ANALYSIS OF MATHEMATICAL PROBLEM WITH GROUP WISE FOR COMPUTERIZED ADAPTIVE TESTING

Rukli[1)]
[1]Universitas Muhammadiyah Makassar, Sulawesi Selatan
e-mail: rukli@unismuh.ac.id

ABSTRACT

This study aims to put the parameters of the problem on a general scale with a group wise design. Placement of questions on a general scale can be interpreted that although the test participants responded to the character of the problem is not the same but the ability of test participants in Competitive Adaptive Testing (CAT) can be compared with each other. Data response question comes from four packages namely package 1, package 2, and package 3 National Exam on mathematics subjects. The total number of questions is 140. The data is analyzed with the Bilog_MG program. The results show that the number of questions meet the Rasch model of 109 questions. The matter can be entered into CAT bank. Although the test participants responded to the problem and the number of questions was not the same but the test scores were comparable.

Keywords: Group Wise, Computerized Adaptive Testing

PENDAHULUAN

Kemajuan teknologi komputer melesat begitu cepat dengan dukungan jaringan nirkabel sehingga mempengaruhi perkembangan bidang lain termasuk psikometri. Pengadministrasian *paper and pencil test* (*P&P*) menyajikan soal yang sama untuk semua peserta tes tanpa memperhatikan kemajuan teknologi komputer kini telah mulai bergeser pada sistem pengujian berbasis komputer. Pengujian berbasis komputer yang dikenal dengan *Computer Based Test* (*CBT*) mengalami satu langkah lebih maju yakni penggunaan dari media kertas ke media komputer namun urutan, jumlah soal serta waktu tes sama untuk semua peserta tes. Kekurangan-kekurangan yang ada terus diperbaiki sehingga muncul sistem pengujian dimana penyajian soal mengadaptasi kemampuan peserta tes secara unik. Artinya, jika kemampuan peserta tes rendah maka diberikan karakteristik soal yang mempunyai tingkat kesukaran yang rendah, sebaliknya jika kemampuan peserta tes tinggi maka diberikan karakteristik soal yang mempunyai tingkat kesukaran yang tinggi. Pola penyajian soal tes yang demikian bila menggunakan komputer disebut sebagai *Computerized Adaptive Testing* (*CAT*).

Berdasarkan hal tersebut, prosedur penyajian soal pada *CAT* berbeda dengan *P&P* maupun *CBT*. Prosedur penyajian soal pada *CAT* mempunyai mekanisme, yakni soal diadaptasikan sesuai dengan kemampuan saat ini, artinya kemampuan peserta tes terus di-*update* sampai kemampuan peserta tes diketahui. Hal tersebut dapat dipahami secara teori bahwa kemampuan setiap peserta tes berbeda sehingga membutuhkan soal yang berbeda juga. Implikasinya, setiap peserta tes menempuh soal yang berbeda baik dari segi karakteristik parameter soal, jumlah soal, dan jumlah waktu tempuh tes. Pertanyaan yang muncul adalah bagaimana membandingkan kemampuan peserta tes kalau setiap peserta tes merespon soal yang tidak sama karakteristiknya? Atau dengan kata lain bagaimana meletakkan beberapa paket tersebut dalam satu skala sehingga soal

apapun yang dikerjakan peserta tes maka skor dapat diperbandingkan? Berkenaan hal tersebut, penelitian ini bertujuan melakukan analisis data respon dari soal matematika (Ujian Nasional) UN dengan desain *group wise* agar parameter soal berada pada skala yang sama pada aplikasi *CAT*.

Perkembangan dan Mekanisme *CAT*

CAT pertama kali diperkenalkan tahun 1908 pada tes inteligensi (Weiss, 1985), selanjutnya tahun 1960 dikembangkan oleh Lord pada *Educational Testing Service* dalam bentuk program (Hambleton *et al.* 1991:145). Perkembangan *CAT* semakin pesat seiring perkembangan kemampuan komputer dengan dukungan jaringan nirkabel.

