



UNIVERSITAS NEGERI YOGYAKARTA  
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT  
PUSAT PENELITIAN KEBIJAKAN DAN SISTEM PENGUJIAN PENDIDIKAN

## *Sertifikat*

No. 1226/UN34.21/TU/2012

Diberikan kepada

*Dr. Samsul Hadi*

atas peran sertanya sebagai

**NARA SUMBER**

dalam acara "WORKSHOP PENENTUAN STANDARD SETTING BAGI GURU SMK KELOMPOK TEKNOLOGI DAN REKAYASA"  
yang diselenggarakan oleh Pusat Penelitian Kebijakan dan Sistem Pengujian Pendidikan LPPM UNY  
pada tanggal 29 Desember 2012

Yogyakarta, 29 Desember 2012

Mengetahui

Ketua LPPM UNY

Prof. Dr. Anik Ghufon

NIP. 19621111 198803 1 001

Ketua,

Prof. Dr. Sudji Munadi, M.Pd.

19530310 197803 1 003



**MENENTUKAN BATAS KELULUSAN (*STANDARD SETTING*)  
DENGAN METODE ANGOFF**

Makalah Disampaikan pada Workshop Penentuan *Standard Setting* Bagi Guru SMK  
Kelompok Teknologi dan Rekayasa

Oleh:

Samsul Hadi

Heri Retnowati

Sudji Munadi

---

**Pusat Penelitian Kebijakan dan Sistem Pengujian Pendidikan  
Lembaga Penelitian dan Pengabdian Kepada Masyarakat  
UNIVERSITAS NEGERI YOGYAKARTA**

**2012**

## **MENENTUKAN BATAS KELULUSAN (*STANDARD SETTING*) DENGAN METODE ANGOFF \*)**

Oleh:

Samsul Hadi, Heri Retnowati, dan Sudji Munadi

**Pusbijadikbangsisjian LPPM UNY**

### **Pendahuluan**

Penilaian pada kurikulum berbasis kompetensi menggunakan acuan kriteria. Asumsi yang digunakan pada acuan kriteria adalah setiap peserta didik dapat belajar pelajaran apa saja, hanya waktunya yang berbeda. Hasil penilaian 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 dipertanggung-jawabkan dari teori pengukuran.

Pemerintah telah menerbitkan Peraturan Pemerintah No 19 tahun 2005 tentang Standar Nasional Pendidikan dalam rangka meningkatkan kualitas pendidikan. PP tersebut mencakup 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 akhir jenjang pendidikan. Kisi-kisi ini digunakan sebagai acuan untuk menulis soal ujian. Siapa saja yang menyusun soal apabila menggunakan kisi-kisi yang sama akan menghasil-

---

\*) Materi disampaikan pada Workshop Penentuan *Standard Setting* Bagi Guru Smk Kelompok Teknologi dan Rekayasa pada tanggal 29 Desember 2012 di LPPM UNY

kan soal yang relatif sama. Soal 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 professional (beberapa professional 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 Angoff saja.

### **Metode Angoff Tradisional (1971)**

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) adalah prosedurnya sederhana dan mudah dilaksanakan, karenanya reliabilitas asesmen intra dan inter *judgement* menjadi tinggi. Metoda ini menggunakan statistik sederhana yang mudah dihitung dan dipahami. Metode Angoff dapat dilaksanakan sebagai berikut (bute-induction.st-andrews.ac.uk/, 2009):

- a. Panelis melihat butir soal pertama dan menilai tingkat kesulitannya.

- b. Setiap panelis secara individu mengestimasi persentase sekelompok peserta tes yang dapat menjawab butir soal dengan benar.
- c. Panelis mendiskusikan hasil estimasi mereka.
- d. Setiap hasil estimasi ditabulasikan dan dihitung rata-ratanya.
- e. Urutan di atas diulang untuk semua butir soal.
- f. Rata-rata hasil estimasi setiap butir dijumlah dan dirata-ratakan kembali untuk memperoleh *cutpoint*.

### **Metode Extended Angoff (1997)**

Untuk menyederhanakan proses pada metode Angoff, Impara dan Plake (1997) mengusulkan agar panelis memutuskan apakah peserta ujian secara individual mampu atau tidak mampu menjawab butir soal dengan benar (Stahl, 2008). Cara ini dinamakan prosedur YA/TIDAK. Jawaban diskor 1 bila ya dan diskor 0 jika tidak. Skor 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.

Rasional dari modifikasi ini adalah akan lebih mudah bagi panelis memperkirakan kemampuan satu peserta ujian dari pada kemampuan sekumpulan peserta ujian, selain itu menjawab YA/TIDAK lebih mudah dilakukan daripada menghitung persentase (Ricker, 2009).

### **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. Kurikulum, yang memuat Standar Kompetensi dan Kompetensi Dasar untuk membuat deskripsi tentang hal-hal/indikator yang diukur oleh suatu butir.
4. Instrumen untuk menuliskan pendapat panelis tentang suatu butir. 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. Pada tiap sel, panelis mengisikan peluang menjawab benar tiap butir yang dapat dicapai siswa (pada metode Angoff) atau menuliskan butir yang dapat dijawab dengan benar (YA, pada metode *Extended* dari Angoff). Rerata dari jumlah peluang menjawab benar setiap butir (pada metode Angoff) atau rerata banyaknya butir yang dapat dijawab peserta (pada metode *Extended* dari Angoff) merupakan hasil dari putaran 1 dan putaran 2. Hasil akhir dari *standard setting* dengan metode Angoff ditentukan dengan rerata hasil putaran 1 dan putaran 2.

Tabel 1. Contoh Instrumen untuk Menuliskan Hasil Panelis

No. Butir	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Deskriptor
1.												
2.												
3.												
.												
.												
.												
n.												

Keterangan :

P : Panelis

## **Simpulan dan Rekomendasi**

Salah satu metode untuk menentukan batas kelulusan yaitu metode Angoff, yang merupakan metode berbasis tes. Metode ini dapat dilaksanakan pada penentuan batas lulus mata pelajaran matematika dengan melibatkan guru matematika berpengalaman, dan menggunakan perangkat tes yang terstandarkan, dan hanya menggunakan instrumen yang sederhana. Metode ini dapat dilaksanakan oleh kelompok guru mata pelajaran (MGMP), 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.
- Glas, 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.
- Goodwin, L. D. (1999). Relations between observed item difficulty levels and Angoff minimum passing levels for a group of borderline examinees. *Applied Measurement in Education*, 12, 13-28.
- 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.

- Impara James C. & Plake Barbara S. 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
- Impara, J.C., & Plake, B.S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement*, 34, 353–366.
- Impara, J.C., Barbara S Plake. 1998. Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. University of Nebraska-Lincoln. *Journal of Educational Measurement Spring*, 1998. Vol.35, No. 1
- Impara, J.C., Barbara S Plake. 2000. A Comparison of Cut Scores using Multiple Standard Setting Methods, Universitas Nebraska- Lincoln, Paper presented at the Large Scale Assessment Conference. Snowbird, UT, June, 2000.
- 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., 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.
- 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.

- 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.
- Stahl, J. A. 2008. Standard Setting Methodologies: Strengths and Weaknesses. Illinois: Pearson VUE. Diambil pada tanggal 28 September 2009 dari [www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502\\_Stahl.pdf](http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502_Stahl.pdf)