

**Seminar Nasional Pendidikan MIPA  
Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Yogyakarta**

**Menentukan Batas Kelulusan (*Standard Setting*) pada Mata Pelajaran  
Matematika dengan Metode *Bookmark***

Heri Retnawati  
(Pendidikan Matematika FMIPA UNY)

**Yogyakarta, 15 Mei 2010**

# **Menentukan Batas Kelulusan (*Standard Setting*) pada Mata Pelajaran Matematika dengan Metode *Bookmark***

Heri Retnawati

(Pendidikan Matematika FMIPA UNY)

[retnawati\\_heriuny@yahoo.co.id](mailto:retnawati_heriuny@yahoo.co.id)

## **Abstrak**

Batas kelulusan (*Standard setting*) atau disebut juga kriteria ketuntasan minimal pada mata pelajaran matematika ditentukan berdasarkan kurikulum yang berbasis indikator yang dapat dicapai oleh siswa setelah proses pembelajaran, atau batas kelulusan yang ditetapkan pemerintah. Pada makalah ini dibahas metode menentukan batas lulus berbasis tes yang telah direspons oleh peserta tes, yaitu metode *Bookmark*. Pada metode ini, butir-butir tes ditentukan tingkat kesulitannya, kemudian butir-butir tersebut diurutkan berdasarkan tingkat kesulitannya yang selanjutnya menjadi nomor halaman. Pelaksanaan metode ini melibatkan guru matematika berpengalaman sebagai panelis yang menentukan pada halaman berapa peserta mulai tidak bisa mengerjakan, dan memerlukan suatu tes/perangkat ujian mata pelajaran matematika yang terstandar, dan instrumen sederhana untuk menuliskan hasil tiap panelis. Tahap pelaksanaan yaitu pelatihan, putaran 1, dan putaran 2. Rerata hasil putaran 1 dan 2 merupakan hasil penentuan batas kelulusan mata pelajaran matematika.

Kata kunci: Batas lulus (*standard setting*), Metode *Bookmark*

## **Pendahuluan**

Sesuai dengan kurikulum yang diberlakukan di Indonesia, 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 *judgement*. Penetapan dengan cara ini memiliki kelemahan, yaitu tidak berdasarkan data empirik dan prosedur yang telah teruji di lapangan. Untuk itu perlu dicari cara yang dapat dipertanggungjawabkan dari teori pengukuran.

Pemerintah berusaha meningkatkan kualitas pendidikan dengan menerbitkan Peraturan Pemerintah No 19 tahun 2005 tentang standar nasional pendidikan. Pada PP ini, pemerintah 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. Kompetensi lulusan adalah kemampuan minimum yang harus dicapai peserta didik.

SKL ini menjadi acuan dalam menyusun kisi-kisi ujian. Selanjutnya kisi-kisi ini digunakan sebagai acuan untuk menulis soal 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. Oleh karena itu dalam menyiapkan soal ujian, kisi-kisi ujian harus jelas, sehingga soal ujian dapat dikembangkan lebih baik.

Ujian Nasional (UN) yang dilaksanakan di Indonesia baik itu bernama Evaluasi Belajar Tahap Akhir (Ebtanas) maupun Ujian Akhir Nasional (UAN) merupakan salah satu proses pengukuran hasil belajar yang telah dilaksanakan secara nasional. Adapun tujuannya sebagai berikut : (1) untuk memperoleh informasi tentang mutu hasil pendidikan secara nasional, (2) mengukur pencapaian hasil belajar siswa baik sekolah/madrasah negeri maupun swasta, (3) memperoleh gambaran perbandingan mutu pendidikan pada sekolah madrasah, antar sekolah/madrasah, dan antar wilayah dari tahun ke tahun, (4) menjadi bahan penentuan kebijakan pembinaan sekolah/madrasah, (5) sebagai bahan pertimbangan dalam memberikan Surat Tanda Tamat Belajar dan seleksi masuk ke jenjang pendidikan yang lebih tinggi.

