketiga pada metode Ebel ini.

Tabel 5. Rekapitulasi *Cutscore* Metode Ebel

No. Panelis	<i>Cutscore</i> Panelis				Perubahan Keputusan	
	Putaran 1		Putaran 2			
1	14,4	36,0%	17,4	43,5%	3,00	7,5%
2	24,6	61,5%	27,5	68,8%	2,90	7,3%
3	24,5	61,3%	23,8	59,5%	0,70	1,8%
4	22,1	55,3%	21,9	54,8%	0,20	0,5%
5	13,7	34,3%	16,34	40,9%	2,64	6,6%
6	14,68	36,7%	14,68	36,7%	0,00	0,0%
7	13,95	34,9%	19,95	49,9%	6,00	15,0%
8	16,35	40,9%	15,5	38,8%	0,85	2,1%
9	18,25	45,6%	18	45,0%	0,25	0,6%
10	25,1	62,8%	22	55,0%	3,10	7,8%
11	35	87,5%	34,75	86,9%	0,25	0,6%
12	16,5	41,3%	18,4	46,0%	1,90	4,8%
Rata-Rata	19,93	49,8%	20,85	52,13%	1,82	4,55%
KKM Final	20,39		50,97%			

Tabel 6 memberi informasi yang lengkap hasil perhitungan KKM dengan menggunakan metode Ebel berikut persentase tingkat ketuntasan siswa di sekolah sampel. KKM sesungguhnya diperoleh dengan mengambil rata-rata skor KKM pada putaran pertama dan kedua. Dari nilai rata-rata itu diperoleh KKM akhir sebesar 50,97%. Dengan menggunakan KKM ini, ada 972 dari 3504 (sekitar 28%) siswa di sekolah sampel yang bisa dikatakan telah tuntas dalam mengikuti materi pembelajaran matematika SMP/MTs kelas VIII.

Tabel 6. Hasil Penentuan KKM Bidang Studi Matematika Jenjang SMP/MTs Kelas VIII di Kabupaten Sleman Menggunakan Metode Ebel

Putaran	Skor KKM	Persentase Skor KKM	Banyak Siswa Tuntas	Persentase Siswa Tuntas
Pertama	19,93	49,82%	1134	32%
Kedua	20,85	52,13%	972	28%
Rata-Rata	20,39	50,97%	972	28%

Metode *Extended* Angoff dan Ebel pada penelitian ini masing-masing menghasilkan nilai *cutscore* 59 dan 50,98 pada skala 100. Ini berarti kedua metode menghasilkan nilai *cutscore* yang berbeda. Perbedaan itu muncul sebagai implikasi logis dari perbedaan konsep yang melatarbelakangi metode-metode tersebut. Beberapa hal yang bisa diuraikan untuk menjelaskan perbedaan itu adalah sebagai berikut.

Pertama, pada metode *Extended* Angoff, baik tingkat kesulitan soal maupun tingkat kemampuan peserta tes berpengaruh secara dependen satu sama lain terhadap keputusan Ya/Tidak panelis. Sementara pada metode Ebel tingkat kesulitan soal dan tingkat kemampuan peserta tes berpengaruh secara independen satu sama lain terhadap keputusan panelis. Selain itu di samping tingkat kesulitan soal, pada metode Ebel ada juga parameter lain yang ikut berpengaruh terhadap keputusan panelis yaitu tingkat relevansi soal. Pada metode Ebel tingkat kesulitan dan tingkat relevansi butir berpengaruh pada keputusan jumlah butir yang tersarang dalam tiap-tiap sel. Sementara tingkat kemampuan peserta tes berpengaruh kepada persentase jawaban benar yang tersarang dalam tiap-tiap sel.

Kedua, metode Ebel membutuhkan lebih banyak estimasi dibanding metode *Extended* Angoff. Jika pada metode *Extended* Angoff panelis hanya diminta membuat satu jenis estimasi, maka tidak demikian halnya pada metode Ebel. Sebab ada tiga parameter yang mesti diestimasi oleh para panelis pada metode Ebel. Parameter itu adalah tingkat kesulitan butir, tingkat relevansi butir, dan tingkat kemampuan peserta tes. Banyaknya parameter yang mesti diestimasi ini akan berbanding lurus dengan banyaknya keraguan yang mungkin muncul dari para panelis. Semakin banyak parameter yang mesti diestimasi, semakin tinggi pula frekuensi munculnya ragu panelis.

Ketiga, nilai proporsi metode *Extended* Angoff bisa dikatakan lebih detail dibanding metode Ebel. Sebab pada metode *Extended* Angoff proporsi ditentukan untuk setiap butirnya sementara pada metode Ebel, nilai proporsi ditentukan untuk

setiap selnya. Hal ini membawa implikasi nilai proporsi pada metode Ebel lebih bias dibanding nilai proporsi pada metode *Extended* Angoff.

Metode *Extended* Angoff pada penelitian ini menghasilkan nilai *cutscore* 59 pada skala 100. Pada nilai *cutscore* 59 ini banyaknya siswa yang tuntas belajarnya adalah 19% dari 3504 siswa sampel. Sementara pada metode Ebel dihasilkan skor *cutscore* 50,98 dan tingkat ketuntasan siswa di angka 28%. Rendahnya persentase ketuntasan siswa ini menunjukkan baik panelis metode *Extended* Angoff maupun Ebel, keduanya telah *over estimate* dalam memberikan penilaian atas kemampuan peserta didiknya. Jika estimasi mereka tepat, maka akan dihasilkan skor KKM yang lebih rendah dan persentase ketuntasan yang lebih tinggi.

#### Validitas Standard Setting

Validitas *cutscore* yang dihasilkan baik oleh metode *Extended* Angoff maupun oleh metode Ebel akan diukur melalui validitas internalnya. Validitas internal yang hendak diukur tersebut meliputi *method consistency*, *decision consistency*, *intra-judge consistency*, dan *inter-judge consistency*.

