PENGARUH TES FORMATIF DAN METODE PEMBELAJARAN BAHASA INGGRIS TERHADAP KEMAMPUAN VERBAL LINGUISTIK

Citra Dewi
Program Studi Sastra Inggris Universitas Dehasen Kota Bengkulu
syifadea@yahoo.com

Abstrak

Penelitian ini bertujuan untuk mengetahui pengaruh tes formatif dan metode pembelajaran terhadap kemampuan verbal linguistik siswa. Penelitian dilaksanakan di SMP Islam Terpadu IQRA' dan SMP Islam Terpadu Al-QALAM di Bengkulu pada semester kedua tahun pelajaran 2011/2012 dengan sampel sebanyak 80 siswa kelas VII yang dipilih secara klaster dan random. Data dikumpulkan menggunakan instrumen kemampuan verbal linguistik dalam bentuk pilihan ganda. Data yang terkumpul selanjutnya dianalisis menggunakan ANAVA dua jalur. Hasil penelitian ini menunjukkan bahwa kemampuan verbal linguistik (1) siswa yang diberi tes uraian lebih tinggi daripada siswa yang diberi tes menjodohkan, (2) siswa yang diajar dengan metode komunikatif kemampuan verbal linguistiknya lebih rendah daripada siswa yang diajar dengan metode audio lingual, (3) terdapat interaksi yang signifikan antara tes formatif dengan metode pembelajaran terhadap kemampuan verbal linguistik siswa, (4) pada kelompok siswa yang diberi tes uraian, siswa yang diajar dengan metode pembelajaran komunikatif lebih tinggi dibandingkan siswa yang diberi tes menjodohkan, (5) pada kelompok siswa yang diberi tes menjodohkan, siswa yang diajar dengan metode pembelajaran komunikatif lebih rendah dibandingkan dengan siswa yang diajar dengan metode audio lingual, (6) pada kelompok siswa yang diajar dengan metode pembelajaran komunikatif, siswa yang diberi tes uraian lebih tinggi dibandingkan dengan siswa yang diajar dengan metode audio lingual, (7) pada kelompok siswa yang diajar dengan metode audio lingual, siswa yang diberi tes uraian lebih rendah dibandingkan dengan siswa yang diberi tes menjodohkan.

Kata kunci: *bentuk tes formatif, metode pembelajaran komunikatif, metode pembelajaran audio lingual, kemampuan verbal linguistik*

THE EFFECTS OF FORMATIVE TEST AND ENGLISH INSTRUCTIONAL METHOD ON THE STUDENTS' VERBAL LINGUISTIC COMPETENCE

Citra Dewi
Program Studi Sastra Inggris Universitas Dehasen Kota Bengkulu
syifadea@yahoo.com

Abstract

This study is aimed to find out the effect of formative test and English instructional method on the student's verbal linguistic competence. This research is done in the second semester of the 7th grades of SMP Islam Terpadu IQRA' and SMP Islam Terpadu Al-Qalam at province of Bengkulu in academic year 2011/2012. The research involves 80 students as the sample chosen through clustering. The data were collected through test instrument in the form of multiple choices. The data obtained is then analyzed by applying two ways ANAVA. The result shows: (1) The students given essay test have higher verbal linguistic competence than those given matching test, (2) The students taught with communicative method have lower verbal linguistic competence than those taught with audio lingual, (3) There is a significant interaction between formative test and English instructional method towards the students' verbal linguistic competence, (4) The verbal linguistic competence of students taught with communicative method and given essay test is higher than that of students taught with audio lingual method, (5) For students measured with matching test, the students taught with communicative method have lower verbal linguistic competence than those taught with the audio lingual method, (6) For the students taught with communicative method, the students measured with essay test have higher verbal linguistic competence than students measured with the matching test, and (7) for the students taught with audio lingual method, the students measured with essay test have lower verbal linguistic competence than the students measured with the matching test.

Keywords: *formative test model, communicative instructional method, audio lingualinstructional method, the students' verbal linguistic competence*

Pendahuluan

Setiap anak di dunia ini memiliki berbagai kecerdasan dalam tingkat dan indikator yang berbeda. Hal ini menunjukkan bahwa semua anak, pada hakikatnya adalah cerdas. Menurut May Lwin dkk (2008:5) kecerdasan atau kemampuan seseorang dapat ditingkatkan melalui latihan dan pajanan. Sementara itu Gardner (1993:12) mengemu-kakan berbagai macam jenis kecerdasan yang dimiliki oleh setiap manusia dalam teorinya yang dikenal dengan teori *Multiple Intelligences* yaitu; kecerdasan verbal-linguistik, kecerdasan logis-matematis, kecerdasan visual-spasial, kecerdasan musikal, kecerdasan kinestetik, kecerdasan interpersonal, dan kecerdasan naturalis. Windura (2008:63) menjelaskan bahwa kemampuan verbal linguistik siswa dapat ditingkatkan dan dikembangkan dengan berbagai cara, salah satunya adalah dengan metode pengajaran yang bervariasi. Pendapat ini juga didukung oleh Tientje (2010:92) yang menyatakan bahwa beberapa faktor yang mempengaruhi keberhasilan dalam mening-katkan dan mengembangkan kemampuan verball inguistik adalah guru, orang tua, materi pelajaran, metode pengajaran, dan lingkungan. Sementara itu Gardner (1993:30) menambah-kan keberhasilan yang dicapai manusia dalam domain yang berbeda akan menuntut wawa-san yang luas lagi tentang proses berfikir. Dalam upaya menggapai perspektif yang lebih luas diperlukan adanya suatu bentuk penilaian kontekstual yang memanfaatkan berbagai kemampuan dan jenis kecerdasan tertentu.

Menurut Nitko (1989:447) penilaian merupakan suatu proses untuk memperoleh informasi yang digunakan untuk membuat keputusan tentang siswa, kurikulum, dan pro-gram serta kebijakan pendidikan. Menurut Grondlund (1982:53), penilaian sebagai proses sistematik pengumpulan, penganalisaan dan penafsiran informasi untuk menentukan se-jauh mana siswa mencapai tujuan. Berdasar-kan uraian di atas maka penilaian merupakan proses pengumpulan berbagai data yang bisa memberikan gambaran perkem-bangan belajar siswa.

Dalam kaitannya dengan penelitian ini, jenis penilaian atau tes yang akan digunakan dalam penelitian ini adalah tes formatif. Menurut

Nurgiyantoro (2010:114) tes formatif merupakan tes dalam proses yang dimaksud-kan untuk mengukur keberhasilan peserta didik mencapai tujuan yang berkaitan dengan pokok bahasan yang baru saja diselesaikan.Woolfolk (2009:409) menjelaskan bahwa penilaian formatif adalah pengetesan yang terjadi sebelum atau selama pengajaran untuk membantu dalam merencanakan dan diagnosis serta membantu membentuk penga-jaran. Dengan demikian dapat disimpulkan bahwa tes formatif adalah tes yang digunakan untuk memantau kemajuan belajar siswa sela-ma proses belajar mengajar berlangsung dalam satu program tertentu.

Sehubungan dengan tujuan penelitian ini, aspek yang diperhatikan dalam tes formatif adalah aspek bentuk tes. Dilihat dari bentuknya, tes formatif dapat dibagi menjadi dua yaitu: (1) tes objektif dan (2) tes uraian. Menurut Kubiszyn dan Borich (2007:135) tes uraian *(essay test)* adalah tes yang digunakan untuk mengembangkan secara penuh kemam-puan siswa dalam memberikan tanggapan /jawaban atas pertanyaan yang diberikan. Selain ingatan dan penerapan akan suatu konsep, ketajaman analisis dan interpretasi sangat diperlukan dalam menjawab tes uraian. Pendapat ini diperkuat oleh Nurgiyantoro (2009:117) yang menyatakan bahwa tes uraian adalah suatu bentuk pertanyaan yang menuntut jawaban peserta didik dalam bentuk uraian dengan mempergunakan bahasa sendiri. Dari penjelasan tersebut di atas maka dapat disimpulkan bahwa tes uraian merupakan tes yang dimaksudkan untuk mengembangkan secara penuh respon siswa, mengharuskan siswa untuk mengingat kembali, meng-interpretasi, atau menganalisis daripada seke-dar mengidentifikasi, mempersiapkan alternatif jawaban seperti yang terjadi pada tes pilihan ganda.

Sementara itu menurut Popham (1995:110) tes menjodohkan *(matching test)* adalah bentuk soal yang terdiri dari dua kelompok pernyataan, kelompok pertama ditulis pada jalur sebelah kiri biasanya merupakan pernyataan soal atau pertanyaan sering juga disebut stimulus atau premis yang berupa kalimat atau frase. Sedangkan kelompok kedua biasanya disebut respon yang ditulis pada lajur sebelah kanan, biasanya merupakan pernyataan jawaban atau pernya-taan respon berupa kata, bilangan, gambar,

atau simbol. Sementara itu Anderson dkk (1995:50) mengemukakan bentuk soal menjodohkan adalah merupakan pertanyaan-pertanyaan dimana siswa diberikan daftar kemungkinan jawaban yang harus dipasangkan dengan daftar kata, frase, kalimat, atau paragraf. Dari pendapat-pendapat di atas, dapat disimpulkan bahwa bentuk tes menjodohkan adalah merupakan bentuk tes dimana siswa diberikan pertanyaan yang homogen dalam bentuk kolom yang berada disebelah kiri untuk dicocokkan dengan pilihan jawaban-jawaban yang terdapat di sebelah kanan.

Salah satu sisi yang sering menjadi perhatian masyarakat dalam pembelajaran bahasa termasuk bahasa asing adalah pendekatan yang digunakan oleh pengajar bahasa yang berpengaruh pada pemilihan metode dan strategi atau teknik pengajarannya. Berhasil tidaknya suatu pengajaran bahasa sering kali dinilai dari pendekatan yang dipilih dan dilakukan oleh guru karena dengan pendekatan inilah kita dapat menentukan isi dan cara pengajaran bahasa. Salah satu upaya untuk meningkatkan keefektifan pembelajaran adalah memilih atau menetapkan teknik pembelajaran yang sesuai dengan kondisi yang diprediksi dapat mempengaruhi kemampuan verbal linguistik siswa. Teknik yang sesuai dengan pembelajaran digunakan oleh guru pengajar bahasa asing di kelas.

Menurut Richard dan Rogers (1986:20) terdapat berbagai macam metode pembelajaran bahasa Inggris, namun dalam penelitian ini peneliti akan mengkaji tentang metode pembelajaran komunikatif dan metode pembelajaran audio lingual. Menurut Richards dan Rogers (1986:66) pembelajaran komunikatif tujuan pengajaran bahasa adalah untuk mengembangkan kemampuan komunikatif serta prosedur pengajaran keempat keterampilan berbahasa yaitu mendengarkan, menulis, membaca, dan berbicara yang mengakui interdependensi atau saling ketergantungan antara bahasa dan komunikasi. Sedangkan menurut Norland (2006:18) pembelajaran komunikatif merupakan kemampuan untuk menghasilkan makna ketika berkomunikasi dengan orang lain dalam mencapai target bahasa. Jadi pada prinsipnya teknik pembelajaran komunikatif adalah cara penyampaian pesan atau informasi kepada sasaran yang dapat disampaikan

melalui berbagai teknik yang dibantu oleh beberapa media berupa gambar, radio, televisi dan media lainnya sesuai dengan pesan yang disampaikan dan tujuan yang ingin dicapai.

Beberapa prosedur dalam pembelajaran komunikatif dikemukakan oleh Tarigan (2009:242), yaitu: (1) menyajikan suatu dialog atau beberapa dialog mini, di dahului oleh suatu motivasi yang berkaitan dengan situasi-situasi dialog terhadap pengalaman-pengalaman masyarakat yang mungkin diperoleh siswa, (2) praktik lisan setiap ucapan dari dialog yang disajikan pada saat itu, (3) pertanyaan-pertanyaan dan jawaban-jawaban tetap berdasarkan topik-topik dialog dan situasi itu sendiri, (4) telaah dan kajian salah satu ekspresi komunikatif dasar dalam dialog itu atau salah satu struktur yang menunjukkan fungsi tersebut, dan (5) memberi contoh tugas pekerjaan rumah secara tertulis.

Sementara itu Norland (2006:45) menjelaskan bahwa metode audio lingual sebagai metode yang paling efektif dan efisien dalam pembelajaran bahasa asing. Kegiatan-kegiatan pembelajaran berdasarkan metode audio lingual adalah; (1) *repetition* siswa mengulang sebuah ujaran dengan keras secepat mungkin setelah dia mendengarkan ujaran tersebut dan siswa melakukannya tanpa melihat teks,(2) *inflection* satu kata yang didengarkan namun dalam pengulangannya diucapkan dalam bentuk yang lain,(3) *replacement* satu kata dalam sebuah ujaran diubah dengan kata ganti ujaran tersebut ,(4) *restatement* siswa memprasekan kembali ujaran tersebut kepada orang lain dalam bentuk yang berbeda sesuai dengan perintahnya,(5) *completion* siswa mendengarkan sebuah ujaran/kalimat yang utuh kecuali ada satu kata yang dihilangkan kemudian melengkapi kata yang dihilangkan sambil mengulang ujaran tersebut,(6) *transposition* perubahan susunan kata sangat diperlukan ketika adanya penambahan kata baru,(7) *expansion* ketika kata tersebut telah ditambahkan pastikan bahwa kata tersebut berada pada letak yang benar, (8) *contraction* satu kata dapat mewakili sebuah prase atau klausa, (9) *transformation* sebuah kalimat diubah menjadi bentuk negatif atau introgatif, (10) *integration* dua ujaran terpisah digabungkan menjadi satu kesatuan, (11) *rejoinders* siswa membuat jawaban atau balasan yang tepat dalam menanggapi sebuah ujaran, dan (12) *restoration* siswa diberikan kata-

kata yang telah dipisah-pisahkan dari sebuah kalimat yang utuh yang pada dasarnya masih mengandung makna dan siswa menggunakan kata-kata tersebut dengan perubahan dan penambahan kalimat yang diminimumkan untuk menjaga bentuk asalnya.

Jadi dari uraian di atas dapat disimpulkan bahwa metode pembelajaran audio lingual adalah merupakan metode pembelajaran yang mendahulukan keteram-pilan mendengarkan dan keterampilan berbicara dengan teknik *drill* atau latihan terhadap struktur-struktur kebahasaan dengan harapan siswa akan menjadikan bahasa sebagai sebuah kebiasaan.

Di dalam dunia pendidikan, salah satu sarana yang dapat digunakan untuk meningkatkan dan mengembangkan kemam-puan verbal linguistik siswa adalah dengan belajar bahasa. Karena dalam pembelajaran bahasa mengandung komponen – komponen dalam linguistik, seperti: tata bahasa, semantik, pragmatik, fonologi, dan juga sintaksis yang dituangkan dalam empat keterampilan berbahasa yaitu mendengar, membaca, berbicara, dan menulis. Namun dalam kenyataannya dalam proses pembelajaran disekolah hasil belajar selalu menjadi tolak ukur keberhasilan siswa-siswanya. Tes kecerdasan inteligensi saja masih jarang dilakukan di sekolah-sekolah untuk mengetahui tingkat inteligensi siswanya, apalagi untuk menguji tes kemampuan verbal linguistik siswa.

Dari penjelasan di atas maka dapat disimpulkan bahwa upaya yang dilakukan untuk meningkatkan kemampuan verbal linguistik adalah dengan melakukan berbagai metode pengajaran bahasa yang bervariasi dan melakukan berbagai bentuk penilaian yang kontekstual yang dapat mengukur kemampuan siswa sehingga metode pengajaran dan bentuk penilaian merupakan komponen-komponen yang mempengaruhi kemampuan verbal linguistik. Untuk itulah menjadi penting untuk diteliti sejauh mana pengaruh bentuk penilaian dan metode pembelajaran bahasa Inggris terhadap kecerdasan linguistik siswa.

Beberapa penelitian yang relevan yang sudah dilakukan antara lain; (1) Sunandar yang melakukan penelitian di SMU di Jakarta menyimpulkan hasil belajar Matematika siswa yang diberi tes formatif esai lebih tinggi daripada hasil belajar Matematika siswa yang diberi tes formatif pilihan

ganda, (2) Sinembela melakukan penelitian mengenai tes esai dan tes pilihan ganda yang dilakukan di SD di Cianjur. Hasil penelitian tersebut adalah sistem pengujian dengan tes esai pemetaan konsep memberikan pengaruh yang lebih baik terhadap hasil belajar IPA dibandingkan dengan tes pilihan ganda.

Metode Penelitian

Penelitian dilaksanakan di SMP Islam Terpadu IQRA Kota Bengkulu dan SMP Islam Terpadu Al-QALAM Kabupaten Bengkulu Selatan Provinsi Bengkulu, dimulai dari bulan Februari sampai dengan April 2012. Penelitian ini menggunakan metode eksperimen ANAVA dua jalur. Variabel bebas terdiri dari satu variabel aktif dan satu variabel atribut. Variabel aktif adalah penilaian kelas yang terdiri dari tes uraian(A_1) dan tes menjodohkan (A_2). Sedangkan variabel atributnya adalah metode pembelajaran Bahasa Inggris yang terdiri dari metode pembelajaran komunikatif (B_1) dan metode pembelajaran audio lingual (B_2). Sedangkan variabel terikat adalah kemampuan verbal linguistik. Desain yang digunakan dalam penelitian ini adalah "Desain Faktorial 2 x 2" dengan ANOVA (*Analysis of Variance*) untuk mengetahui perbedaan varians antara kelompok perlakuan. Desainnya dapat dilihat dalam matriks sebagai berikut.

Tabel 1. Desain Eksperimen Faktorial 2 x 2

Tes formatif / Metode	Uraian (A_1)	Menjodohkan (A_2)
Metode Pembelajaran Bahasa Inggris (B)		
Metode pembelajaran komunikatif (B_1)	A_1B_1	A_1B_2
Metode pembelajaran audio lingual (B_2)	A_1B_2	A_2B_2

Keterangan:

A_1	Kelompok siswa yang diberi tes uraian	A_1B_1	Kelompok siswa yang diberi tes uraian yang diajarkan dengan metode pembelajaran komunikatif
A_2	Kelompok siswa yang diberi tes menjodohkan	A_2B_1	Kelompok siswa yang diberi tes menjodohkan yang diajarkan dengan Metode pembelajaran komunikatif
B_1	Kelompok siswa yang diajar dengan metode pembelajarn komunikatif	A_1B_2	Kelompok siswa yang diberi tes uraian yang diajarkan dengan metode pembelajaran audio lingual
B_2	Kelompok siswa yang diajar dengan metode pembelajaran audio lingual	A_2B_2	Kelompok siswa yang diberi tes menjodohkan yang diajarkan dengan metode pembelajaran audio lingual

Populasi yang digunakan dalam penelitian ini adalah seluruh siswa kelas VII di SMP Islam terpadu IQRA Kota Bengkulu dan SMP Islam Terpadu AL-QALAM Kabupaten Bengkulu Selatan provinsi Bengkulu tahun ajaran 2010-2011 yang berjumlah 300 orang siswa yang tersebar di sembilan kelas. Dari dua sekolah tersebut akan dipilih secara random satu sekolah sebagai kelas eksperimen dan satu sekolah sebagai kelas kontrol. Sampel penelitian ini ditentukan dengan teknik random sampling, yaitu memilih secara acak dari kelas-kelas paralel di setiap sekolahyang akan dipilih dua kelas. Pada sekolah yang menjadi kelompok eksperimen akan dipilih secara random cluster satu kelas terpilih sebagai kelompok siswa yang diberi tes uraian dengan menggunakan metode pembelajaran komunikatif, satu kelas terpilih sebagai kelompok siswa yang diberi tes menjodohkan dengan menggunakan metode pembelajaran komunikatif. Sementara itu satu sekolah yang menjadi kelompok kontrol juga akan

dipilih secara random satu kelas sebagai kelompok siswa yang diberi tes uraian dengan menggunakan metode audio lingual dan satu kelas sebagai kelompok siswa yang diberi tes menjodohkan dengan menggunakan metode pembelajaran audio lingual.

Dalam rancangan tersebut, seluruh sampel adalah 80 siswa yang masing-masing terdiri atas 20 siswa sebagai kelompok eksperimen yang diberi tes uraian dengan menggunakan metode pembelajaran komunikatif, 20 siswa sebagai kelompok eksperimen yang diberi tes menjodohkan dengan menggunakan metode pembelajaran komunikatif, 20 siswa sebagai kelompok kontrol yang diberi tes uraian dengan menggunakan metode pembelajaran audio lingual serta 20 siswa sebagai kelompok kontrol yang diberi tes menjodohkan dengan menggunakan metode pembelajaran audio lingual.

Hasil Penelitian dan Pembahasan

Berikut disajikan skor kemampuan verbal linguistik siswa berdasarkan pengam-bilan data di lapangan dengan instrumen tes kemampuan verbal linguistik.

Tabel 2. Skor Kemampuan Verbal Linguistik Siswa

	A_1				A_2				n
B_1	39	29	34	36	29	29	27	27	40
	28	33	32	34	28	27	25	30	
	31	31	37	36	30	28	22	25	
	35	37	39	35	29	31	30	21	
	32	35	33	34	31	29	30	33	
B_2	34	30	29	33	29	29	30	34	40
	27	29	24	31	27	31	35	30	
	29	33	27	29	31	32	31	32	
	34	31	29	29	33	34	33	31	
	27	30	31	31	37	26	35	32	
N	40				40				80

Berdasarkan hasil eksplorasi data tersebut, dilakukan uji persyaratan analisis untuk mengetahui tingkat normalitas dan homogenitas data penelitian. Uji normalitas data menggunakan metode Lilliefors, dengan syarat: H_0 ditolak apabila $L_{hitung} > L_{tabel}$ pada taraf signifikansi $\alpha = 0{,}05$. Perumusan hipotesisnya:

H_0 = Sampel berasal dari populasi berdistribusi normal

H_1 = Sampel berasal dari populasi berdistribusi tidak normal.

