

KOMPARASI *STANDARD SETTING* METODE *GROUP CONTRAST* DAN *BOOKMARK* PADA MATA PELAJARAN AKUNTANSI

Natalina Premastuti B.
Pendidikan Akuntansi FKIP USD
natalina.premastuti@yahoo.com

Abstrak

Penelitian ini bertujuan untuk menentukan *cut score* (batas lulus) ujian nasional mata pelajaran akuntansi dengan menggunakan metode *Group Contrast* dan *Bookmark*. Penelitian ini merupakan penelitian deskriptif kuantitatif. Data berupa perangkat UN kompetensi kejuruan teori akuntansi tahun 2009 dan respon peserta dari dokumen sebanyak 8 SMK Bisnis dan Manajemen di wilayah Kota Yogyakarta sebanyak 314 peserta didik. Analisis data butir tes dilakukan secara klasik dengan program microcat Iteman, sedangkan analisis secara modern (Rasch Model) menggunakan program AcerConQuest. Judgment ahli digunakan dalam penentuan *cutscore*. *Cut score* yang dihasilkan dengan metode *Group Contrast* adalah sebesar 63,09 (skala 100) dan dengan metode *Bookmark* adalah sebesar 68,96 (Respon Probability/RP 0,67) dan 67,36 (RP 0,5). Metode *Bookmark* relatif lebih akurat dalam menentukan *cut score* karena butir diungkap parameter indeks kesukarannya kemudian dilakukan analisis kemampuan peserta tes dengan menggunakan model Rasch atau Teori Respon Butir, dan adanya judgment ahli.

Kata kunci: *batas lulus, metode group contrast, metode bookmark*

STANDARD SETTING COMPARATION USING CONTRAST GROUP AND BOOKMARK METHODS ON ACCOUNTING SUBJECT

Natalina Premastuti B.

Pendidikan Akuntansi FKIP USD
natalina.premastuti@yahoo.com

Abstract

The objective of the research is to determine national examination cut score on accounting subject using contrast group and bookmark methods. The research data was in the form a set of accounting theory for national examination for vocational competence in 2009 and examinee's responses toward examination set of 314 students of business management vocational schools in Yogyakarta. Data analyzed was quantitative qualitatively. The methods were contrast group and bookmark. Item test data was analyzed classically by microcat program while the modern analysis (Rasch Model) was done by Acer Conquest Program. The qualitative data is the expert judgment about the cut score determination process. The research result shows that the cut score resulted from contrast group method is 63.09 on a 1-100 point scale. On the other hand, the cut score resulted from bookmark method is 68.96 with an RP of 0.67 and 67.36 with an RP 0.5. Bookmark method is relatively more accurate in cut score determination because the items difficulty index parameter was analyzed then the examinee's ability was analyzed using Rasch model or butir response theory, and it is supported with expert judgment.

Key words: cut score, group contrast method, bookmark method

Pendahuluan

Kualitas lulusan SMK dibentuk dengan kualitas pembelajaran yang baik. Sementara untuk mengetahui seberapa baik kualitas pembelajaran diperlukan informasi yang diperoleh melalui kegiatan pengukuran. Pemerintah menyelenggarakan ujian nasional dengan berlandaskan pada Peraturan Pemerintah No 19 Tahun 2005 tentang Standar Nasional Pendidikan khususnya bab X tentang Standar Penilaian Pendidikan.

Kompetensi minimal yang harus dikuasai oleh peserta didik identik dengan *cut score* yang telah ditetapkan. Peserta didik yang telah melampaui *cut score* dapat dimaknai peserta didik tersebut telah lulus kompetensi minimal. Harapannya tentu peserta didik dapat lulus 100%, sehingga sekecil apapun persentase ketidaklulusan, perlu ditindaklanjuti dengan berbagai bentuk perbaikan. Ujian teori akuntansi merupakan mata uji prasyarat bagi materi yang lain dimana *cut score* (batas lulus) yang ditetapkan pemerintah adalah 4,00 (skala 10) pada ujian nasional kompetensi kejuruan tahun 2008/2009.

Crocer dan Algina (1986: 410) menyatakan bahwa kegiatan untuk mensetting *cut score* pada umumnya disebut sebagai *standard setting*. Terdapat berbagai model *standard setting* namun kebanyakan merujuk pada *expert judgment*. Secara garis besar pendekatan *standard setting* mengerucut pada dua pendekatan besar, yakni pendekatan yang berdasarkan pada peserta tes dan pendekatan yang berdasarkan pada tes. Pendekatan yang berdasar pada tes melahirkan model Angoff, Yes No, Bookmark, Nedelsky, Ebel, dll. Sementara model yang berdasar pada peserta tes antara lain *contrasing group* dan *borderline*.

Berbagai penelitian tentang *standard setting* telah banyak dilakukan oleh pemerhati di dunia pendidikan maupun di bidang lain. Heri Retnowati (2008: 52) melakukan penelitian untuk menentukan *cut score* ujian nasional pada mata pelajaran matematika. Ternyata secara empirik, *cut score* ujian nasional mata pelajaran matematika adalah sebesar 71, 6 (dengan metode grup kontras) dan sebesar 72, 25 (dengan metode Ebel) semuanya pada skala 100. Sementara *cut score* yang ditentukan oleh pemerintah adalah sebesar 4,26 (untuk tahun 2006 dan 2007). Artinya, terjadi perbedaan *cut*

score menurut *judgment* pemerintah dan berdasarkan kemampuan riil peserta didik.