Tabel 7. Kesalahan Baku Penjurian (SEj) Metode *Extended* Angoff dan Metode Ebel

Putaran	Kesalahan Baku (SEj)		Referensi <sup>i</sup>
	Ext. Angoff	Ebel	
Pertama	1,148	1,870	≤ 1.3885
Kedua	1,225	1,657	≤ 1.3885
Putaran 1 & 2	0,828	1,225	≤ 1.3885

Konsistensi metode dapat diukur melalui *standard error of judgement* (SEj) atau kesalahan baku penjurian. Sebagaimana ditunjukkan oleh Tabel 7, baik pada putaran pertama maupun pada putaran kedua, nilai kesalahan baku (*standard error*) pada metode Ebel selalu lebih tinggi dibanding pada metode *Extended* Angoff. Tabel 7 juga menunjukkan bahwa nilai Se<sub>j</sub> metode Ebel baik pada putaran 1 maupun pada putaran 2 selalu di atas nilai referensi yang disyarat-

kan yaitu setengah kali nilai SEM. Berangkat dari fenomena ini dapat dikatakan bahwa *method consistency* pada metode Ebel putaran 1 dan putaran 2 belum memenuhi syarat. Syarat konsistensi metode Ebel ini baru terpenuhi ketika rating *cutscore* individual digabung antara putaran 1 dan putaran 2. Itupun juga masih lebih besar dibanding metode *Extended* Angoff.

Berangkat dari premis-premis tersebut dapat disimpulkan bahwa *method consistency* metode *Extended* Angoff lebih bagus dibanding metode Ebel. Atau dengan kata lain dapat diungkapkan bahwa metode *Extended* Angoff mampu memberikan konsistensi penjurian yang lebih baik dibanding metode Ebel. Hal ini bisa menjadi petunjuk bahwa bias estimasi pada metode Ebel lebih tinggi dibanding pada metode *Extended* Angoff.

Metode Ebel memungkinkan munculnya lebih banyak kesalahan dibanding metode *Extended* Angoff. Hal tersebut disebabkan metode Ebel membutuhkan lebih banyak estimasi dibanding metode *Extended* Angoff. Banyaknya parameter yang harus diestimasi ini berbanding lurus dengan banyaknya keraguan yang mungkin muncul dari para panelis. Semakin banyak parameter yang mesti diestimasi, semakin tinggi pula frekuensi munculnya ragu panelis. Keraguan inilah yang sering membawa implikasi pada munculnya bias estimasi. Bias estimasi inilah yang terbaca pada nilai *standard error* sebagaimana ditunjukkan Tabel 7.

Selain itu, tingkat akurasi nilai proporsi pada metode *Extended* Angoff bisa dikatakan lebih tinggi dibanding pada metode Ebel. Sebab pada metode *Extended* Angoff proporsi ditentukan untuk setiap

butirnya sementara pada metode Ebel, nilai proporsi ditentukan untuk setiap selnya. Hal ini membawa implikasi nilai proporsi pada metode Ebel lebih bias dibanding nilai proporsi pada metode *Extended* Angoff.

Konsistensi keputusan ditentukan dengan cara menghitung nilai harga mutlak Z terlebih dahulu. Harga mutlak Z kemudian digunakan untuk menentukan estimasi koefisien kesepakatan ( $p_0$ ) dan nilai kappa ( $k$ ) dari dua tabel Subkoviak. Tabel 8 menunjukkan harga  $|Z|$ , alpha,  $p_0$ , dan  $k$  untuk masing-masing metode dan masing-masing putaran. Berdasarkan tabel Subkoviak nilai maksimum  $p_0$  adalah 0,98 dan nilai maksimum untuk  $k$  adalah 0,71. Oleh karena itu dapat disimpulkan bahwa berdasar Tabel 8, konsistensi keputusan kedua metode pada penelitian ini sudah cukup memuaskan seperti ditunjukkan oleh harga  $p_0$  dan  $k$ .

Sebagaimana diketahui bahwa semakin tinggi koefisien kesepakatan ( $p_0$ ) maka semakin tinggi pula proporsi keputusan konsisten yang diharapkan jika prosedur pengujian ulang dilakukan. Oleh karena itu, harga  $p_0$  yang di atas 0,8 (maksimum 0,98) menunjukkan bahwa munculnya proporsi keputusan konsisten yang diharapkan cukup tinggi jika prosedur pengujian ulang dilakukan baik pada metode *Extended* Angoff maupun pada metode Ebel. Tabel 8 juga menunjukkan bahwa nilai  $p_0$  metode *Extended* Angoff selalu lebih tinggi dibanding Ebel baik itu pada putaran 1, putaran 2, maupun gabungan putaran 1 dan 2. Hal ini mengindikasikan bahwa harapan munculnya proporsi keputusan yang konsisten jika pengujian ulang dilakukan selalu lebih tinggi pada metode *Extended* Angoff dibanding pada Metode Ebel.

Tabel 8. Nilai Koefisien Kesepakatan ( $p_0$ ) dan Kappa ( $k$ ) Metode *Extended* Angoff dan Metode Ebel

	<i>Extended</i> Angoff			Ebel		
	Put. 1	Put. 2	Total	Put. 1	Put. 2	Total
$ Z $	0,67	0,81	0,74	0,25	0,38	0,31
Alpha	0,85	0,85	0,85	0,85	0,85	0,85
$p_0$	0,86	0,88	0,87	0,83	0,84	0,83
Kappa ( $k$ )	0,64	0,62	0,63	0,65	0,65	0,65

Jika diperbandingkan nilai kappa, maka metode *Extended Angoff* selalu menghasilkan nilai kappa yang lebih rendah dibandingkan metode Ebel, baik pada putaran 1, putaran 2, maupun gabungan putaran 1 dan 2. Hal ini menunjukkan bahwa *cutscore* yang dihasilkan metode *Extended Angoff* selalu terletak lebih jauh dari area distribusi skor dengan tingkat densitas tertinggi dibandingkan dengan *cutscore* yang dihasilkan metode Ebel.

Sebagaimana telah diuraikan di depan, metode *Extended Angoff* dan Ebel pada penelitian ini masing-masing menghasilkan nilai *cutscore* 59 dan 50,98 pada skala 100. Di samping itu, berdasarkan statistik deskriptif dapat diketahui bahwa nilai *skew* skor tes yang dihasilkan berharga positif. Nilai *skew* yang berharga positif mengindikasikan bah-

wa sebagian besar skor peserta tes terletak di bagian bawah (skor rendah) dari distribusi skor. Selain itu juga diketahui bahwa nilai kurtosis skor tes berharga positif. Nilai kurtosis yang berharga positif menunjukkan distribusi yang lebih lancip (memuncak) yang berarti sebagian besar skor peserta mengumpul di sekitar nilai modus (modus = 37,5) yang condong ke sebelah kiri dari kurva distribusi. Oleh karena itu, bisa dipahami jika nilai kappa yang dihasilkan metode Ebel selalu lebih tinggi dibanding nilai kappa yang dihasilkan metode *Extended Angoff*, karena *cutscore* yang dihasilkan metode Ebel selalu lebih dekat ke nilai modus (modus = 37,5) skor peserta tes dibanding *cutscore* yang dihasilkan metode *Extended Angoff* sebagaimana ditunjukkan oleh Tabel 9.