Meskipun pelaksanaan UN ada yang pro dan kontra, batas lulus (*cut of score*) UN ditentukan menggunakan kebijakan. Pada tahun 2005, batas lulus yang ditentukan 4,01. Untuk tahun 2006 dan 2007, batas lulus selanjutnya ditingkatkan menjadi 4,26. Batas lulus ini diberlakukan untuk tiga mata pelajaran, yakni bahasa Indonesia, matematika dan bahasa Inggris. Nilai 4,01 maupun 4,26 merupakan batas yang relatif rendah dibandingkan batas lulus negara-negara lainnya, namun demikian masyarakat meresponnya dengan penuh kecemasan dan keresahan, dan batas ini dianggap terlalu tinggi. Pada batas lulus untuk mata pelajaran matematika, bilangan-bilangan akan menjadi sangat sensitif. Hal ini terkait dengan mata pelajaran ini yang dianggap sulit, sehingga menjadi penyebab siswa tidak lulus di suatu jenjang sekolah.

Dalam pelaksanaan pendidikan, selalu ada harapan peserta didik yang menempuh UN tetap dapat lulus sesuai dengan tingkat kemampuannya dan memenuhi standar kompetensi lulusan yang telah ditentukan. Terkait dengan hal ini, batas lulus perlu ditentukan tidak hanya berdasarkan kebijakan (*judgement*) dari pemerintah semata, namun juga harus berdasarkan data empiris yang ada atau berdasarkan kemampuan siswa. Kemampuan siswa ini dapat diestimasi berdasarkan pola respons peserta didik yang diberikan terhadap UAN. Selanjutnya permasalahan yang timbul yakni “berapakah batas lulus ujian matematika berdasarkan tes matematika yang terstandar (UN misalnya)?” Untuk menjawab permasalahan ini, diperlukan penelitian untuk menentukan batas lulus (*cut of score*) Ujian nasional atau kriteria ketuntasan minimal. Pada makalah ini akan disajikan metode menentukan batas lulus dengan metode Angoff yang berbasis tes.

### **Pengertian Kriteria Kelulusan**

Definisi tentang *standard* telah banyak dikemukakan para pakar dan juga definisi menurut kamus. *Standard* dapat diartikan sebagai ukuran atau patokan yang disepakati. *Standard setting* adalah proses menentukan *cut score* 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).

Komponen esensial dari *standard setting* melalui *judgment* seperti yang dikemukakan oleh Angoff (1971), Ebel (1972), Jaeger (1982), and Nedelsky (1954) adalah panelis atau penilai ahli (Plake, Melican, & Mills, 1991). Jaeger (1991) 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; dan (8) mempunyai daya ingat semantik yang lebih kompleks.

*Standard setting* adalah proses yang digunakan untuk menentukan atau memilih suatu *passing score* pada suatu ujian. Dari semua langkah-langkah di dalam proses pengembangan tes, *standard setting* merupakan tahapan yang lebih dekat pada seni daripada sains (ilmu pengetahuan); sedang metode statistik yang sering digunakan di dalam pelaksanaan suatu *standard setting*, juga lebih banyak melalui pertimbangan dan atau kebijakan. Hattie & Brown (2003) menyatakan bahwa *setting performance standard* merupakan suatu proses meminta pertimbangan rasional dari para ahli yang (a) memiliki pengetahuan tentang kebutuhan akan tes dan asesmen yang ingin ditetapkan standarnya; (b) memahami makna skor pada level yang bervariasi pada skala yang digunakan untuk menyimpulkan performansi peserta tes; dan (c) memahami sepenuhnya batasan tentang prestasi yang berhubungan dengan standar performansi yang dimintakan kepada mereka untuk ditetapkan.

### **Metode *Standard Setting***

Terdapat lebih dari 30 metode menentukan *standard setting* yang berbeda dan telah diuraikan di dalam berbagai literatur pengukuran (Glass, 1978; Hambleton,

1980; Jaeger, 1979). Metode dalam *standard setting* dibedakan menjadi empat, *standard setting* berdasarkan pada materi, butir/tes yang digunakan, berdasarkan pada peserta tes (*examinee*) dan berdasarkan kebijakan (*judgement*).

a. Metode berpusat pada item/tes

Tes dianggap sebagai sekumpulan butir (*item pool*). Metode yang berpusat pada tes ini lebih cenderung menggunakan pendekatan klasik.