Berdasarkan latar belakang yang telah diungkap, maka permasalahan penelitian akan dirumuskan sebagai berikut: berapakah batas lulus (*cut score*) ujian nasional kompetensi kejuruan teori akuntansi dengan menggunakan metode *Group Contrast* dan metode *Bookmark*?

Terminologi "*standard*" dapat dimaknai sebagai skor yang menunjukkan pengetahuan dan *skill* yang relevan. Dalam dunia pendidikan, jumlah *standard* kinerja memiliki variasi dari satu (untuk membedakan lulus dan tidak lulus) sampai empat atau lima kategori. Ahli lain mengungkap *standard* sebagai jawaban dari "Seberapa baik yang disebut baik?" (Livingstone and Zieki, 1982: 10). Banyak istilah yang digunakan dalam *literature* pengukuran dan pengujian, untuk menunjuk pada istilah standar kinerja. Istilah-istilah tersebut adalah (Hamelton & Pitoniak, 2006: 435): *passing scores, cut scores, cutoff scores, performance scores, achievement levels, mastery levels, proviciency levels, thresholds levels, dan standard*. Dalam kasus yang sederhana, terdapat satu *cut score* dengan dua kategori, yaitu kategori lulus dan gagal.

Mehrens & Lehmann (1991) dalam (Cizek 2001: 6) menyatakan bahwa pengambilan keputusan sangat berhubungan dengan prosedur-prosedur pengukuran dan praktik dalam dunia pendidikan:

Decision making is daily task. Many people make hundreds of decisions daily; and to make wise decisions, one needs information. The role of measurement is to provide decision makers with accurate and relevant information....The most principle of the text is that measurement and evaluation are essential to sound educational decision making.

Pendekatan acuan norma dapat pula disebut sebagai pendekatan *relatif*, dan diperkenalkan oleh Nedelsky yang menggambarkan prestasi peserta didik yang memadai "*relatif* untuk kelasnya atau kelompok peserta didik lain yang khusus". Pada metode ini, *standard* berdasarkan pada prestasi peserta tes. Standar *relatif* dapat didasarkan pada rata-rata dari peserta tes, atau berdasarkan penentuan unit-unit standar deviasi dari rata-rata, sehingga dapat dipastikan standar tersebut akan berubah.

Pada tahun 1970an, dengan perkembangan tes beracuan kriteria, maka pendekatan *relatif* untuk *standard setting* diganti dengan pendekatan kriteria (*absolute*). Pada metode ini, standar ditentukan terlebih dahulu terhadap kemampuan (*skill*) yang harus dikuasai oleh peserta tes sehingga peserta tes dapat dikatakan lulus. Dengan demikian, *standard* kriteria selalu sama dari tahun ke tahun.

Pendekatan relatif tersebut melahirkan berbagai metode yang dapat membantu dalam penentuan *cut score*. Dua diantaranya adalah metode *Group Contrast* dan metode *Bookmark*. Metode *Group Contrast* merupakan metode yang berpusat pada peserta didik. Metode *Group Contrast* (Kontras Grup) dikenalkan oleh Berk pada tahun 1976, yang menyarankan tentang sebuah perluasan dari sebuah prosedur validasi kelompok. Prosedur kelompok digunakan untuk menentukan perbedaan *cut score* antara siswa-siswa yang terlatih dan tidak terlatih, atau siswa yang telah menguasai materi dan yang tidak menguasai materi (Cizek & Bunch, 2007: 106). Metode ini melibatkan pelaksanaan ujian dua kelompok siswa dimana mereka menerima pengajaran secara efektif mencakup materi tes yang akan diujikan. Dua distribusi dari prestasi peserta tes dipotongkan untuk menemukan *cut score* dengan probabilitas kebenaran maksimal dan minimal. Kebenaran maksimal misalnya menentukan kelompok *master* dan *nonmaster*, sementara kebenaran minimal misalnya mengidentifikasi kekeliruan *master* dan *nonmaster*.

Metode *Group Contrast* mempunyai kelebihan, yaitu mudah untuk diterapkan dan memberikan hasil yang akurat karena berdasarkan pada kondisi real peserta didik. Secara lebih spesifik, tes dengan bentuk pilihan ganda cocok jika penentuan *cut score* dilakukan dengan model *contrasting group* karena ahli akan menentukan kelompok *master* dan *nonmaster* dengan lebih mendasar (Livingstone dan Zieky, 1982: 53). Namun demikian kekurangan metode ini adalah kesulitan memperoleh evaluasi yang sebanding untuk wilayah yang lebih luas semisal tingkat nasional.

Metode *Bookmark* merupakan salah satu metode untuk menentukan *cut score*, yang dikembangkan oleh para peneliti CTB/McGraw-Hill. Metode ini merupakan salah satu dari metode yang berpusat pada butir. Prosedur *Bookmark* dikembangkan untuk mengatasi kekurangan pada metoda Angoff

yang dimodifikasi, yang merupakan metode paling populer dalam penentuan *cut score*.

Metode *bookmark* merupakan model yang populer karena beberapa alasan (Cizek & Bunch, 2007: 157-160) yakni: (1) dari sisi kepraktisan perspektif, metode ini dapat digunakan untuk berbagai situasi yang kompleks, penggabungan dari format pengukuran, dan ahli menggunakan metode dengan mempertimbangkan secara bersama-sama antara *SR (Selected Response)* dan *CR (Constructed Response)*, (2) dari perspektif tersebut maka keputusan dapat diambil melalui metode yang dipilih, (3) metode *bookmark* merupakan metode yang mudah untuk diterapkan, (4) dari perspektif psikometri, metode ini merupakan metode yang sangat menguntungkan karena analisis yang digunakan berdasar pada *butir response teory (IRT)*.