Tabel 3. Rekapitulasi Hasil Uji Normalitas Sampel dengan Uji Lilliefors

Variabel	**Jumlah**	L_0	L_t	**Kesimpulan**
A_1	40	0,1056	0,1400	Normal
A_2	40	0,0869	0,1400	Normal
B_1	40	0,0991	0,1400	Normal
B_2	40	0,1218	0,1400	Normal
A_1B_1	20	0,0891	0,1900	Normal
A_1B_2	20	0,1430	0,1900	Normal
A_2B_1	20	0,1392	0,1900	Normal
A_2B_2	20	0,0891	0,1900	Normal

Keterangan:

L_0	Nilai hitung Lilliefors	B_2	Kelompok siswa dengan metode pembelajaran audio lingual
L_t	Nilai kritis Lilliefors	A_1B_1	Kelompok siswa yang diberi tes uraian dan Diajardengan metode pembelajaran komunikatif

A_1	Kelompok siswa yang diberi tes uraian	A_1B_2	Kelompok siswa yang diberi tes uraiandan diajar denganmetode pembelajaran audio lingual
A_2	Kelompok siswa yang diberi tes menjodohkan	A_2B_1	Kelompok siswa yang diberi tes menjodohkandan diajar denganmetode pembelajaran komunikatif
B_1	Kelompok siswa dengan metode pembelajaran komunikatif	A_2B_2	Kelompok siswa yang diberi tes menjodohkandan diajar denganmetode pembelajaran audio lingual

Hasil perhitungan uji Lilliefors menunjukkan bahwa nilai L_{hitung} untuk semua kelompok data lebih kecil dari pada L_{tabel} pada taraf signifikansi α = 0,05, dengan demikian H_0 diterima, artinya data sampel berasal dari populasi yang berdistribusi normal.

Tabel 4. Uji Homogenitas Varians Data Kemampuan Verbal Linguistik pada Dua kelompok Perlakuan yaitu Kelompok Tes uraian (A_1) dan kelompok Tes menjodohkan (A_2)

Sampel	db	s	s^2	$db(s^2)$	log s^2	$db.logs^2$
A1	39	3,317	11,0025	429,0971	1,041	40,599
A2	39	3,278	10,745284	419,0661	1,031	40,209
Jumlah	78		21,748	848,1603		80,808

Hasil pengujian homogenitas kelompok A_1 dan A_2 dengan uji Barlett menunjukkan bahwa χ^2_{hitung} = 0,0688, sedangkan χ^2_{tabel} pada taraf signifikansi α = 0,05 adalah 3,84. Hal ini menunjukkan bahwa $\chi^2_{hitung} < \chi^2_{tabel}$ maka H_0 diterima, artinya kedua kelompok data

mempunyai varians yang sama atau skor kemampuan verbal linguistik dari kedua kelompok data adalah homogen.

Tabel 5. Uji Homogenitas Varians Data Kemampuan Verbal Linguistik pada Dua kelompok Perlakuan yaitu pada kelompok Perlakuan Metode Pembelajaran Komunikatif (B_1) dan Kelompok Metode Pembelajaran Audio Lingual (B_2)

Sampel	db	s	s^2	db(s^2)	log s^2	db.logs^2
B_1	39	4,03	16,0278	625,119	1,2049	46,991
B_2	39	2,75	7,5378	293,974	0,8872	36,2108
Jumlah	78			919,093		83,2019

Hasil pengujian homogenitas kelompok B_1 dan B_2 dengan uji Barlett menunjukkan bahwa χ^2_{hitung} = 0,82, sedangkan χ^2_{tabel} pada taraf signifikansi α = 0,05 adalah 3,84. Hal ini menunjukkan bahwa $\chi^2_{hitung} < \chi^2_{tabel}$ maka H_0 diterima, artinya kedua kelompok data mempunyai varians yang sama atau skor kemampuan verbal linguistik dari kedua kelompok data adalah homogen.

Tabel 6. Uji homogenitas Keempat Kelompok Sel Rancangan Eksperimen yaitu A_1B_1, A_1B_2, A_2B_1, dan A_2B_2

Sampel	db	s	s^2	db(s^2)	log s^2	db.(logs^2)
A1B1	19	2,777	7,7132	146,5500	0,887	16,857
A1B2	19	2,540	6,45	122,5500	0,810	15,382
A2B1	19	2,864	8,2	155,8000	0,914	17,362
A2B2	19	2,722	7,411	140,8000	0,870	16,527
Jumlah	76		29,774	565,700		66,129

Hasil pengujian homogenitas kelompok A_1B_1, A_1B_2, A_2B_1, dan A_2B_2 dengan uji Barlett menunjukkan bahwa $\chi^2_{hitung} = 0{,}289$, sedangkan χ^2_{tabel} pada taraf signifikansi $\alpha = 0{,}05$ adalah 7,81. Hal ini menunjukkan bahwa $\chi^2_{hitung} < \chi^2_{tabel}$ maka H_0 diterima. Dengan demikian keempat kelompok data mempunyai varians yang sama atau skor kemampuan verbal linguistik dari keempat kelompok data adalah homogen.

Pengujian hipotesis dalam penelitian ini menggunakan analisis varian (ANAVA) dua jalur yang dilanjutkan dengan uji-t Dunnet. Untuk menguji hipotesis nol (H_o), hipotesis statistika dirumuskan sebagai berikut.

1. $H_o : \mu A_1 \leq \mu A_2$
   $H_1 : \mu A_1 > \mu A_2$
2. $H_o : \mu B_1 \leq \mu B_2$
   $H_1 : \mu B_1 > \mu B_2$
3. H_o : Interaksi A x B = 0
   H_1 : Interaksi A x B ≠ 0
4. $H_o : \mu A_1B_1 \leq \mu A_1B_2$
   $H_1 : \mu A_1B_1 > \mu A_1B_2$
5. $H_o : \mu A_2B_1 \geq \mu A_2B_2$
   $H_1 : \mu A_2B_1 < \mu A_2B_2$
6. $H_o : \mu_2 A_1B_1 \leq \mu A_2B_1$
   $H_1 : \mu A_1B_1 > \mu A_2B_1$
7. $H_o : \mu A_1B_2 \geq \mu A_1B_2$
   $H_1 : \mu A_1B_2 < \mu A_2B_2$

Keterangan:

μA_1 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberikan tes uraian.

μA_2 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberikan tes menjodohkan.

μB_1 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diajar dengan metode pembelajaran komunikatif.

μB_2 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diajar dengan metode audio lingual.

μA_1B_1 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberikan tes uraian dan yang diajar dengan metode pembelajaran komunikatif.

μA_1B_2 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberi tes uraian dan diajar dengan metode audio lingual.

μA_2B_1 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberikan tes menjodohkan dan diajar dengan metode pembelajaran komunikatif.

μA_2B_2 : Rerata skor kemampuan verbal linguistik kelompok siswa yang diberi tes menjodohkan dan diajar dengan metode audio lingual.

Ringkasan hasil perhitungan analisis data Uji ANAVA dapat dilihat pada tabel 7 berikut.

Tabel 7. Analisis Varian Keseluruhan

Sumber Varian	JK	db	RJK	F_0	F_{tabel} $\alpha = 0{,}05$
Antar A	80	1	80	10,75 *	3,96
Antar B	1,25	1	1,25	$0{,}17^{ns}$	
Interaksi AB	281,25	1	281,25	37,79**	
Dalam (D)	565,7	76	7,4434		
Total (T)		79			

Keterangan:

* = Signifikan pada $\alpha = 0{,}05$

** = Sangat signifikan

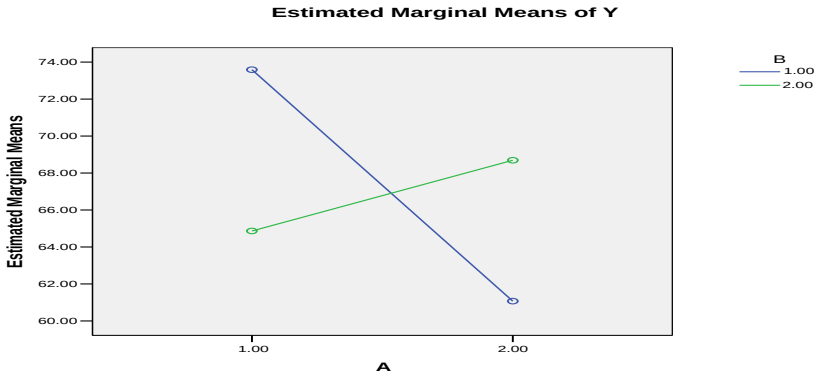
ns = Non signifikan

db = Derajat kebebasan
JK = Jumlah kuadrat
RJK = Rerata jumlah kuadrat

Berdasarkan hasil perhitungan ANAVA dua jalur di atas, dapat dijelaskan beberapa hal sebagai berikut: Berdasarkan hasil analisis data dengan menggunakan ANAVA dua jalur di atas, diperoleh nilai F_{hitung} = 10,75, sedangkan nilai F_{tabel} pada taraf signifikansi $\alpha = 0,05$ adalah 3,96. Jadi nilai $F_{hitung} > F_{tabel}$, dengan demikian H_0 ditolak sebagai konsekuensinya maka H_1 diterima. Hal ini berarti bahwa terdapat perbedaan rata-rata skor kemampuan verbal linguistik antara kelompok siswa yang diberi tes uraian dengan kelompok siswa yang diberi tes menjodohkan. Berdasarkan hasil analisis data dengan menggunakan ANAVA dua jalur di atas, diperoleh nilai F_{hitung} = 0,17, sedangkan nilai F_{tabel} pada taraf signifikansi $\alpha = 0,05$ adalah 3,96. Jadi nilai $F_{hitung} < F_{tabel}$, dengan demikian H_0 diterima sebagai konsekuensinya maka H_1 ditolak. Hal ini berarti bahwa tidak terdapat perbedaan rata-rata skor kemampuan verbal linguistik antara kelompok siswa yang diajar dengan metode pembelajaran komunikatif dengan kelompok siswa yang diajar dengan metode audio lingual. Berdasarkan hasil analisis data dengan menggunakan ANAVA dua jalur di atas, diperoleh nilai F_{hitung} = 37,79, sedangkan nilai F_{tabel} pada taraf signifikansi $\alpha = 0,05$ adalah 3,96. Jadi nilai $F_{hitung} > F_{tabel}$, dengan demikian H_0 ditolak sebagai konsekuensinya maka H_1 diterima. Dengan demikian dapat disimpulkan bahwa terdapat pengaruh interaksi antara bentuk tes formatif dan metode pembelajaran terhadap kemampuan verbal linguistik. Bentuk interaksi antara bentuk tes formatif dan metode pembelajaran terhadap kemampuan verbal linguistik dapat dilihat pada gambar1.

Perpotongan dua garis di atas menunjukkan bahwa terdapat interaksi di antara dua variabel bebas terhadap variabel terikat. Dengan kata lain interaksi dapat diartikan bahwa pengaruh bentuk tes formatif terhadap kemampuan verbal linguistik siswa tergantung pada metode pembelajaran yang diberikan, atau sebaliknya. Sebagai konsekuensi terjadinya interaksi

secara signifikan, maka perlu dilakukan uji rata-rata (ANAVA dua jalan) yang dilanjutkan dengan uji t-Dunnet.



Gambar 1. Interaksi Antara Tes Formatif dengan Metode Pembelajaran

Berdasarkan hasil pengujian hipotesis yang telah dilakukan dalam penelitian ini, ternyata efek utama *(main effect)* dari bentuk tes formatif yang digunakan berpengaruh terhadap kemampuan verbal linguistik siswa. Hasil analisa data dengan menggunakan ANAVA dua jalur, diperoleh nilai F_{hitung} = 10,75, sedangkan nilai F_{tabel} pada taraf signifikansi $\alpha = 0{,}05$ adalah 3,96. Jadi nilai $F_{hitung} > F_{tabel}$ dengan demikian H_0 ditolak sebagai konsekuensinya maka H_1 diterima. Hal ini berarti bahwa terdapat perbedaan rata-rata skor kemampuan verbal linguistik antara kelompok siswa yang diberi tes uraian dengan kelompok siswa yang diberi tes menjodohkan.

Hasil analisis data dengan menggunakan uji t-Dunnet diperoleh nilai t_{hitung} sebesar 14,7, sedangkan nilai t_{tabel} pada taraf signifikansi $\alpha = 0{,}05$ adalah 1,69. Jadi nilai $t_{hitung} > t_{tabel}$, dengan demikian H_0 ditolak maka sebagai konsekuensi H_1 diterima. Hal ini berarti rata-rata skor kemampuan verbal linguistik siswa yang diberi tes uraian lebih tinggi daripada rata-rata skor kemampuan verbal linguistik siswa yang diberi tes menjodohkan.

Data yang diperoleh dari tes kemampuan verbal linguistik menunjukkan bahwa skor rata-rata kemampuan verbal linguistik siswa yang

diberi tes uraian (A_1) = 31,85, sedangkan skor rata-rata kemampuan verbal linguistik siswa yang diberi tes menjodohkan (A_2) = 29,85. Jika dibandingkan rata-rata skor kemampuan verbal linguistik siswa yang diberi tes uraian dengan siswa yang diberi tes menjodohkan, terbukti bahwa rata-rata skor kemampuan verbal linguistik siswa yang diberi tes uraian lebih tinggi daripada kemampuan verbal linguistik siswa yang diberi tes menjodohkan. Hal ini menunjukkan kemampuan verbal linguistik yang menggunakan bentuk tes formatif uraian terbukti lebih tinggi daripada menggunakan bentuk tes formatif menjodohkan. Ini berarti bahwa tes uraian yang digunakan dalam pembelajaran bahasa Inggris dapat meningkatkan kemampuan verbal linguistik siswa.

Secara teoretik tes uraian adalah bentuk tes yang membutuhkan satu struktur jawaban dari seseorang siswa dalam menguraikan, menjelaskan, membandingkan serta memberi alasan yang sesuai dengan tuntutan pertanyaan dengan menggunakan bahasanya sendiri dengan menggunakan ketajaman analisis dan interpretasi terhadap pertanyaan yang diajukan. Hal tersebut mengandung pengertian bahwa setiap siswa yang diberi tes uraian secara individu maupun kelompok dapat menggunakan daya imajinasinya secara utuh sehingga ide-ide yang ada dalam pikiran siswa dapat digali dengan secara maksimal. Melalui bentuk tes formatif uraian, siswa termotivasi dan kreatif dalam menemukan dan mengembangkan pengetahuan secara ilmiah sesuai dengan perkembangan jaman. Sementara bentuk soal menjodohkan sangat baik untuk mengukur kemampuan peserta didik yang sangat sederhana dan kemampuan mengidentifikasikan kemampuan menghubungkan antara dua hal. Di dalam tes menjodohkan siswa hanya menentukan pilihan jawaban yang tersedia, sehingga siswa tidak dapat mengembangkan imajinasi, kosa kata yang dimiliki dalam kalimat-kalimat yang dihasilkan. Oleh karena itu, dapat disimpulkan bahwa kemampuan verbal linguistik siswa yang diberi tes uraian lebih tinggi daripada kemampuan verbal linguistik yang diberi tes menjodohkan.

Pembahasan selanjutnya adalah bagi siswa yang diajar dengan menggunakan metode pembelajaran komunikatif dan siswa yang diajar dengan menggunakan metode pembelajaran audio lingual tidak terdapat

perbedaan yang signifikan dalam pencapaian kemampuan verbal linguistik. Hasil analisa data menunjukkan F_{hitung} sebesar 0,17 sedangkan nilai F_{tabel} = 3,69. Karena $F_{hitung} < F_{tabel}$ maka H_0 diterima. Dilihat dari perolehan skor siswa yang diajar dengan metode pembelajaran komunikatif dengan metode pembelajaran audio lingual. Hasil penelitian menunjukkan bahwa kelompok siswa yang diajar dengan metode pembelajaran komunikatif memiliki rata-rata skor sebesar 30,98 yang lebih tinggi daripada kelompok siswa yang diajar dengan metode pembelajaran audio lingual yaitu sebesar 30,73.

Pada prinsipnya metode pembelajaran komunikatif adalah cara penyampaian pesan atau informasi kepada sasaran yang dapat disampaikan melalui berbagai teknik yang dibantu oleh beberapa media berupa gambar, radio, televisi dan media lainnya sesuai dengan pesan yang disampaikan dan tujuan yang ingin dicapai. Dalam metode ini siswa banyak diberikan ilustrasi gambar dan juga simulasi-simulasi dalam proses pembelajarannya, dimana siswa dibentuk dalam kelompok-kelompok kecil dalam melaksanakan simulasi. Pada kenyataannya di lapangan waktu yang telah di buat tidak sesuai dengan yang telah direncanakan sebelumnya, hal ini disebabkan oleh bebarapa faktor seperti: pembentukan kelompok yang membutuhkan waktu yang lama, siswa lebih fokus dengan ilustrasi gambar yang diberikan bukan pada konten materi yang diberikan, di dalam pelaksanaan pembuatan simulasi di dalam kelompok membutuhkan waktu yang lama sehingga target-target pengajaran selanjutnya tidak tercapai.

Sementara pada kelas yang diajar dengan metode pembelajaran audio lingual waktu yang ditargetkan cenderung lebih sesuai dengan yang telah direncanakan. Pada pelaksanaannya di lapangan, banyaknya latihan/*drill* yang diberikan membuat siswa memudahkan untuk mengingat materi yang diberikan. Dari uraian di atas, maka disimpulkan bahwa kemampuan verbal linguistik pada kelompok yang diajar dengan metode pembelajaran komunikatif lebih rendah daripada kemampuan verbal linguistik pada kelompok siswa yang diajar dengan metode pembelajarn audio lingual.

Pada pembahasan berikutnya analisa data dengan menggunakan ANAVA dua jalur untuk melihat interaksi antara bentuk tes formatif dan

metode pembelajaran bahasa Inggris, diperoleh nilai F_{hitung} = 37,79, sedangkan nilai F_{tabel} pada taraf signifikansi α = 0,05 adalah 3,69. Ini berarti terdapat pengaruh interaksi antara bentuk tes formatif dan metode pembelajaran untuk mempengaruhi skor kemampuan verbal linguistik. Interaksi mengandung pengertian bahwa adanya kerjasama antara dua atau lebih variabel bebas dalam mempengaruhi variabel terikat. Metode pembelajaran bahasa Inggris merupakan suatu proses atau langkah dari penjabaran sebuah pendekatan yang dikuasai oleh guru dalam mengajar dan menyajikan bahan pelajaran di kelas dengan tujuan agar proses belajar mengajar lebih efektif sehingga pelajaran tersebut dapat dipahami dan digunakan oleh siswa secara baik.

Dalam pembahasan selanjutnya, diuraikan hasil analisis data dengan menggunakan uji t-Dunnet untuk melihat perbedaan kemampuan verbal linguistik antara siswa yang diajar dengan metode pembelajaran komunikatif dengan siswa yang diajar dengan metode pembelajaran audio lingual pada kelompok siswa yang diberi tes uraian diperoleh nilai t_{hitung} sebesar 14,81, sedangkan nilai t_{tabel} pada taraf signifikansi α = 0,05 adalah 1,69. Jadi nilai $t_{hitung} > t_{tabel}$, dengan demikian H_0 ditolak maka sebagai konsekuensi H_1 diterima. Dengan demikian maka dapat disimpulkan bahwa rata-rata skor kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif lebih tinggi daripada rata-rata skor kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran audio lingual pada kelompok siswa yang diberi tes uraian.

Pembahasan selanjutnya diuraikan hasil analisis data dengan menggunakan uji t-Dunnet untuk melihat perbedaan kemampuan verbal linguistik antara siswa yang diajar dengan menggunakan metode pembelajaran komunikatif dengan siswa yang diajar dengan menggunakan metode pembelajaran audio lingual pada kelompok siswa yang diberi tes menjodohkan diperoleh nilai t_{hitung} sebesar 12,50, sedangkan nilai t_{tabel} pada taraf signifikansi α = 0,05 adalah 1,69. Jadi nilai $t_{hitung} > t_{tabel}$, dengan demikian H_0 ditolak maka sebagai konsekuensi H_1 diterima. Hal ini berarti rata-rata skor kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif lebih tinggi daripada rata-rata skor

kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran audio lingual pada kelompok siswa yang diberi tes menjodohkan. Hal ini dapat dilihat pula pada kelompok siswa yang diberi tes menjodohkan, skor kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif memperoleh skor rata-rata 28,1 lebih rendah daripada kelompok siswa yang diajar dengan metode pembelajaran audiolingual sebesar 31,6. Oleh karena itu tes menjodohkan ini kurang cocok dipergunakan untuk mengevaluasi penguasaan kemampuan verbal linguistik pada kelompok siswa yang diberi metode pembelajaran komunikatif karena pembelajaran komunikatif mengembangkan daya kreativitas siswa dalam belajar dan peran aktif siswa sangat memiliki peranan yang penting. Sehingga dapat disimpulkan bahwa untuk kelompok siswa yang diberi tes menjodohkan kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif lebih tinggi daripada kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran audio lingual.

Hasil analisis data dengan menggunakan uji t-Dunnet untuk menguji perbedaan kemampuan verbal linguistik antara siswa yang diberi tes uraian dengan siswa yang diberi tes menjodohkan pada kelompok siswa yang diajar dengan metode pembelajarn komunikatif diperoleh nilai t_{hitung} sebesar 19,83, sedangkan nilai t_{tabel} pada taraf signifikansi $\alpha = 0,05$ adalah 1,69. Jadi nilai $t_{hitung} > t_{tabel}$, dengan demikian H_0 ditolak maka sebagai konsekuensi H_1 diterima. Hal ini berarti rata-rata skor kemampuan verbal linguistik siswa yang diberi tes uraian lebih tinggi daripada rata-rata skor kemampuan verbal linguistik siswa yang diberi tes menjodohkan pada kelompok siswa yang diajar dengan metode pembelajaran komunikatif. Dengan metode ini, jika siswa diberikan tes uraian menulis paragraf tentang suatu tema, maka akan dapat dilihat sejauh mana perubahan prilaku yang ada dalam diri siswa yang dituangkannya kedalam tulisan. Sementara itu untuk metode pembelajaran komunikatif ini apabila siswa diberikan tes dalam bentuk menjodohkan maka akan sulit terlihat perubahan prilaku yang ada dalam diri siswa, karena siswa hanya bersifat memilih jawaban yang tertera denagn pernyataan yang ada sehingga bersifat hafalan karena yang merupakan ciri

khusus bentuk soal menjodohkan adalah terbatas mengukur kemampuan mengidentifikasi informasi berdasarkan hubungan yang sederhana. Kalaupun hendak digunakan untuk mengukur kemampuan yang lebih tinggi, paling tidak dapat digunakan untuk kemampuan menghubungkan antara dua hal yang homogen antara premis dan respon. Oleh karena itu, dapat disimpulkan bahwa untuk kelompok siswa yang diajar dengan metode pembelajaran komunikatif kemampuan verbal linguistik siswa yng diberi tes uraian lebih tinggi daripada kemampuan verbal linguistik siswa yang diberi tes menjodohkan.