Metode berpusat pada tes misalnya :

- 1). metode Nedelsky (berdasarkan banyaknya pilihan (*option*) esensial yang mempunyai fungsi),
- 2). penilaian profesional (beberapa profesional diminta menilai, kemudian secara intuitif menetapkan peserta mana yang sudah dan belum menguasai wilayah *criteria*),
- 3). metode Angoff (beberapa *rater* menetapkan *cut of score* berdasarkan perkiraan peluang menjawab benar terhadap butir yang dikelompokkan berdasarkan tingkat kesulitan).
- 4). Metode Ebel (merupakan perbaikan dari metode Angoff, dengan mempertimbangkan tingkat kesukaran butir dan relevansi isi).

Untuk metode berpusat pada item, lebih cenderung menggunakan pendekatan teori respons butir, yakni :

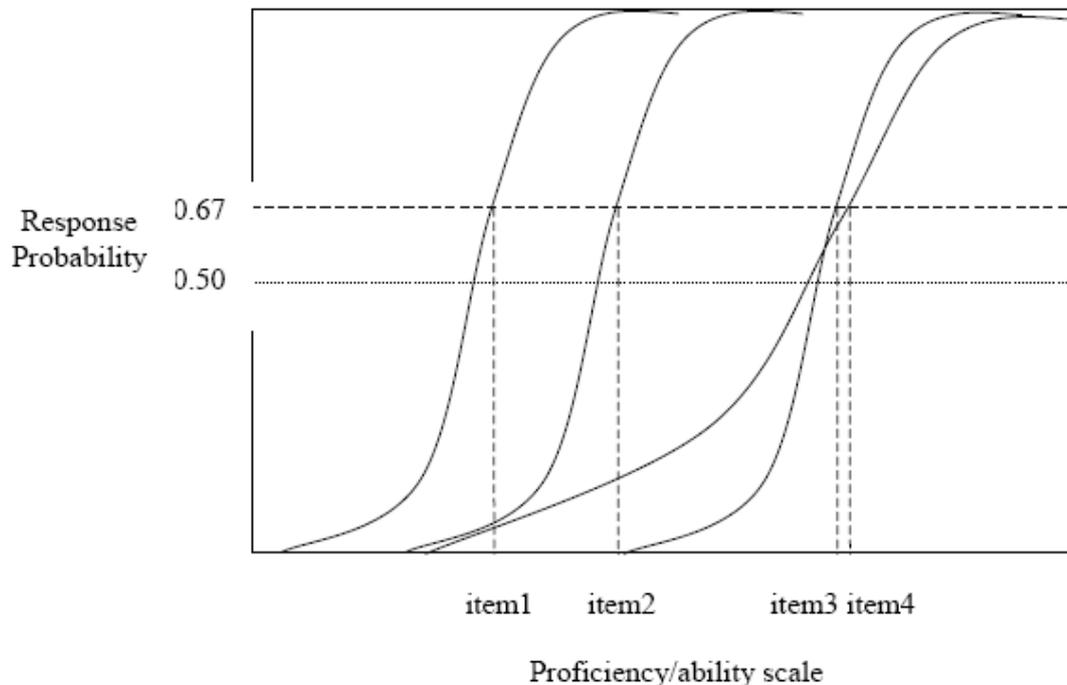
- 1). Metode *Bookmark* (perbaikan metode Angoff, dengan mempertimbangkan parameter butir)
- 2). metode pemetaan butir (*item mapping*) yang mempertimbangkan parameter-parameter butir hasil estimasi.

Pada makalah ini akan dibahas tentang metode *Bookmark* saja.