Tabel 9. KKM Metode *Extended Angoff* dan Metode Ebel vs Modus Skor Tes

	KKM <i>Extended Angoff</i>		KKM Ebel		Modus Skor Tes	
	Skor	n benar	Skor	n benar	Skor	n benar
Put. 1	60	24	49,825	20		
Put. 2	57,5	23	52,025	21	37,5	15
Total	58,75	23,5	50,925	20,5		

*Intra-judge consistency* diukur melalui *intra-judge consistency* baik di dalam maupun antar putaran. *Intra-judge consistency across rounds* atau konsistensi *intra-judge* antarputaran dapat dilihat dari magnitudo perubahan rating *cutscore* individual panelis dari putaran ke putaran dan banyaknya rating butir yang berubah dari putaran ke putaran (Loomis et al., 2000, p.13). Tabel 10 menunjukkan magnitudo perubahan rating *cutscore* individual panelis dari putaran 1 ke putaran 2 sementara Tabel 11 menunjukkan banyaknya butir yang rating *cutscore* butirnya berubah dari putaran 1 ke putaran 2. Kedua tabel tersebut menggunakan basis perhitungan banyaknya butir tes yaitu 40.

Tabel 10 menunjukkan kepada kita bahwa magnitudo perubahan rating *cutscore* individual panelis dari putaran 1 ke putaran 2 metode *Extended Angoff* lebih kecil dibanding metode Ebel. Rata-rata perubahan untuk metode *Extended Angoff* adalah 1,5 (3,75%) sementara untuk metode Ebel adalah 1,82 (4,54%). Ini berarti pada metode *Extended Angoff* setiap panelis rata-rata membuat perubahan keputusan sebesar 3,75% dari keputusan sebelumnya. Sementara pada metode Ebel setiap panelis rata-rata membuat perubahan keputusan sebesar 4,54% dari keputusan sebelumnya. Angka-angka ini menunjukkan bahwa *intra-judge consistency across rounds* pada metode *Extended Angoff* lebih baik dibanding pada

metode Ebel. Dengan kata lain panelis pada metode *Extended* Angoff lebih konsisten dalam membuat keputusan atau penjurian antar putaran dibanding panelis pada metode Ebel.

Tabel 10. Magnitudo Perubahan Rating *Cutscore* dari Putaran 1 ke Putaran 2 Metode *Extended* Angoff dan Metode Ebel

	Magnitudo Perubahan Rating <i>Cutscore</i> (Skala 40)	
	Ext. Angoff	Ebel
Panelis 1	1	3
Panelis 2	1	2,9
Panelis 3	1	0,7
Panelis 4	1	0,2
Panelis 5	0	2,64
Panelis 6	5	0
Panelis 7	3	6
Panelis 8	2	0,85
Panelis 9	2	0,25
Panelis 10	0	3,1
Panelis 11	0	0,25
Panelis 12	2	1,9
Mean	1,5	1.82

Tabel 11 menunjukkan banyaknya butir yang rating *cutscore* butirnya berubah dari putaran 1 ke putaran 2. Sebagaimana Tabel 10, Tabel 11 juga menggunakan basis perhitungan banyaknya butir tes yaitu 40. Secara umum seluruh panelis baik pada metode *Extended* Angoff maupun metode Ebel dapat dikatakan tidak konsisten karena melakukan perubahan terhadap keputusan butir. Hanya ada seorang panelis di metode *Extended* Angoff yang bisa dika-

takan konsisten yaitu Panelis 10 karena dia sama sekali tidak melakukan perubahan atas keputusan butir.

Sementara itu, dibanding metode *Extended* Angoff konsistensi panelis pada metode Ebel bisa dikatakan lebih buruk. Hampir seluruh panelis pada metode Ebel melakukan perubahan di atas 20% dari butir yang ada. Hanya ada seorang panelis yaitu Panelis 1 yang melakukan perubahan sebanyak 1 digit. Sementara ada 4 panelis yang melakukan perubahan 50% atau lebih yaitu Panelis 3, Panelis 4, Panelis 7, dan Panelis 10. Bahkan ada 2 panelis yang melakukan perubahan terhadap lebih dari 60% butir yang ada, yaitu Panelis 2 dan Panelis 9.

Tabel 11. Banyak Butir yang Berubah dari Putaran 1 ke Putaran 2 Metode *Extended* Angoff dan Metode Ebel

	Banyak Butir yang Berubah	
	Ext. Angoff	Ebel
Panelis 1	15	8
Panelis 2	9	30
Panelis 3	7	21
Panelis 4	1	20
Panelis 5	4	14
Panelis 6	5	12
Panelis 7	3	21
Panelis 8	2	10
Panelis 9	10	30
Panelis 10	0	22
Panelis 11	2	13
Panelis 12	10	15
<b>Mean</b>	<b>5,7</b>	<b>18</b>

Jika dilihat rata-ratanya banyak keputusan butir yang berubah untuk metode *Extended* Angoff adalah 5,7 butir (14%) sementara untuk metode Ebel adalah 18 butir (45%). Ini berarti pada metode *Extended* Angoff banyaknya butir yang berubah keputusannya pada putaran 2 untuk setiap panelis adalah 14% dari butir yang ada. Sementara pada metode Ebel banyaknya butir yang berubah keputusannya pada putaran 2 untuk setiap panelis adalah 45% dari butir yang ada. Angka-angka ini menunjukkan bahwa *intra-judge consistency across rounds* pada metode *Extended* Angoff lebih baik dibanding pada metode Ebel. Dengan kata lain panelis pada metode *Extended* Angoff lebih konsisten dalam membuat keputusan atau penjurian antar putaran dibanding panelis pada metode Ebel.

Fenomena ini sekali lagi membuktikan premis yang telah dituliskan di depan bahwa banyaknya estimasi yang mesti dilakukan oleh panelis berbanding lurus dengan besarnya ragu yang mungkin muncul dari para panelis. Oleh karena itu angka-angka tersebut bisa dijelaskan dengan cara menelusuri tahapan-tahapan yang ada pada masing-masing metode. Disparitas yang cukup tinggi antara kedua metode (14% *versus* 45%) menunjukkan perbedaan potensi ragu yang cukup signifikan. Angka itu bisa mengatakan bahwa metode Ebel memiliki potensi ragu 3 kali lebih besar dari potensi ragu yang mungkin timbul pada metode *Extended* Angoff. Meski masih perlu dibuktikan lagi, tapi angka itu sebanding dengan proporsi banyaknya estimasi antara kedua metode. Jika metode *Extended* Angoff hanya membutuhkan satu jenis estimasi maka pada metode Ebel dibutuhkan tiga jenis estimasi.