Buklet butir yang diurutkan (*The Ordered Butir Booklet/OIB*) merupakan hal yang esensial dalam prosedur *Bookmark*. *OIB* dikonstruksi dengan menggunakan butir-butir tes. Butir-butir diurutkan dari butir yang mudah dan kemudian semakin sulit, pengurutan tersebut ditentukan oleh prestasi siswa pada saat mengerjakan tes.

Kelebihan dari metode *Bookmark* adalah pada metode ini berfokus pada level data butir. Kemudian butir-butir tersebut tidak hanya dalam bentuk pilihan ganda tapi juga dalam bentuk esai. Keuntungan lain adalah metode ini sangat efisien untuk menentukan satu atau lebih *cut score* dengan tes yang sama (Zieky, Perie, dan Livingston, 2008: 69). Pada sisi lain, untuk menggunakan metode ini harus membuat buku butir terurut (*ordered butir booklet*). Selanjutnya, semua butir dikalibrasi dengan IRT. Kelemahan metode ini adalah penilai kesulitan untuk memahami dan menggunakan kemungkinan jawaban (*Response Probability*).

Seperti yang telah diungkapkan pada poin-poin terdahulu bahwa hasil akhir dari prosedur *standard setting* adalah *passing score/cut-off score* (Cizek, 1996: 21). *Cut-off Score* (yang dikenal juga sebagai *passing score, cut score, cutting score*) merupakan kriteria yang digunakan untuk menganalisis tingkat keberhasilan dalam penilaian. Lebih lanjut *cut-off score* digunakan untuk menggolongkan peserta tes dalam kelompok *master* atau *nonmaster*, atau *advance, proficient, basic, dan below basic* (Hambleton, 2001: 89). Peserta tes

digolongkan sebagai *master* jika *score* peserta tes tersebut sama atau lebih besar dibanding dengan *passing point*. Sementara peserta tes digolongkan sebagai *nonmaster* jika skor yang dimiliki berada di bawah *passing point*, dan seterusnya untuk empat kategorial berbeda.

Namun demikian, sering terjadi juga kesalahan pengklasifian yang disebut sebagai *misclassified*. Misklasifikasi tipe pertama diistilahkan dengan *false* negatif yang merupakan kesalahan penerimaan (Crocer & Algina, 1986: 421). Sebagai contoh, seorang peserta tes yang tidak memiliki kompetensi minimal, tetapi dapat lulus ujian. Misklasifikasi tipe kedua disebut dengan *false*-positif yakni suatu kesalahan penolakan. Dalam klasifikasi jenis ini, seorang peserta tes yang memiliki kemampuan, ternyata harus gagal dalam ujian. Dampaknya adalah pada tindak lanjut terhadap keputusan tersebut. Peserta tes kompeten tapi tidak lulus pada dasarnya tidak memerlukan pembelajaran tambahan, sementara bagi peserta tes yang tidak kompeten sebenarnya membutuhkan pembelajaran tambahan tersebut (Mehrens & Cizek, 2001: 483).

Terkait dengan hal di atas, penelitian ini dilakukan untuk menentukan *cut score* (batas lulus) ujian nasional mata pelajaran akuntansi dengan menggunakan metode *Group Contrast* dan *Bookmark*.

Metode Penelitian

Penelitian ini merupakan penelitian deskriptif kuantitatif yang didukung oleh data kualitatif. Dalam penelitian ini diperlukan data kuantitatif berupa respon peserta dan soal ujian teori akuntansi, serta data kualitatif berupa *judgment* ahli yang digunakan untuk menentukan *cut score*.

Sebelum dilakukan penentuan *cut score*, maka perangkat tes perlu dianalisis karakteristiknya. Karakteristik butir akan dikaji secara kuantitatif melalui analisis empirik. Pada penelitian ini dilakukan analisis butir soal menurut teori tes klasik dengan menggunakan program Iteman versi 3.0 dan teori modern (*Rasch Model*) dengan menggunakan program Acer ConQuest Version 2.0.

Untuk menentukan *cut score*, dalam penelitian ini digunakan dua metode, yaitu:

a. Metode *Group Contrast*

Prosedur/langkah-langkah yang dilakukan pada metode ini adalah:

1) Mengidentifikasi ahli (*expert*).

Expert dalam penelitian ini adalah guru. Pemilihan guru sebagai ahli karena guru menguasai kompetensi mata pelajaran yang diujikan serta memiliki pengetahuan dalam mengenal karakteristik peserta tes. Kualifikasi yang lain adalah guru telah memiliki pengalaman mengajar selama minimal 5 tahun.

2) Ahli mendiskusikan indikator *master* dan *nonmaster*.

Kelompok *master* merupakan kelompok peserta tes yang memiliki skor melebihi batas lulus (*cut score*). Sementara kelompok *nonmaster* merupakan kelompok peserta tes yang memiliki skor di bawah batas lulus (*cut score*).

3) Menghitung skor hasil tes (ahli tidak diperkenankan mengetahui skor tersebut).

Pada tahap ini peneliti melakukan penghitungan terhadap skor-skor peserta tes. Skor-skor tersebut berasal dari respon/jawaban peserta tes dalam mengerjakan soal ujian tes teori akuntansi.

4) Menyiapkan distribusi frekuensi.

Berdasar skor-skor capaian peserta tes, maka dibuat tabel distribusi frekuensi untuk masing-masing kelompok.