CAT telah diimplementasikan dalam berbagai bidang psikologi, teknik rekayasa, dan pendidikan dalam melakukan sertifikasi dan lisensi. Beberapa aplikasi *CAT* yang telah diimplementasikan secara *online* tersebar di beberapa negara. Misalnya, *Graduate General Examination* (*GRE*), yaitu sistem pengujian yang dikembangkan pada tahun 1993 di Amerika Serikat (Mills & Stefen, 2000) dan *MATHCAT* yaitu *CAT* di Belanda. (Verschoor & Straetmans, 2000). Sebaliknya, perkembangan *CAT* di Indonesia masih dalam tahap rancangan dan uji coba secara lokal. Misalnya, Rukli (2009) telah merancang *CAT* berbasis web dan Rukli (2010) telah melakukan uji coba aplikasi *CAT* pada mata pelajaran matematika, fisika, dan bahasa Indonesia di SMA berbasis desktop

Perkembangan *CAT* semakin meningkat dalam berbagai variasi mekanisme pengembangan dengan melibatkan fitur-fitur tambahan. Misalnya, Rukli dan Hartati (2011) menggunakan fitur Sistem Pendukung Keputusan (SPK) pada *CAT* dalam menjaring kemampuan peserta tes sesuai dengan preferensi *user*. Pengembangan *CAT* tersebut pada prinsipnya memiliki beberapa aturan pokok walaupun berbeda dalam mekanisme teknisnya.

Mekanisme pengembangan aplikasi *CAT* membutuhkan beberapa aturan pokok, yakni: **Pertama**, penentuan awal kemampuan peserta tes atau soal pertama yang direspon peserta tes. **Kedua**, aturan penghentian pengujian pada *CAT* . Aturan tersebut bervariasi, antara lain dapat menggunakan respon sebuah soal untuk mendapatkan galat baku penaksiran parameter kemampuan lebih kecil atau sama dengan 0,43 (Reshetar dalam Hambleton, *et al*, 1991: 149) atau lebih kecil atau sama dengan 0,3 (Gustha, 2003: 10). Respon peserta tes terhadap dua soal berturut-turut untuk mendapatkan selisih galat baku penaksiran parameter kemampuan lebih kecil atau sama dengan 0,01 (Rukli, 2010). **Ketiga**, metode penyajian soal dalam memilih soal yang adaptif terhadap kemampuan peserta. **Keempat**, bank soal merupakan tempat soal yang direspon peserta tes. *CAT* memerlukan soal dalam jumlah besar yang disimpan dalam bank *CAT*. Soal tersebut dianalisis terlebih dahulu secara kuantitatif untuk menentukan parameter soal dalam ruang lingkup teori pengukuran.

Secara psikometri, ada beberapa model dalam menentukan karakteristik soal, misalnya model Rasch. Model Rasch merupakan model pengukuran. Model tersebut berbeda dengan model statistik, misalnya model logistik satu parameter (1PL). Dalam kaitan dengan pengembangan bank *CAT* soal pada aplikasi *CAT*,

memilih soal yang memenuhi model memiliki peran yang penting dalam kaitan dengan penentuan soal tersebut pada skala yang sama. Oleh karena itu, data respon peserta tes dianalisis untuk menentukan apakah data tersebut *fit* dengan model Rasch dalam menempatkan soal pada skala umum dengan desain *group wise*.

Desain *Group Wise*

Analisis soal dengan data respon peserta tes memerlukan dua kondisi tertentu yang terpadu. Pertama, kondisi data UN SD sama seperti data UN SMP-SMA yang terdiri dari beberapa paket soal dimana setiap paket pada tahun yang sama mempunyai soal *anchor* sebanyak 10 soal sehingga data respon soal tersebut dan kuncinya perlu di_*setting* dalam desain analisis. Cara *setting* desain demikian dapat dilakukan pada beberapa program analisis data. Misalnya, penelitian Chuesathucchon (2009) menganalisis data respon soal matematika menggunakan program RUMM dengan desain skala *linking* terhadap 10 *common item*. Thomson dan Weiss (2011) mengatakan bahwa soal tes di dalam bank *CAT* perlu ditempatkan dalam satu skala umum jika di dalam bank soal tersebut sudah mempunyai soal sebelumnya sehingga mempunyai metrik skala baru, jika bank soal masih baru maka jarang dilakukan link pada skala umum namun cukup dilakukan kalibrasi apakah soal sesuai dengan model walaupun tebakan soal akan berpengaruh jelek sehingga tidak stabil digunakan dalam *CAT*.