*Intra-judge consistency within rounds* ditentukan berdasarkan pemetaan keputusan rating setiap butir pada tabel tingkat kesulitan *item*. Pemetaan itu akan diuji melalui nilai korelasinya. Nilai korelasi yang tinggi mengindikasikan tingkat konsistensi yang memuaskan, sebaliknya nilai korelasi yang

rendah mengindikasikan rendahnya konsistensi *intra-judge* dalam putaran.

Tabel 12 menunjukkan hasil pemetaan antara tingkat kesulitan butir statistik dan representasi tingkat kesulitan butir empirik metode *Extended* Angoff dan metode Ebel. Pada tabel tersebut, pemetaan dilakukan dengan terlebih dahulu mengurutkan butir berdasar tingkat kesulitannya ( $p$ ) secara statistik dari mudah ke sulit. Urutan representasi tingkat kesulitan butir empirik kedua metode dibuat mengikuti urutan tingkat kesulitan butir secara statistik. Pengurutan dengan cara itu memungkinkan kita mengukur masing-masing nilai korelasinya.

Tingkat kesulitan butir statistik merupakan proporsi siswa yang menjawab benar butir tersebut yang diperoleh dari *Item Analysis*. Sementara representasi tingkat kesulitan butir empirik untuk metode *Extended* Angoff diperoleh dari banyaknya jawaban panelis yang bernilai 1 untuk tiap butirnya atau bisa juga disebut rating *cutscore* butir. Untuk mendapatkan representasi tingkat kesulitan butir empirik metode Ebel, terlebih dahulu butir mudah diberi skor 3, butir sedang diberi skor 2, dan butir sulit diberi skor 1. Jumlah skor tingkat kesulitan dari semua panelis untuk setiap butirnya itulah yang merupakan representasi tingkat kesulitan butir empirik untuk metode Ebel.

Tabel 12 menunjukkan adanya nilai korelasi yang cukup tinggi antara tingkat kesulitan butir statistik dan representasi tingkat kesulitan butir empirik baik pada metode *Extended* Angoff maupun pada metode Ebel. Keseluruhan nilai korelasi tersebut sudah jauh di atas angka yang dipersyaratkan Alderson yaitu 0,3. Ini menunjukkan bahwa para panelis baik pada metode *Extended* Angoff maupun metode Ebel, baik pada putaran 1 maupun putaran 2 telah memiliki nilai *intra-judge consistency within rounds* yang memuaskan. Atau dengan kata lain, tingkat konsistensi penjurian para panelis dalam tiap putaran sudah memuaskan.

Tabel 12. Pemetaan Tingkat Kesulitan Butir Statistik dan Representasi Tingkat Kesulitan Butir Empirik Metode *Extended Angoff* dan Metode Ebel

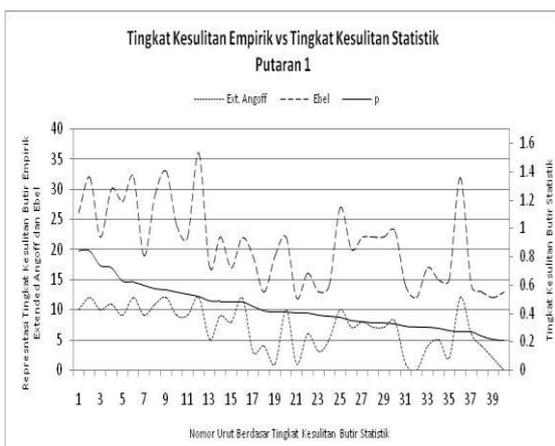
No Urut	No Butir	P	Kategori	Representasi Tingkat Kesulitan Butir Empirik			
				Putaran I		Putaran II	
				Ext. Angf	Ebel	Ext. Angf	Ebel
1	30	0,844	Mudah	10	26	10	27
2	2	0,840	Mudah	12	32	12	32
3	12	0,737	Sedang	10	22	11	24
4	21	0,720	Sedang	11	30	10	31
5	29	0,627	Sedang	9	28	11	29
6	1	0,623	Sedang	12	32	12	35
7	13	0,598	Sedang	9	19	8	15
8	11	0,572	Sedang	11	29	12	27
9	10	0,563	Sedang	12	33	12	28
10	3	0,550	Sedang	9	24	11	25
11	33	0,536	Sedang	9	22	8	18
12	9	0,517	Sedang	12	36	12	36
<b>13</b>	<b>25</b>	<b>0,484</b>	<b>Sedang</b>	<b>5</b>	<b>17</b>	<b>5</b>	<b>17</b>
14	16	0,483	Sedang	9	22	9	21
15	7	0,451	Sedang	8	17	10	18
16	20	0,478	Sedang	12	22	10	22
17	5	0,451	Sedang	3	19	5	20
18	8	0,416	Sedang	4	13	1	13
<b>19</b>	<b>17</b>	<b>0,408</b>	<b>Sedang</b>	<b>1</b>	<b>19</b>	<b>2</b>	<b>19</b>
20	4	0,407	Sedang	10	22	10	22
<b>21</b>	<b>34</b>	<b>0,405</b>	<b>Sedang</b>	<b>1</b>	<b>12</b>	<b>2</b>	<b>13</b>
22	32	0,402	Sedang	6	16	6	17
23	19	0,384	Sedang	3	13	4	14
24	36	0,377	Sedang	5	14	7	13
25	14	0,369	Sedang	10	27	8	27
26	26	0,349	Sedang	7	20	8	20
27	23	0,339	Sedang	8	22	8	19
28	15	0,333	Sedang	7	22	7	24
29	6	0,332	Sedang	7	22	9	19
30	24	0,325	Sedang	8	23	10	26
<b>31</b>	<b>22</b>	<b>0,305</b>	<b>Sedang</b>	<b>1</b>	<b>14</b>	<b>3</b>	<b>13</b>
<b>32</b>	<b>40</b>	<b>0,303</b>	<b>Sedang</b>	<b>0</b>	<b>12</b>	<b>0</b>	<b>13</b>
33	27	0,300	Sedang	4	17	7	17
34	35	0,295	Sulit	5	15	3	14
35	31	0,276	Sulit	2	15	2	16
<b>36</b>	<b>28</b>	<b>0,270</b>	<b>Sulit</b>	<b>12</b>	<b>32</b>	<b>12</b>	<b>34</b>
37	39	0,268	Sulit	6	14	6	14
38	38	0,236	Sulit	4	13	2	12
39	37	0,216	Sulit	2	12	3	12
40	18	0,209	Sulit	0	13	0	13
<b>Correlation</b>				<b>0,63</b>	<b>0,608</b>	<b>0,621</b>	<b>0,591</b>