5) Membuat grafik distribusi frekuensi pada tiap grup (*master* dan *nonmaster*).

Tahap selanjutnya, membuat grafik distribusi frekuensi dari masing-masing grup, dimana perpotongan dari dua kelompok tersebut merupakan *cut score*.

Namun demikian, perlu dilakukan pengkroscekan hasil *cut score* yang diperoleh berdasarkan titik potong dengan hasil dari persamaan *regresi logistic*.

b. Metode *Bookmark*

Prosedur/langkah-langkah yang dilakukan adalah sebagai berikut:

- 1) Analisis butir dengan menggunakan Model Rasch.
Analisis ini dilakukan untuk mengetahui butir-butir mana saja yang cocok dengan Model Rasch, yang diperlukan dalam penyusunan buklet. Berdasarkan pada analisis ini terdapat 37 butir yang cocok dengan Model Rasch, 3 diantaranya tidak cocok dengan Model Rasch. Selain itu, pada analisis ini dapat diketahui indeks kesukaran dari masing-masing butir soal.
- 2) Menyusun butir dari yang paling mudah sampai yang paling sulit kemudian disusun dalam *Ordered Butir Buklet* (OIB/Buklet Butir Terurut).
Pada tahap ini, butir yang telah cocok dengan Model Rasch selanjutnya disusun berdasarkan indeks kesukarannya dalam sebuah Buklet Butir Terurut. Butir-butir tersebut disusun dari butir yang termudah sampai butir yang paling sulit. Informasi lain yang perlu ditampilkan pada buklet adalah besarnya kemampuan yang diperlukan (θ) untuk menjawab butir dengan tingkat kesukaran tertentu dengan tepat pada *Respon Probability* (RP) 0,67 dan RP 0,5.
- 3) Ahli mendeskripsikan mengapa butir tersebut mudah atau sulit.
Setelah dibuat buklet, maka ahli mendeskripsikan dari tiap butir mengapa butir tersebut sulit dan mengapa mudah, yang dimaksudkan adalah pendeskripsian tentang kemampuan yang harus dimiliki untuk menjawab butir tersebut.
- 4) Ahli diminta untuk meletakkan *bookmark* pada batas antara butir yang memiliki indeks kesukaran yang mudah ke butir yang sulit.
Ahli pada tahap ini diminta untuk mencermati tiap halaman buklet dan memastikan apakah peserta tes dapat menjawab soal tersebut dengan tepat. Ahli akan berhenti dan meletakkan *bookmark* pada halaman dimana soal tersebut tidak dapat dijawab dengan tepat oleh peserta tes kelompok sedang.

- 5) Menentukan rerata *bookmark* sebagai *cut score* yang direkomendasikan. Bisa dipastikan ada perbedaan letak *bookmark* dari kedelapan ahli, dengan demikian perlu diperhitungkan rerata *bookmark* yang menghasilkan *cut score* tersebut.

Hasil Penelitian dan Pembahasan

1. Karakteristik Perangkat UN Teori Akuntansi Tahun Ajaran 2008/2009

Data set yang digunakan pada penelitian pendahuluan berupa respon siswa SMK di Kota Yogyakarta. Data ini merupakan data dokumentasi yang diperoleh peneliti dari Pusat Penilaian Pendidikan Balitbang Depdiknas Jakarta. Pada penelitian pendahuluan ini, peneliti menganalisis 314 respon siswa yang berasal dari 8 SMK Bisnis dan Manajemen di wilayah Kota Yogyakarta. Perangkat UN Teori Akuntansi tahun ajaran 2008/2009 terdiri atas 40 butir soal pilihan ganda dengan 5 alternatif jawaban. Terlebih dahulu disusun kisi-kisi soal sesuai dengan standar kompetensi, kompetensi dasar beserta indikatornya hal ini dilakukan untuk mengelompokkan butir-butir butir.

Berdasarkan teori tes klasik dapat diungkap bahwa 52,5 % (21 butir) termasuk dalam kategori mudah, 40% (16 butir) termasuk kategori sedang (tidak mudah maupun tidak sulit), dan 7,5% (3 butir) termasuk kategori sukar. Ditinjau dari daya pembeda hanya terdapat 2 butir soal yang memiliki daya beda di bawah kriteria ideal, artinya terdapat 38 butir (95%) merupakan butir yang dapat membedakan kemampuan peserta tes. Sementara dengan menggunakan model Rasch, ujian termasuk dalam kualifikasi baik. Hal ini dikarenakan 92,5% atau 37 butir soal cocok dengan model Rasch.

2. *Cut score* berdasarkan Metode Bookmark

a. Proses Pelaksanaan

Pelaksanaan metode ini terdiri dari pemaparan materi dilanjutkan dengan pelatihan kepada para ahli (*judge*), kemudian dilakukan *Bookmark* putaran I dan putaran II. Masing-masing ahli diberikan materi yang telah

disusun oleh peneliti, buklet, kisi-kisi soal dan standar kelulusan kompetensi kejuruan SMK Bisnis dan Manajemen.