Sebagai pembahasan terakhir, dari hasil analisis data dengan menggunakan uji t-Dunnet diperoleh nilai t_{hitung} sebesar 6,65, sedangkan nilai t_{tabel} pada taraf signifikansi $\alpha = 0{,}05$ adalah 1,69. Jadi nilai $t_{hitung} > t_{tabel}$, dengan demikian H_0 ditolak maka sebagai konsekuensi H_1 diterima. Hal ini berarti bahwa rata-rata skor kemampuan verbal linguistik siswa yang diberi tes uraian lebih rendah dengan rata-rata skor kemampuan verbal linguistik siswa yang diberi tes menjodohkan pada kelompok siswa yang diajar dengan metode pembelajaran audio lingual. Metode pembelajaran ini banyak memberikan pengulangan sehingga kemampuan yang didapatkan siswa berupa hafalan. Kemampuan kognitif yang didapatkan berupa pengetahuan dan pemahaman. Sementara itu bentuk soal menjodohkan sangat baik untuk mengukur kemampuan peserta didik yang sangat sederhana dan kemampuan mengidentifikasikan kemampuan meng-hubungkan antara dua hal. Oleh karena itu tes menjodohkan dianggap cocok dengan ciri-ciri pembelajaran audio lingual. Sementara itu kemampuan verbal linguistik juga merupakan kemampuan dalam memahami informasi melalui kata-kata dan bahasa yang meliputi kemampuan semantik berupa makna kata. Dalam metode audiolingual menggunakan teknik latihan yang beruntun maka siswa akan cepat menghafal kosakata yang diajarkan dengan makna benar. Sedangkan pada tes uraian siswa harus menjadikan sebuah kata menjadi kalimat yang baik dan efektif yang dapat dipahami oleh pembaca ataupun pendengar.

Dari penjelasan di atas maka dapat disimpulkan bahwa untuk kelompok siswa yang diajar dengan metode pembelajaran audio lingual,

kemampuan verbal linguistik siswa yang diberi tes uraian lebih rendah daripada kemampuan verbal linguistik siswa yang diberi tes menjodohkan.

Kesimpulan

Berdasarkan hasil pengujian hipotesis, maka dikemukakan beberapa kesimpulan berikut.

1. Kemampuan verbal linguistik kelompok siswa yang diberi tes uraian lebih tinggi dari kemampuan verbal linguistik kelompok siswa yang diberi tes menjodohkan.
2. Kemampuan verbal linguistik kelompok siswa yang diberi metode pembelajaran komunikatif lebih rendah dari kemampuan verbal linguistik kelompok siswa yang diberi metode pembelajaran audio lingual.
3. Terdapat pengaruh interaksi di antara bentuk tes formatif dan metode pembelajaran terhadap kemampuan verbal linguistik.
4. Khusus kelompok siswa yang diberi tes uraian, kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif lebih tinggi daripada kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran audio lingual.
5. Khusus kelompok siswa yang diberi tes menjodohkan, kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran komunikatif lebih rendah daripada kemampuan verbal linguistik siswa yang diajar dengan metode pembelajaran audio lingual.
6. Khusus kelompok siswa yang diajar dengan metode komunikatif, kemampuan verbal linguistik siswa yang diberi tes uraian lebih tinggi daripada kemampuan verbal linguistik siswa yang diberi tes menjodohkan.
7. Khusus kelompok siswa yang diajar dengan metode pembelajran audio lingual, kemampuan verbal linguistik siswa yang diberi tes uraian lebih rendah daripada kemampuan verbal linguistik siswa yang diberi tes menjodohkan.

Daftar Pustaka

Anderson, Charles J., C Clapham, dan Dianne Wall. 1995. *Language Test and Evaluation.*:Cambridge University Press, New york.

Braine, George. 2005. *Teaching English to The World.* New Jersey: Lawrence Erlbaum Associates.

Brown, Douglas. 2004. *Language Assessment. Principle and Classroom Practice.* New York: Pearson Education Inc.

Gardner, Howard. 1993. *Frames of Minds. The Theory of Multiple Intelligences.* New York: Collins.

Grounlund, Norman E. 1982. *Constructing Achievement Test.* London: Prentical Hall.Inc.

Kubiszyn, Tom, dan G Borich. 2007. *Educational testing and Measurement.* New York: John Wiley & Sons.

Lwin, May, Adam Khoo, Kenneth Lyen, dan Caroline Sim. 2008. *How to Multiply Your Child's Intelligence.* Jakarta: PT. Indeks.

Nitko, Anthony J. 1989. *Designing test That are Integrated with Instruction Educational Measurement.* New York: MacMillan.

Norland, Deborah L., dan Terry Pruet Said. 2006. *A Kaleidescope of Models and Strtegies for Teaching English to Speaker of Other language.* New York: Teacher Ideas Press.

Nurgiyantoro, Burhanudin. 2010. *Penilaian Pembelajaran Bahasa.* Yogyakarta: BPFE-Yogyakarta.

Popham, W. James. 1995. *Classroom Asessment. What Teacher Needs to Know.* Boston, Allyn & Bacon.

Richard, Jack, dan Theodore Rogers. 1994. *Approaches and Method in Language Teaching.* Australia: Cambridge University Press.

Richard, Jack, dan Theodore Rogers. 1986. *Approach and Method in Language Teaching. Description and Analysis.* Australia: Cambridge University Press.

Richards, Jack C., dan Willy A Renandya. 2002. *Methodology in Language Teaching.* New York: Cambridge University press.

Sinembela, Ida. 1994. *Tes Esai Pemetaan Konsep Sebagai Alat Ukur Dalam Belajar Bermakna.* Disertasi. Jakarta: Universitas Negeri Jakarta.

Sunandar. 2000. *Pengaruh Metode Mengajar dan Bentuk Tes Formatif Terhadap Hasil Belajar Matematika.* Disertasi. Jakarta: Universitas Negeri Jakarta.

Tarigan, Henry Guntur. 2009. *Metodologi Pengajaran Bahasa.* Bandung: Angkasa.

Tientje, Nurlaila Nqm. 2010. *Multiple Intelligensi Pendidikan Usia Dini.* Bogor: Rekatama.

Wahab, Abdullah A. 2008. *Metode dan Model-Model Mengajar.* Bandung: Alfabeta.

Windura, Sutanto. 2008. *Brain Management Series.* Jakarta: PT Elex Media Komputindo.

Woolfolk, Anita. 2009. *Educational Psychology.* Boston: Pearson education Inc.

MODEL EVALUASI PRAKTIKUM KIMIA DI LEMBAGA PENDIDIKAN TENAGA KEPENDIDIKAN

Endang Susilaningsih
Jurusan Kimia FMIPA UNNES
susilaningsihe@yahoo.com

Abstrak

Penelitian ini bertujuan untuk memperoleh model evaluasi program yang teruji, untuk mengevaluasi kualitas praktikum kimia. Pengembangan model evaluasi program terdiri atas empat variabel sebagai komponen model yaitu: (1) *environmemt context* yang dieksplanasi menjadi *environment, laboratory work*, *technology and society;* (2) evaluasi *Input;* (3) evaluasi *process*; dan (4) evaluasi *result* yang disesuaikan dengan praktikum kimia analisis dasar. Penetapan konstruk instrumen terdiri atas instrumen penilaian model dan instrumen penilaian kinerja mahasiswa dilakukan melalui pendapat: (1) pakar evaluasi; (2) pakar psikometri, dan (3) praktisi laboratorium. Subjek coba penelitian ini terdiri atas mahasiswa dan praktisi. Uji model evaluasi program dilakukan dengan uji: (1) *utility*; (2) *feasibility*; (3) *proprietly*; dan (4) *accuracy*. Uji kecocokan model untuk mengetahui *reliability* komponen program dan *goodness of fit index* menggunakan program Lisrel. Hasil penelitian menunjukkan bahwa evaluasi program yang dikembangkan diberi nama model ECIPR sudah teruji memenuhi kriteria: (1) *utility*; (2) *feasibility*; (3) *proprietly*; dan (4) *accuracy*. *reliability* untuk komponen *Environment context*=0,8649, *input*=0,8429, *process*=0,7828, *result*=0,9560. *Goodness of fit index:* p = 0,107 *RMSEA*=0,051, GFI=0,910. Dampak implentasi model yang dikembangkan terbukti dapat meningkatkan kinerja mahasiswa, keterampilan dan pengalaman laboratorium mahasiswa sebagai sasaran program.

Kata kunci: *model evaluasi program, praktikum kimia, keterampilan laboratorium*

THE EVALUATION MODEL OF THE CHEMISTRY PRACTICUM IN TEACHER EDUCATION INSTITUTIONS

Endang Susilaningsih
Jurusan Kimia FMIPA UNNES
susilaningsihe@yahoo.com

Abstract

This study aims to produce a program evaluation model which is valid to evaluate the quality of chemistry practicum. The program evaluation model consists of four variables as model components,i.e:(1) the context environment which is elaborated into environment, laboratory work, technology, and society; (2) the input evaluation; (3) the process evaluation; and (4) the result evaluation adjusted to basic analytic chemistry practicum.The establishment of the instrument constructs, consisting of a model evaluation instrument and a student performance evaluation instrument, was based on the opinions of (1) evaluation experts, (2) psychometrics experts, and (3) laboratory practitioners. The subjects for tryout in this research consisted of students and practitioners. The test of the program evaluation model was carried out on the basis of: (1) utility, (2) feasibility, (3) propriety, and (4) accuracy. The model fit test to find out the reliability of the program components and the goodness of fit index used the LISREL program package. The calculation of the reliability coefficient of the instrument for the laboratory performance evaluation used the SPSS program based on the Generalizability coefficient model developed by Thorndike. The components of the variance were the person, rater, item, interactions of the person and rater, the person and item, the rater and item, and the interaction of the person, rater and item.The results of the study show that the ECIPR program evaluation model is valid and satisfies the criteria for: (1) utility, (2) feasibility, (3) propriety, and (4) accuracy. The indicator reliability of the component of environment context is 0.8649, that of the input is 0.8429, that of the process is 0.7828, and that of the result is 0.9560. The goodness of fit index is indicated by p = 0.107, RMSEA = 0.051, and GFI = 0.910. The generalizability coefficient of the qualitative analytic instrument is 0.8726 and that of the quantitative analytic instrument is 0.9384. The implementation of the ECIPR model has been proved capable of improving the students' performances, which have effects on the improvement of their laboratory skills and experiences as the program target.

Keywords: *program evaluation model, chemistry practicum, laboratory skills*

Pendahuluan

Selama ini evaluasi program praktikum Kimia di Lembaga Pendidikan Tenaga Kependidikan, belum pernah dilakukan secara simultan, sehingga pelaksanaan praktikum dari dulu sampai sekarang selalu sama. Mulai dari buku petunjuk praktikum, materi acara praktikum, model pendekatan sampai model asesmen praktikum. Bahkan materi acara praktikum yang dipraktikkan hanya 50%, dan Mata kuliah praktikum kimia kini cenderung dipadatkan (dari 15 menjadi 9 mata kuliah praktikum kimia). Bermula dari adanya penurunan jumlah mata kuliah praktikum kimia dan materi acara praktikum yang berimplikasi pada penggunaan waktu yang lebih singkat, materi acara praktikum yang lebih sedikit, peneliti ingin mengetahui apakah praktikum kimia sekarang ini benar-benar dapat mencapai tujuan esensial yang sudah ditetapkan.

Sementara ini lima kompetensi guru mata pelajaran Kimia pada SMA dan sekolah lain yang sederajat dari empat belas kompetensi yang dituntut seperti yang tercantum dalam Lampiran Peraturan Menteri Pendidikan Nasional nomor 16 tahun 2007, diperoleh dari praktikum Kimia di laboratorium. Praktikum merupakan ciri khusus pembelajaran kimia, sehingga praktikum tidak bisa lepas dari pembelajaran kimia untuk memperoleh pengalaman laboratorium, keterampilan proses sains, dan pengalaman untuk investigasi.

Evaluasi program sangat diperlukan dalam konteks untuk mengevaluasi praktikum kimia, karena evaluasi program praktikum merupakan aktivitas yang dimaksudkan untuk mengukur kelayakan suatu desain praktikum dan metode implementasi desain tersebut. Aktivitas evaluasi digunakan untuk mencapai tujuan yang telah ditentukan, baik yang bersifat spesifik maupun tujuan umum. Evaluasi program dapat digunakan untuk mengukur dampak program, dan faktor-faktor yang dapat mempengaruhi tingkat kemanfaatan program, serta distribusi manfaat program yang dihasilkan (Valades & Bamberger, 1994: 13). Manfaat dari suatu evaluasi program yaitu: (1) Memperbaiki metode yang digunakan untuk mengidentifikasi suatu program agar selalu konsisten dengan tujuan yang telah ditentukan, (2) Meyakinkan tingkat keberhasilan yang mungkin

diperoleh dari program itu, dan membantu memilih strategi yang paling cocok dalam rangka mencapai tujuan yang telah ditentukan, (3) Menentukan apakah implementasi program itu dapat menjangkau orang-orang yang menjadi sasaran *(beneficiary)* program, (4) Mengukur apakah implementasi program itu mampu mancapai tujuan yang telah ditentukan, dan (5) Mengukur dampak program terhadap pembangunan (pendidikan, ekonomi dan pertahanan) dalam skala luas.

Evaluasi program praktikum kimia diharapkan bermanfaat untuk meningkatkan efektivitas praktikum kimia. Menjangkau mahasiswa sebagai sasaran program dalam upaya meningkatkan keterampilan laboratorium *(laboratory skill),* pengalaman laboratorium *(laboratory experience),* pengalaman investigasi *(investigation experience)*, dan peningkatan sikap terhadap kimia (*atittudes toward chemistry*).

Kriteria model evaluasi program yang baik harus memenuhi empat standar seperti yang dikemukakan oleh Stuffebeam, (2002: 2); Guskey, (2000: 60-63); dan The Joint commite (1994) yaitu: (1) standar kegunaan *(utility standard)*, (2) standar kelayakan *(feasibility standard),* (3) standar kesopanan *(propriety standard)*, dan (4) standar kecermatan atau ketelitian *(accuracy standard)*. Selanjutnya Stufflebeam (2003:10) dan The Joint Committe (2003) menyatakan standar evaluasi program yang lengkap sebagai berikut:

> *These require to meet conditions of utility (serving the information needs of intended users), feasibility (keeping evaluation operations realistic, prudent, viable, and frugal), propriety (conducting evaluations legally, ethically, and with due regard for the welfare of participants and those affected by result), and accuracy (revealing and conveying technically sound information about the features that determine the evaluand's merit, worth, probity, and/or significance).*

Kriteria model evaluasi program untuk praktikum Kimia Analisis Dasar yang baik dalam penelitian ini mengacu pada standar model evaluasi di atas, disesuaikan untuk mengevaluasi Praktikum Kimia Analitik Dasar. Model evaluasi program Praktikum Kimia Analitik Dasar yang baik memenuhi persyaratan sebagai berikut: (1) Komprehensip: objek evaluasi

mencakup komponen-komponen program praktikum secara menyeluruh mulai dari *environment context, input, process,* dan *result*; (2) Bermanfaat: penggunaan model bermanfaat untuk meningkatkan kinerja praktisi, meningkatkan kualitas praktikum, dan meningkatkan kompetensi mahasiswa sebagai sasaran program; (3) Praktis: model bersifat sederhana dan mudah dalam penggunaanya serta mempunyai tingkat keterlaksanaan yang cukup tinggi; (4) Ekonomis: penggunaan model tersebut membutuhkan biaya yang dapat dijangkau, tidak membutuhkan waktu yang lama, dan tidak banyak melibatkan tenaga di luar praktisi; dan (5) Ketepatan instrumen pengumpul data harus memberikan gambaran nyata kinerja di laboratorium.

Berdasarkan hasil kajian dari model-model evaluasi program yang tersedia, ternyata didesain untuk mengevaluasi program pelatihan, program-program yang berhubungan dengan kebijakan, dan program pembelajaran sosial. Untuk memperoleh model evaluasi yang cocok untuk mengevaluasi praktikum kimia, yang akurat, valid dan reliabel bukan suatu hal yang mudah dilakukan, diperlukan modifikasi model dan adaptasi metode dari beberapa model evaluasi program yang sudah tersedia, yang mendekati program praktikum dan dilakukan uji kelayakan model.

Penelitian ini bertujuan untuk 1) menghasilkan produk berupa model evaluasi program yang cocok, akurat, valid, dan reliabel untuk mengevaluasi mata kuliah Praktikum Kimia Analisis Dasar. Penemuan model dilakukan dengan cara modifikasi dan adaptasi prosedur dari model evaluasi program yang sudah tersedia. 2) menguji model evaluasi program yang ditemukan untuk memperoleh model evaluasi program yang benar-benar teruji dan cocok untuk mengevaluasi mata kulian Praktikum Kimia Analisis Dasar 3) mengimplementasikan model evaluasi program yang ditemukan dan teruji untuk mengevaluasi Praktikum Kimia Analisis Dasar.

Metode Penelitian

Penelitian disertasi ini termasuk penelitian pengembangan yang bertujuan untuk menghasilkan produk berupa: Evaluasi program model ECIPR yang komponennya meliputi *Environment Context* untuk

mengevaluasi kesesuaian antara isi dan tujuan praktikum dengan lingkungan, *Input* untuk mengevaluasi rencana pelaksanaan praktikum, *Process* untuk mengevaluasi aktivitas praktikum, dan *Result* untuk mengevaluasi hasil yang telah dicapai selama praktikum. Model evaluasi program yang dihasilkan digunakan untuk mengevaluasi Praktikum Kimia Analisis Dasar di Jurusan Kimia FMIPA UNNES. Evaluasi tersebut bertujuan untuk mengetahui kualitas Praktikum Kimia Analisis Dasar, dalam upaya untuk meningkatkan kompetensi akademik, kinerja mahasiawa di laboratorium, pengalaman laboratorium, keterampilan laboratorium, dan meningkatkan sikap terhadap kimia mahasiswa sebagai sasaran program. Dampak tidak langsung dari evaluasi program Praktikum Kimia Analisis Dasar, mahasiswa dapat mengembangkan potensi keterampilan yang dimilikinya melalui investigasi dan pengalaman laboratorium yang diperoleh selama mengikuti praktikum.

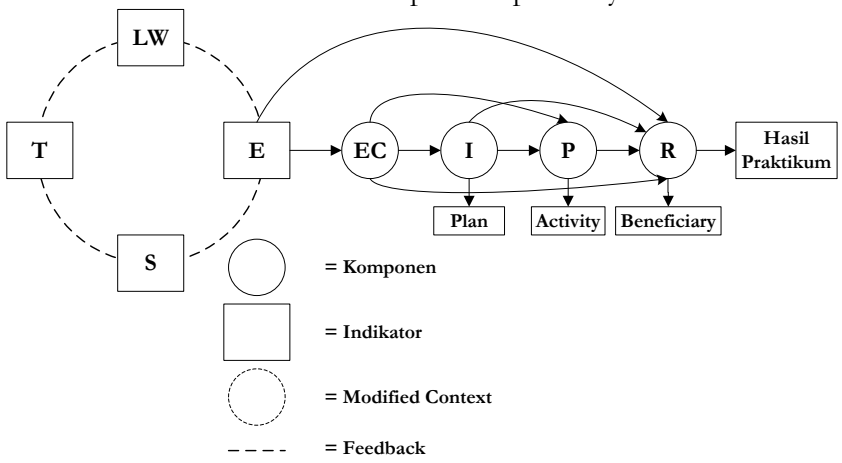
Model pengembangan yang akan dilakukan mengikuti langkah kerja pengembangan yang telah dikemukakan oleh Borg & Gall (1983:775), yang dimodifikasi dengan langkah pengembangan pembelajaran yang dikemukakan oleh Cenamo dan Kalk (2005). Prosedur penelitian ini mengikuti model pengembangan yang dilaksanakan sesuai langkah-langkah pengembangan yaitu: Pendahuluan *(Pralimenary Investigation),* Pendefinisian (*define phase),* Perencanaan prototipe *(design phase)*, uji coba desain prototipe, analisis data uji coba pengembangan prototipe *(develop phase),* uji coba prototipe, analisis data uji coba prototipe, dan implementasi model *(implentation phase)*.

Hasil Penelitian

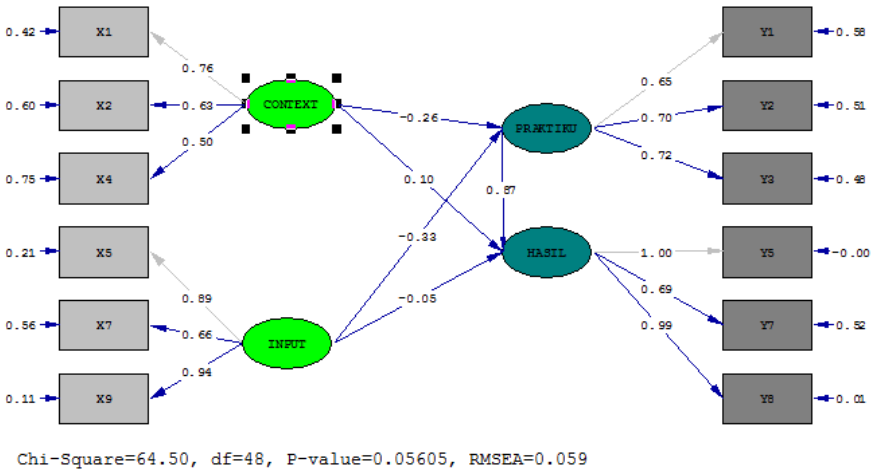
Produk pertama ditemukan informasi yang berkaitan dengan potret dan karakter Praktikum Kimia Analisis Dasar yang dilaksanakan di FMIPA UNNES Semarang (2008-2011) yang akan dievaluasi dengan evaluasi program model ECIPR hasil pengembangan. Desain analisis uji kelayakan model dilakukan dengan program LISREL.

Evaluasi program model ECIPR hasil pengembangan dengan cara modifikasi model, adaptasi metode, dan prosedur dari evaluasi program

model Kirkpatrick dan evaluasi program model CIPP 2003 yang sudah tersedia. Bentuk model ECIPR dapat dilihat pada Gambar 1. Evaluasi program model ECIPR terdiri atas empat komponen yaitu



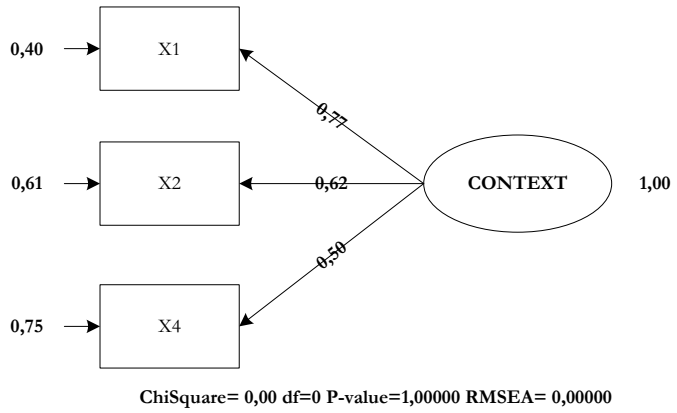
Gambar 1. Evaluasi Program Model ECIPR Teruji



Gambar 2. Desain Uji Kelayakan Evaluasi program model ECIPR

a. Evaluasi Environment Context

Berdasarkan analisis hasil evaluasi pada tahap *Environment Context* diperoleh *construct indicator reliability* = 0,8649, validitas indikator *environment*=0,77 *laboratory work*=0,62 *society*=0,5. Analisis deskripsi jawaban responden pada kesesuaian isi dan tujuan program terhadap lingkungan 74% menjawab sangat sesuai, 25% menjawab sesuai dan 1% menjawab cukup sesuai. Ini berarti dapat dipahami bahwa bukti empirik dari evaluasi *environment context* sesuai dengan evaluasi *environment context* secara teoretis.