### Penentuan Standard Setting dengan Metode Bookmark

Prosedur bookmark didasarkan pada IRT (Lord, 1980) suatu framework simultan antara karakteristik kemampuan peserta dan tingkat kesulitan butir. 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 tersebut pada suatu lokasi dalam skala IRT sedemikian hingga siswa dengan skor skala dekat pada butir spesifik dapat disimpulkan memiliki pengetahuan ketrampilan dan kemampuan yang diperlukan untuk merespon secara sukses pada butir dengan probabilitas khusus.

*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 prosedur bookmark, probabilitas khusus kesuksesan diset 0,67, siswa dengan suatu skor skala pada titik potong akan memiliki probabilitas 0,67 dari jawaban butir pada titik potong dengan benar. Kegunaan 0,67 sebagai probabilitas respon (*response probability*, RP) didukung oleh penelitian Huynh tahun 1998 (Lin, tth). Huynh menyatakan bahwa model 3 PL, fungsi informasi butir termaksimumkan ketika  $\theta$  berada pada  $P(\theta) = (c + 2)/3$ . Sebagai akibatnya pada model 2 PL, ketika *guessing* (c) tidak ada, nilai  $RP = 2/3$ .

### **1. Material Standard Setting dengan Bookmark**

Material utama yang sering digunakan pada penentuan standar seting dengan bookmark yakni buku tes, jawaban peserta dan petunjuk penskoran, buku tes dengan butir terurut dan peta butir terhubungkan. Menggunakan parameter b, butir diurutkan dari yang mudah ke yang sulit dalam buku tes. Seperti diilustrasikan pada gambar 2 buku tes dengan butir terurut memiliki satu butir tiap halaman dengan halaman pertama berisi butir termudah dan yang terakhir butir tersulit. Pada titik skor bersama-sama, kedua jenis butir ditempatkan pada suatu buku tes dengan butir terurut sendiri-sendiri dan dibandingkan bersama-sama oleh panelis (Mitzel, dkk, 2001). Tujuan dari buku tes dengan butir terurut dinyatakan oleh Lewis dkk (1998), yakni untuk membantu partisipan menyusun suatu konsep terintegrasi dari apa yang diukur oleh tes sebaik menyajikan alat untuk membuat keputusan *cutscore*.

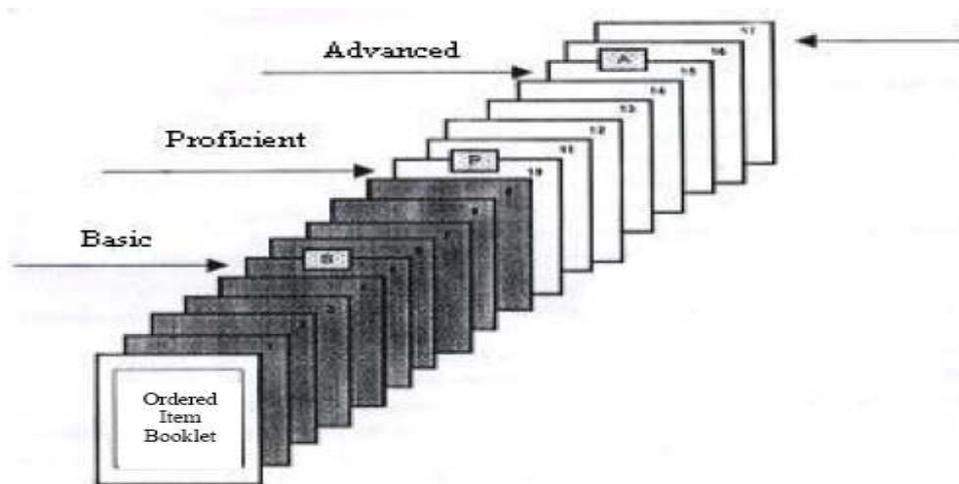
### **2. Mengeset *cutscore***

Mengeset dengan bookmark secara khusus melibatkan tiga ronde atau tiga iterasi. Setiap ronde difokuskan untuk meningkatkan konsensus dan mengurangi perbedaan diantara panelis.