Ahli diminta untuk mendeskripsikan mengapa butir-butir tersebut sulit maupun mudah. Selanjutnya ahli menentukan apakah kelompok sedang dapat menjawab butir tersebut dengan benar dan berhenti pada butir yang diyakini tidak dapat dijawab dengan benar oleh peserta tes kelompok sedang. *Respon Probability* (RP) yang digunakan adalah sebesar RP 0,67 dan RP 0,5. Penggunaan kedua RP tersebut dikarenakan beberapa ahli yang mengungkapkan rekomendasinya berkaitan dengan RP yang tepat untuk digunakan ketika Rasch Model sebagai alat dalam mengestimasi kemampuan peserta tes. Wang (2003) dalam Cizek & Bunch (2007: 162) mengungkapkan bahwa dalam konteks Rasch Model maka RP 0,5 lebih akurat untuk digunakan. Namun demikian menurut Zieky, Perie & Livingston (2008: 113) menyatakan bahwa RP sebesar 0,67 lebih umum digunakan untuk metode *Bookmark*.

b. Batas Kelulusan

Pada Tabel 1 berikut tampak bahwa batas lulus terletak pada skor 68,96 (pada RP 0,67) baik pada putaran 1 maupun 2 dengan tingkat kelulusan sebesar 68,15% (214 peserta tes).

Tabel 1. *Cut score* dengan RP 0,67

Putaran	Skala Kemampuan		Skala		% Kelulusan
	Rerata	St. Dev	Rerata	St. Dev	
1	1,52	0,88	68,96	11,05	68,15
2	1,52	0,88	68,96	11,05	68,15

Pada Tabel 2 berikut diungkap bahwa batas lulus terletak pada skor 67,36 (pada 0,5) baik pada putaran 1 maupun 2. Pada batas lulus tersebut tingkat kelulusan sebesar 72,61% (228 peserta tes).

Tabel 2. *Cut score* dengan RP 0,5

Putaran	Skala Kemampuan		Skala		% Kelulusan
	Rerata	St.Dev	Rerata	St.Dev	
1	1,39	1,17	67,36	14,61	72,61
2	1,39	1,17	67,36	14,61	72,61

c. Kemampuan Lulusan

Terdapat 28 deskriptor kompetensi yang harus dikuasai oleh peserta tes pada batas lulus 68,96 maupun 67,36 yang terurai pada Tabel 3 berikut ini.

Tabel 3. Deskripsi Ketuntasan Minimal

No	Deskriptor Ketuntasan Minimal
1.	Menganalisis efek pengambilan prive terhadap ekuitas
2.	Mampu mendefinisikan pengertian buku besar
3.	Menentukan format neraca lajur 10 kolom
4.	Menyusun jurnal pembelian peralatan
5.	Menyusun jurnal penyesuaian gaji yang belum dibayar
6.	Menentukan berbagai akun yang disajikan dalam laporan neraca
7.	Menentukan definisi akuntan public
8.	Menganalisis efek penjualan kredit terhadap jurnal khusus
9.	Menentukan fungsi kertas kerja
10.	Menentukan komponen yang terdapat dalam bukti transaksi
11.	Menentukan jurnal akibat penjualan dengan memberikan cek
12.	Menganalisis efek pengambilan prive terhadap modal dan harta
13.	Menentukan definisi posting
14.	Menentukan jurnal untuk menutup akun retur penjualan
15.	Menentukan dokumen sumber dalam penyusunan jurnal penutup
16.	Menentukan jurnal penerimaan dengan nota kredit dan potongan
17.	Menentukan kelompok rekening yang mempunyai saldo normal kredit
18.	Menentukan perbedaan pokok akuntansi perusahaan dagang dan perusahaan manufaktur
19.	Menentukan definisi transaksi akuntansi
20.	Menentukan fungsi rekapitulasi jurnal khusus
21.	Menentukan definisi akuntansi
22.	Sumber data pencatatan jurnal pembelian

No	Deskriptor Ketuntasan Minimal
23.	Menentukan kolom-kolom dalam jurnal umum
24.	Menganalisis formula persamaan dasar akuntansi
25.	Menentukan jurnal penutup rugi bersih perusahaan
26.	Menganalisis laba/rugi pada neraca lajur dan pengaruhnya terhadap ekuitas
27.	Menentukan ayat jurnal penyesuaian dengan materi persekot sewa.
28.	Menentukan bukti transaksi akibat penyeteroran modal oleh pemilik.

3. *Cut score* berdasarkan metode Grup Kontras

a. Proses pelaksanaan

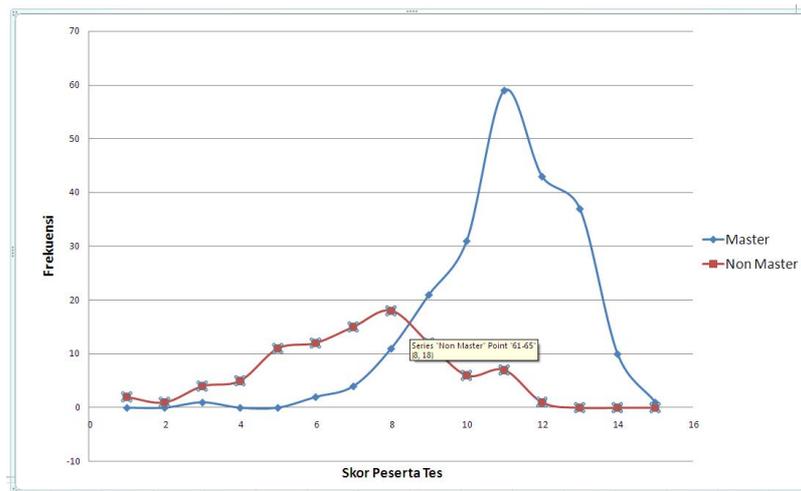
Pengelompokan peserta menjadi kelompok master dan nonmaster berdasarkan pada respon peserta masing-masing sekolah. Informasi yang didapat dari *Focus Group Discussion* (FGD) bahwa tiap sekolah menggunakan Kriteria Ketuntasan Minimal (KKM) sebesar 70 (pada skala 100). Berdasarkan kriteria ini, sekolah dikategorikan sebagai kelompok master jika mencapai rerata skor UN minimal 70, sedangkan dikategorikan nonmaster jika belum mencapai 70. Berikut ini adalah rerata tiap sekolah yang dijadikan tempat penelitian.