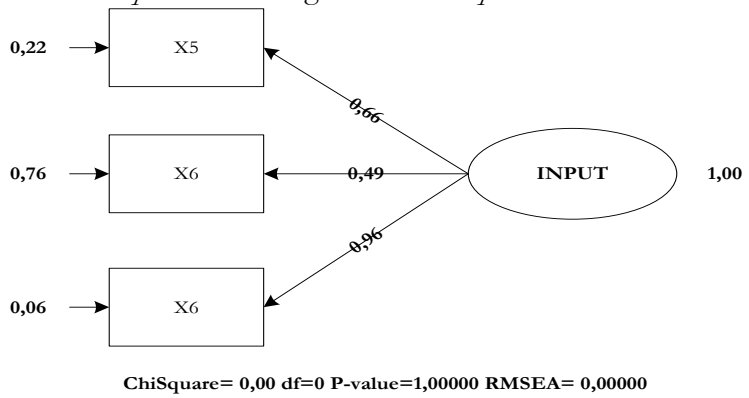
Gambar 3. Diagram Bagian Analisis Variabel *Environment Context*

Tabel 1. Reliabilitas Konstruk Indikator *Environment Context*

No	Indikator *Manifest*	*λ Standar*	*Error*	*Construct indicator Reliability*
1	Lngkungan	0,77	0,40	0,8649
2	Kerja laboratorium	0,62	0,41	
3	Masyarakat	0,50	0,75	
4	Jumlah	$\sum^2$ = 3,5721	1,56	

b. Evaluasi *Input*

Berdasarkan analisis hasil evaluasi pada tahap *input construct indicator reliability*=0,8429 validitas indikator rasio alat laboratorium dengan jumlah mahasiswa=0,88 rasio bahan kimia dengan jumlah mahasiswa=0,49 rasio buku petunjuk praktikum dengan jumlah mahasiswa=0,96. Berdasarkan analisis diskripsi jawaban responden pada keidealan rasio fasilitas laboratorium terhadap jumlah mahasiswa 40% responden menjawab sangat ideal, 55% menjawab ideal dan 5% menjawab cukup ideal. Berarti data empirik evaluasi *input* sesuai dengan evaluasi *input* teoretik.



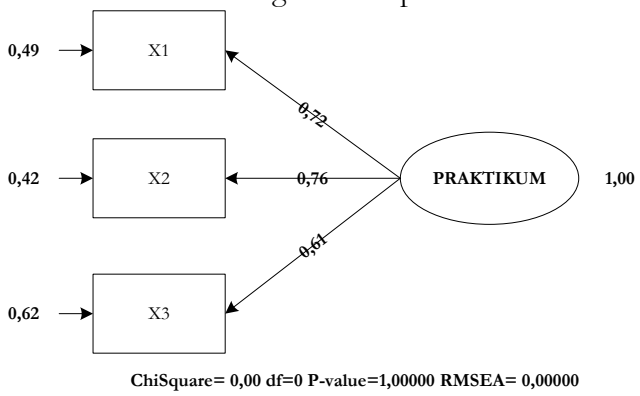
Gambar 4. Diagram Bagian Analisis Variabel *Input*

Tabel 2. Reliabilitas Konstruk Indikator *Input*

No	Indikator *Manifest*	*λ Standar*	*Error*	*Construct indicator Reliability*
1	Rasio alat lab dengan mahasiswa	0,88	0,22	0,8429
2	Rasio bahan prak dengan mahasiswa	0,49	0,76	
3	Rasio buku petunjuk prak dengan mahasiswa	0,96	0,04	
4	Jumlah	$\sum^2$ = 5,4289	1,02	

c. Evaluasi *Process*

Berdasrkan analisis hasil evaluasi *process* diperoleh *construct indicator reliability*=0,7666, validitas indikator persiapan praktikum= 0,72 keterampilan proses=0,76 membuat laporan sementara=0,61. Hasil analisis diskripsi kualitatif 99,90% responden menyatakan praktikum berjalan baik dan lancar sesuai dengan yang direncanakan. Ini berarti bahwa evaluasi *process* secara teoretis didukung data empirik.



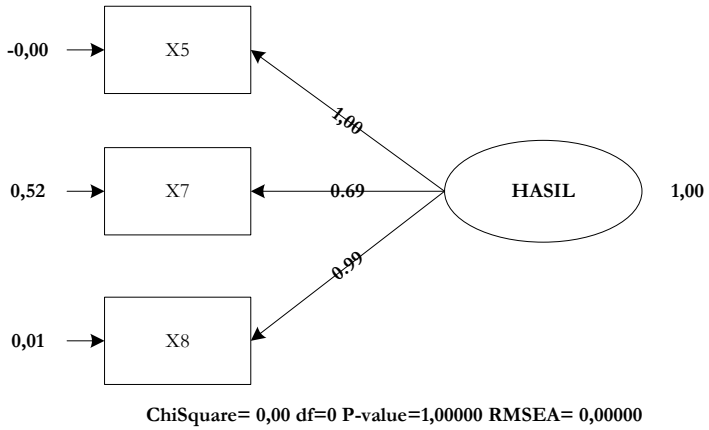
Gambar 5. Diagram Bagian Analisis Variabel Process (Praktikum)

Tabel 3. Reliabilitas Konstruk Indikator *Process* (praktikum)

No	Indikator *Manifest*	*λ Standar*	*Error*	*Construct Indicator Reliability*
1	Rasio alat lab dengan mahasiswa	0,72	0,49	0,7416
2	Rasio bahan prak dengan mahasiswa	0,76	0,42	
3	Rasio buku petunjuk prak dengan mahasiswa	0,61	0,62	
4	Jumlah	$\sum^2$ = 4,3681	1,53	

d. Evaluasi *Result*

Berdasarkan analisis hasil evaluasi pada tahap hasil diperoleh *construct indicator reliability* = 0,9560, validitas indikator kinerja mahasiswa dilaboratorium = 1,00 keterampilan laboratorium = 0,69, keterampilan investigasi = 0,99. Ini dapat dipahami bahwa evaluasi pada tahap hasil sesuai dengan kondisi hasil yang dicapai di laboratorium, Evaluasi *result* teoretik didukung oleh data empirik hasil penelitian.



Gambar 6. Diagram Bagian Analisis Variabel Hasil *(Result)*

Tabel 4. Reliabilitas konstruk hasil (*Result)*

No	Indikator *Manifest*	*λ Standar*	*Error*	*Construct Indikator Reliability*
1	Rasio alat lab dengan mahasiswa	1,00	0,00	0,9650
2	Rasio bahan prak dengan mahasiswa	0,69	0,32	
3	Rasio buku petunjuk prak dengan mahasiswa	0,99	0,01	
4	Jumlah	$\sum^2$ = 7,1824	0,33	

Berdasarkan analisis hasil evaluasi secara keseluruhan *Goodness of Fit index* dari evaluasi program model ECIPR, RMSEA=0,051, indeks P_{value} =0,1071. GFI 0,91 secara lengkap dapat dilihat pada tabel 5. Dapat disimpulkan bahwa evaluasi program model ECIPR cocok dan teruji untuk mengevaluasi praktikum Kimia Analisis Dasar.

Evaluasi program model ECIPR dapat dibuktikan telah memenuhi kriteria *utility*, *feasibility, propriety,* dan *accuracy*. Ini berarti evaluasi program model ECIPR telah terbukti memberikan informasi yang menyeluruh mulai dari penetapan komponen model, indikator, instrumen, validasi instrumen, uji kriteria model, cara analisis hasil evaluasi, ketelitian, dan kebermanfaatan, sehingga evaluasi program model ECIPR dapat dilaksanakan secara mudah, realistik, praktis dan akurat.

Tabel 5. Hasil (Goodness of Fit Index) Model ECIPR

No	***Index***	***Cut of value***	**Hasil**	**Keterangan**
1	χ^2 kecil, p	> 0,05	0,1071	Terpenuhi
2	NCP	<<<	0,000	Terpenuhi
3	RMSEA	< 0,08	0,051	Terpenuhi
4	ECVI.	< ECVI sat = 1,58	1,09 – 1,46	Terpenuhi
5	AIC	< AIC Sat = 156,00	120,45	Terpenuhi
6	CAIC	< tCAIC s = 437.20	228,61	Terpenuhi
7	sRMR	< 0,05	0,07	Tidak terpenuhi
8	GFI	> 0,9	0,91	Terpenuhi
9	AGFI	>0,9	0,85	Tidak terpenuhi
10	PGFI	> 0,9	0,56	Tidak terpenuhi
11	NFI	> 0,9	0,94	Terpenuhi
12	NNFI	> 0,9	0,98	Terpenuhi
13	PNFI	> 0,9	0,68	Tidak terpenuhi
14	CFI	> 0,9	0,98	Terpenuhi
15	IFI	> 0,9	0,99	Terpenuhi
16	RFI	> 0,9	0,91	Terpenuhi

Simpulan dan Saran

Evaluasi program model ECIPR adalah salah satu produk penelitian yang cocok dan teruji untuk mengevaluasi praktikum Kimia Analisis Dasar. Keunggulan evaluasi program model ECIPR: (1) Model ECIPR sederhana, hanya terdiri atas empat komponen, mencakup semua indikator Praktikum Kimia Analisis Dasar yang telah teruji secara empirik dan didukung oleh data observasi langsung, (2) Praktis artinya mudah dilakukan hanya melibatkan praktisi laboratorium saja, (3) Dapat digeneralisasi untuk mengevaluasi pembelajaran di laboratorium selain Kimia Analisis Dasar. Artinya semua pembelajaran di laboratorium (praktikum) dapat dievaluasi dengan model ECIPR Hal ini disebabkan komponen-komponen semua praktikum sama, aspek yang dinilai untuk praktikum sama, dan hanya sedikit modifikasi pada isi/materi acara praktikum disesuaiakn dengan mata kuliah praktikum yang akan dievaluasi, (4) Ekonomis: tidak membutuhkan beaya yang tinggi, dapat dijangkau pada tingkat jurusan, karena instrumen yang digunakan sudah standar dan teruji, (5) Instrumen penilaian praktikum dapat digunakan terus menerus selama pendekatan praktikum dan tujuan praktikum tidak berubah, (6) Dampak implementasi model dapat meningkatkan kinerja mahasiswa dan praktisi laboratorium, meningkatkan pengalaman dan keterampilan laboratorium, pengalaman mengivestigasi, dan sikap mahasiswa.

Saran untuk implementasi evaluasi program model ECIPR dapat dirinci sebagai berikut: (1) Memaparkan isi dan tujuan praktikum yang ingin dicapai mealui evaluasi *environment context*. (2) Membuat perencanaan metode, strategi, fasilitas praktikum, sumber daya manusia untuk mencapai tujuan yang telah ditetapkan. Aktifitas ini dilakukan pada evaluasi tahap *input*. (3) Aktivitas pembelajaran di laboratorium merupakan proses untuk mencapai tujuan program, maka langkah terpenting dalam evaluasi *process* ini adalah menyusun alat ukur untuk mengetahui tingkat ketercapaian program, validasi alat ukur tersebut, dan (4) Hasil evaluasi dapat dilihat pada evaluasi *result*.

Evaluasi program model ECPR ini baru bisa mencermati peningkatan kinerja mahasiswa selama praktikum, belum mengungkap

penilaian praktikum yang tujuannya berbeda atau pendekatan praktikumnya berbeda. Berdasarkan model pendekatan praktikum yang ada tergantung dari tujuan praktikum yang akan dicapai oleh mahasiswa.

Daftar Pustaka

Cennamo, K., & Kalk, D. 2005. *Real World Instructional Design*. Canada: ThomsomWadsworth.

Guskey, T.R. 2000. *Evaluating Professional Development*. California: Corwin Press

Kirkpatrick, D.L. 1996. *Evaluating Training Programs*. San Fransisco: Berrett-koehler publisher.

Lagowski, J. J. 2002. *The Role of The Laboratory in Chemical Education*. Paper presented in Departement of Chemistry and Biochemistry The University of Texas at Austin.

Nitco, A.J., & Brookhart, S.M. 2001. *Educational Assessment of Student* (5th ed), Upper Saddle River NJ: Pearson Education, Inc

Ogle, G.J. 2002. *Towards A Formatif Evaluation Tool*. Dissertation: Virginia Polytechnic Institute and State University – USA.

Popham, W.J. 2005. *Classroom Assessment What Teachers Need to Know*, (4th ed).Boston : Pearson Education, Inc.

Stufflebeam, D. L., & Madaus, G. F. 2000. *Evaluation Models*: *Viewpoints on Educational and Human Services Evaluation*. (Eds). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L. 2002. CIPP Model Checklist, Retrieved on 8 th November 2004 [versi elektronik] http://www.wmich.edu/evalctr/checklist.

Stufflebeam, D. L. 2003.*The CIPP Model For Evaluation*. Paper presented at the annual Conference of the Oregon Program Evaluators Network, Portland, Oregon, Retrived on 10th March 2004 [versi elektronik]:

http://www.wmich.edu/evalctr/pubs/CIPP-ModelOregon10-03.pdf

Thiagarajan, S., Dorothy, S.S., & Melvyn, I. S. 1974. *Instructional development for training teachers of exeptional children.* Bloomington Indiana: Indiana University.

Thorndike, R.L. 1982. *Applied Psychometrics*. Boston: Houghton Mifflin Company.

Valadez, J., & Bamberger, M. 1994. *Monitoring and Evaluating Social Program in Developing Countries*. Washington: the World Bank.

KEMAMPUAN KOMUNIKASI MATEMATIS DAN PEMECAHAN MASALAH MELALUI PROBLEM BASED-LEARNING

Fatia Fatimah
UPBJJ-UT Padang
fatia@ut.ac.id

Abstrak

Tujuan penelitian untuk mengetahui kemampuan komunikasi matematis dan pemecahan masalah melalui *Problem Based Learning* (PBL). Populasi penelitian adalah mahasiswa STKIP PGRI Sumatera Barat yang mengambil mata kuliah Statistika Elementer tahun akademik 2008/2009. Jenis penelitan adalah eksperimen semu dengan *pretest-postest control group design*. Kelompok eksperimen diberi perlakuan berupa pembelajaran dengan model *Problem Based Learning* dan kelompok kontrol dengan pembelajaran biasa. Data dikumpulkan melalui *worksheet*, rubrik presentasi, assesmen investigasi dan tes hasil belajar Statistika Elementer. Hasil analisis data menunjukkan bahwa 1) kemampuan komunikasi matematis mahasiswa dengan menerapkan model PBL dalam pembelajaran Statistika Elementer tidak lebih baik dibandingkan dengan pembelajaran biasa, 2) kemampuan pemecahan masalah mahasiswa dengan menerapkan model PBL dalam pembelajaran Statistika Elementer lebih baik dibandingkan dengan pembelajaran biasa.

Kata kunci: *komunikasi matematis, pemecahan masalah, problem based learning*

COMMUNICATION SKILLS AND PROBLEM SOLVING THROUGH MATHEMATICAL PROBLEM-BASED LEARNING

Fatia Fatimah
UPBJJ-UT Padang
fatia@ut.ac.id

Abstract

The purpose of this research is to know the mathematic communicative competence and problem solving competence through problem based learning (PBL). This research was done to students of STKIP PGRI West Sumatera who were studying Statistica Elementer subject in 2008/2009 academic year. This research was quasi experimental research. The design of this research was pretest-postest control group. PBL model was given to experimental group and conventional learning was done on control group. Data were collected through worksheet, presentation rubric, investigation assessment, and achievement test of Statistica Elementer subject. Data analysis shows that 1) students implementing PBLbetter do not show mathematic communicative competence than those implementing conventional learning. 2) students who used PBL model show better ability on problem solving than those using conventional learning.

Keywords: *communication of mathematic, problem solving, problem based learning*

Pendahuluan

Kemampuan komunikasi matematis dan pemecahan masalah penting untuk dikuasai oleh mahasiswa. Hal ini dikarenakan masyarakat membutuhkan kaum intelektual yang mampu menyelesaikan masalah secara sistematis dan mampu untuk menginterpretasikan ke dalam bahasa lisan maupun tulisan yang mudah dipahami. Perguruan tinggi sebagai wadah pendidikan formal terakhir diharapkan dapat memfasilitasi berkembangnya kemampuan komunikasi matematis dan pemecahan masalah mahasiswa.

Problem based learning (PBL) sebagai salah satu model pembelajaran memiliki ciri khas yaitu selalu dimulai dan berpusat pada masalah. Di dalam PBL para mahasiswa dapat bekerja di dalam kelompok-kelompok kecil dan harus mengidentifikasi apa yang mereka ketahui serta apa yang mereka tidak ketahui dan harus belajar untuk memecahkan suatu masalah. Peran utama dari dosen untuk memudahkan proses kelompok dan belajar, bukan untuk menyediakan jawaban secara langsung.

Beberapa penelitian tentang PBL dalam pembelajaran matematika/statistika diperoleh hasil yang positif (Wagiran, 2007; Dina, 2008). Akan tetapi penelitian yang sekaligus melihat kemampuan komunikasi matematis dan pemecahan masalah melalui PBL pada statistika belum dilakukan. Sementara itu, penelitian yang terkait dengan kemampuan komunikasi matematis masih terdapat hasil yang kurang memuaskan (Firdaus, 2005).

Berdasarkan observasi awal terhadap mahasiswa matematika STKIP PGRI Sumbar yang mengambil matakuliah Statistika Elementer diperoleh temuan bahwa mahasiswa kurang mampu dalam memecahkan masalah matematika yang terkait dengan dunia nyata dan belum terbiasa menuangkan pemikiran dalam bentuk lisan maupun tulisan. Mereka kesulitan dalam menentukan masalah, tahapan yang harus dipilih untuk mencari solusi serta menentukan pola yang dapat digunakan. Mahasiswa lebih senang jika diberikan soal berbentuk simbol dan angka-angka sehingga langsung tahu apa yang akan dicari tanpa harus menginterpretasikan soal terlebih dahulu. Pertanyaan tentang "kenapa"

paling sulit untuk dijawab. Mahasiswa beralasan, ”tahu jawabannya namun susah untuk diutarakan”.

Tujuan penelitian adalah untuk mengetahui apakah kemampuan komunikasi matematis mahasiswa melalui PBL dalam pembelajaran Statistika Elementer lebih baik dari kemampuan komunikasi matematis mahasiswa dengan pembelajaran biasa dan mengetahui apakah kemampuan pemecahan masalah mahasiswa melalui PBL dalam pembelajaran Statistika Elementer lebih baik dari kemampuan pemecahan masalah mahasiswa dengan pembelajaran biasa.

Metode Penelitian

Penelitian menggunakan *pretest-postest control group design*. Populasi penelitian adalah mahasiswa program studi pendidikan matematika STKIP PGRI Sumatera Barat yang mengambil mata kuliah Statistika Elementer pada tahun akademik 2008/2009. Teknik pengambilan sampel adalah *purposive sampling*. Pengukuran kemampuan komunikasi matematis dan pemecahan masalah diambil dari *worksheet* dan tes hasil belajar serta menggunakan rubrik untuk menilai presentasi, *assesmen* investigasi.

PBL dalam penelitian ini dilakukan secara berkelompok dengan teknik pertanyaan studi terbuka dan tertutup. Tahap-tahap pelaksanaan PBL menerapkan prosedur yang dikemukakan oleh Ronnis (2000) yaitu: 1. *Determining whether a problem exists*, 2. *Creating an exact statement of the problem*, 3. *Identifying information needed to understand the problem*, 4. *Identifying resources for gathering information*, 5. *Generating possible solutions*, 6. *Analyzing the solutions*, 7. *Presenting the solution, orally and/or in writing*.

Kemampuan komunikasi matematis dan pemecahan masalah diukur tidak hanya menggunakan soal tes essay akan tetapi juga melalui lembar kerja (*worksheet*) dan menggunakan assesmen presentasi dan investigasi. Soal tes essay dibuat dengan mengacu pada indikator komunikasi matematis dan pemecahan masalah.

Indikator komunikasi matematis pada soal tes adalah menyajikan pernyataan matematika secara lisan, tertulis, gambar dan diagram, mengajukan dugaan, melakukan manipulasi matematika, menarik

kesimpulan, menyusun bukti, memberikan alasan atau bukti terhadap kebenaran solusi. Indikator kemampuan pemecahan masalah pada soal tes adalah menunjukkan pemahaman masalah /memformulasi masalah, mengorganisasi data dan memilih informasi yang relevan dalam pemecahan masalah, menyajikan masalah secara matematik dalam berbagai bentuk, memilih pendekatan dan metode pemecahan masalah secara tepat, mengembangkan strategi pemecahan masalah, membuat dan menafsirkan model matematika dari suatu masalah.

Indikator *worksheet* yang digunakan adalah kemampuan menuliskan permasalahan dengan bahasa sendiri, kemampuan mengembangkan solusi terkait permasalahan, dan kemampuan mengkaji ulang solusi. Indikator rubrik presentasi adalah kebenaran konsep yang dijelaskan mahasiswa, mendiskripsikan dan menjelaskan secara detail, kedalaman argumentasi mahasiswa, dan kebenaran membuat kesimpulan. Indikator rubrik assesmen investigasi adalah kemampuan memahami permasalahan dan konsep pembelajaran, kemampuan menggunakan strategi investigasi, kemampuan mengemukakan ide dalam bentuk laporan tertulis, dan kebenaran membuat kesimpulan. Pedoman penskoran *worksheet* dan rubrik menggunakan kategori sebagai berikut, 4: sangat baik, 3: baik, 2: cukup, dan 1: kurang.

Hasil dan Pembahasan

Berdasarkan hasil olahan data worksheet untuk setiap kelompok selama lima kali pertemuan (Tabel 1) diperoleh hasil bahwa tiga kelompok terkategori baik untuk kemampuan komunikasi matematis dan selebihnya masih kurang. Kemampuan pemecahan masalah untuk keempat kelompok terkategori baik dan sisanya kurang.