#### ***Ronde/Putaran 1***

Tujuan utama dari ronde 1 yakni agar panelis kenal dengan buku tes dengan butir terurut, mengeset awal bookmark, dan mendiskusikan penempatan. Pada ronde ini panelis bekerja pada kelompok kecil, mendiskusikan setiap ukuran parameter butir

dan apa yang menyebabkan butir ini lebih sulit dari butir sebelumnya. Diskriptor kemampuan umum untuk level yang berbeda (misalnya : dasar, menengah, dan lanjut) juga dipresentasikan dan diskusikan. Panelis kemudian diminta untuk mendiskusikan dan menentukan isi yang seharusnya dikuasai siswa untuk suatu level kemampuan yang diberikan. Keputusan bebas mereka tentang *cutscore* dinyatakan dengan menyederhanakan tempat suatu *bookmark* antar item yang dianggap merepresentasikan suatu titik potong (*cut-point*) suatu *bookmark* ditempatkan pada setiap *cut-point* yang diperlukan. **Butir awal dari bookmark milik partisipan menyatakan isi bahwa semua siswa pada level kemampuan yang diberikan diharapkan untuk mampu dan mengetahui memperoleh kesuksesan dengan probabilitas 0,67.** Sebaliknya, siswa diharapkan memperoleh kesuksesan pada butir tersebut yang diberi tanda dengan probabilitas kurang dari 0,67.



Gambar 2 Ilustrasi buku tes dengan butir terurut pada prosedur Bookmark (Diadaptasi dari Mitzel, Lewis, Patz, & Green (2001), p. 263)

### Ronde/Putaran 2

Aktivitas awal di ronde 2 melibatkan kepemilikan *bookmark* yang ditempatkan oleh setiap anggota pada buku tes yang memuat butir terurut dimana

setiap panelis lain dalam grup kecilnya membuat penempatan bookmark mereka. Untuk suatu grup dengan 6 orang, setiap buku tes terurut milik panelis akan memiliki 6 *bookmark* untuk satu *cut-point*. Diskusi kemudian difokuskan pada butir antara bookmark yang pertama dan terakhir pada setiap level kemampuan. Sementara menyelesaikan diskusi mereka, panelis secara independen mengeset *bookmark* mereka. Median pada *bookmark* ronde 2 untuk setiap *cut-point* diperoleh sebagai rekomendasi grup untuk *cut-point* yang akan ditentukan.

#### *Finalisasi Standar Kemampuan*

Berdasarkan pada himpunan *cutscore*, diskriptor level kemampuan ditulis oleh panelis. Diskriptor kemampuan menggambarkan kemampuan ketrampilan dan pengetahuan khusus yang dimiliki siswa pada level kemampuan yang diberikan. Butir-butir utama pada *bookmark* merefleksikan isi dari level kemampuan siswa yang diharapkan menjawab paling tidak disekitar 0,67. Pengetahuan dan ketrampilan diperlukan untuk merespon dengan benar butir-butir tersebut kemudian disintesis untuk merumuskan diskriptor dari level kemampuan. Diskriptor level kemampuan tersebut menjadi suatu tambahan alami yang diperoleh dari penentuan *cutscore*.

#### **Pelaksanaan Penentuan *Standard Setting***

Pada pelaksanaan *standard setting* dengan metode Angoff, ada beberapa hal yang diperlukan yaitu:

1. Panelis, yaitu guru matematika yang ahli di bidang tersebut dan mempunyai pengalaman mengajar yang cukup lama, misalnya 10 tahun. Panelis yang diperlukan minimal sejumlah 11 orang yang mewakili sekolah dengan prestasi rendah, sedang, dan tinggi, juga mewakili keterwakilan daerah, yakni desa atau pinggiran, dan kota.
2. Tes matematika yang terstandar, misalnya perangkat UN mata pelajaran matematika.

3. Respons peserta tes terhadap tes matematika yang terstandar, yang digunakan untuk menentukan parameter tingkat kesulitan butir.
4. Booklet, yang berupa kumpulan dari butir-butir tes yang telah diurutkan berdasarkan tingkat kesulitannya. Contoh format halaman tiap booklet disajikan pada Gambar 3.