Lebih penting dari semua itu adalah informasi dimensional tes sudah terjamin di dalam bank soal. Cara yang dikemukakan oleh Thomson dan Weiss (2011) dalam menguji dimensional tes berdaya guna jika *CAT* tes menggunakan rancangan paket *fixed.* Aplikasi *CAT* menggunakan paket *free* diartikan setiap peserta tes menempuh soal bervariasi karakteristik dan jumlah diadaptasikan dengan kemampuan saat ini sehingga menggunakan rancangan paket *free* suatu hal yang tidak mungkin diterapkan. Kalau dikaitkan dengan dimensional tes lebih tepat untuk sementara mengasumsikan bahwa mata pelajaran yang diujikan sudah dominan mengukur satu unjuk kerja. Djemari Mardapi (1996: 15-19) mengemukakan bahwa pembuatan kisi-kisi tes mengacu pada konsep bahwa materi pokok bahasan pada materi tes dapat tidak sama, namun tetap memiliki substansi pokok yang sama, dan penilaian pada ranah kognitif cenderung berdimensi satu. Walaupun unidimensi suatu tes dapat ditentukan dan digambarkan secara sederhana dan lebih mudah dengan *fuzzy c-means* (Rukli, 2011) namun secara teori semakin kecil ruang lingkup mata pelajaran atau mata pelajaran tersebut dibagi dalam beberapa sub pokok bahasan maka soal dari sub pokok bahasan tersebut lebih dominan dalam mengukur suatu unjuk kerja. Hal tersebut tidak bertentangan teori psikometri dalam kaitan pengembangan bank soal dalam *CAT*.

Kedua, kondisi data yang ada memerlukan desain yang lebih rasional namun memiliki dasar teori yang cukup. Jika data respon peserta tes terdiri dari beberapa paket dengan tahun yang berbeda maka memerlukan desain yang tepat. Penempatan soal *anchor* ke dalam satu kelompok kolom terpisah suatu hal tidak mungkin secara praktek sebab paket yang lain yang berasal dari tahun berbeda juga perlu ditempatkan dengan cara yang sama. Soal *anchor* yang tersebar pada

beberapa tahun memerlukan desain multi kelompok dan multi subtes. Desain tersebut sesuai dengan desain *group wise* dalam Program Bilog-MG.

METODE PENELITIAN

Data respon soal berasal dari empat paket yakni paket 1, paket 2, dan paket 3 UN mata pelajaran matematika di Propinsi Sulawesi Selatan ditambah data respon soal dari paket 2 dari Kabupaten Sidoarjo Propinsi Jawa Timur. Tiap paket tersebut diambil secara acak 500 data respon soal sehingga jumlah respon soal sebanyak 2000. Tiap paket mempunyai 40 soal dengan 10 soal *anchor* sehingga jumlah keseluruhan soal dianalisis sebanyak 140 soal.

Data respon peserta tes setiap paket diambil dari daerah tingkat kabupate/kota dalam bentuk file berekstensi dbf. Data tersebut ditransformasikan beberapa tahap dengan program adobe acrobat 9 pro extended sehingga dapat di_*copy-paste* ke microsoft office excel. Data yang berekstensi xls diedit sesuai rancangan desain analisis. Data tersebut kemudian di_*copy paste* ke notepad dikemudian di_*save* dalam bentuk eksistensi dat untuk di_*RUN* di Program Bilog-MG.

Analisis data soal menggunakan program Bilog_MG. Program tersebut disetting sesuai desain *group wise* dan model Rasch.

Rancangan Analisis Data

Data respon UN yang terdiri atas empat paket, dimana paket1, paket2, dan paket3 masing-masing mempunyai 40 soal dengan 10 soal *anchor* terdiri dari soal 3, 12, 16, 21, 27, 30, 32, 36, 38, dan 39, sedangkan satu paket Sidoarjo sehingga tidak memiliki soal *anchor* dalam analisis ini sebanyak 40 soal.