Skor rubrik setiap kelompok untuk lima kali pertemuan dan rata-ratanya dapat dilihat pada Tabel 2. Berdasarkan Tabel 2 dapat diketahui bahwa kemampuan komunikasi secara berkelompok mengalami peningkatan dan secara rata-rata terkategori baik. Kemampuan pemecahan masalah secara berkelompok terkategori baik.

Tabel 1. Skor *worksheet* untuk setiap kelompok

Kel.	Komunikasi matematis					rata-rata	Pemecahan Masalah					rata-rata
	1	2	3	4	5		1	2	3	4	5	
1	1	1	2	2	3	1,8	2	3	4	4	4	3,4
2	3	1	3	3	3	2,6	1	2	3	3	3	2,4
3	1	3	4	4	4	3,2	1	2	3	3	3	2,4
4	3	3	3	3	4	3,2	2	2	3	4	4	3
5	1	3	3	4	4	3	2	3	3	3	4	3
6	1	1	4	4	4	2,8	2	3	4	4	4	3,4

Tabel 2. Skor rubrik setiap kelompok

Kelompok	Presentasi		Investigasi
	I	II	
1	2	3	4
2	2	2	3
3	2	3	3
4	3	4	3
5	3	4	4
6	3	4	3

Pretest dan *postest* yang diberikan dianalisis menggunakan uji t. Berdasarkan hasil uji normalitas diperoleh hasil bahwa *gain score* komunikasi matematis dan pemecahan masalah berdistribusi normal (Tabel 3). Hasil uji

homogenitas varians diperoleh bahwa *gain score* komunikasi dan pemecahan masalah mempunyai kesamaan varians (Tabel 4).

Tabel 3. Hasil uji normalitas untuk setiap kelompok

Kelompok	**Perlakuan**	$\mathbf{L_{hitung}}$	$\mathbf{L_{tabel}}$
Gain score Komunikasi Matematis	PBL	0,102	0,1367
	Biasa	0,097	0,1497
Gain score Pemecahan Masalah	PBL	0,102	0,1367
	Biasa	0,097	0,1497

Tabel 4. Hasil uji homogenitas varians

Kelompok	χ^2_{hitung}	χ^2_{tabel}
Gain score komunikasi matematis	1,06	3,84
Gain score pemecahan masalah	3,63	3,84

Berdasarkan hasil perhitungan uji hipotesis untuk tiap kelompok diperoleh bahwa kemampuan komunikasi matematis dengan menggunakan PBL tidak lebih baik daripada pembelajaran biasa dan kemampuan pemecahan masalah pada PBL lebih baik daripada pembelajaran biasa (Tabel 5).

Tabel 5. Hasil uji hipotesis untuk komunikasi matematis dan pemecahan masalah

Kelompok	$\mathbf{t_{hitung}}$	$\mathbf{t_{tabel}}$
Gain score komunikasi matematis	0,0712	1,96
Gain score pemecahan masalah	5,1452	1,96

Hal ini bertentangan dengan hipotesis awal penelitian yaitu kemampuan komunikasi mahasiswa yang pembelajarannya menggunakan model PBL lebih baik dari pembelajaran secara biasa. *Problem Based Learning* memang memfasilitasi kemampuan komunikasi mahasiswa akan tetapi lebih banyak memberikan sarana untuk melatih kemampuan pemecahan masalah.

Sesuai teori yang menyatakan bahwa "pembelajaran yang menggunakan model PBL mempunyai dua tahap inti, yaitu analisis pemecahan masalah secara kolabotarif dan belajar mandiri" (Paulina dkk, 2001). Seperti definisi PBL yang dikemukakan oleh Ronnis (2000) yaitu

> *PBL is a curriculum development and instructional system that simultaneously develops both problem solving strategies and disciplinary knowledge bases and skills by placing students in the active role of problem-solvers confronted with an ill-structured problem that mirros real-world problems".*

Soal komunikasi pada topik Statistika Deskriptif dalam bentuk interpretasi data ke penyajian grafik. Soal ini cukup mudah untuk dipahami oleh mahasiswa baik yang pembelajarannya menerapkan PBL maupun pembelajaran biasa. Alasan pembuatan soal komunikasi matematis dalam bentuk penyajian secara grafik berdasarkan definisi kemampuan komunikasi menurut Hari Suderajat (2003) yaitu "kecakapan menyampaikan informasi pada orang lain melalui bahasa lisan atau simbol-simbol tertulis, termasuk chart, peta atau alat demonstrasi lainnya".

Pada pembelajaran dengan model PBL proses komunikasi masih menunggu arahan dan bimbingan dosen. Sikap individual mahasiswa masih dominan, karena mahasiswa belum terbiasa mengemukakan pendapat dan menghargai pendapat teman dan cenderung bertahan walau belum tentu benar jawabannya.

Hal ini telah diungkapkan oleh (Paulina dkk, 2001) sebagai salah satu kelemahan dari PBL yaitu "sejauh ini, setiap mahasiswa berasumsi bahwa mereka hanya penerima pasif dari informasi yang disampaikan dosen. Asumsi ini tumbuh berdasarkan pengalaman belajar yang dialami mereka dalam jenjang pendidikan sebelumnya. Ketika mereka berpartisipasi dalam

PBL, peran yang dituntut adalah berbeda. Seringkali hal ini menjadi kendala bagi mahasiswa pemula". Jadi dapat dikatakan bahwa hal-hal di ataslah yang merupakan penyebab terjadinya hasil belajar untuk kemampuan komunikasi mahasiswa yang diajar dengan model PBL tidak lebih baik daripada dengan pembelajaran secara biasa.

Proses pelaksanaan pembelajaran dengan PBL, yaitu: Tahap 1 mahasiswa berdiskusi di dalam kelompoknya untuk menentukan permasalahan yang ada pada *worksheet*., Tahap 2 mahasiswa menuliskan permasalahan dengan bahasa sendiri. Tahap 3 mahasiswa mengidentifikasi informasi dengan melakukan percobaan atau membaca literatur yang relevan dengan masalah, Tahap 4 mahasiswa mencari informasi tambahan mengenai pembelajaran pada hari itu melalui berbagai sumber. Tahap 5 mahasiswa mengembangkan solusi yang mungkin dari masalah melalui kegiatan penemuan secara berkelompok serta merujuk pada bacaan yang terkait dengan penyelesaian masalah. Tahap 6 masing-masing kelompok mengkaji ulang solusi yang diperoleh dan membuat laporan pemecahan masalahnya. Tahap 7 dosen meminta kelompok untuk menyajikan laporannya ke depan kelas. Kelompok lain boleh memberikan pendapat dan menceritakan hasil temuan kelompok masing-masing.

Setiap pembelajaran dengan PBL selalu diawali dengan permasalahan. Tahapan proses pemecahan masalah digunakan sebagai kerangka atau panduan dalam proses pembelajaran PBL (Paulina dkk, 2001). Pada PBL, mahasiswa terlatih untuk memecahkan masalah. Sesuai teori "l*atihan memegang peranan penting, lebih banyak latihan dan ulangan maka akan lebih lama pengalaman dan pengetahuan itu tinggal dalam kesadaran dan ingatan seseorang" (Oemar Hamalik, 1995).*

*Pada PBL mahasiswa memecahkan masalah dengan melakukan percobaan dan invesitigasi melalui bacaan. Hal ini sejalan dengan "i*mpelementasi empat pilar pendidikan UNESCO dalam pembelajaran matematika terlihat dalam pembelajaran dan penilaian yang sifatnya *learning to know* (fakta, skills, konsep, dan prinsip), *learning to do* (*doing mathematics*), *learning to be* (*enjoy mathematics*), dan *learning to live together* (*cooperative learning in mathematics*)" (Depdiknas, 2007).

Pada pembelajaran biasa dosen menyajikan permasalahan dalam bentuk contoh soal dan cara untuk menyelesaikan masalah dijelaskan oleh dosen kemudian masalah lain disajikan kembali dalam bentuk latihan. Mahasiswa terkendala untuk menyelesaikan masalah jika yang ditanya berbeda dengan contoh soal. Pembelajaran biasa menurut Depdiknas (2003) mempunyai ciri-ciri pembelajaran tidak memperhatikan pengalaman siswa, pembelajaran abstrak dan teoritis serta penilaiannya hanya ditentukan oleh tes bukan pada penilaian proses belajar.

Dalam pembelajaran yang mengakomodir pemecahan masalah, mahasiswa dilatih untuk berpikir divergen/kreatif dalam usaha mencetuskan sebanyak mungkin gagasan terhadap suatu masalah. Selain itu mahasiswa dilatih untuk berpikir secara konvergen dengan menggunakan penalaran logis-kritis dalam mempertimbangkan atau merumuskan jawaban yang paling tepat. Sehingga dengan berkembangnya kemampuan pemecahan masalah mahasiswa diharapkan dapat membangun pengetahuan matematis baru, memecahkan masalah baik yang terdapat dalam matematika, maupun konteks lain dengan menerapkan berbagai strategi yang cocok serta mampu merefleksi proses-proses yang telah dilakukan dalam memecahkan masalah. Jadi dapat dikatakan bahwa hal-hal di ataslah yang merupakan penyebab terjadinya hasil belajar untuk pemecahan masalah mahasiswa yang diajar dengan model PBL lebih baik daripada dengan pembelajaran secara biasa.

Kesimpulan

Model *problem based learning* lebih sesuai untuk meningkatkan kemampuan pemecahan masalah dan kurang tepat untuk kemampuan komunikasi matematis. Instrumen penilaian selama pembelajaran seperti *worksheet*, rubrik selain untuk mengukur kinerja kelompok agar dapat juga mengukur kemampuan mahasiswa secara individu. Soal tes untuk mengukur kemampuan komunikai matematis dan pemecahan masalah dapat diperbanyak item soalnya agar dapat lebih detil mengukur setiap indikator.

Daftar Pustaka

_____. 2007. *Kajian kebijakan kurikulum mata pelajaran matematika*. Badan Penelitian dan Pengembangan Pusat Kurikulum.

Departemen Pendidikan Nasional. 2003. *Kurikulum berbasis kompetensi*. Jakarta: Depdiknas.

Dina Mellita. 2008. Metode pembelajaran *peer teaching* dan *problem based learning* untuk memotivasi sosialisasi dalam kelas (pada pembelajaran statistika). Jurnal Ilmiah Bina Edukasi, 1(2), 87-98.

Firdaus. 2005. *Melalui pembelajaran dalam kelompok kecil tipe team assisted individualization (TAI) dengan pendekatan berbasis masalah (eksperimen pada salah satu SMA di Bandung)*. Tesis tidak dipublikasikan. Bandung: UPI

Hari Suderajat. 2003. *Implementasi kurikulum berbasis kompetensi (KBK)*. Jakarta: Depdiknas.

Oemar Hamalik. 1995. Kurikulum dan pembelajaran. Jakarta: Bumi Aksara.

Paulina Pannen dkk. 2001. *Konstruktivisme dalam pembelajaran*. Jakarta: PAU-PPAI, Universitas Terbuka.

Ronnis, Diane. 2000. *Problem-based learning for math and science:integrating inquiry and the internet*. Illinois: Skylight Professional Development.

Wagiran. 2007. Peningkatan keaktivan mahasiswa dan reduksi miskonsepsi melaluipendekatan *problem based-learning*. Jurnal Kependidikan, 37(1), 1-22.

EVALUASI PROGRAM PEMBELAJARAN ILMU PENGETAHUAN SOSIAL SMP/MTS DI KOTA BIMA

Ida Waluyati
Sekolah Tinggi Ilmu Tarbiyah (STIT) Sunan Giri Bima NTB
ida.waluyati@yahoo.com

Abstrak

Penelitian ini bertujuan untuk mengetahui kesesuaian pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima, yang meliputi: 1) perencanaan, 2) pelaksanaan, dan 3) penilaian serta tindak lanjutnya, 4) dan hasil pembelajaran kognitif IPS siswa dengan standar Kriteria Ketuntasan Minimal belajar. Penelitian ini merupakan penelitian evaluasi dengan menggunakan *Model Countenance Stake,* yang meliputi evaluasi *Antecedents, Transactions, dan Outcomes*. Subjek penelitian adalah 78 guru IPS yang diambil secara populasi dan siswa sejumlah 390 orang yang dipilih secara *Quota Stratified Random Sampling*. Teknik pengumpulan data yang digunakan adalah angket, tes, dan dokumentasi. Instrumen yang digunakan adalah kuesioner, tes, dan daftar cek. Analisis data dilakukan dengan analisis deskriptif kuantitatif, dengan membandingkan hasil penelitian dengan kriteria yang telah ditentukan. Hasil penelitian menunjukkan bahwa kesesuaian antara pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan adalah sebagai berikut.(1) Perencanaan pembelajaran berada pada kategori baik (sesuai), (2) Proses pembelajaran berada pada kategori cukup baik (cukup sesuai), (3) Penilaian hasil belajar serta tindak lanjut hasil penilaian berada pada kategori baik (sesuai), (4) dan hasil belajar kognitif IPS siswa berada pada kategori baik (sesuai) standar Kriteria Ketuntasan Minimal belajar.

Kata kunci*: evaluasi program, pembelajaran IPS*

EVALUATION OF SOCIAL STUDIES INSTRUCTIONAL PROGRAM FOR JUNIOR HIGH SCHOOL/MADRASAH TSANAWIYAH IN BIMA REGENCY

Ida Waluyati
Sekolah Tinggi Ilmu Tarbiyah (STIT) Sunan Giri Bima NTB
ida.waluyati@yahoo.com

Abstract

This study aims to evaluate the implementation of social studies instructional program for Junior High School/Madrasah Tsanawiyah in Bima regency, which involves: (1) planning, (2) implementation process, (3) learning outcome assessments and its follow-up with education process standard, (4) and students cognitive learning outcomes with learning minimal criteria standard. This is an evaluation research study employing the *Countenance Stake's* model consisting of the evaluation of antecedents, transactions, and outcomes. The subjects of this study were 78 social studies teachers of Junior High School/Madrasah Tsanawiyah in Bima regency selected through population techniques and 390 students through quota stratified random sampling. The data were collected by using questionnaires, test and documentation techniques.. The validity of the Instruments was done through logical. The questionnaire reliability was calculated by using Alpha Cronbach, while the questionnaires validity used content validity technique from expert and it was calculated by using Product Moment Corellation. Reliability coefficient of the questionnaires in terms of social studies instructional implementation process is 0.842 and learning outcome assessments and its follow-up is 0.772. Based on test results, all the instruments are valid and reliable. The data are analyzed by using the quantitative descriptive analysis by comparing the results of this research with the predetermined criteria. Data analysis result shows that the implementation of the social studies instructional program for Junior High School/Madrasah Tsanawiyah in Bima regency with the education process standard, is as follows. (1) The instructional plan is in *good (compatible)*category. (2) The instructional process is in *quite good (quite compatible)* category. (3) The learning outcome assessments and its follow-up is in *good (compatible)* category. (4) students cognitive learning outcomes is in *good (compatible)* category with the learning minimal criteria standard.

Keywords:*program evaluation, social studies instructional*

Pendahuluan

Inti dari pendidikan adalah pembelajaran sebagai proses interaksi manusiawi antara peserta didik (siswa) dengan pendidik (guru). Interaksi tersebut terjalin secara kompleks antar berbagai aspek dalam diri pribadi yang terlibat dalam proses pembelajaran. Dari sisi siswa, interaksi memberikan jaminan bahwa proses akan berjalan dan dapat menghasilkan keluaran (*output)* yang diharapkan, manakala siswa memiliki minat, motivasi, dan kemampuan untuk belajar. Teori belajar dan pembelajaran mengemukakan bahwa apabila siswa terlibat secara aktif dalam pembelajaran, maka *output* akan berkualitas. Teori pembelajaran inilah yang melahirkan pendekatan dan metode pembelajaran yang mengedepankan pada aktifitas siswa dalam belajar (*student active learning)*.

Dari sisi guru, proses pembelajaran akan menjamin *output* yang berkualitas apabila guru menguasai materi, metode penyajian, menjalin hubungan yang akrab dengan siswa, serta kemampuan untuk menjadikan dirinya menarik bagi siswa. Dengan kemampuannya inilah, maka proses pembelajaran menjadi menarik, mencerdaskan, dan membangkitkan (*joyfull)*.

Dalam Permendiknas Nomor 22 Tahun 2006 tentang Standar Isi disebutkan bahwa substansi mata pelajaran sosial di SD/MI, SMP/MTs, dan SMK dikemas dalam mata pelajaran IPS. Untuk menunaikan amanah tersebut, maka perlu dicarikan metode, teknik, serta cara pandang yang tepat untuk dapat memfasilitasi pelaksanaan pembelajaran yang sifatnya padu, agar dapat membantu peserta didik mencapai standar kompetensi yang ditentukan.

Saat ini mata pelajaran IPS SMP/MTs telah menyatukan seluruh kompetensi ilmu sosial dalam satu mata pelajaran.Model pembelajaran terpadu merupakan salah satu model implementasi kurikulum yang dianjurkan untuk diaplikasikan.Melalui pembelajaran terpadu, peserta didik dapat memperoleh pengalaman yang sesuai dengan kenyataan di masyarakat, sehingga dapat menambah kekuatan untuk menerima, menyimpan dan mereproduksi kesan-kesan tentang hal-hal yang dipelajarinya. Dengan demikian, peserta didik terlatih untuk dapat

menemukan sendiri berbagai konsep yang dipelajari secara holistik, bermakna, otentik, dan aktif.

Berdasarkan Standar Proses Pendidikan dalam Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan Pasal 19 Ayat (3) dan pasal 20 disebutkan bahwa:

> *Setiap satuan pendidikan melakukan perencanaan proses pembelajaran, pelaksanaan proses pembelajaran, penilaian hasil pembelajaran, dan pengawasan proses pembelajaran yang efektif dan efisien….Perencanaan proses pembelajaran meliputi silabus dan rencana pelaksanaan pembelajaran yang memuat sekurang-kurangnya tujuan pembelajaran, materi ajar, metode pengajaran, sumber belajar, dan penilaian hasil belajar.*

Memahami amanah peraturan pemerintah tersebut, IPS sebagai bagian integral dari kurikulum sekolah, dinilai memerlukan guru-guru yang mampu melaksanakan pembelajaran secara efektif, dalam arti menguasai materi, mampu memilih permasalahan yang layak diangkat sebagai bahan belajar, serta mengembangkan strategi pembelajaran yang mampu mengoptimalkan pencapaian kompetensi. Melalui pembelajaran terpadu dalam IPS, siswa belajar memahami kenyataan hidup di masyarakat dengan berbagai ragam masalahnya yang tidak dapat dipecahkan menggunakan kata kerja keilmuan secara terpisah.Pembelajaran terpadu dinilai sebagai alternatif untuk menjawabnya.

Ada tiga model pengintegrasian materi dalam pembelajaran IPS yaitu: model integrasi berdasarkan topik, model integrasi berdasarkan potensi utama, dan model integrasi berdasarkan masalah. Selanjutnya, langkah yang harus dilakukan dalam merancang pembelajaran terpadu adalah pengembangan atau pemilihan tema-tema sebagai pemersatu. Tema dikembangkan berangkat dari cabang atau disiplin yang ada dalam IPS atau dari konsep umum mata pelajran IPS, kemudian disoroti dari berbagai disiplin lainnya, atau tema dikembangkan dengan melibatkan berbagai mata pelajaran (Panduan Pengembangan Pembelajaran IPS Terpadu SMP/MTs, Depdiknas: 2006). Hal terpenting dalam mengembangkan tema-tema ini adalah bahwa tema itu mampu menumbuhkan prinsip pembelajaran yang

berpusat pada siswa (*student centred)* dan pembelajaran siswa aktif (*student active learning).*

Dalam sebuah program, pasti memiliki kelebihan dan kekurangan dalam pelaksanaannya. Oleh karena itu, penelitian ini dilakukan untuk dapat mengetahui kelemahan, kemudian dilakukan perbaikan dan penyempurnaan, sehingga diharapkan akan dapat meningkatan mutu pendidikan untuk masa mendatang. Banyak bukti yang menunjukkan bahwa sejumlah SMP/MTs tidak bisa melaksanakan kurikulum dengan baik. Hal ini dipengaruhi oleh beberapa aspek, yaitu perencanaan konsep kurikulum yang kurang sesuai dengan kondisi sekolah dan peserta didik, *stakeholders* sekolah yang tidak memahami metode pelaksanaannya, manajemen yang tidak profesional, serta keterbatasan kapabilitas (sumber daya manusia, perangkat lunak, perangkat keras, dan biaya).

Berdasarkan hasil wawancara dengan Kepala Seksi Kurikulum Pendidikan Menengah Kota Bima pada tanggal 23 September 2010, bahwa pelaksanaan pembelajaran IPS pada dua puluh enam SMP/MTs di Kota Bima, telah dimulai sejak tahun 2006/2007. Namun sampai saat ini, masih dijumpai adanya berbagai permasalahan yang terjadi di lapangan antara lain,guru kurang memahami konsep pembelajaran IPS yang padu, guru belum menyusunrencana pembelajaran IPS secara tepat, guru belum melaksanakan proses pembelajaran IPS secara padu, guru belum melaksanakan penilaian hasil belajar serta tindak lanjut hasil penilaian belajar IPS secara komprehensif, serta evaluasi terhadapprogram pembelajaran IPS SMP/MTs di Kota Bima belum pernah dilaksanakan, sehingga tidak diketahui tingkat kesesuaian antara pelaksanaan program pembelajaran IPS dengan standar proses pendidikan.

Berdasarkan latar belakang masalah, maka penelitian ini dibatasi pada permasalahan evaluasi terhadap program pembelajaran IPS SMP/MTs di Kota Bima yang belum pernah dilaksanakan, sehingga tidak diketahui tingkat kesesuaian antara pelaksanaan program pembelajaran IPS dengan standar proses pendidikan. Oleh karena itu, penelitian ini difokuskan pada evaluasi pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima, yang meliputi: perencanaan pembelajaran, pelaksanaan proses pembelajaran, dan penilaian hasil belajar dan tindak lanjut hasil penilaian.

Hal ini juga dijadikan input berharga bagi para guru IPS, kepala sekolah, serta seluruh *stakeholders* pendidikan dalam upaya perbaikan dan pengembangan program pembelajaran IPS pada masa yang akan datang.

