

**PENENTUAN STANDARD SETTING MATA
PELAJARAN KIMIA DENGAN METODE ANGOFF,
IRT (ITEM RESPONSE THEORY), DAN SPLINES
CUBIC HERMITE FUNCTIC N**

Suwahono

Mahasiswa Pascasarjana S3 PEP Universitas Negeri Yogyakarta

Abstrak

Batas kelulusan (Standard setting) atau disebut juga kriteria ketuntasan minimal pada mata pelajaran kimia ditentukan berdasarkan kurikulum yang berbasis indicator yang dapat dicapai oleh siswa setelah proses pembelajaran, atau batas kelulusan yang ditetapkan pemerintah.

Pada penelitian ini dibahas metode menentukan batas lulus berbasis tes, yaitu: *Angoff*, *item response theory* (IRT), dan *splines cubic Hermite function*. Pada metode diatas, butir-butir tes ditentukan tingkat kesulitannya, kemudian butir-butir tersebut diurutkan berdasarkan tingkat kesulitannya yang selanjutnya menjadi nomor halaman. Pelaksanaan metode ini melibatkan guru kimia berpengalaman sebagai panelis yang menentukan pada halaman berapa peserta mulai tidak bisa mengerjakan, dan memerlukan suatu tes/perangkat ujian mata pelajaran kimia 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 kimia

PENDAHULUAN

Berdasarkan kurikulum berbasis kompetensi (KBK) yang dilanjutkan kurikulum tingkat satuan pendidikan (KTSP), 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.

Untuk melakukan peningkatan pendidikan, pemerintah 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.

Standar kompetensi lulusan digunakan sebagai acuan dalam menyusun kisi-kisi ujian. Selanjutnya kisi-kisi ini digunakan sebagai acuan untuk menulis soal ujian. Apabila menggunakan kisi-kisi yang sama, siapa saja yang menyusun soal 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 Nasional (UN) merupakan salah satu proses

pengukuran hasil belajar yang telah dilaksanakan secara nasional. Tujuan UN 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 perkembangan 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.

UN terus dilaksanakan setiap tahun, meskipun pelaksanaannya ada yang pro dan kontra, batas lulus (*cut of score*) UN ditentukan menggunakan kebijakan. Pada tahun 2005/2006, batas lulus yang ditentukan 4,01. Setiap tahun batas lulus ditingkatkan pada tahun 2009/2010 dan 2010/2011 batas lulus berturut-turut dinaikkan menjadi 4,25 dan 4,26. 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 Kimia, bilangan-bilangan 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 ter-

Pelaksanaan Penentuan *Standard Setting*

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

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

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.

Langkah-langkah Penentuan Cut Score

1. **Langkah-langkah Metode Angoff untuk menentukan cut score**
 - a. Instrumen diberikan kepada beberapa ahli sebagai rater
 - b. Kemudian rater memperkirakan proporsi sampel minimal calon yang menjawab setiap item dalam tes dengan benar.

- c. Kemudian ditabulasi jumlah cek pada tabel kolom (Mudah dan Sulit),
 - d. Mereka memberi cek pada item yang sulit
 - e. Jika pada sebuah item tidak ada cek dianggap kategori item mudah, jika ada kategori sulit
 - f. Menghitung perbandingan banyak item sulit dan banyak item semuanya (item mudah dan sulit) nilai ini merupakan cut off
- 2. Langkah-langkah Metode IRT untuk menentukan cut score:**
- a. Instrumen di diberikan pada peserta (examinee)
 - b. Hasil kerja peserta dianalisis dengan metode IRT
 - c. Indeks kesulitan item yang akhir kalibrasi digunakan untuk mengidentifikasi item yang sulit dan mudah.
 - d. Item mudah memiliki indeks kesulitan negatif,
 - e. Item sulit memiliki indeks kesulitan bukan negatif (dapat nol atau positif)
 - f. Menghitung perbandingan banyak item sulit dan banyak item semuanya (item mudah dan sulit) nilai ini merupakan cut off
- 3. Langkah-langkah Metode Splines untuk menentukan cut score**
- a. Mengkonversi skor awal (x) menjadi skor akhir (y) dengan fungsi splines Hermit kubik
 - b. Splines menghasilkan hasil titik-titik yang melewati semua poin (x_i, y_i) , mengecek perubahan kemiringan garis singgung, di mana $i=1, \dots, n$, n adalah jumlah poin yang dimasukkan ke dalam prosedur ($n=6$) dengan langkah:
 - 1) Lereng garis menghubungkan dua poin berturut-turut dihitung sebagai:

$$\Delta_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \quad \text{untuk } k=1, 2, \dots, n-1$$

- 2) m garis singgung awal pada setiap titik data adalah:

$$m_k = \frac{\Delta_{k-1} + \Delta_k}{2} \text{ untuk } k=2, \dots, n-1;$$

$$m_1 = \Delta_1 \text{ dan } m_n = \Delta_{n-1}$$

- 3) Ketika $y_k = y_{k+1}$ untuk $k=1, 2, \dots, n-1$, maka $x_k = x_{k+1}$ diatur ke nol untuk membuat spline datar dalam rangka melestarikan kemonotonan dari fungsi. Langkah 4) dan 5) untuk k mereka akan dilewati.
- 4) Masukkan $\alpha_k = \frac{m_k}{\Delta_k}$ dan $\beta_k = \frac{m_{k+1}}{\Delta_k}$, jika α atau β adalah nol (Yaitu input data titik-titik tersebut tidak monoton), kemudian m_k dan m_{k+1} ditetapkan menjadi nol untuk memastikan bahwa kurva monoton masih bisa dihasilkan.
- 5) Dalam rangka mencapai kemonotonan dari *Spline*

$$\alpha_k^2 + \beta_k^2 > 9$$

kemudian mengatur

$$m_k = \tau_k \alpha_k \Delta_k, m_{k+1} = \tau_k \beta_k \Delta_k \text{ dan}$$

$$\tau_k = \frac{3}{\sqrt{\alpha_k^2 + \beta_k^2}}$$

Hasil

1. Contoh Hasil *Cut Score* dengan *Angoff*

Berdasarkan hasil judgment rater dinyatakan dalam Tabel 1.

Tabel 1. Hasil judgment rater item instrumen

Komponen Tes Kimia	Mudah	Sulit
Stokiometri	6	6
Ikatan Kimia	6	4
Kimia Hidrokarbon	5	7
Sistem periodik unsur	4	2

Selanjutnya dihitung cut scorenya seperti dinyatakan Tabel 2.

Tabel 2. Cut score berdasarkan judgment rater

Komponen Tes Kimia	Mudah	Sulit	Cut Score
Stokiometri	6	6	6/12
Ikatan Kimia	6	4	4/10
Kimia Hidrokarbon	5	7	7/12
Sistem periodik unsur	4	2	2/6
Total	21	19	19/40

2. Contoh Hasil Cut Score dengan IRT

Dari hasil analisis dengan IRT, misalnya BILOG, dihasikan item sulit dan mudah sebagai berikut:

Tabel 3. Hasil analisis dengan IRT item sulit dan mudah

Komponen Tes Kimia	Mudah	Sulit
Stokiometri	6	6
Ikatan Kimia	6	4
Kimia Hidrokarbon	5	7
Sistem periodik unsur	3	3