Berdasarkan kondisi soal yang terdiri dari beberapa paket dengan tahun berbeda maka rancangan analisis data dideskripsikan seperti Gambar berikut. Soal ke- 3, 12, 16, 18, 19, ..., 110, 112, 116, 118, dan 119 berada pada ketiga paket sebagai soal *anchor*, sedangkan paket 4 (Sidoarjo) tidak memiliki soal *anchor* sehingga tidak ada link dengan paket lain. Berdasarkan rancangan tersebut, data respon, kunci, dan peserta tes dibuat dalam satu format yang bisa dibaca dan sesuai dengan *syntax* Program Bilog-MG.

Paket1	Paket2	Paket3	Paket Sidoarjo
T001	T001	T001	T001
T002	T002	T002	T002
T003	T003	T003	T003
T004	T004	T004	T004
T005	T005	T005	T005
T006	T006	T006	T006
T007	T007	T007	T007
T008	T008	T008	T008
T009	T009	T009	T009
T010	T010	T010	T010
T011	T011	T011	T011
T012	T012	T012	T012
T013	T013	T013	T013
T014	T014	T014	T014
T015	T015	T015	T015
T016	T016	T016	T016
T017	T017	T017	T017

Gambar 1. Rancangan Sub Tes 4 Paket pada Soal *Anchor*

Selanjuntya, kunci masing-masing kunci paket 1, 2, 3, dan 4 (Sidoarjo) di tempatkan pada baris pertama sampai baris empat. Baris 5-504 menunjukkan jawaban peserta yang merespon paket 1, baris 505-1004 menujukkan jawaban peserta tes yang merespon paket 2, baris 1005-1504 menujukkan jawaban peserta tes yang merespon paket 3, dan baris 1505-2004 menujukkan jawaban peserta tes yang merespon paket 4 (Sidoarjo). Khusus soal *anchor* dikondisikan sesuai dengan desain Gambar 1 dimana soal tersebut direspon semua oleh kelompok 1, 2, dan 3 kecuali 4 (Sidoarjo). Berdasarkan aturan tersebut *syntax* fortran analisis pertama dibuat, diimplementasikan, dan di_ *RUN* dalam program Bilog_MG.

Program Bilog_MG mengeluarkan empat *output*, yakni: Ph1, Ph2, Ph3, dan kurva. *Output* Ph1 mengenai statistik tes klasik, sedangkan Ph2 menyangkut teori model pengukuran Rasch, dan Ph3 menyangkut karakteristik kemampuan peserta namun tidak ditampilkan pada penelitian ini.

Hasil keluaran *output Ph1* memuat informasi nilai statistik soal tiap kelompok soal yakni nilai korelasi soal terhadap tes. Korelasi tersebut menggunakan rumus Pearson dan korelasi biserial. Jika nilai kedua statistik pada suatu soal kurang dari -0,15 maka soal tersebut dihilangkan pada tahap *output Ph2* karena diasumsikan salah kunci. Selajutnya *output Ph2* memuat informasi parameter tingkat kesukaran soal dan nilai Kai_Kuadrat menginformasikan tingkat kecocokan data dengan model Rasch. Soal yang tidak memenuhi model selanjutnya dikeluarkan dari analisis berikutnya. Khusus soal *anchor* jumlahnya sepuluh soal jika ada yang tidak memenuhi sub tes salah satu dari kelompok tes maka soal tersebut didrop sedangkan yang lebih besar atau sama dengan 0,05 soal tersebut akan dianalisis lebih lanjut pada putaran kedua. Hal tersebut dilakukan secara terus menerus pada putaran berikutnya sampai semua soal memenuhi model dalam suatu putaran.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

Analisis data soal UN mata pelajaran matemaika terdiri 4 paket menggunakan desain *group wise* dengan dua putaran sebagai berikut:

Putaran Pertama

Berdasarkan *syntax* dan desain analisis yang telah dibuat kemudian program Bilog-MG di *RUN* maka diperoleh *output Ph1* pada masing-masing sub sebagai berikut. Sub 1 menunjukkan semua soal mempunyai nilai statistik Pearson dan korelasi biserial positif kecuali soal15 namun masih di atas -0,15 sehingga pada subtest 1 semua soal tidak ada yang dibuang dalam analisis berikutnya. Sub tes 2 menunjukkan semua soal mempunyai nilai statistik Pearson dan korelasi biserial positif sehingga tidak ada soal dibuang. Sub tes 3 menunjukkan semua soal mempunyai nilai statistik Pearson dan korelasi biserial masih di atas -0.15 kecuali soal ke 82 dibuang dan tidak diikutkan dalam analisis selanjutnya pada sub tes 3. Sub tes 4 menunjukkan semua soal mempunyai nilai statistik Pearson dan korelasi biserial positif kecuali soal ke 132 namun masih di atas -0.15 sehingga tidak ada yang dibuang dalam analisis selanjutnya.