Tabel 12 juga memberi informasi tentang kekurangtepatan estimasi beberapa butir oleh para panelis. Kekurangtepatan estimasi butir oleh para panelis itu terlihat kongruen baik pada metode *Extended* Angoff maupun pada metode Ebel. Butir-butir tersebut adalah butir nomor 13, 19, 21, 31, 32, dan 36. Para panelis baik pada metode *Extended* Angoff maupun metode Ebel menganggap butir nomor 13, 19, 21, 31, dan 32 sulit sementara secara statistik butir tersebut termasuk dalam tingkat kesulitan sedang. Begitu pula sebaliknya, para panelis baik pada metode *Extended* Angoff maupun metode Ebel menganggap butir nomor 36 mudah sementara secara statistik butir tersebut termasuk dalam butir yang sulit.

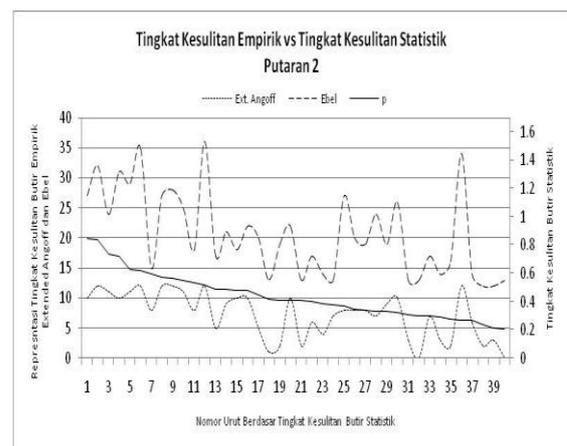
Jika kita bandingkan nilai korelasinya, terlihat bahwa metode *Extended* Angoff memiliki nilai korelasi lebih tinggi dibanding metode Ebel baik pada putaran 1 maupun putaran 2. Ini menunjukkan bahwa konsistensi *intra-judge* dalam masing-masing putaran pada metode *Extended* Angoff lebih baik dibanding metode Ebel. Atau dengan kata lain, panelis pada metode *Extended* Angoff lebih konsisten dalam melakukan penjurian di dalam setiap putaran dibanding panelis pada metode Ebel.

Kesimpulan di atas juga terlihat jelas pada Gambar 1 dan Gambar 2. Pada kedua gambar tersebut sumbu aksis menunjukkan nomor urut berdasar tingkat kesulitan butir secara statistik. Sumbu ordinat mayor menunjukkan representasi tingkat kesulitan butir secara empirik untuk metode *Extended* Angoff dan metode Ebel, sementara sumbu ordinat minor menunjukkan tingkat kesulitan butir secara statistik (p).

Pada kedua gambar, baik Gambar 1 maupun Gambar 2 terlihat bahwa grafik tingkat kesulitan butir secara statistik bergerak secara tenang (garis tidak putus-putus) sementara grafik representasi tingkat kesulitan butir secara empirik baik pada metode *Extended* Angoff maupun metode Ebel bergerak secara fluktuatif (garis putus-putus pendek dan panjang). Namun demikian, dari kedua gambar tersebut terlihat juga bahwa grafik nilai representasi tingkat kesulitan butir secara empirik pada metode Ebel selalu lebih fluktuatif dibanding metode *Extended* Angoff. Hal inilah yang menjelaskan mengapa nilai korelasi pada metode Ebel selalu lebih rendah dibanding nilai korelasi pada metode *Extended* Angoff. Fluktuasi nilai ini berbanding terbalik dengan tingkat konsistensi penjurian. Semakin fluktuatif berarti semakin tidak konsisten.



Gambar 1. Pemetaan Tingkat Kesulitan Butir Statistik dan Representasi Tingkat Kesulitan Butir Empirik Metode *Extended* Angoff dan Metode Ebel Putaran 1



Gambar 2. Pemetaan Tingkat Kesulitan Butir Statistik dan Representasi Tingkat Kesulitan Butir Empirik Metode *Extended* Angoff dan Metode Ebel Putaran 2

*Inter-judge consistency* dilihat dari simpangan baku *cutscore*. Tabel 13 memperlihatkan simpangan baku *cutscore* panelis pada metode *Extended* Angoff dan metode Ebel. Tabel 13 tersebut menunjukkan bahwa simpangan baku *cutscore* metode *Extended* Angoff selalu lebih kecil dibanding metode Ebel baik pada putaran 1, putaran 2, maupun gabungan putaran 1 dan 2. Hal itu mengindikasikan bahwa *inter-judge consistency* metode *Extended* Angoff selalu lebih baik dibanding metode Ebel. Dengan kata lain rating *cutscore* individual masing-masing panelis pada metode *Extended* Angoff selalu lebih konsisten dibanding rating *cutscore* individual masing-masing panelis pada metode Ebel baik pada putaran 1 maupun putaran 2.

Tabel 13. Simpangan Baku Penjurian Metode *Extended* Angoff dan Metode Ebel

Putaran	Simpangan Baku	
	Ext. Angoff	Ebel
Pertama	3,977	6,476
Kedua	4,243	5,740
Total	4,054	6,003

Jika kedua metode tersebut dibandingkan, manakah yang lebih disarankan atau direkomendasikan untuk digunakan? Penelitian ini menyimpulkan metode *Extended* Angoff lebih disarankan dibanding metode Ebel. Argumen yang mendasari rekomendasi itu adalah sebagai berikut:

Pertama, metode *Extended* Angoff memiliki prosedur yang relatif lebih mudah dibanding metode Ebel. Kemudahan ini di antaranya berkenaan dengan estimasi yang harus dilakukan oleh para panelis. Bagi panelis tentunya lebih mudah membuat estimasi untuk tiap butirnya dibanding membuat estimasi untuk tiap sel yang di dalamnya bersarang beberapa butir. Kemudahan lainnya adalah berkenaan dengan prosedur setelah estimasi. Pada metode *Extended* Angoff, untuk memperoleh *cutscore* panelis tinggal menjumlahkan banyaknya skor 1 yang dituliskan. Sementara pada metode Ebel, para panelis terlebih dahulu harus mengalikan banyaknya butir dan proporsi

yang tersarang pada tiap selnya, baru kemudian menjumlahkan hasil perkalian dari 12 sel yang ada tersebut.