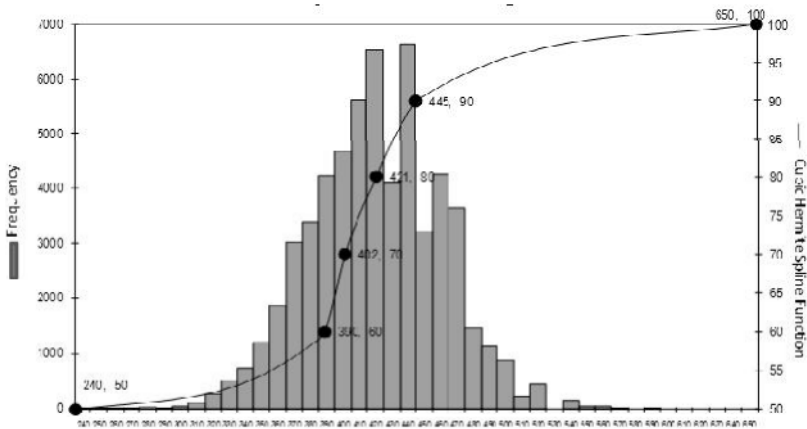
Selanjutnya dihitung cut score dinyatakan yang dinyatakan pada Tabel 4.

Tabel 4. Cut score berdasarkan analisis IRT

Komponen Tes Kimia	Mudah	Sulit	Cut Score
Stokiometri	6	6	6/12
Ikatan Kimia	6	4	4/10
Kimia Hidrokarbon	5	7	7/12
Sistem periodik unsur	3	3	3/6
Total	20	20	20/40

3. Contoh Hasil Cut score dengan Splines

Dengan skor awal skala yang ada (240-650) ke skor akhir dengan skala baru (50-100) dengan menggunakan fungsi kubik spline Hermite untuk mengembangkan konversi. Hasilnya diperoleh grafik seperti Gambar dengan cut core, yakni: F (240,50), D (390,60), C (402,70), B (421,80), dan A (445, 90).



Gambar 2. Splines Fungsi kubik Hermite

PEMBAHASAN

Cut Score yang ditentukan dengan menggunakan metode Angoff dan metode IRT hampir sama hasilnya, bahkan dapat dikatakan tidak berbeda. Hal ini seperti dinyatakan bahwa tidak ada perbedaan yang signifikan ditampilkan dalam analisis chi-kuadrat antara skor cut off Angoff dan skor cut off IRT (Nielie L. Tiratira,). Untuk mendapatkan cut score metode Angoff lebih sederhana karena tidak perlu software sehingga segera dapat diketahui, sementara dengan IRT diperlukan software dan yang jelas setelah ada hasil dari responden. Karena dengan hasil yang diperoleh akhirnya sama, maka metode Angoff lebih efektif dari pada metode IRT.

Tampaknya penting bahwa para rater ahli dalam konstruk tes dan memahami peserta ujian. Semakin yakin mereka tentang keahlian mereka, semakin kecil kemungkinan mereka akan terlalu terpengaruh oleh data empiris. Busch dan Jaeger (1990) menyarankan bahwa data dampak penyajian ini diperlukan agar tes standar yang wajar dalam kaitannya dengan perilaku peserta tes, meskipun fakta bahwa menetapkan standar kinerja dan nilai koresponden dipotong menyiratkan bahwa menilai tidak memihak

dan terpisah dari data. Namun, karena menilai tidak pernah sempurna, untuk itu diperlukan tes standar dan skor cut off untuk bersikap realistis agar mereka dapat diterima oleh stakeholder.

Ada beberapa kelemahan menggunakan estimasi parameter untuk pengaturan IRT skor dipotong (cut score).

1. menghitung estimasi parameter untuk model IRT adalah suatu proses rumit dan membutuhkan sebuah program komputer khusus.
2. ukuran sampel yang besar (minimal 500-1500 peserta ujian, tergantung pada model IRT yang digunakan) diperlukan untuk menentukan estimasi parameter yang stabil untuk item.