Metode Penelitian

Penelitian ini merupakan penelitian evaluasi dengan menggunakan *Model Countenance Stake,* yang meliputi evaluasi *Antecedents, Transactions,* dan *Outputs* (Worthen & Sanders, 1973: 121). Berdasarkan model evaluasi Stake penelitian ini membandingkan kesesuaian antara pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan sebagai kriteria standar untuk menilai tingkat keberhasilan program IPS. Menurut Stake (Madaus, Schriven, & Stufflebeam, 1986: 287-308), Penekanan model ini adalah bahwa suatu evaluasi menekankan adanya pelaksanaan dua hal pokok, yaitu melakukan penggambaran (*description)* dan pertimbangan (*judgement)* mengenai sesuatu yang dievaluasi. Dengan demikian, tujuan pemilihan model evaluasi *Countenance Stake* dalam penelitian ini adalah untuk mengetahui kesesuaian antara pelaksanaan program pembelajaran IPS dengan standar proses pendidikan secara menyeluruh, yang meliputi perencanaan pembelajaran, pelaksanaan proses pembelajaran, penilaian hasil belajar dan tindak lanjut hasil penilaian, serta rerata hasil belajar kognitif IPS siswa pada 26 SMP/MTs di Kota Bima, tercantum pada tabel 1.

Berdasarkan tabel model evaluasi Stake program pembelajaran IPS SMP/MTs di atas, maka tahapan evaluasi dapat dijabarkan sebagai berikut:

1. Tahap perencanaan (*Antecedents).* Pada tahapan ini dilakukan evaluasi terkait kesesuaian antara perencanaan pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan.
2. Tahap pelaksanaan (*Transactions).* Pada tahapan ini dilakukan evaluasi terkait kesesuaian antara pelaksanaan proses pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan.
3. Tahap hasil (*Outputs).* Pada tahapan ini dilakukan evaluasi yang berfokus pada aspek *outputs* terkait kesesuaian antara penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS SMP/MTs di Kota Bima dengan

standar proses pendidikan, serta penilaian terhadap rerata hasil belajar IPS siswa pada 26 SMP/MTs di Kota Bima dengan standar Kriteria Ketuntasan Minimal belajar IPS SMP/MTs di Kota Bima.

Kesimpulan dari ketiga tahapan model evaluasi *Stake* tersebut adalah: (1) standar relatif (*relative standard)* yang menjelaskan pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima, dan (2) standar absolut (*absolute standard*), yakni mendasarkan pada standar atau kriteria standar proses pendidikan. Selanjutnya, dapat disimpulkan terkait kesesuaian antara pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan.

Tabel 1. Model Evaluasi *Countenance Stake* Program Pembelajaran IPS diSMP/MTs Kota Bima.

No		*Description Matrix*		*Judgement Matrix*
1.	*Antecedents*	Perencanaan pembelajaran IPS SMP/MTs di Kota Bima	← Kesesuaian →	Perencanaan pembelajaran IPS sesuai standar proses pendidikan.
2.	*Transactions*	Pelaksanaan Proses pembelajaran IPS SMP/MTs di Kota Bima	← Kesesuaian →	Pelaksanaan Proses pembelajaran IPS sesuai standar proses pendidikan
3.	*Outputs*	Penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS SMP/MTs di Kota Bima dan rerata hasil belajar kognitif IPS siswa	← Kesesuaian →	Penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS sesuai standar proses pendidikan dan KKM Belajar IPS SMP/MTS di Kota Bima

Subjek penelitian adalah guru IPS sejumlah 78 orang yang diambil secara populasi dan siswa sejumlah 390 orang yang dipilih secara *Quota Stratified Random Sampling*. Teknik pengumpulan data yang digunakan adalah angket, tes, dan dokumentasi. Instrumen yang digunakan adalah kuesioner, tes, dan daftar cek. Untuk memperjelas deskripsi teknik pengumpulan data dalam evaluasi program pembelajaran IPS SMP/MTs di Kota Bima ini, maka dapat disajikan dalam tabel 2 sebagai berikut.

Tabel 2. Deskripsi Teknik Pengumpulan Data Evaluasi Program Pembelajaran IPS SMP/MTs di Kota Bima

No	Tahapan Model Evaluasi Countenance Stake	Komponen Evaluasi	Teknik dan Alat Pengumpulan Data	Sumber Data
1.	*Antecedents*	Perencanaan pembelajaran IPS	Dokumentasi (Daftar Cek)	Dokumen Rencana Pelaksanaan Pembelajaran (RPP)
2.	*Transactions*	Pelaksanaan Proses pembelajaran IPS	Angket (Kuesioner)	Siswa Kelas VII, VIII, IX
3.	*Outputs*	Penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS	1. Angket (Kuesioner) 2. Tes	Guru IPS Siswa

Validasi instrumen menggunakan validasi logis dari pertimbangan ahli dan validari empiris dengan menggunakan perhitungan *Korelasi Product Moment*. Sedangkan perhitungan reliabilitasinstrumen menggunakan rumus

Alpha Cronbach. Koefisien reliabilitasangket pada aspek pelaksanaan proses pembelajaran IPS yaitu sebesar 0,842 dan penilaian hasil belajar dan tindak lanjut hasil penilaian sebesar 0,772. Berdasarkan hasil uji coba, seluruh instrumen dinyatakan valid dan reliabel.

Penelitian ini menggunakan analisis data dengan teknik deskriptif kuantitatif yang bertujuan untuk memberikan gambaran realitas tentang pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima. Data dari hasil kuesioner, tes, dan daftar cek dianalisis secara deskriptif kuantitatif dengan cara membandingkan persentase perolehan skor pada tiap variabel yang diukur dengan kriteria penilaian yang telah ditentukan. Besarnya persentase dari hasil perhitungan menunjukkan kategori data yang terungkap, sehingga dapat diketahui posisi masing-masing variabel dalam keseluruhan maupun bagian dari variabel yang diteliti. Data dideskripsikan dengan mentabulasikan menurut masing-masing variabel. Berdasarkan perhitungan, maka diperoleh harga: mean, median, modus, skor minimum, dan skor maksimum.

Dalam mendeskripsikan aspek pelaksanaan program pembelajaran IPS, digunakan skor kriteria sebagai norma perbandingan yang digunakan sebagai kriteria perbandingan dalam menyusun kriteria empiris, yaitu kriteria yang disusun atau dikembangkan berdasarkan kondisi lapangan, yang diukur atau mengacu pada aspek program pembelajaran IPS, yang meliputi: perencanaan pembelajaran, pelaksanaan proses pembelajaran, penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS, serta rerata hasil belajar kognitif siswa pada 26 SMP/MTs di Kota Bima. Adapun perhitungan kriteria data evaluasi yang diperoleh melalui teknik angket, tes, dan dokumentasi untuk masing-masing tahapan evaluasi dapat disajikan sebagai berikut:

Evaluasi tahapan *antecedents*: Kriteria aspek perencanaan pembelajaran IPS

Kriteria penilaian pada aspek perencanaan pembelajaran IPS ini diukur dari 20 butir pernyataan dengan pilihan jawaban menurut skala Likert dengan penskoran yang digunakan adalah skala 4, 3, 2, dan 1. Skor terendah yang diperoleh dari tiap-tiap respon adalah sebesar 20, sedangkan

skor tertinggi adalah sebesar 80. Skor minimum dan maksimum dari datanya adalah sebesar 20 (1 x 20) sampai dengan 80 (4 x 20), sehingga luas jarak sebenarnya adalah 80 – 20 = 60. Dengan demikian, Mean idealnya (Mi) sebesar 50, yakni (1/2 (80 + 20). Hasil perhitungan kriteria tersebut disajikan dalam tabel 3 sebagai berikut.

Tabel 3. Hasil Perhitungan Kriteria Perencanaan Pembelajaran IPS

Interval Skor	Interpretasi/Kriteria
65 – 80	Sangat Baik/ Sangat Sesuai
50 – 64	Baik/Sesuai
35 – 49	Cukup Baik/Cukup Sesuai
20 - 34	Kurang Baik/Kurang Sesuai

Berdasarkan acuan kriteria yang dijadikan patokan penilaian pada tabel 2 tersebut, maka skor yang berada pada rentang skor 20-34 dikelompokkan pada kategori kurang baik/kurang sesuai, berarti aspek perencanaan pembelajaran IPS yang dilakukan oleh guru dinilai kurang baik/kurang sesuai. Rentang skor 35-49 dikelompokkan pada kategori cukup baik/cukup sesuai. Rentang skor 50-64 dikelompokkan pada kategori baik/sesuai dan rentang skor 65-80 dikelompokkan pada kategori sangat baik/sangat sesuai.

Evaluasi tahapan *Transactions:* Kriteria aspek pelaksanaan proses pembelajaran IPS

Kriteria penilaian pada aspek pelaksanaan proses pembelajaran IPS ini diukur dari 42 butir pernyataan dengan pilihan jawaban menurut skala Likert dengan penskoran yang digunakan adalah skala 4, 3, 2, dan 1. Skor terendah yang diperoleh dari tiap-tiap respon adalah sebesar 42, sedangkan skor tertinggi adalah sebesar 168. Skor minimum dan maksimum dari datanya adalah sebesar 42 (1 x 42) sampai dengan 168 (4 x 42), sehingga luas jarak sebenarnya adalah 168 – 42 = 126. Dengan demikian, Mean

idealnya (Mi) sebesar 105, yakni (1/2 (168 + 42). Hasil perhitungan kriteria tersebut dapat dilihat pada tabel 4 sebagai berikut.

Tabel 4. Hasil Perhitungan Kriteria Pelaksanaan Proses Pembelajaran IPS

Interval Skor	Interpretasi/Kriteria
136,50 – 168	Sangat Baik/Sangat Sesuai
105 – 135,50	Baik/Sesuai
73,50 – 104	Cukup Baik/Cukup Sesuai
42 – 72,50	Kurang Baik/Kurang Sesuai

Berdasarkan acuan kriteria yang dijadikan patokan penilaian pada tabel 3 tersebut, maka skor yang berada pada rentang skor 42-72,5 dikelompokkan pada kategori kurang baik/kurang sesuai, berarti aspek proses pembelajaran IPS yang dilaksanakan oleh guru dinilai kurang baik/kurang sesuai. Rentang skor 73,5-104 dikelompokkan pada kategori cukup baik/cukup sesuai. Rentang skor 105-135,5 dikelompokkan pada kategori baik/sesuai dan rentang skor 136,5-168 dikelompokkan pada kategori sangat baik/sangat sesuai.

Evaluasi tahapan *Outputs:* Kriteria penilaian aspek penilaian hasil belajar dan tindak lanjut hasil penilaian, serta rerata hasil belajar kogntif siswa

Kriteria penilaian pada aspek penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS ini diukur dari 24 butir pernyataan dengan pilihan jawaban menurut skala Likert dengan penskoran yang digunakan adalah skala 4, 3, 2, dan 1. Skor terendah yang diperoleh dari tiap-tiap respon adalah sebesar 24, sedangkan skor tertinggi adalah sebesar 96. Skor minimum dan maksimum dari datanya adalah sebesar 24 (1 x 24) sampai dengan 96 (4 x 24), sehingga luas jarak sebenarnya adalah 96 – 24 = 72.

Dengan demikian, Mean idealnya (Mi) sebesar 60, yakni (1/2 (96 + 24). Hasil perhitungan kriteria tersebut dapat dilihat pada tabel 5.

Tabel 5. Hasil Perhitungan Kriteria Penilaian Hasil Belajar dan Tindak Lanjut Hasil Penilaian Belajar IPS

Interval Skor	Interpretasi/Kriteria
78 – 96	Sangat Baik/Sangat Sesuai
60 – 77	Baik/Sesuai
42 – 59	Cukup Baik/Cukup Sesuai
24 – 41	Kurang Baik/Kurang Sesuai

Berdasarkan acuan kriteria yang dijadikan patokan penilaian pada tabel 4 tersebut, maka skor yang berada pada rentang skor 24-41 dikelompokkan pada kategori kurang baik/kurang sesuai, berarti aspek penilaian hasil belajar serta tindak lanjut hasil penilaian belajar IPS yang dilakukan oleh guru dinilai kurang baik/kurang sesuai. Rentang skor 42-59 dikelompokkan pada kategori cukup baik/cukup sesuai. Rentang skor 60-77 dikelompokkan pada kategori baik/sesuai dan rentang skor 78-96 dikelompokkan pada kategori sangat baik/sangat sesuai.

Adapun kriteria penilaian untuk mengukur *outputs* dari rata-rata hasil belajar kognitif IPS siswa pada 26 SMP/MTs di Kota Bima, diukur dari 25 butir soal jenis objektif pilihan ganda dengan pilihan 4 pilihan jawaban. Adapun teknik penskoran yang digunakan adalah skor 1 untuk jawaban yang benar dan skor 0 untuk jawaban yang tidak benar. Skor terendah yang diperoleh dari tiap-tiap respon adalah sebesar 1, sedangkan skor tertinggi adalah sebesar 25. Dengan demikian, Mean idealnya (Mi) sebesar 13. Hasil perhitungan kriteria tersebut dapat dilihat pada (tabel 5) sebagai berikut.

Tabel 6. Hasil Perhitungan Kriteria ideal Penilaian Hasil belajar Kognitif IPS Siswa SMP/MTs

Interval Skor Rerata	Interpretasi/Kriteria
19 – 25	Sangat Baik/Sangat Sesuai
13 – 18	Baik/Sesuai
7 – 12	Cukup Baik/Cukup Sesuai
1 – 6	Kurang Baik/Kurang Sesuai

Berdasarkan acuan kriteria yang dijadikan patokan penilaian pada tabel 5 tersebut, maka skor yang berada pada rentang skor 1-6 dikelompokkan pada kategori kurang baik/kurang sesuai. Rentang skor 7-12 dikelompokkan pada kategori cukup baik/cukup sesuai. Rentang skor 13-18 dikelompokkan pada kategori baik/sesuai dan rentang skor 19-25 dikelompokkan pada kategori sangat baik/sangat sesuai. Dari hasil perhitungan rerata skor hasil belajar kognitif IPS siswa pada tiap sekolah tersebut, selanjutnya dibandingkan dengan Kriteria Ketuntasan Minimal (KKM) skor pada mata pelajaran IPS SMP/MTs di Kota Bima, yakni sebesar 68.

Berdasarkan hasil analisis data dari berbagai instrumen di atas dapat dirangkum dan disimpulkan secara deskriptif kuantitatif menjadi data yang akurat untuk menjawab pertanyaan-pertanyaan penelitian terkait evaluasi program pembelajaran IPS SMP/MTs di Kota Bima.

Hasil Penelitian dan Pembahasan

Deskripsi Data Hasil Penelitian

Penelitian ini merupakan penelitian evaluasi dengan menggunakan Model Countenance Stake, yang meliputi evaluasi *Antecedents, Transactions,* dan *Outputs* dalam membandingkan kesesuaian antara pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan sebagai kriteria standar untuk menilai tingkat keberhasilan

program IPS. Penekanan model ini adalah bahwa suatu evaluasi menekankan adanya pelaksanaan dua hal pokok, yaitu melakukan penggambaran (description) dan pertimbangan (judgement) mengenai sesuatu yang dievaluasi. Dengan demikian, tujuan pemilihan model evaluasi Countenance Stake dalam penelitian ini adalah untuk mengetahui kesesuaian antara pelaksanaan program pembelajaran IPS dengan standar proses pendidikan secara menyeluruh, yang meliputi perencanaan pembelajaran, pelaksanaan proses pembelajaran, penilaian hasil belajar dan tindak lanjut hasil penilaian, serta rerata hasil belajar kognitif IPS siswa pada 26 SMP/MTs di Kota Bima.

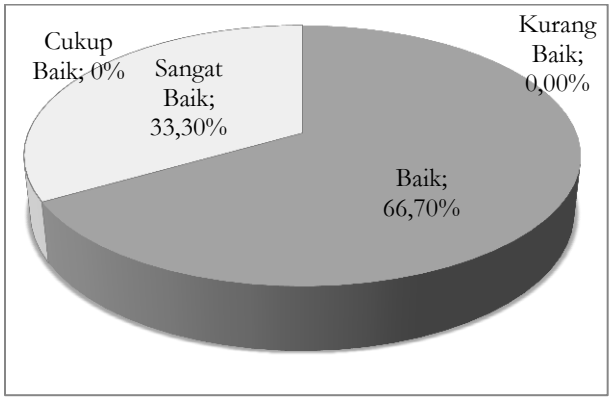
Pada bagian ini disajikan deskripsi data hasil penelitian dari masing-masing variabel dari objek penelitian program pembelajaran IPS SMP/MTs di Kota Bima. Deskripsi data yang disajikan meliputi rata-rata (mean), median, modus, standar deviasi, skor minimum, skor maksimum, dan distribusi frekuensi. Hasil perhitungan skor tersebut diperoleh dengan pengolahan data statistik berdasarkan program SPSS versi 17,0.

a. Perencanaan Pembelajaran IPS SMP/MTs di Kota Bima

Menurut Smith & Ragan (Shambaugh & Magliaro, 2006:35), perencanaan pembelajaran sebagai proses sistematis dalam menenjemahkan prinsip-prinsip belajar dan pembelajaran ke dalam rencana kegiatan dan bahan pembelajaran. Hal ini menunjukkan bahwa guru harus mempersiapkan perangkat yang harus disusun dalam merencanakan program pembelajaran. Hal tersebut menunjukkan bahwa perencanaan pembelajaran IPS memainkan peran penting dalam memandu guru untuk melaksanakan tugas sebagai pendidik dalam melayani kebutuhan belajar siswa.

Hasil analisis data membuktikan, kesesuaian antara perencanaan pembelajaran IPS SMP/MTs di Kota Bima dengan standar proses pendidikan berada pada kategori baik/sesuai. Kategori ini diperoleh berdasarkan analisa perhitungan skor rerata empiris data dari teknik dokumentasi Rencana Pelaksanaan Pembelajaran (RPP) yang dibuat oleh 78 orang guru IPS dari dua puluh enam SMP/MTs se-Kota Bima berdasarkan indikator-indikator variabel perencanaan pembelajaran, yakni

sebesar 63,91, dengan persentase sebesar 66,7%. Hasil penelitian ini menunjukkan bahwa guru telah menyusun RPP secara lengkap, tepat, dan penjelasan yang terperinci. Di samping itu, hasil penelitian ini juga menunjukkan bahwa 33,3% perencanaan pembelajaran IPS SMP/MTs di Kota Bima termasuk dalam kategori sangat baik/sangat sesuai, namun tidak ada guru yang termasuk dalam kategori cukup baik/cukup sesuai dan kurang baik/kurang sesuai. Untuk menunjukkan persentase skor penilaian pada variabel perencanaan pembelajaran IPS tersebut dapat dilihat pada gambar 1 di bawah ini.



Gambar 1. Diagram LingkaranVariabelPerencanaanPembelajaran IPS

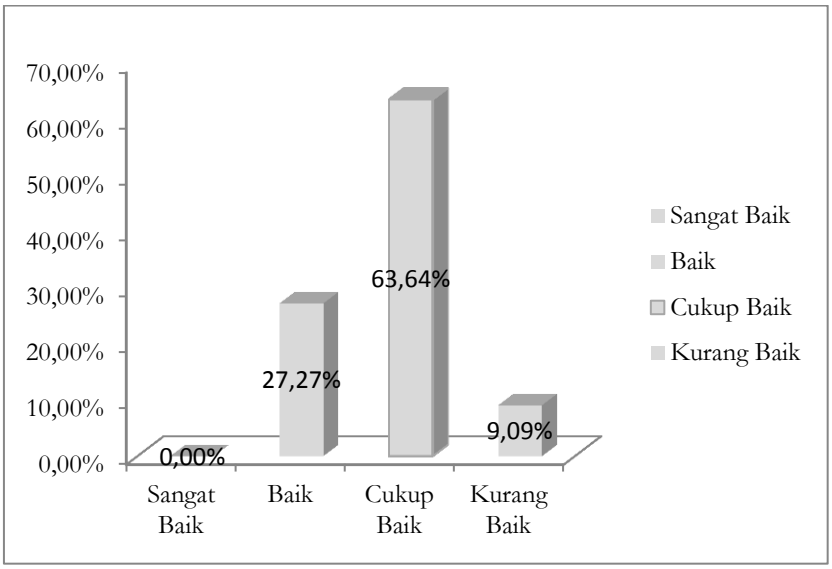
Berdasarkan diagram lingkaran, terdapat 33,3 % guru IPS SMP/MTs di Kota Bima menyusun perencanaan pembelajaran IPS dengan kategori baik/sesuai; 66,7 % termasuk dalam kategori cukup baik/cukup sesuai dan tidak ada guru IPS SMP/MTs di Kota Bima yang termasuk dalam kategori kurang baik/kurang sesuai dan tidak baik/tidak sesuai.

b. Pelaksanaan Proses Pembelajaran IPS SMP/MTs di Kota Bima

Pelaksanaan proses pembelajaran menjadi sesuatu yang sangat penting dalam upaya mewujudkan *output* pendidikan yang berkualitas. Konsekuensinya, proses pembelajaran harus dilaksanakan secara tepat,

ideal, dan proporsional sehingga guru harus memiliki kemampuan dalam melaksanakan teori pembelajaran ke dalam realitas pembelajaran yang kontekstual antara lain, mendesain, memfasilitasi, dan mengkondisikan situasi pembelajaran agar dapat dijadikan sarana untuk membimbing dan mengembangkan potensi siswa.

Pendekatan terpadu dalam mata pelajaran IPS lebih menekankan keterlibatan siswa dalam belajar, membuat siswa secara aktif terlibat dalam proses pembelajaran. Pendekatan pembelajaran terpadu dapat dipandang sebagai suatu upaya untuk memperbaiki kualitas pendidikan di tingkat dasar terutama dalam rangka penjejalan materi yang sering terjadi dalam pelaksanaan proses pembelajaran di sekolah dan lingkungan kehidupan sekitar siswa.



Gambar 2. Histogram Variabel Pelaksanaan Proses Pembelajaran IPS

Berdasarkan hasil analisis data, kesesuaian antara pelaksanaan proses pembelajaran IPS di Kota Bima dengan standar proses pendidikan berada pada kategori cukup baik/cukup sesuai. Kategori ini diperoleh berdasarkan

analisa perhitungan skor rerata empiris dari teknik angket yang diberikan kepada 390 responden siswa dari dua puluh enam SMP/MTs se-Kota Bima berdasarkan indikator-indikator variabel proses pembelajaran IPS, yakni sebesar 104,75, dengan persentase sebesar 63,64%. Hasil penelitian ini menunjukkan bahwa pelaksanaan proses pembelajaran IPS dilaksanakan secara bertahap, cukup sistematis, dan cukup menggunakan pembelajaran IPS yang padu. Untuk menunjukkan persentase skor penilaian pada variabel pelaksanaan proses pembelajaran IPS tersebut dapat dilihat pada (Gambar 3) di bawah ini.