Kedua, metode *Extended* Angoff membutuhkan waktu yang relatif lebih sedikit dibandingkan metode Ebel. Pada metode *Extended* Angoff para panelis hanya diminta membuat estimasi satu kali banyak butir soal. Estimasi itu hanya mengenai mampu atau tidaknya siswa dalam menjawab benar suatu butir. Sementara pada metode Ebel para panelis diminta setidaknya membuat estimasi dua kali jumlah butir soal. Dua jenis estimasi itu mengenai tingkat kesulitan dan tingkat relevansi butir. Di samping itu panelis juga masih harus menentukan estimasi proporsi butir terjawab benar untuk 12 sel yang ada. Selain itu tahapan sesudah estimasi pada metode Ebel membutuhkan waktu yang lebih banyak dibanding metode *Extended* Angoff.

Ketiga, tingkat akurasi nilai proporsi metode *Extended* Angoff cenderung lebih tinggi dibanding metode Ebel. Sebab pada metode *Extended* Angoff proporsi ditentukan untuk setiap butirnya sementara pada metode Ebel, nilai proporsi ditentukan untuk setiap selnya. Hal ini membawa implikasi nilai proporsi pada metode Ebel kemungkinan lebih bias dibanding nilai proporsi pada metode *Extended* Angoff.

#### Umpan Balik Panelis

Dengan menggunakan metode *Extended* Angoff diperoleh *cutscore* 59 sementara dengan menggunakan metode Ebel diperoleh *cutscore* 50,98 pada skala 100. Hasil ini menunjukkan bahwa batas KKM kelas VIII jenjang SMP/MTs untuk mata pelajaran Matematika di Kabupaten Sleman yang diperoleh dalam penelitian ini lebih rendah dibanding skor KKM yang biasanya ditetapkan oleh sekolah. Sebagaimana pernyataan para panelis, skor KKM mata pelajaran Matematika kelas VIII tahun ajaran 2011/2012 yang ditetapkan oleh sekolah sampel ada di kisaran nilai 70–80. Temuan ini sangat menarik karena perbedaan skor KKM yang dihasilkan cukup signifikan. Ada beberapa penyebab yang bisa diuraikan untuk menginterpretasi perbedaan tersebut.

Penyebab pertama adalah baik metode *Extended* Angoff maupun metode Ebel, keduanya menggunakan perangkat tes sebagai dasar penentuan *cutscore*. Hal ini membawa implikasi terhadap tinggi rendahnya nilai *cut-score* yang dihasilkan. Pada metode Ebel tingkat kesulitan dan tingkat relevansi butir akan berpengaruh langsung pada keputusan panelis dalam menuliskan jumlah butir yang tersarang dalam tiap-tiap sel. Sementara pada metode *Extended* Angoff tingkat kesulitan butir akan berpengaruh langsung terhadap keputusan Ya/Tidak panelis untuk setiap butirnya. Jika instrumen tes cenderung sulit maka nilai *cutscore* yang dihasilkan akan cenderung rendah. Mengacu pada rata-rata proporsi jawaban benar (rata-rata P) yang terletak di nilai 0,441 dapat disimpulkan bahwa tingkat kesulitan instrumen ini masih dalam kategori normal, namun cenderung ke sulit. Oleh karenanya menjadi logis dan rasional jika nilai *cutscore* yang dihasilkan dalam penelitian ini menjadi cenderung rendah.

Penyebab yang kedua adalah kemampuan dari peserta tes. Kemampuan peserta tes ini juga berimplikasi langsung terhadap nilai *cutscore* yang dihasilkan. Pada metode Ebel kemampuan peserta tes akan berpengaruh langsung pada keputusan panelis dalam menuliskan persentase butir yang bisa dijawab dengan benar oleh siswa yang tersarang dalam tiap-tiap sel. Sementara pada metode *Extended* Angoff kemampuan peserta tes akan berpengaruh langsung terhadap keputusan Ya/ Tidak panelis untuk setiap butirnya. Jika kemampuan peserta tes cenderung tinggi maka nilai *cutscore* yang dihasilkan pun akan cenderung tinggi. Mengacu kepada hasil statistik deskriptif skor peserta tes dapat disimpulkan bahwa kemampuan peserta tes cenderung rendah. Oleh karenanya menjadi beralasan jika *cutscore* yang dihasilkan pun menjadi cenderung rendah.

Sebab yang ketiga adalah posisi independen para panelis. Sebagaimana pernyataan para panelis, posisi mereka yang independen dalam penelitian ini memungkinkan mereka bisa melepaskan diri dari kepentingan di luar kepentingan nonakademis. Inde-

pendensi ini berimplikasi langsung terhadap ke-putusan-keputusan panelis dalam memberikan jawaban Ya/Tidak pada metode *Extended* Angoff maupun dalam pengisian jumlah butir dan persentase sel pada metode Ebel. Posisi independen panelis ini menjadi penting sebab sebagaimana diketahui, para guru di sekolah biasanya diminta menetapkan skor KKM tidak untuk kepentingan akademis semata namun lebih untuk kepentingan nonakademis. Sudah menjadi rahasia umum bahwa skor KKM yang ditetapkan guru di sekolah biasanya merupakan pesanan birokrasi pendidikan. Pesan itu berupa tuntutan skor KKM minimal 70 agar sekolah bisa dianggap telah melaksanakan Standar Nasional Pendidikan terutama Standar Penilaian. Jadi adanya patokan itu menyebabkan guru sering sekedar mematu-matu skor pada tiga aspek penentu KKM agar skor KKM lebih atau sama dengan 70. Itulah mengapa ketika guru diberikan kebebasan untuk memberikan penilaian maka muncul nilai *cutscore* yang lebih rendah dari KKM yang telah mereka tetapkan di sekolah.

Kualitas penyelenggaraan pertemuan *standard setting* diukur melalui survei dengan membagikan angket kepada para panelis setelah penentuan *cutscore* selesai. Angket itu digunakan untuk mendapat umpan balik peserta menyangkut 4 komponen utama yaitu fasilitas, penyelenggaraan, pemahaman materi, dan refleksi atas pertemuan *standard setting* yang telah dilakukan. Setiap panelis diminta memberikan tanggapan sikapnya atas setiap pernyataan yang ada di dalam angket.