Matematika	Kemampuan yang diperlukan untuk mempunyai peluang menjawab benar (RP) 0.67		Tingkat Kesulitan	Halaman
Butir 03	-3.05	11.84	-3.851	1

Diketahui premis-premis seperti di bawah ini:

- I. Jika ada kerusakan mesin maka mobil tidak dapat bergerak.
- II. Mobil dapat bergerak.

Kesimpulan yang sah dari kedua premis di atas adalah ...

- A. Ada kerusakan mobil.
- B. Ada kerusakan pada mobil.
- C. Tidak ada kerusakan mesin pada mobil.
- D. Tidak ada kerusakan roda.
- E. Masih banyak bahan bakar.

**KUNCI: C**

5. Instrumen untuk menuliskan pendapat panelis pada butir yang mana siswa berhenti mengerjakan.
6. Instrumen yang dapat dibuat berupa tabel, misalnya seperti yang disajikan pada Tabel 1.

Penentuan *Standard Setting* dilakukan dengan 3 tahap, yaitu pelatihan kepada panelis, putaran 1 dan putaran 2. Tiap panelis mengisikan pada butir mana siswa berhenti menjawab benar pada *Booklet*. Rerata kemampuan siswa hasil panelis merupakan rerata kemampuan hasil dari putaran 1 dan putaran 2. Hasil akhir dari *standard setting* dengan metode *Bookmark* ditentukan dengan rerata hasil putaran 1 dan putaran 2.

Tabel 1. Contoh Instrumen untuk Menuliskan Hasil Panelis

**RANGKUMAN BOOKMARK PANELIS**

Petunjuk:

Isikanlah halaman yang menjadi batas siswa dapat menjawab benar dengan peluang 0,67 (bookmark) pada booklet butir terurut yang dipilih oleh panelis pada kolom yang tersedia

Panelis	Halaman Bookmark	Kemampuan @ Rp=0,67	Skala 100
P1	5	-2,046	24,420
P2	13	-1,052	36,845
P3	15	-0,954	38,070
P4	14	-1,047	36,908
P5	15	-0,954	38,070
P6	13	-1,052	36,845
P7	15	-0,954	38,070
P8	10	-1,167	35,408
P9		#N/A	#N/A
P10		#N/A	#N/A
P11		#N/A	#N/A
Rerata (Kriteria Kelulusan Minimal)		-1,154	35,580
Standar Deviasi		0,368	4,600

Persentase Kelulusan = 88,6%

Keterangan :

P : Panelis

## **Simpulan dan Rekomendasi**

Salah satu metode untuk menentukan batas kelulusan yaitu metode Bookmark, yang merupakan metode berbasis tes dan peserta tes. Metode ini dapat dilaksanakan pada penentuan batas lulus mata pelajaran matematika dengan melibatkan guru matematika berpengalaman, ahli psikometri dan menggunakan perangkat tes yang terstandarkan. Metode ini dapat dilaksanakan oleh kelompok guru mata pelajaran (MGMP) bekerjasama dengan perguruan tinggi atau Pusat Penilaian Pendidikan, yang hasilnya dapat digunakan sebagai pembanding kriteria kelulusan mata pelajaran matematika yang selama ini digunakan oleh guru di sekolah.

## **Referensi**

- 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. et al.(1997).A Step Model to Analyze Partial Credit. In Hambleton, K.R & vander Linden W.J. (eds, 1997), *Handbook of Modern Item Response Theory*, Springer.
- Hambleton, K.R. et al. (1991). *Fundamentals of Item Response Theory*, Sage Publications.
- Hattie, J.A., & Brown, G. T. L. (2003, August). Standard setting for asTTle reading: A comparison of methods. asTTle Technical Report #21, University of Auckland/Ministry of Education.
- Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 485–514). New York: American Council on Education/Macmillan.

Jaeger, R. M. (1991). Selection of judges for standard-setting. *Educational Measurement: Issues and Practice*, 10(2), 3-6, 10.

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.

Plake, B. S., Melican, G. J., & Mills, C. N. (1991). Factors influencing intrajudge consistency during standard-setting. *Educational measurement: Issues and Practice*, 10(2), 15-16, 22, 25.