Analisis menunjukkan bahwa tidak ada nilai statistik soal lebih kecil -0,15 kecuali soal ke-82 pada sub tes 3 sedangkan soal ke 132 pada sub tes 4 atau soal 12 pada paket Sidoarjo tidak dirop karena hanya mempunyai nilai korelasi biserial lebih kecil dari -0,15. Oleh karena itu, soal 2 pada paket 3 didrop sedangkan soal lain di analisis lebih lanjut dalam menentukan karakteristik soal berdasarkan model Rasch.

Output Ph2 pada setiap sub tes menunjukkan hasil sebagai berikut. Pertama, sub tes 1 paket 1 beberapa soal yang tidak memenuhi model yaitu soal 1, 2, 7, 10, 15, 22, 27, 32, 36, dan 38. Selanjutnya, sepuluh soal tersebut terdapat 4 soal *anchor* yakni 27, 32, 36, dan 38. Kedua, sub tes 2 paket 2 beberapa soal tidak memenuhi model yaitu soal 1, 3, 10, 14, 15, 32, dan 38. Selanjutnya, 7 soal soal tersebut terdapat 3 soal *anchor* yaitu 3, 32, dan 38. Ketiga, sub tes 3 paket 3 beberapa soal yang tidak memenuhi model yaitu soal 1, 2, 3, 11, 13, 27, 29, 35, dan 37. Selanjutnya, 9 soal tersebut terdapat 2 soal *anchor* yakni 3 dan 27. Keempat, sub tes 4 paket 2 beberapa soal yang tidak memenuhi model yaitu soal 1, 2, 9, 11, 13, 19, 20, 21, dan 23. Berdasarkan hal tersebut, jumlah soal yang didrop pada putaran pertama sebanyak 26 soal ditambah 5 soal *anchor*. Jadi pada putaran pertama terdapat 31 soal yang tidak memenuhi model Rasch sedangkan sisanya dianalisis lebih lanjut pada putaran kedua.

Putaran Kedua

Hasil analisis putaran pertama menjadi dasar dalam melakukan analisis putaran kedua dengan jumlah soal sebanyak 129 soal. Analisis putaran kedua dilakukan dengan memodifikasi *syntax* program Bilog_MG pada putaran pertama.

Hasil analisis menunjukkan pada sub tes 1 tidak terdapat nilai-nilai statistik soal lebih kecil dari -0,15. Sub tes 2 semua soal mempunyai nilai statistik Pearson dan korelasi biserial positif. Sub tes 3 menunjukkan semua soal mempunyai nilai statistik Pearson dan korelasi biserial positif kecuali soal ke 70 dan 89 namun masih di atas -0.15 sehingga subtest 3 pada putaran kedua semua

soal tidak ada yang dibuang, dan demikian halnya subset 4. Oleh karena itu, semua soal termasuk baik berdasarkan analisis teori tes klasik.

Output Ph2 memuat informasi parameter tingkat kesukaran soal dan nilai Kai_Kuadrat menginformasikan tingkat kecocokan data dengan model Rasch. Analisis menunjukkan semua soal memenuhi model setelah dilakukan analisis tahap kedua sehingga tidak perlu lagi dilakukan putaran ketiga.

Subtes 1 dimana soal 4, 5, 6, 8, 9, 11, 12, 13, 14, dan 8 memenuhi model Rasch dengan nilai *prob* tiap soal lebih besar dari 0,05. Artinya data respon peserta pada soal tersebut cocok dengan model. Subtes 2 dimana soal ke 42, 44, 45, 46, 47, 48, 49, 51, 52, dan 53 memenuhi model Rasch dengan nilai *prob* tiap soal lebih besar dari 0,05. Artinya data respon peserta pada soal tersebut cocok dengan model Rasch. Subtes 3 dimana soal ke 84, 85, 86, 87, 89, 90, 92, 94, dan 95 memenuhi model Rasch dengan nilai *prob* tiap soal lebih besar dari 0,05. Artinya data respon peserta pada soal tersebut cocok dengan model Rasch.