Berdasarkan gambar di atas, dalam pelaksanaan proses pembelajaran tersebut terdapat 27,27% guru IPS yang termasuk dalam kategori baik/sesuai dan 9,09% termasuk dalam kategori kurang baik/kurang sesuai. Namun tidak ada guru yang termasuk dalam kategori sangat baik/sangat sesuai dalam pelaksanaan proses pembelajaran IPS.

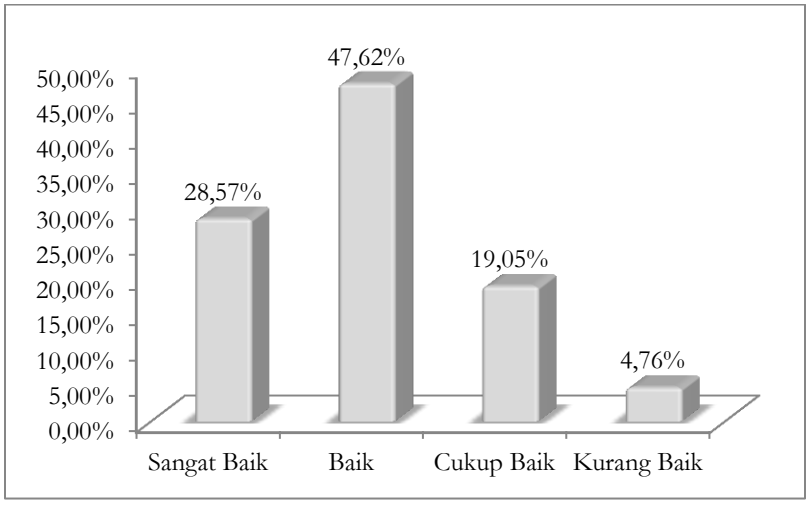
c. Penilaian Hasil Belajar dan Tindak Lanjut Hasil Penilaian Belajar IPS SMP/MTs di Kota Bima.

Evaluasi program pembelajaran IPS merupakan kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang proses dan hasil belajar siswa yang dilakukan secara sistematis dan berkesinambungan sehingga menjadi informasi yang bermakna dalam pengambilan keputusan, memberikan umpan balik bagi siswa dalam mengetahui kemampuan dan kekurangannya dan akhirnya menimbulkan motivasi untuk memperbaiki hasil belajarnya.

Berdasarkan hasil analisis data, kesesuaian antara penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS di Kota Bima dengan standar proses pendidikan termasuk dalam kategori cukup baik/cukup sesuai. Kategori ini diperoleh berdasarkan analisa perhitungan skor rerata empiris data dari teknik angket yang diberikan kepada 78 orang guru dari dua puluh enam SMP/MTs se-Kota Bima berdasarkan indikator-indikator variabel Penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS, yakni sebesar70,94, dengan persentase sebesar 47,62%. Hasil penilitian ini menunjukkan bahwa penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS telah dilaksanakan oleh guru IPS SMP/MTs di Kota Bima

secara bertahap dan komprehensif. Di samping itu, terdapat 28,57% guru yang termasuk dalam kategori sangat baik/sangat sesuai; 19,05% termasuk dalam kategori cukup baik/cukup sesuai; dan 4,76% termasuk dalam kategori kurang baik/kurang sesuai.

Untuk menunjukkan persentase skor penilaian pada variabel penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS tersebut dapat dilihat pada gambar 4.



Gambar 3. Diagram Kerucut Variabel Penilaian Hasil Belajar dan Tindak Lanjut Hasil Penilaian Belajar IPS

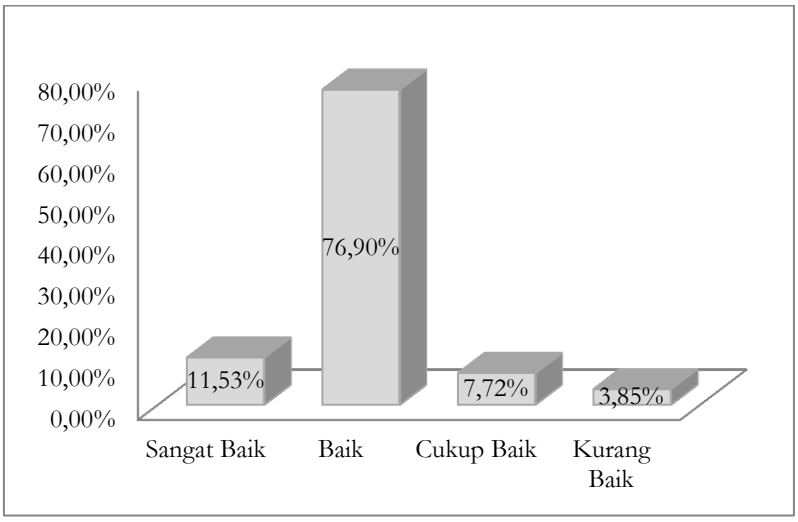
Berpedoman pada kategori penilaian analisis data, rerata empiris variabel penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS adalah sebesar 70,94 dan termasuk dalam kategori baik/sesuai.

d. Hasil Belajar Kognitif IPS siswa SMP/MTs di Kota Bima

Hasil belajar sebagai pencapaian kompetensi yang diperoleh siswa setelah mengikuti kegiatan belajar.Siswa yang berhasil dalam belajar adalah siswa yang berhasil mencapai tujuan-tujuan pembelajaran yang telah

ditentukan.Tes digunakan untuk mengukur sejauh mana siswa telah menguasai materi yang telah disampaikan dalam pembelajaran.Tes merupakan himpunan pertanyaan yang harus dijawab, ditanggapi, atau berupa tugas yang harus dilaksanakan oleh orang yang dites.Berkaitan dengan hal tersebut, tes terhadap domain kognitif berupa kemampuan aspek pengetahuan siswa tentang materi pembelajaran.Dalam penelitian ini, penilaian hanya dibatasi pada evaluasi terhadap aspek kognitif IPS siswa SMP/MTs di Kota Bima.

Hasil analisis data hasil belajar kognitif IPS siswa pada 26 SMP/MTs di Kota Bima diperoleh rerata (*mean)* sebesar 81,05. Berpedoman pada kategorisasi interval rerata skor ideal yang telah ditentukan, maka rerata hasil analisis data hasil belajar kognitif IPS siswa berada pada kategori baik/sesuai berdasarkan standar Kriteria Ketuntasan Minimal (KKM) yang berlaku di Kota Bima. Untuk menunjukkan persentase perolehan skor rerata hasil belajar kognitif IPS siswa SMP/MTs di Kota Bima tersebut dapat dilihat pada (Gambar 5) sebagai berikut.



Gambar 4. Diagram Batang Perolehan Skor Rerata Hasil Belajar Kognitif IPS Siswa SMP/MTs di Kota Bima

Berdasarkan gambar tersebut, secara umum perolehan skor rerata hasil belajar kognitif IPS siswa SMP/MTs di Kota Bima termasuk dalam kategori baik/sesuai dengan persentase sebesar 76,90% dari 26 sekolah. Di samping itu, terdapat 11,53% sekolah yang termasuk dalam kategori sangat baik/sangat sesuai; 7,72% termasuk dalam kategori cukup baik/cukup sesuai; dan 3,85% termasuk dalam kategori kurang baik/kurang sesuai.

Hasil analisis data skor hasil belajar kognitif IPS siswa tersebut menunjukkan bahwa penilaian hasil belajar dan tindak lanjut hasil penilaian yang dilaksanakan oleh seluruh guru IPS SMP/MTs di Kota Bima telah sesuai standar proses pendidikan dan kondisi serta karakteristik peserta didik pada masing-masing satuan pendidikan.

Simpulan

Hasil analisis data menunjukkan bahwa kesesuaian antara pelaksanaan program pembelajaran IPS SMP/MTs di Kota Bima dapat disimpulkan sebagai berikut:

1. Perencanaan pembelajaran IPS SMP/MTs di Kota Bima berada pada kategori baik/sesuai dengan standar proses pendidikan. Hal tersebut menunjukkan bahwa perencanaan pembelajaran IPS SMP/MTs di Kota Bima disusun dengan lengkap, tepat, dan penjelasan yang terperinci.
2. Pelaksanaaan proses pembelajaran IPS SMP/MTs di Kota Bima berada pada kategori cukup baik/cukup sesuai dengan standar proses pendidikan. Hal tersebut menunjukkan bahwa proses pembelajaran IPS SMP/MTs di Kota Bima dilaksanakan secara bertahap, cukup sistematis dan menggunakan pendekatan pembelajaran IPS yang padu.
3. Penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS SMP/MTs di Kota Bima berada pada kategori baik/sesuai dengan standar proses pendidikan. Hal tersebut menunjukkan bahwa penilaian hasil belajar dan tindak lanjut hasil penilaian belajar IPS dilaksanakan secara bertahap dan komprehensif.
4. Hasil belajar kognitif IPS siswa berada pada kategori baik/sesuai standar Kriteria Ketuntasan Minimum (KKM) belajar SMP/MTs di Kota Bima yang ditetapkan pada tahun ajaran 2010/2011.

Berdasarkan hasil penelitian ini diharapkan dapat mengembangkan pembelajaran IPS sesuai KTSP yang berlandaskan pada standar proses pendidikan pada tingkat satuan pendidikan masing-masing melalui upaya pelatihan dan monitoring secara berkala dan berkesinambungan pada satuan pendidikan masing-masing melalui wadah Musyawarah Guru Mata Pelajaran (MGMP) IPS oleh pihak kepala sekolah maupun Dikpora Kota Bima selaku pengambil kebijakan pendidikan lokal.

Di samping itu, disarankan kepada pihak satuan pendidikan maupun Dikpora Kota Bima agar memberikan penghargaan bagi para guru IPS yang berprestasi dalam pelaksanaan program pembelajaran IPS, sehingga dapat memacu kreatifitas dan semangat diri untuk terus mengembangkan pembelajaran yang kreatif, aktif, menantang, inovatif, dan menyenangkan bagi para peserta didik.

Daftar Pustaka

Depdiknas. 2006. Panduan Pengembangan Pembelajaran IPS Terpadu SMP/MTs

Depdiknas. 2005.Peraturan Pemerintah RI Nomor 19, Tahun 2005, tentang Standar Nasional Pendidikan.

_____. 2006. Permendiknas RI Nomor 22, Tahun 2006, tentang Standar Isi.

Madaus, G. Scriven, M. & Stufflebeam, L. Daniel. 1986. Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff Publishing.

Shambaugh, N. & Magliaro, G. Susan. 2006. Instructional design a systematic approach for reflective practice. United States of America: Pearson Education, Inc.

Worthen, R. & Sanders, J. 1973. Educational evaluation theory and practice.California: Wardsworth Publishing Company.

MENGGUNAKAN MODEL DINA DALAM PENGEMBANGAN TES DIAGNOSTIK UNTUK MENDETEKSI SALAH KONSEPSI

Kusaeri
Institut Agama Islam Negeri Sunan Ampel Surabaya
erinda_1@yahoo.co.id

Abstrak

Penelitian ini bertujuan 1) menemukan cara mengembangkan tes diagnostik dengan model DINA, sehingga mampu memberikan informasi salah konsepsi dalam aljabar, 2) mengidentifikasi karakteristik tes diagnostik yang baik yang dikembangkan dengan model DINA. Penelitian pengembangan ini merupakan penelitian empirik, dengan pendekatan deskriptif eksploratori. Pendekatan tersebut digunakan untuk mendeskripsikan tahapan pengembangan tes dan mencari karakteristik item. Subjek penelitian adalah siswa kelas VIII SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul dan SMPN 1 Panjatan Kulon Progo. Data dianalisis dengan *software* CDM, Mplus dan R. Hasil penelitian menunjukan 1) tahapan pengembangan tes meliputi: identifikasi kompetensi dasar dan merumuskan indikator, menyusun *learning continuum*, menyusun hierarki materi, merumuskan atribut, mengonstruk soal, validasi ahli dan uji empirik. Setelah melalui tujuh tahapan, dikembangkan 37 item tes diagnostik, 2) dari 37 item tes diagnostik, 15 item di antaranya harus dihilangkan/dihapus dari paket tes, karena tidak memenuhi uji fit model dan kualitas item tersebut jelek (indeks daya beda kurang dari 0,2).

Kata kunci: *DINA, latent class, atribut, salah konsepsi, dan aljabar*

DINA IN DEVELOPMENT MODEL USING DIAGNOSTIC TESTS FOR DETECTING WRONG CONCEPTION

Kusaeri
Institut Agama Islam Negeri Sunan Ampel Surabaya
erinda_1@yahoo.co.id

Abstract

This research aims to 1) find out a way to develop the diagnostic test using the DINA model, so that it can give information about misconceptions in algebra, 2) Identify the characteristics of the good diagnostic test developed by using the DINA model. This research and development is an empirical research with the descriptive explanatory used to describe the development stages of the test. The subjects of this research were year VIII students of SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul, and SMPN 1 Panjatan, Kulon Progo. The data were analyzed by using the CDM, Mplus, and R software. The results of the study are as follows 1) The stages of the test development in this research were: identifying basic competence and formulating indicators, constructing the learning continuum, constructing the material hierarchy, formulating the attributes, constructing the problems, conducting validation by expert judgment, and administering an empirical test. Through those seven stages, 37 items of the diagnostic test were developed, 2) Of the 37 items, 15 items must be eliminated/discarded from the test. The items were eliminated because their quality was low and they did not meet the requirements of the model fit test because their discrimination indexes were less than 0.2.

Keywords: *DINA, latent class, attribute, misconceptions, and algebra*

Pendahuluan

Ketika anak memasuki bangku kelas VII SMP, mereka dihadapkan pada rumpun matematika baru, yaitu aljabar. Materi ini sebagai pengembangan dari aritmetika yang dipelajarinya sewaktu di sekolah dasar. Menurut Rakes (2010:44-45), anak yang awal dalam belajar aljabar akan menghadapi berbagai kendala, paling tidak berasal dari dua sumber.

Pertama, mempelajari aljabar menuntut anak mempelajari bahasa simbol matematika yang benar-benar asing dengan pengalaman sebelumnya. Pada aritmetika, anak banyak menghadapi dan memanipulasi simbol berupa angka. Dengan angka, anak dapat segera membayangkan seberapa besar yang disimbolkan. Namun, aljabar tidak hanya menggunakan simbol angka, melainkan juga huruf ataupun kombinasi angka dan huruf. Berbagai cara simbol aljabar digunakan dan dideskripsikan oleh guru selama pembelajaran, sehingga sering kali menyebabkan anak kesulitan mengaitkan simbol aljabar dengan makna simbol yang dimaksud.

Kedua, aljabar merupakan pelajaran yang menuntut anak mengembangkan penalaran abstrak dan pemecahan masalah. Sifat aljabar yang demikian membuat aljabar lebih sulit bagi anak SMP dibandingkan aritmetika. Kenyataan ini mempengaruhi kemampuan anak mengkonstruk berbagai representasi objek aljabar, sehingga berdampak pada banyak konsep aljabar dipahaminya secara salah.

Chi (2008:16) mengaitkan kendala yang dihadapi anak di atas dengan kemampuan awal (berupa aritmetika yang telah dimiliki) dan mengklasifikasikan mereka dalam tiga kelompok. *Pertama*, anak yang tidak memiliki pengetahuan awal tentang aljabar. Walaupun mereka memiliki beberapa pengetahuan relevan yang dipelajarinya di aritmetika, namun pengetahuan itu tidak hadir (*missing*). Hal itu disebabkan objek yang dipelajari dalam aritmetika berbeda dengan objek aljabar. Dalam konteks demikian, belajar aljabar merupakan kegiatan menambah pengetahuan baru. *Kedua,* anak mungkin memiliki beberapa pengetahuan awal yang cukup tentang konsep aljabar yang dipelajari, namun pengetahuan awal itu tidak lengkap (*incomplete*). Dalam kondisi demikian, belajar aljabar dianggap

sebagai proses pengisian celah (*gap*) terhadap konsep yang ada. *Ketiga,* anak memiliki bekal pengetahuan awal yang memadai terkait dengan aritmetika, namun pengetahuan awal itu berbeda dengan konsep yang sedang dipelajari. Dalam kondisi seperti ini, sangat mungkin terjadi perubahan konsep (*conceptual change*) yang telah dimilikinya, namun juga sangat mungkin terjadi salah konsepsi akibat adanya konflik antara konsep lama dengan konsep baru.

Salah konsepsi terjadi apabila anak gagal menghubungkan pengetahuan baru dengan pengetahuan sebelumnya (Russel & O'dwyer, 2009:414). Salah konsepsi muncul ketika anak secara salah menerapkan strategi pengetahuan yang dipelajari sebelumnya guna menyelesaikan permasalahan baru. Salah konsepsi mempengaruhi secara mendasar bagaimana anak memahami konsep tertentu, sehingga menyebabkan terjadinya bentuk kesalahan lainnya. Dengan demikian, salah konsepsi pada anak harus dicegah, dieliminir dan bila perlu dihilangkan. Untuk mencegah atau mendeteksi terjadinya salah konsepsi, Booth (Xiabao Li, 2006) menyarankan agar diidentifikasi jenis kesalahan yang dilakukan anak serta dilakukan investigasi alasan-alasan munculnya kesalahan tersebut. Dalam kaitan inilah pentingnya dilakukan kegiatan diagnosis.

Kegiatan diagnosis bagi anak yang mengalami masalah belajar berbasis pada teori tes klasik (*Classical Test Theory*, CTT) dan teori respon butir (*Item Response Theory*, IRT) telah banyak dilakukan. Sejumlah penelitian diagnosis berbasis keduanya telah dilakukan, seperti penelitian Sugiharto (2003), Fauzan (2010), Waskito (2010), Suwarto (2011) dan Awal Isgiyanto (2011). Penelitian ini biasanya dilakukan melalui dua tahapan, yakni: (a) menetapkan skor masing-masing anak dan (b) menetapkan *cut-off score* (skor batas) yang tepat. Skor yang diperoleh masing-masing anak, selanjutnya dibandingkan dengan *cut-off score* yang telah ditetapkan. Kedua prosedur di atas, menurut Templin (2011) rawan muncul kesalahan sehingga mempengaruhi hasil yang didapat. Sumber kesalahan umumnya terletak pada penetapan *cut-off score,* karena hanya didasarkan pada estimasi semata.

Sebaliknya, penelitian diagnosis berbasis *latent class* belum banyak dilaku-kan. Prinsip dasar yang digunakan model ini adalah menempatkan peserta tes ke dalam satu dari dua kelompok, yakni kelompok menguasai

(*mastery*) atau kelompok tidak menguasai (*non-mastery*). Hal ini menurut Templin (2011:43) lebih mudah dilakukan dibandingkan dengan menempatkan peserta tes pada sebuah skala. Salah satu model diagnosis berbasis *latent class* yang saat ini sedang berkembang dan banyak menarik kalangan peneliti di Amerika Serikat dan Eropa adalah model DINA.

DINA berasal dari kata "*deterministic input*, *noisy,* dan "AND" *gate."* Komponen "*deterministic input*" menggambarkan kemampuan seorang anak dalam menjawab suatu item (benar atau salah) didasarkan pada penguasaan mereka terhadap atribut yang diukur (Rupp et al., 2010:319). Atribut merupakan kemampuan atau kompetensi yang harus dimiliki anak agar mampu menyelesaikan suatu item (Kusaeri, 2012:17). Penguasaan terhadap atribut tersebut direpresentasi-kan dalam sebuah matriks Q, yaitu matriks dengan M baris dan N kolom yang unsur-unsur di dalamnya terdiri atas bilangan 0 dan 1. Unsur pada matriks Q akan bernilai 1 apabila seorang anak menguasai semua atribut yang diperlukan menyelesaikan item tertentu. Sebaliknya, bernilai 0 bila anak tidak menguasai salah satu atribut yang dipersyaratkan (Ying Liu et al., 2009:58; de la Torre et al., 2010:229).

Komponen "*noisy"* berkaitan dengan parameter *slip* dan *guessing.* Artinya, seorang anak yang menguasai seluruh atribut pada item tertentu dapat *slip* dan menjawab salah. Sebaliknya, anak yang tidak menguasai atribut dapat menebak (*guessing*), dan menjawab item secara benar dengan probablilitas yang tidak nol (de la Torre, 2008:117; de la Torre & Karelitz, 2009:453).

Komponen terakhir adalah "AND *gate,"* merujuk pada proses konjungtif dalam menentukan jawaban benar pada suatu item memerlukan seluruh kemampuan yang dipersyaratkan item itu (de la Torre, 2008:117). Artinya, agar anak dapat menjawab benar suatu item, maka mereka harus menguasai seluruh atribut.

Berdasarkan uraian dan karakteristik yang dimiliki Model DINA tersebut, menarik untuk mengembangkan tes diagnostik dengan menggunakan model DINA. Oleh karena itu, fokus tulisan ini adalah: (a) menguraikan tahap pengembangan tes diagnostik dengan model DINA; dan (b) mengidentifikasi karakteristik tes diagnos-tik yang baik yang dikembangkan dengan model DINA.

Metode Penelitian

Jenis penelitian ini adalah penelitian pengembangan, yakni mengembangkan tes diagnostik. Dalam prosesnya, dilakukan penelitian empirik. Penelitian empirik yang digunakan adalah penelitian deskriptif eksploratori. Penelitian deskriptif digunakan untuk menguraikan tahapan-tahapan pengembangan tes diagnostik. Eksploratori digunakan untuk mencari informasi berupa karakteristik item, meliputi item yang fit dengan model, tingkat kesukaran dan indeks daya beda item.

Tes diagnostik dikembangkan melalui tujuh tahapan. Ketujuh tahapan itu dimulai dengan mengidentifikasi kompetensi dasar (KD) dan merumuskan indikator, menyusun *learning continuum*, menyusun hierarki materi, merumuskan atribut, mengonstruk soal, validasi ahli dan uji empirik. Langkah keempat yaitu merumuskan atribut merupakan langkah penting dalam proses pengembangan tes diagnostik berbasis model DINA. Langkah ini sebagai dasar dan acuan dalam mengembangkan item maupun membangun matriks Q. Sementara itu, langkah pertama hingga ketiga merupakan tahapan awal yang harus dilakukan untuk mendapatkan atribut yang diperlukan.

Uji coba dilakukan pada siswa kelas VIII SMP di 3 sekolah, yakni SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul dan SMPN 1 Panjatan Kulon Progo. Pada uji coba ini diikuti 180 siswa, sehingga dapat dijaring 180 data berupa respon siswa. Tes diagnostik yang diuji cobakan berbentuk pilihan ganda dengan 4 pilihan jawab (*option*), 1 kunci jawab dan 3 pengecoh.