Penentuan Cut score dengan splines memiliki beberapa keunggulan, antara lain:

1. transformasi hampir linear untuk kira-kira setengah siswa di pusat daerah masing-masing distribusi (mudah dicatat dalam grafik untuk semua kelas konten- kombinasi).
2. perbedaan harus dibuat paling hati-hati (Lebar bersyarat standar-kesalahan band terkait dengan nilai ekstrim dengan demikian akan agak berkurang, yang mungkin membuat penggunaan mereka lebih menarik)
3. Skore mudah diterjemahkan ke dalam nilai huruf (tingkat prestasi) yang memiliki makna dalam hal keputusan lulus atau tingkat yang sebanding kinerja.

PENUTUP

1. Ada tiga metode untuk menentukan cut score kelulusan yaitu metode Angoff, IRT, dan Splines yang merupakan metode berbasis tes. Metode ini dapat dilaksanakan pada penentuan batas lulus mata pelajaran Kimia dengan melibatkan ahli atau praktisi kimia berpengalaman, dan menggunakan perangkat tes yang standar.
2. Penggunaan metode Angoff dalam pengaturan cut score memerlukan pertimbangan rater pada tingkat

- item, apakah ada kemungkinan bahwa siswa dapat menjawab dengan benar.
3. Penggunaan metode Angoff dalam pengaturan cut score proses lebih sederhana dan cepat
 4. Penggunaan metode IRT memerlukan software dan banyak perhitungan untuk sampai pada cut score.
 5. Metode IRT dapat untuk menegaskan metode Angoff karena frekuensi tanggapan akan memberitahu jika item yang benar-benar sulit karena hanya sedikit siswa dapat menjawab item tertentu.
 6. Metode splines dalam menentukan cut score moderat dan transparan
 7. Metode splines menentukan beberapa cut score F, D, C, B, dan A sebagai mana nilai di perguruan tinggi.
 8. Metode splines memerlukan fungsi yang rumit dan software untuk menentukan cut score

DAFTAR PUSTAKA

- Angoff, W. H. (1971). Scale, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, I C: American Council on Education.
- Ebel, Robert L. (1972). *Essentials of educational measurement*. Englewood Cliffs: Prentice- Hall.
- Gary Skaggs, Serge F. Hein, and Risper Awuor. 2007. *Setting Passing Scores on Passage-Based Tests: A Comparison of Traditional and Single-Passage Bookmark Methods*. *Journal of APPLIED MEASUREMENT IN EDUCATION*, 20(4), 405-426
Copyright © 2007, Lawrence Erlbaum Associates, Inc.
- 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 23 Desember 2010.

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*. Ed-

ucational Measurement: Issues and Practice, 10(2), 3-6, 10.

Kurt F. Geisinger and Carina M. McCormick . 2010. Adopting Cut Scores: Post-Standard-Setting Panel Considerations for Decision Makers. *Educational Measurement: Issues and Practice Spring 2010, Vol. 29, No. 1, pp. 38-44*

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.

Niclie L. Tiratira. 2009. *CutOff Scores: The Basic Angoff Method And the Item Response Theory Method* .The International Journal of Educational and Psychological Assessment April 2009, Vol. 1, Issue 1, pp. 27-35

Paul Nichols, Pearson, Jon Twing, Pearson, Canda D. Mueller.2010. Standard-Setting Methods as Measurement Processes. *Educational Measurement: Issues and Practice Spring 2010, Vol. 29, No. 1, pp. 14-24*

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, Cana-

da): Centre for Research in Applied Measurement and Evaluation University of Alberta.

Robert G. MacCann and Gordon Stanley. 2006. *The Use of Rasch Modeling To Improve Standard Setting* . Practical Assessment Research & Evaluation Volume 11 Number 2, January 2006 ISSN 1531-7714

Stahl, J. A. 2008. Standard Setting Methodologies: Strengths and Weaknesses. Illinois: Pearson VUE. Diambil pada tanggal 30 Desember 2010 dari www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502_Stahl.pdf

Schafer and Xiaodong Hou. 2011. *Test Score Reporting Referenced to Doubly-Moderated Cut Scores Using Splines*. Practical Assessment Research & Evaluation Volume 16, Number 13, October 2011 ISSN 1531-7714