Pada bagian pertama angket, panelis diminta memberikan tanggapannya terhadap tempat pertemuan, suasana pertemuan, dan fasilitas modul kegiatan. Berdasarkan rekap hasil angket, semua panelis baik pada metode *Extended* Angoff maupun pada metode Ebel menyatakan bahwa tempat dan suasana pertemuan *standard setting* cukup kondusif, nyaman, dan memadai. Selain itu, meskipun ada beberapa panelis yang merasa ragu namun mayoritas panelis menyatakan bahwa modul kegiatan yang diberikan mampu memberikan pemahaman yang memadai

tentang *standard setting* serta metode yang digunakan. *Feedback* ini berguna untuk memperbaiki kualitas modul yang digunakan jika dilakukan pertemuan sejenis pada waktu yang akan datang.

Tanggapan panelis memperlihatkan bahwa kegiatan Pemaparan Materi mampu memberi pemahaman yang memadai terhadap pengertian *standard setting* berikut metode *Extended Angoff* dan Ebel yang digunakan. Begitu pula halnya dengan kegiatan Pelatihan. Seluruh panelis baik pada metode *Extended Angoff* maupun metode Ebel menyatakan bahwa kegiatan Pelatihan mampu memberi pemahaman dan pengalaman yang memadai tentang *standard setting* berikut metode *Extended Angoff* dan metode Ebel yang digunakan. Tanggapan ini memberi informasi tentang pentingnya kegiatan Pemaparan Materi berikut Pelatihan dalam membangun pemahaman yang utuh dan pengalaman yang memadai tentang metode *standard setting* yang hendak digunakan. Kedua kegiatan ini merupakan bagian dari validitas prosedural yang dapat berimplikasi langsung terhadap validitas internal *cutscore* yang dihasilkan.

Tahap selanjutnya setelah dua kegiatan di atas adalah kegiatan Mengerjakan Soal. Kegiatan ini dimaksudkan untuk memberi gambaran yang memadai kepada panelis tentang tingkat kesulitan butir yang digunakan dalam penentuan *cutscore*. Tujuan ini ternyata sejalan dengan pernyataan para panelis yang merasa sangat terbantu dalam mendapatkan gambaran tentang tingkat kesulitan butir melalui kegiatan Mengerjakan Soal. Pada metode Ebel, kegiatan Mengerjakan Soal ternyata juga sangat membantu panelis dalam memperoleh gambaran yang memadai tentang tingkat relevansi soal.

Rekapitulasi tanggapan panelis juga menunjukkan bahwa instruksi-instruksi yang diberikan dalam penentuan *cutscore* cukup jelas dan mudah dipahami oleh para panelis baik pada metode *Extended Angoff* maupun metode Ebel. Hal ini kemungkinan disebabkan oleh pemahaman yang bagus dari para panelis tentang *standard setting* berikut metode yang digunakan.

Selain itu, dari sisi waktu, para panelis menyatakan bahwa mereka merasa cukup puas dengan alokasi waktu dalam kegiatan pertemuan *standard setting* ini. Sebagian besar panelis baik pada metode *Extended Angoff* maupun Ebel menyatakan mendapat alokasi waktu yang cukup memadai baik untuk kegiatan diskusi maupun untuk menyelesaikan tugas-tugas individual mereka.

Berdasarkan tanggapan panelis diketahui bahwa setelah mengikuti kegiatan Pemaparan Materi dan Pelatihan, para panelis memiliki pemahaman yang memadai tentang *standard setting* berikut tahapan-tahapan yang ada dalam metode *Extended Angoff* dan metode Ebel. Karena mereka memiliki pemahaman yang memadai maka mereka tidak merasa kesulitan mengikuti tahapan-tahapan yang ada dalam metode *Extended Angoff* dan metode Ebel.

Sebagai guru senior yang memiliki pengalaman mengajar matematika lebih dari 10 tahun, para panelis menyatakan memiliki pemahaman yang memadai tentang kemampuan akademis siswa-siswinya di sekolah. Selain itu mereka juga merasa cukup familiar dengan jenis-jenis soal berikut deskriptor masing-masing butir soal yang ada dalam instrumen tes yang digunakan dalam penentuan *standard setting*. Hampir seluruh panelis pada metode *Extended Angoff* menyatakan tidak mengalami kesulitan dalam menentukan keputusan-keputusan butir. Begitu pula para panelis pada metode Ebel, yang sebagian besar menyatakan tidak mengalami kesulitan dalam menentukan tingkat kesukaran maupun tingkat relevansi butir serta tidak mengalami kesulitan dalam menentukan proporsi jawaban butir pada setiap selnya. Pernyataan itu mengindikasikan bahwa para panelis memang benar-benar telah memiliki pemahaman yang memadai tentang *standard setting* berikut tahapan-tahapan yang ada dalam metode *Extended Angoff* dan metode Ebel. Oleh karena itu, sebagian besar panelis menyatakan memiliki kepercayaan yang tinggi terhadap keputusan-keputusan yang diambil yang digunakan untuk menentukan *cutscore* individual mereka masing-masing.

Bagian keempat dari angket ini adalah angket refleksi. Melalui angket refleksi ini hendak dijangkau pendapat panelis mengenai penentuan KKM dengan menggunakan metode konvensional dan penentuan KKM dengan menggunakan metode *standard setting*. Rekapitulasi terhadap hasil angket tersebut menginformasikan bahwa kegiatan penentuan *cutscore* dengan menggunakan metode *standard setting* ini merupakan pengalaman pertama mereka. Tak satu pun panelis yang menyatakan pernah melakukan hal serupa sebelumnya. *Standard setting* ini merupakan hal yang baru bagi mereka dan merupakan pengalaman pertama mereka, namun mereka tidak merasa kesulitan mengikuti dan menyelesaikan tahapan-tahapannya. Itulah mengapa sebagian besar mereka berpendapat bahwa guru-guru senior lain di luar kegiatan ini akan mampu menggunakan metode *Extended Angoff* maupun metode Ebel jika diberi pemahaman terlebih dahulu. Pernyataan ini mengindikasikan bahwa kedua metode tersebut cukup kompatibel bagi para guru untuk digunakan dalam penentuan KKM.

Selain itu, hasil rekapitulasi tersebut juga memberikan informasi bahwa semua guru pernah menentukan KKM dengan menggunakan metode konvensional di sekolah masing-masing. Sebagian besar dari mereka baik pada metode *Extended Angoff* maupun metode Ebel menyatakan mengalami kesulitan dalam menentukan skor untuk intake, daya dukung, dan kompleksitas. Itulah mengapa sebagian besar dari mereka setuju bahwa patokan keputusan yang digunakan metode *Extended Angoff* dan Ebel **lebih pasti** dibanding patokan skor yang digunakan pada metode konvensional.