Tabel 4. Rerata kemampuan sekolah

No	Sekolah	Banyaknya Siswa	Rerata Skor	Standar Deviasi	Kelompok
1	SMK N 1 Yogyakarta	74	80,61	6,35	Master
2	SMK Muhammadiyah 1	23	54,57	12,61	Nonmaster
3	SMK Koperasi	57	74,12	7,87	Master
4	SMK BOPKRI 1	37	61,22	12,95	Nonmaster
5	SMK PIRI 3	21	62,02	8,61	Nonmaster
6	SMK Muhammadiyah 2	13	58,85	8,39	Nonmaster
7	SMK 7	73	82,71	8,4	Master
8	SMK Marsudi Luhur 1	16	71,09	9,99	Master
	Total	314	61,27	9,39	

b. Batas Kelulusan

Setelah diketahui rerata masing-masing sekolah, kemudian dibuat grafik distribusi frekuensi pada kelompok master dan nonmaster. Perpotongan grafik pada kelompok master dan nonmaster ini merupakan *cut score* dari nilai UN teori akuntansi. Perpotongan kedua distribusi tersebut terletak pada titik antara 61-65 (Gambar 1).



Gambar 1. Distribusi Frekuensi Kelompok Master dan Nonmaster

Berdasarkan pada gambar, perpotongan distribusi frekuensi kelompok master dan nonmaster belum tampak begitu jelas titik perpotongan tersebut (Gambar 1). Dengan demikian diperlukan persamaan yang dihasilkan dari regresi logistik berdasarkan output sebagai berikut: $y^* = 12,046 - 0,183 (x)$; dengan $y^* = 0,5$ maka didapat x sebesar 63,09 nilai ini merupakan *cut score* yang direkomendasikan menurut metode grup kontras. Sebanyak 247 peserta tes (78,66%) lulus dengan *cut score* sebesar 63,09 tersebut.

d. Kemampuan Lulusan

Pada metode grup kontras, *cut score* yang dihasilkan sebesar 63,09. Pada batas lulus tersebut peserta tes memiliki 27 deskriptor kemampuan yang dijabarkan pada Tabel 5 berikut ini.

Tabel 5. Deskripsi Ketuntasan Minimal

No	Deskriptor Ketuntasan Minimal
1.	Menganalisis efek pengambilan prive terhadap ekuitas
2.	Mampu mendefinisikan pengertian buku besar
3.	Menentukan format neraca lajur 10 kolom
4.	Menyusun jurnal pembelian peralatan
5.	Menyusun jurnal penyesuaian gaji yang belum dibayar
6.	Menentukan berbagai akun yang disajikan dalam laporan neraca
7.	Menentukan definisi akuntan public
8.	Menganalisis efek penjualan kredit terhadap jurnal khusus
9.	Menentukan fungsi kertas kerja
10.	Menentukan komponen yang terdapat dalam bukti transaksi
11.	Menentukan jurnal akibat penjualan dengan memberikan cek
12.	Menganalisis efek pengambilan prive terhadap modal dan harta
13.	Menentukan definisi posting
14.	Menentukan jurnal untuk menutup akun retur penjualan
15.	Menentukan dokumen sumber dalam penyusunan jurnal penutup
16.	Menentukan jurnal penerimaan dengan nota kredit dan potongan
17.	Menentukan kelompok rekening yang mempunyai saldo normal kredit
18.	Menentukan perbedaan pokok akuntansi perusahaan dagang dan perusahaan manufaktur
19.	Menentukan definisi transaksi akuntansi
20.	Menentukan fungsi rekapitulasi jurnal khusus
21.	Menentukan definisi akuntansi
22.	Sumber data pencatatan jurnal pembelian
23.	Menentukan kolom-kolom dalam jurnal umum
24.	Menganalisis formula persamaan dasar akuntansi
25.	Menentukan jurnal penutup rugi bersih perusahaan
26.	Menganalisis laba/rugi pada neraca lajur dan pengaruhnya terhadap ekuitas
27.	Menentukan ayat jurnal penyesuaian dengan materi persekot sewa.

Cut score yang diperoleh dengan menggunakan metode *Bookmark* diperoleh hasil sebesar 68,96 (skala 100) pada RP 0,67 dan sebesar 67,36 (skala 100) pada RP 0,5. Sementara dengan menggunakan metode grup kontras diperoleh hasil sebesar 63,09. Hal ini menunjukkan bahwa batas kelulusan berdasarkan data empirik ternyata lebih tinggi dari yang ditetapkan oleh pemerintah sebesar 40,00 (skala 100 tahun 2009).

Mencermati hasil ini, ada beberapa penyebab yang dapat digunakan untuk menginterpretasikan tingginya batas lulus. Pada metode dengan menggunakan grup kontras, hasil perhitungan *cut score* dipengaruhi oleh distribusi skor peserta tes. Pada penelitian ini baik kelompok master maupun nonmaster rata-rata memiliki kurva skor juling kiri (rerata lebih tinggi dibandingkan dengan nilai tengah) sehingga menyebabkan perolehan *cut score* hasil perhitungan menjadi lebih tinggi. Hal ini sejalan dengan hasil analisis karakteristik butir soal, dimana sebagian besar butir soal yakni sebesar 52,5 % (21 butir) termasuk dalam kategori mudah, 40% (16 butir) termasuk kategori sedang (tidak mudah maupun tidak sulit).