Tabel 1 Jumlah Soal Fit dan Tidak Fit Model

Tempat/Tahun	**Jumlah Soal**	**Analisis Tahap I**		**Analisis Tahap II**		**Bank Soal**
		Fit Model	**Tidak Fit Model**	**Fit Model**	**Tidak Fit Model**	
Sidoarjo 2008/2009	40	31	9	31	0	31
Sulawesi-Selatan 2009/2010	120-20	78	22	78	0	78
Jumlah Soal	140	109	31	109	0	109

Selanjutnya, subtes 4 dimana soal ke 123, 124, 125, 126, 127, 128, 130, 132, 134, dan 135 memenuhi model Rasch dengan nilai *prob* tiap soal lebih besar dari 0,05. Artinya data respon peserta pada soal tersebut cocok dengan model Rasch. Berdasarkan sub tes 1, 2, 3, dan 4 terdapat 109 soal yang memenuhi model Rasch sedangkan 31 soal yang tidak memenuhi model. Hasil analisis data respon menemukan bahwa jumlah soal yang memenuhi model sebanyak 109 soal seperti pada Tabel 1.

Pembahasan

Soal yang memenuhi model ditempatkan pada skala masing-masing dengan memperhatikan nilai rerata dan deviasi standar. Nilai rerata dan deviasi standar keempat sub tes masing-masing pada Tabel 2.

Tabel 2 Rerata dan Deviasi Standar Empat Sub Tes

Sub Tes	Rerata	Deviasi Standar
Pertama	0,000	1,381
Kedua	0,019	1,274
Ketiga	0,017	1,572
Keempat	0,007	1,243

Soal yang memenuhi model tersebut kemudian di tempatkan pada satu skala umum yang menjadi referensi dari yang lain. Hal tersebut dapat ditempuh dengan memperhatikan rerata dan simpangan baku dari setiap sub tes dari kelompok yang ada. Menurut Thomson dan Weiss (2011) soal tes di dalam bank soal perlu ditempatkan dalam satu skala umum jika di dalam bank *CAT* tersebut sudah mempunyai soal sebelumnya sehingga mempunyai metrik skala baru. Berdasarkan hal tersebut, dimana data penelitian berasal dari empat paket, Sumadi Suryabrata (1987) menyarankan pemilihan yang mana menjadi referensi sama saja yang penting kelompok pembakuan didefinisikan. Namun Tabel 2 menunjukkan rerata dan deviasi standar ke empat kelompok sub tes hampir sama maka yang dapat dijadikan patokan yang lebih mendekati kurva normal adalah sub kelompok pertama dengan rerata 0 dan deviasi standar 1,381. Selanjutnya, soal yang telah diketahui parameternya dikondisikan dengan skala umum tersebut pada saat penskoran skala yaitu skor *normal deviate*.

Hal tersebut menjadi informasi lebih lanjut bahwa keempat sub tes mengacu pada skala yang sama dimana rerata dan deviasi standar hampir sama. Artinya, soal tersebut dapat langsung dimasukkan dalam bank soal tanpa dipilih salah satu sub kelompok yang menjadi preferensi. Menurut Thomson dan Weiss (2011) hal tersebut dapat ditempuh jika bank *CAT* dalam keadaan kosong sehingga tidak perlu ada link ke skala umum. Soal yang memenuhi model dan keempat sub tes mengacu pada skala yang sama maka jumlah soal sebanyak 109 memenuhi syarat untuk bank soal *CAT* dimana soal tersebut dapat direspon peserta tes dimana masing-masing kemampuan tiap peserta tes dapat dibandingkan walaupun mempunyai jumlah dan karakteristik soal direspon tidak sama.

SIMPULAN DAN SARAN

Simpulan

Jumlah soal memenuhi model Rasch berdasarkan desain *group wise* sebanyak 109 soal. Soal tersebut dapat digunakan untuk bank *CAT* dengan memperhatikan metrik umum yang sesuai jika soal di dalam bank tersebut sudah ada soal sebelumnya. Jika soal belum tersedia di bank tersebut maka soal tersebut dapat langsung dimasukkan tanpa melakukan link dengan skala umum.