Jika dilihat dari sisi penggunaan metode, para panelis menyatakan bahwa penggunaan metode *Extended Angoff* dan Ebel lebih memudahkan mereka dalam penentuan KKM dibanding saat menggunakan metode konvensional. Pada saat menggunakan metode konvensional para panelis menyatakan sering merasa bingung ketika diminta menentukan skor masing-masing komponen. Kebingungan itu kemungkinan bersumber dari dua hal. Pertama, dari belum adanya

acuan penilaian untuk masing-masing komponen yang digunakan. Kedua, bersumber dari terlalu banyaknya faktor yang harus dipertimbangkan ketika memberikan skor untuk masing-masing komponen tersebut. Oleh karena itu, mereka menganggap bahwa hasil KKM dengan menggunakan metode *Extended Angoff* dan metode Ebel lebih argumentatif dibanding KKM yang dihasilkan dengan menggunakan metode konvensional.

## Simpulan dan Saran

### Simpulan

Kesimpulan yang bisa diperoleh dari penelitian ini adalah sebagai berikut.

Pertama, analisis butir soal dengan menggunakan teori tes klasik menunjukkan bahwa **semua** butir dalam soal UKK jenjang SMP/MTs kelas VIII mata pelajaran Matematika Kabupaten Sleman tahun pelajaran 2011/2012 **bisa diterima** atau diikutkan dalam penentuan *cutscore* menggunakan metode *Extended Angoff* dan metode Ebel.

Kedua, *cutscore* yang diperoleh metode *Extended Angoff* adalah **59** sedangkan metode Ebel adalah **50,98** pada skala **100**. *Cutscore* ini berbeda secara signifikan dengan KKM sekolah yang ditentukan dengan menggunakan metode konvensional.

Ketiga, KKM metode *Extended Angoff*, menghasilkan persentase ketuntasan **19%** sementara KKM metode Ebel menghasilkan persentase ketuntasan **28%**. Kecilnya angka persentase ketuntasan ini menunjukkan bahwa para panelis memiliki estimasi yang terlalu tinggi terhadap kemampuan akademis siswa-siswinya.

Keempat, penelitian ini menunjukkan bahwa ada 38 deskriptor kemampuan Matematika siswa kelas VIII jenjang SMP/MTs di kabupaten Sleman yang telah dikatakan tuntas jika menggunakan skor KKM berdasarkan metode *Extended Angoff*.

Kelima, berdasarkan analisis terhadap *internal validity of standard setting* dapat disimpulkan bahwa metode *Extended Angoff* memberikan hasil yang relatif lebih valid dibanding metode Ebel.

Keenam, metode *Extended Angoff* dan metode Ebel cukup kompatibel bagi

para guru untuk digunakan dalam penentuan KKM karena guru mampu menggunakan kedua metode tersebut tanpa mengalami kesulitan dan kedua metode tersebut tidak membutuhkan dukungan analisis butir soal yang rumit.

Ketujuh, berdasarkan pendapat panelis, metode *Extended* Angoff dan metode Ebel **lebih mudah** digunakan untuk menentukan KKM dibandingkan dengan metode konvensional yang telah digunakan oleh para guru selama ini karena memiliki patokan estimasi yang **lebih pasti**.

#### Saran

Berdasarkan hasil penelitian dan mempertimbangkan beberapa keterbatasan dalam penelitian ini maka dapat disampaikan saran sebagai berikut.

Pertama, penggunaan cara konvensional dalam penentuan KKM perlu ditinjau dan dikaji ulang karena penentuan KKM dengan menggunakan metode *Extended* Angoff dan metode Ebel menghasilkan skor yang lebih argumentatif dan lebih bisa dipertanggungjawabkan.

Kedua, perlu dilakukan upaya untuk mendapatkan skor KKM tingkat kabupaten/kota menggunakan metode yang ada dalam *standard setting*. Skor KKM tingkat kabupaten/kota itu bisa digunakan sebagai acuan bagi sekolah ketika menentukan skor KKM sekolah. Penelitian ini menyimpulkan metode *Extended* Angoff lebih disarankan untuk digunakan dibandingkan dengan metode Ebel.

Ketiga, perlu dilakukan diseminasi di kalangan guru mengenai penentuan KKM dengan menggunakan metode yang ada dalam *standard setting* terutama metode *Extended* Angoff.

Keempat, perlu digunakannya KKM dalam pertimbangan penyusunan perencanaan, desain, dan strategi pembelajaran pada suatu jenjang pendidikan untuk membantu mencapai target kelulusan yang akan ditetapkan.

Kelima, perlu mulai dipikirkan digunakannya beberapa level *cutscore* sekaligus pada suatu kelas tertentu untuk membantu memetakan hasil pembelajaran.

#### Daftar Pustaka

- Alderson, J.C. (1993). *Judgements in language testing*. In D. Douglas C. Chapelle (Eds), *A new decade of language testing research* (pp.46-57). Alexandria, VA: TESOL
- Athanasou, J. A., Lamprianou, I. (2009). *A teacher's guide to educational assessment*. Rotterdam: Sense Publishers.
- Cizek, G. J. & Bunch, B. B. (2007). *Standard setting: a guide to establishing and evaluating performance standard on test*. ND: Sage Publications
- Cohen, A. S., Kane, M. T., & Crooks, T. J. (1999). *A generalized examinee-centered method for setting standard on achievement test* [Versi Elektronik]. *Applied Measurement in Education*, 12(4),
- Ebel, Robert L. (1979). *Essentials of educational measurement*. New Jersey: Prentice-Hall.
- Fowell, S.L., Fewtrell, R., & McLaughlin, P.J. (2008) *Estimating the minimum number of judges required for test-centered standard setting on written assessments. Do discussion and iteration have an influence?* *Advances in Health Sciences Education: Theory and Practice*, 13 (1): 11-24.
- Loomis, S.C., Hanick, P.L., Bay, L., & Crouse, J.D. (2000). *Developing achievement levels for the 1998 NAEP in civics interim report: field trials*. ACT, Inc.
- Norcini, J.J., Lipner, R.S., Langdon, L.O., & Strecker, C.A. (1987). A comparison of three variations on standard-setting method. *Journal of Educational Measurement*, 24 (1), 56-64.
- Reynolds, C. R., Livingston, R. B., Willson, V. (2010). *Measurement and assessment in education (2<sup>nd</sup> ed)*. New Jersey: Pearson Education, Inc.
- Stahl, John A. (2008). *Standard setting methodologies: Strengths and weaknesses.*, [http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502\\_Stahl.pdf](http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502_Stahl.pdf).
- Wells, C. S. (2007). *Makalah kuliah umum*. Yogyakarta.