Karakteristik parameter butir secara tidak langsung berpengaruh pada hasil perhitungan *cut score* yang dihitung dengan menggunakan metode grup kontras. Butir-butir yang memiliki tingkat kesulitan rendah akan dijawab peserta tes dengan benar sehingga skor siswa menjadi tinggi dan skor-skor ini mempengaruhi distribusi peserta.

Kemampuan ahli dalam mengenal karakteristik peserta tes mempengaruhi penentuan letak *bookmark* pada buklet, yang pada akhirnya menghasilkan *cut score*. Seperti yang telah dijelaskan pada paragraph terdahulu, buklet tersebut berisi butir-butir yang telah diurutkan dari indeks kesukaran yang paling mudah sampai yang paling sulit. Indeks kesukaran yang telah diketahui berdasarkan analisis butir soal membantu ahli dalam menentukan *bookmark* (batas) yang pada akhirnya direkomendasikan sebagai *cut score*. Secara keseluruhan karakteristik soal berdasarkan teori respon butir menyimpulkan 80% butir merupakan butir dengan kualifikasi indeks kesukaran sedang. Hal ini mengakibatkan *cut score* yang dihasilkan dalam metode *Bookmark* lebih tinggi. Pada pihak lain penggunaan RP 0,67 dan RP 0,5 menghasilkan *cut score* yang berbeda sekalipun menuntut kompetensi minimal yang sama. Hal ini dikarenakan ahli pada RP 0,5

cenderung meletakkan batas buklet lebih mundur (pada lembar halaman yang lebih banyak) dibandingkan pada penggunaan RP 0,67. Ahli, pada sisi lain terlalu *over estimates* terhadap kemampuan siswa sehingga dalam meletakkan batas tidak sejalan dengan kemampuan siswa kelompok sedang yang sesungguhnya.

Ahli mengungkapkan bahwa adakalanya dijumpai butir-butir yang dimungkinkan dapat dijawab oleh peserta tes kelompok sedang, kemungkinan pula tidak dapat dijawab. Hal ini mengakibatkan ahli bingung dalam menempatkan *bookmark*. Pada sisi lain, penggunaan metode *Bookmark* menjadi sesuatu yang tampak sulit untuk diaplikasikan pada level guru karena karakteristik butir harus diungkap berdasarkan model Rasch atau IRT yang *notabene* masih awam bagi guru. Kesulitan lain adalah perlunya peserta tes dalam jumlah yang banyak (misalnya 300) sementara guru di sekolah maksimal hanya memiliki 80 peserta didik.

Penggunaan metode kontras grup relatif lebih aplikatif bagi guru karena tidak adanya pengungkapan parameter instrumen. Prosedur dalam metode kontras grup memungkinkan guru menyusun *cut score*/KKM sekalipun pada tingkat kelas. Dalam arti kata lain, metode grup kontras dapat digunakan sebagai alternatif dalam menyusun *cut score*/KKM selain dengan prosedur penentuan KKM yang selama ini dilakukan yakni dengan mempertimbangkan kompleksitas materi, daya dukung, dan *intake* peserta didik.

Terkait dengan tingkat keakuratan *cut score* yang dihasilkan, metode *bookmark* dalam prosedurnya mengungkap terlebih dahulu parameter kesukaran butir kemudian dilanjutkan dengan estimasi ahli pada kelompok sedang dalam menjawab butir-butir soal berdasarkan hasil parameter butir tersebut. Artinya, metode *Bookmark* lebih mengakomodasi berbagai pertimbangan sehingga menghasilkan *cut score* yang akurat. Sekalipun pada sisi lain, terdapat RP yang memerlukan pemahaman mendalam bagi ahli.

Pada dasarnya terdapat perbedaan prosedur yang digunakan untuk menentukan *cut score* antara metode grup kontras dengan metode *Bookmark*. Penggunaan metode *Bookmark* memiliki beberapa kelebihan yakni: 1) penggunaan buklet yang berisi butir-butir yang diurutkan membantu guru dalam menentukan letak *bookmark*; 2) butir-butir tersebut dapat diurutkan

berdasarkan *p-value* (berdasarkan pada teori klasik) sehingga tidak memerlukan jumlah peserta tes yang besar, setidaknya tidak seperti penggunaan analisis dengan menggunakan IRT atau Rasch Model.

Berdasarkan hasil penelitian, maka tampak bahwa terdapat perbedaan *cut score* antara *cut score* yang ditetapkan berdasarkan kebijakan pemerintah dengan data empirik. Pemerintah dalam menentukan *cut score* mempertimbangkan berbagai hal, sebagai contoh kualitas sekolah dan kondisi daerah yang beragam. Dengan demikian, pemerintah sudah selayaknya mengakomodasi berbagai keadaan tersebut sehingga standar yang ditetapkan dapat diterima pada berbagai kondisi sekolah di seluruh Indonesia. Berdasarkan data empirik, *cut score* lebih tinggi dari kebijakan pemerintah sehingga dapat dimungkinkan pemerintah dapat menaikkan *cut score*. Tentu kebijakan ini harus dibarengi dengan berbagai perbaikan pada kualitas guru dan sarana prasarana penunjang serta sistem yang terjamin baik sehingga tingkat kelulusan mendekati 100%.