Saran

(1) Desain *group wise* dapat digunakan dalam pengembangan soal pada peletakan pada skala yang sama. Oleh karena itu, disarankan untuk menggunakan desain tersebut dalam pengembangan bank *CAT*. (2) Desain *group wise* merupakan salah satu desain yang ada pada program Bilog_MG sehingga terbuka untuk melakukan kajian yang lebih luas untuk membandingkan antar desain dengan program lain, misalnya RUMM, Winstep, atau Rascal. (3) Soal yang tidak memenuhi kriteria statistik, misalnya model Rasch selayaknya tidak langsung didrop namun soal tersebut dapat direvisi secara kualitatif dengan melibatkan *user*. Hal tersebut dapat dilakukan dengan pendekatan model pengukuran Rasch. Oleh karena itu, keterlibatan guru dalam merakit soal khususnya soal UN lebih didesentraliasikan

ke tingkat sekolah walaupun perlu ada pendampingan dari pakar sehingga karakteristik soal *link* dengan karakteristik siswa dan konteks sekolah.

DAFTAR PUSTAKA

Chuesathucchon, C., (2009). *Item Banking and Computerized Adaptive Testing with Rasch Measurement: An Example for Primary Mathematics in Thailand*. Dalam Waugh. R. (Ed). Application of Rasch Measurement in Education. New York, Nova Science Publishers, Inc.

Djemari Mardapi (1991). *Penilaian unjuk kerja sebagai usaha meningkatkan kemampuan sumber daya manusia*. Pidato Dies Natalis XXXII IKIP Yogyakarta, Yogyakarta.

Gushta, M. M. (2003). *Standard setting issues in computerized adaptive testing*. Artikel. Diambil pada tanggal 12 Juli 2009, dari http://www.education. ualberta.ca/ educ/psych/crame/files/GushtaCSSE2003.pdf.

Hambleton, R. K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Mills, C.N., & Steffen M. (2000). *The gre computer adaptive test: Operational issues*. Dalam Linden W.J., & Glas, G.A.W. (Eds.). Computerized adaptive testing: theory and practice. Bonston, PA: Kluwer.

Rukli. (2009). Prosedur rancangan tes adaptif menggunakan komputer berbasis web. Prosiding. Seminar Nasional V Jurusan Matematika FMIPA UNS, Semarang. ISBN: 978-602-8467-12-14.

_____,(2010). *Penerapan model logistik satu parameter pada Computerized Adaptive Testing ujian masuk Perguruan Tinggi*. Prosiding Seminar Nasional HEPI, Jakarta. ISBN 978-602-96343-0-3.

_____,(2011). *Studi Simulasi Menggunakan Fuzzy C-Means dalam Mengklasifikasi Konstruk Tes*. Jurnal Penelitian dan Evaluasi Pendidikan. ISSN 1410-4725. Tahun 15, Nomor 1.

_____,(2012). *Simulasi Pola Inisialisasi Kemampuan dengan Metode Penalaran Fuzzy Tsukamoto dalam Mengidentifikasi Cheating*. Prosiding Seminar Nasional PPS UNY, Yogyakarta. ISBN: 978-602-96172-8-3.

Rukli & Hartati, S. (2011). *Penerapan Sistem Pendukung Keputusan dalam Sistem Ujian Computerized Adaptive Testing*. Indonesian Journal of Computing and Cybernetic Systems, ISSN 1978-1520, Vol. 5 No. 3.

Sumadi Suryabrata, 1987, *Pengembangan Tes Hasil Belajar*. Rajawali Pers., Jakarta.

Thomson, N. A., & Weiss, D. J. (2011). A Framework for Development of Computerized Adaptive Tests. Pracrical Asssessment, Research & Evaluation, ISSN 1531-7714. Volume 16, Number 1, January.

Verschoor. A. J., & Straetmans, G. J. J. M. (2002). *Mathcat: A flexible testing system in mathematics education for adults*. Dalam Linden W.J., & Glas, G.A.W. (Eds.). Computerized adaptive testing: theory and practice: Bonston, PA: Kluwer