Penyelenggaraan Ujian Nasional (UN) merupakan barometer untuk mencermati kualitas pembelajaran yang tercermin pada kemampuan lulusan. Sudah barang tentu yang diharapkan adalah kemampuan lulusan memenuhi standar kompetensi yang telah ditetapkan. *Cut score* (batas lulus) menjadi sesuatu yang esensial untuk ditetapkan sehingga siswa dapat ditempatkan sesuai dengan kemampuannya (kompeten/tidak kompeten). Sekalipun terdapat konsekuensi persentase kelulusan masih di luar harapan, bukan berarti *cut score* harus rendah sehingga banyak siswa dapat lulus. Hal ini justru menjadi tantangan bagaimana mengembangkan kapasitas peserta didik dan memberdayakan segala sumberdaya yang ada sehingga *cut score* tetap realistis untuk dicapai. Berdasarkan penelitian ini, *cut score* teori Akuntansi masih cukup beralasan untuk dapat dinaikkan dengan harapan deskripsi kemampuan siswa akan semakin baik dan terstandar sebagai lulusan yang berkualitas.

Simpulan dan Saran

1. Simpulan:

- a. Karakteristik soal ujian teori akuntansi berdasarkan model Rasch termasuk dalam kualifikasi baik hal ini dikarenakan 92,5% atau 37 butir soal cocok dengan model Rasch.
- b. Penentuan *cut score* dengan metode *bookmark* diperoleh hasil 68,96 pada *Respon Probability* (RP) 0,67 dan sebesar 67,36 pada RP 0,5 sementara pada metode *group contrast* menghasilkan *cut score* sebesar 63,09. *Cut score* yang dihasilkan oleh berbagai metode tersebut kesemuanya pada skala 100.
- c. Metode *Bookmark* lebih memberikan hasil *cut score* yang relatif lebih akurat karena dalam prosedurnya mempertimbangkan setidaknya dua aspek, yakni parameter tingkat kesukaran butir dan estimasi expert dalam mengestimasi tiap butir dengan respon probabilitas tertentu. Respon peserta diperlukan untuk mengungkap parameter tingkat kesukaran butir tersebut.
- d. Metode *Group Contrast* relatif lebih mudah diterapkan bagi guru karena tidak membutuhkan jumlah peserta tes dengan jumlah tertentu, tidak memerlukan pengungkapan tingkat kesukaran butir dengan model Rasch atau IRT, dan tidak memerlukan estimasi kemampuan peserta tes.

2. Saran

- a. Mencermati hasil penelitian ini, dapat diungkapkan bahwa *cut score* untuk ujian nasional teori akuntansi terlalu rendah, sehingga perlu peninjauan ulang. Penelitian yang terkait dengan batas lulus dan penggunaan metode yang berbeda sangat disarankan sehingga dapat diperbandingkan hasil *cut score* yang diperoleh, lebih jauh hasil tersebut dapat digunakan sebagai bahan masukan kepada pemerintah.
- b. Perlu adanya berbagai pelatihan penentuan *cut score* bagi guru sehingga guru memiliki tambahan pengetahuan yang pada akhirnya dapat menerapkan metode-metode tersebut.
- c. Sekalipun *cut score* dapat dinaikkan ternyata persentase kelulusan masih jauh dari angka 100%, pada pihak guru perlu adanya peningkatan dan

variasi dari berbagai aspek pembelajaran. Sebagai contoh penerapan berbagai model pembelajaran, strategi pembelajaran yang lebih kontekstual, dan lain-lain.

Daftar Pustaka

- Cizek, G., J. (2001). *Conjuncture on the rise and call of standard setting: An introduction to context and practice, in setting performance standards: Concepts, methods, and perspectives*, Gregory J. Cizek ed., New Jersey: Lawrence Erlbaum Associated, Pub. Mahwah.
- Cizek, G. J., Bunch, B., B.(2007). *Standard setting: a guide to establishing and evaluating performance standard on test*. ND: Sage Publications.
- Cizek, G. J. (1996). *Setting passing scores* [Versi Elektronik]. *Educational Measurement Issues and Practice*, 20-31.
- Depdiknas. (2005). *Peraturan Pemerintah No.19 Tahun 2005, tentang standar nasional pendidikan*.
- Crocer, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: The Dryden Press Saunders College Publishing.
- Hambleton, R. K., Pitoniak, M. J. (2006). *Educational measurement (4th Ed)*, R.L Brenan ed., USA: American Council on Education and Praeger Publisher.
- Hambleton, R. K. & (2001). *Setting performance standards on educational assessments and criteria for evaluating the process, in setting performance standards: Concepts, methods, and perspectives*, Gregory J. Cizek ed., New Jersey: Lawrence Erlbaum Associated, Pub. Mahwah.
- Heri Retnawati. (2008). *Penentuan batas lulus (Standard Setting) ujian nasional mata pelajaran matematika di DIY*. Laporan Penelitian: PKPSP LP UNY.

- Livingston, S. A., & Zieky, M. J. (1982). *Passing score: a manual for setting standards of performance on educational and occupational test*. US: Educational Testing Service.
- Mehrens, W. A., & Cizek, G. J. (2001). *Standard setting and the public good: benefit accrued and anticipated, in setting performance standards: Concepts, methods, and perspectives*, Gregory J. Cizek ed. New Jersey: Lawrence Erlbaum Associated, Pub. Mahwah.
- Zieky, M. J., Perie, M., & Livingstone, S. A. (2008). *Cut score: a manual for setting standard of performance on educational and occupational test*. ND: Sage Publications.