Data hasil uji coba dianalisis dengan menggunakan dua *software* komputer, yakni program Mplus dan R. Program Mplus yang digunakan adalah Mplus versi 6 dari Muthen & Muthen (2010) untuk uji fit model. Dalam pelaksanaanya, untuk membangkitkan *syntax* Mplus perlu dukungan program CDM (*Cognitive Diagnostic Model)* yang dikembangkan oleh Templin (2008). Sementara itu, program R versi 2.14.1 dari Venables & Smith (2011) digunakan untuk mengestimasi parameter item (berupa *guessing* dan *slip*).

Level parameter *guessing* dan *slip* yang digunakan dalam penelitian ini dimodifikasi de la Torre et al. (2010:234) yakni: (a) rendah, bila terletak pada interval 0,00 – 0,15; (b) sedang, bila terletak pada interval 0,16 – 0,25; (c) tinggi, bila terletak pada interval 0,26 – 0,40 dan (d) sangat tinggi, bila terletak pada interval 0,41 – 1. Level parameter *slip* dan *guessing* (rendah, sedang, tinggi, dan sangat tinggi) selanjutnya digunakan untuk menentukan tingkat kesukaran item, dengan mengacu pada pendapat Zhang (2006:108). Menurut Zhang (2006) sebuah item memiliki tingkat kesukaran mudah apabila item itu memiliki parameter *guessing* tinggi/sangat tinggi dan *slip* rendah. Item yang sulit memiliki parameter *guessing* rendah dan *slip* tinggi/sangat tinggi. Item dengan tingkat kesukaran sedang, memiliki parameter slip dan guessing rendah.

Hasil estimasi parameter *guessing* dan *slip* digunakan pula untuk menentukan indeks daya beda item (δ_i), dihitung dengan menggunakan formula: $\delta_i = (1 - s_i) - g_i$. Simbol s_i menunjukan parameter *slip* dan g_i parameter *guessing*. Indeks daya beda memberikan gambaran tentang kualitas item. Kriteria yang digunakan untuk menentukan kualitas item, mengikuti pendapat Crocker & Algina (1986:315), yakni suatu item bila memiliki indeks daya beda: (a) baik, bila nilainya lebih dari sama dengan 0,40, (b) cukup bila nilainya terletak di antara 0,30 sampai dengan 0,39, (c) kurang, bila nilainya di antara 0,20 sampai dengan 0,29, dan (d) jelek, bila nilainya kurang dari sama dengan 0,19.

Hasil-hasil Pembahasan

1. Hasil Pengembangan Tes Diagnostik

a. Hasil Identifikasi KD dan Perumusan Indikator

Tahapan ini dimulai dengan mengkaji Standar Isi (SI) mata pelajaran matematika SMP Edisi 2006 (Depdiknas, 2006). Identifikasi dilakukan pada semua kompetensi aljabar di kelas VII. Hasil identifikasi itu digunakan untuk menyusun buram (*draft*) KD beserta indikatornya. Buram KD beserta indikator yang telah tersusun, selanjutnya dilakukan penelaahan melalui dua tahap kegiatan: teknik Delphi dan *focus group discussion* (FGD).

Teknik Delphi digunakan untuk menjaring saran terkait dengan draft yang telah dirumuskan. Teknik ini dilakukan dengan cara mengirimkan draft kepada 8 orang guru matematika dan 2 orang praktisi. Setelah dilakukan pemilahan, didapatkan saran dan dapat diklasifikasikan dalam tiga kelompok, yaitu agar: (a) melengkapi indikator yang telah ada; (b) menata ulang urutan antar indikator; dan (c) mengganti beberapa indikator agar lebih selaras dengan KD.

Untuk mendapatkan gambaran tentang perubahan yang dilakukan terkait dengan saran di atas, diambil contoh pada KD pertama. KD ini berbunyi: "Mengenali bentuk aljabar dan unsur-unsurnya," dan semula dirumuskan empat indikator. Setelah dilakukan perbaikan dengan mempertim-bangkan berbagai saran, maka: (a) ada penambahan dua indikator baru, yakni: menjelaskan pengertian bentuk aljabar dan menjelaskan pemfaktoran bentuk aljabar; (b) dilakukan perbaikan pada rumusan indikator nomor 4. Indikator ini semula berbunyi: "Menuliskan bentuk aljabar dari suatu ungkapan," diubah menjadi: "Menuliskan bentuk aljabar (model matematika) dari suatu ungkapan verbal/kehidupan sehari-hari;" (c) ada usulan untuk menukar urutan indikator dengan nomor urut 1 agar ditempatkan pada urutan 3. Sebaliknya indikator nomor urut 2 dipindah ke posisi 1, sedangkan urutan nomor 2 ditempati dengan indikator baru: "Menjelaskan pengertian bentuk aljabar."

Keseluruhan itu dilakukan agar diperoleh rumusan indikator yang rinci, urut dan lengkap sehingga memudahkan dalam menyusun *learning continuum*. Hasil perubahan yang dilakukan pada tahapan ini memberikan dampak banyaknya indikator bertambah dari 55 menjadi 63.

Selanjutnya dilakukan FGD. Forum FGD I dilakukan pada 27 Juni 2011 berhasil menyelesaikan dua tugas, yakni menulis KD beserta indikator untuk kelas VI dan VIII dan menulis materi prasyarat. Data yang diperoleh dari kegiatan ini berupa: (a) data hasil penulisan KD beserta indikator untuk kelas VI dan VIII; dan (b) data hasil penulisan materi prasyarat. Untuk kelas VI, berhasil dirumuskan 8 KD dan 46 indikator. Kesemuanya diturunkan dari 2 standar kompetensi (SK). Untuk kelas VIII juga berhasil dirumuskan 55 indikator yang mengacu pada 2 SK dan 10 KD.

Berkaitan dengan penulisan materi prasyarat, masing-masing indikator diidentifikasi materi yang terkait dan bersifat mendasarinya. Materi-materi itu dipetakan dari kelas VI atau dari materi di kelas VII yang memiliki tingkat kesulitan lebih rendah. Aspek yang diperhatikan dalam penetapan materi prasyarat adalah urutan materi yang tercermin dari rumusan indikator yang diselesaikan pada teknik Delphi. Dari kegiatan ini berhasil diidentifikasi 2 hingga 15 materi prasyarat untuk masing-masing indikator kompetensi.

b. Hasil Perumusan Learning Continuum

Dasar perumusan *learning continuum* adalah indikator-indikator kompetensi yang telah disepakati dan ditulis bersama dalam kegiatan FGD I. Dengan demikian, semua indikator yang dihasilkan dari forum FGD I langsung dipindahkan secara berurutan mulai dari kelas VI hingga kelas VIII. Dari langkah ini didapatkan 166 rumusan indikator yang dijadikan sebagai *learning continuum*. Namun setelah dibahas lebih lanjut, diperoleh usulan agar menambahkan beberapa butir indikator agar rumusan learning *continuum* menjadi lebih baik.

Misalkan, pada rumusan awal tidak dijumpai: "Pengertian lawan penjumlahan (*invers additive*) suatu bilangan," dan "Pengertian lawan perkalian (*invers multiplikative*) suatu bilangan." Kedua rumusan ini dianggap penting, karena sebagai dasar dalam menyelesaikan persamaan atau pertidaksamaan linier satu variabel. Dengan mengakomodir berbagai usulan, dihasilkan *learning continuum* baru dengan 184 pernyataan dari semula 166.

Hasil telaah praktisi, masih ditemukan adanya tumpang tindih pada rumusan *learning continuum*. Demikian pula, dijumpai beberapa rumusan indikator yang muncul pada nomor tertentu di awal, muncul kembali pada nomor berikutnya. Artinya, ada beberapa indikator yang muncul secara berulang atau memiliki rumusan berbeda namun memiliki makna hampir sama. Sebagai contoh, pada rumusan *learning continuum* nomor 30 sampai dengan 34 ditemukan pernyataan-pernyataan: "(30) Menjumlahkan pecahan biasa, cam-puran dan desimal; (31) mengurangkan pecahan biasa, campuran dan desimal; (32) mengalikan pecahan biasa, campuran dan

desimal; (33) membagi pecahan biasa, campuran dan desimal, dan (34) melakukan operasi hitung campuran yang melibatkan pecahan biasa dan pecahan campuran."

Pernyataan hampir sama ditemukan pada rumusan *learning continuum* nomor 47 dan 48: "(47) Menyelesaikan operasi hitung: tambah, kurang, kali dan bagi yang melibatkan pecahan biasa; dan (48) menyelesaikan operasi hitung: tambah, kurang, kali dan bagi yang melibatkan pecahan desimal." Kedua kelompok pernyataan di atas, memiliki makna yang tidak jauh berbeda. Oleh karena itu, menurut saran mereka harus ada yang dihilangkan. Kasus serupa ditemukan pada komponen *learning continuum* untuk rumusan lainnya.

Berdasarkan temuan di atas, selanjutnya dijadikan dasar melakukan perbaikan. Dalam melakukan perbaikan diupayakan menghindari penggunaan istilah yang kurang tepat (seperti masih ditemukan banyaknya penggunaan kata "dan" sebagai kata sambung antara dua kalimat, padahal seharusnya menggunakan kata "atau"), dan menghilangkan pernyataan-pernyataan yang memiliki makna sama. Perbaikan ini menghasilkan rumusan *learning continuum* baru yang terdiri atas 182 butir pernyataan.

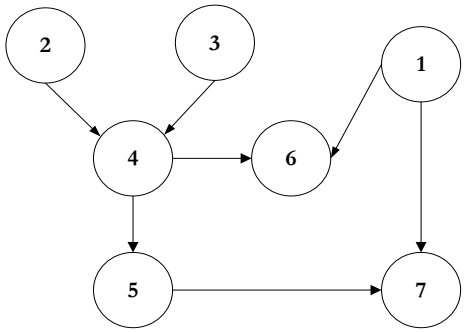
c. Hasil Penyusunan Hierarki Materi

Kegiatan penyusunan hierarki materi, diawali dengan pemetaan terhadap sejumlah kompetensi pada *learning continuum*. Pemetaan difokuskan pada kompetensi-kompetensi yang memiliki potensi diukur oleh soal. Pada masing-masing kompetensi yang hendak diukur, diidentifikasi sejumlah kompetensi yang harus dikuasai anak terlebih dahulu. Kompetensi yang harus dikuasai merupakan kompetensi atau kemampuan prasyarat.

Dari langkah tersebut, dibuat hubungan antar kompetensi satu dengan lainnya yang hendak diukur atau antar materi satu dengan lainya. Dengan demikian, tersusunlah hierarki materi. Hierarki ini menunjukkan prasyarat ketergantungan langsung antar materi yang teridentifikasi.

Pada tahap awal, hasil identifikasi peserta FGD I terhadap semua materi prasyarat untuk masing-masing indikator, langsung dibuat menjadi sebuah urutan dalam bentuk diagram. Sebagai contoh, agar anak mampu

menyederhanakan bentuk aljabar yang memiliki suku-suku sejenis, maka materi prasyarat yang mendasarinya adalah: (1) operasi hitung bilangan bulat dan pecahan, (2) pengertian variabel, (3) pengertian koefisien, (4) pengertian bentuk aljabar, (5) suku sejenis dan tidak sejenis, (6) penjumlahan suku-suku sejenis pada bentuk aljabar, dan (7) pengurangan suku-suku sejenis pada bentuk aljabar. Ketujuh materi prasyarat itu, selanjutnya disusun dalam bentuk hierarki berikut.

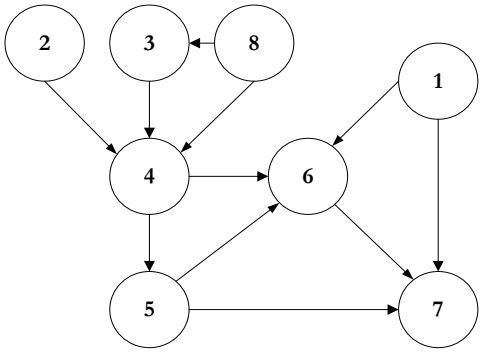


Gambar 1. Diagram Hirarki

Menurut praktisi, untuk mencapai kompetensi di atas diperlukan satu materi prasyarat lagi yakni pengertian konstanta. Pengertian bentuk aljabar tidak menjadi prasyarat dalam memahami suku-suku sejenis dan tidak sejenis. Sementara itu, pemahaman tentang suku sejenis dan tidak sejenis tidak hanya mendasari materi pengurangan suku-suku sejenis pada bentuk aljabar, namun juga diperlukan dalam melakukan penjumlahan suku-suku sejenis pada bentuk aljabar. Dengan memperhatikan pendapat di atas, akhirnya dilakukan beberapa perubahan sehingga mempengaruhi bentuk diagram hierarki materi, dan diagram hierarki materi berubah menjadi:

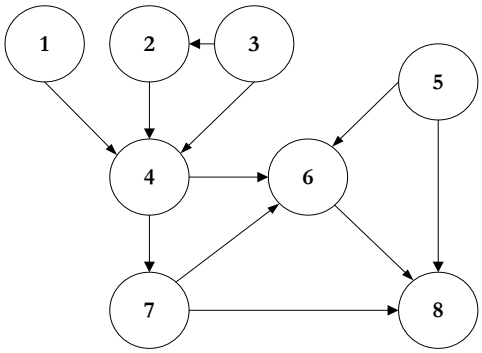
Hasil perbaikan berdasarkan saran praktisi, selanjutnya dikonsultasikan dengan promotor. Menurut promotor, ada yang kurang tepat dalam penempatan nomor urut pada diagram. Seharusnya, nomor urut di puncak diagram diawali dari yang terkecil, berlanjut ke nomor urut yang lebih besar. Artinya, materi atau konsep yang lebih sederhana diberi nomor urut kecil, sedangkan materi atau konsep yang lebih kompleks

diberi nomor urut lebih besar. Dengan demikian, urutan materi perlu ditata ulang agar memenuhi kaidah itu.



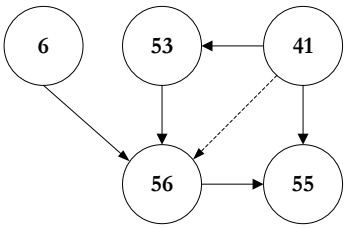
Gambar 2.

Atas dasar itulah, dilakukan penataan ulang terhadap materi atau konsep yang telah dirumuskan. Oleh karena itu, urutan kedelapan materi menjadi: (1) pengertian variabel; (2) pengertian koefisien; (3) pengertian konstanta; (4) pengertian bentuk aljabar; (5) operasi hitung bilangan bulat dan pecahan; (6) penjumlahan suku-suku sejenis pada bentuk aljabar; (7) suku sejenis dan tidak sejenis; dan (8) pengurangan suku-suku sejenis pada bentuk aljabar. Langkah yang sama dilakukan pada diagram lainnya. Diagram urutan kedelapan materi tersebut tampak seperti berikut.



Gambar 3.

Hasil perubahan ini selanjutnya dikonsultasikan pula ke kopromotor, dan ada yang kurang tepat dengan sistem penomoran pada diagram. Artinya, tidak semua diagram diawali dengan nomor urut 1. Nomor pada diagram seharusnya mengacu pada nomor urut yang ada pada *learning continuum*. Berdasarkan saran ini, dilakukan penyesuaian nomor urut dengan mengacu nomor urut pada *learning continuum.* Dengan demikian, nomor urut pada diagram menjadi: (6) sifat distributif perkalian terhadap pengurangan; (41) operasi hitung campuran pada bilangan bulat; (53) suku-suku sejenis pada bentuk aljabar; (55) operasi hitung penjumlahan, pengurangan, perkalian, atau pembagian bentuk aljabar; dan (56) menyederhanakan bentuk aljabar yang memiliki suku-suku sejenis. Dengan demikian, diagramnya berubah menjadi:



Gambar 4.

d. Hasil Penyusunan Atribut

Rumusan *learning continuum* dan hierarki materi telah memberi panduan dalam proses penyusunan atribut. Atribut-atribut ini selajutnya menjadi dasar dalam menyusun matriks Q. Mengingat pentingnya peranan atribut dalam penyusunan matriks Q, maka penentuan atribut ini dilakukan secara hati-hati.

Penentuan atribut juga terkait dengan item yang dikonstruk, karena keduanya akan diletakkan secara bersama-sama dalam matriks Q. Memperhatikan kondisi demikian, keseluruhan hirarki materi yang telah terbentuk dicari kedekatan materi prasyaratnya. Hal ini untuk menghindari banyaknya atribut dalam matriks Q lebih banyak dibandingkan banyaknya item. Berdasarkan hal tersebut, berhasil disusun 62 atribut. Setiap atribut

diberi kode tertentu, dengan tujuan memudahkan dalam proses penyusunan matriks-Q. Dalam konteks ini diberi kode A_i, dengan i = 1, 2, 3,… 62.

e. Hasil Konstruksi Soal

Tahap membangun soal juga dilakukan melalui dua rangkaian kegiatan, yaitu teknik Delphi dan FGD. Berdasarkan teknik Delphi diperoleh 48 item yang ditulis oleh 8 orang guru. Komposisi ke-48 item itu adalah: 5 item untuk materi bentuk aljabar; 19 item materi persamaan dan pertidaksamaan linier satu variabel; 6 item materi aritmetika sosial; 3 item materi perbandingan dan 14 item materi himpunan.

Item-item itu berbetuk pilihan ganda, dilengkapi dengan empat pilihan jawab (*option*). Keempat pilihan jawaban terdiri atas 1 kunci jawab dan 3 pengecoh. Pada kunci jawab juga disertai dengan uraian tentang cara mendapatkan jawaban. Pada masing-masing pengecoh juga diberikan uraian langkah kemungkinan salah konsepsi yang terjadi pada anak sehingga menyebabkan anak terjebak dan memilih pengecoh itu.

Hasil penulisan item, selanjutnya dibawa dan dibahas dalam forum FGD II yang dilaksanakan pada 27 Juli 2011. Pembahasan diawali dengan paparan hasil telaah terhadap item yang ditulis para guru. Dari paparan itu ditemukan hal-hal berikut: (1) beberapa item dirumuskan secara monoton, sehingga kurang menarik. Misalkan, ada 6 item yang dirumuskan secara mirip karena rumusannya diulang-ulang seperti: "Penyelesaian dari persamaan… adalah…" Kelompok item seperti ini, disarankan untuk dimodifikasi atau diganti; (2) pada item untuk menguji materi aritmetika sosial, ditemukan aspek aljabarnya yang belum tampak. Item-item yang digunakan untuk menguji topik ini lebih cenderung ke aspek aritmetika. Padahal kompetensi yang diinginkan untuk topik ini kesemuanya berbunyi "Menggunakan konsep aljabar untuk menghitung…;" (3) beberapa item terlalu sulit bila diujikan pada siswa SMP. Item seperti ini juga dipandang kurang tepat untuk tes diagnostik.

Berdasarkan temuan yang disampaikan oleh praktisi, selanjutnya dijadikan dasar oleh penulis item dalam melakukan revisi. Hasil revisi dirangkum oleh peneliti dan jumlah item mengalami perubahan menjadi 42

item. Berkurangnya jumlah item disebabkan karena ada penghilangan beberapa item yang memiliki kesamaan. Distribusi ke-42 item hasil revisi juga mengalami perubahan, yakni: 9 item untuk bentuk aljabar, 10 item untuk persamaan dan pertidaksamaan linier satu variabel, 6 item untuk aritmetika sosial, 3 item untuk materi perbandingan, dan 14 item untuk himpunan. Ke-42 item ini selanjutnya dilakukan penelaahan ahli.

f. Hasil Validasi Ahli

Validasi dilakukan terhadap semua perangkat instrumen yang dikembang-kan, mencakup rumusan SK, KD dan indikator, rumusan *learning continuum*, susunan hierarki materi, item tes diagnostik, dan analisis pengecoh. Jadi, validasi tidak semata-mata fokus pada item tes yang berhasil dikonstruk. Hasil validasi ahli dalam bentuk saran tertulis, dan saran itu dapat dikelompokkan ke dalam empat kategori, yakni terkait dengan: (a) teknik penulisan, (b) kelengkapan suatu materi, (c) urutan materi, dan (d) subtansi isi.

Teknik penulisan mencakup: konsistensi dalam menggunakan istilah tertentu, misalnya dalam penulisan lambang atau simbol matematika seperti tanda kurang (–) atau kali (×). Dalam hal kelengkapan suatu materi, perlu dipastikan tidak adanya materi yang tertinggal, padahal materi itu mendasari materi berikutnya. Urutan materi juga perlu dipastikan agar materi tersusun secara hierarkis, ditandai dengan materi yang lebih mudah ke materi yang kompleks. Selain itu, tidak terjadi tumpang tindih antar materi. Subtansi isi mencakup ketepatan suatu soal dalam mengukur indikator tententu.

g. Hasil Hasil Uji Empirik

Hasil uji coba empirik yang dilakukan di SMPN 16 Yogyakarta, menunjukan bahwa ada satu item yang sangat jelek. Dikatakan jelek karena tidak banyak anak yang memilih pengecoh yang disediakan. Hasil analisis dengan Program ITEMAN, dari 102 anak yang mengikuti tes, 94 anak (92%) dapat menjawab benar (memilih kunci C). Pengecoh A hanya dipilih seorang anak (1%), pengecoh B dipilih 4 anak, dan pengecoh D hanya dipilih 3 anak (3%). Atas pertimbangan ini, item nomor 1 dihilangkan dari paket tes. Dengan dihapusnya item nomor 1, keseluruhan item tes menjadi 37. Ke-37 item tersebut, selanjutnya diujicobakan pada siswa SMPN 1

Panjatan Kulon Progo dilakukan dengan maksud untuk melengkapi data kerangka pikir pada pengecoh yang belum didapat pada waktu uji coba di SMPN 16 Yogyakarta. Hasil uji coba dari sekolah ini, selanjutnya dipadukan.

Uji coba pada dua sekolah (SMPN 16 Yogyakarta dan SMPN 1 Panjatan Kulon Progo) melibatkan 166 anak dan mendapatkan beberapa temuan menarik. Temuan itu dapat dipilah menjadi 3 kelompok, yakni: (a) perlu dilakukan modifikasi terhadap beberapa pilihan jawaban, (b) perlu dilakukan perubahan pada beberapa kerangka pikir pengecoh, dan (c) didapatkannya alur pikir terjadinya salah konsepsi pada item-item baru, sehingga menjadi dasar dalam menyusun pengecoh.

Untuk kelompok pertama, ada empat item yang memenuhi kriteria itu yakni item nomor 10, 24, 29 dan 30. Perubahan pada keempat item itu, selain mempertimbangkan jawaban anak juga memperhatikan homogenitas pengecoh. Dengan disediakannya pengecoh yang semakin homogen, diharapkan akan benar-benar menguji pemahaman konsep anak.

Beberapa pengecoh yang dirumuskan oleh para guru, ternyata juga berbeda dengan yang dipikirkan oleh kebanyakan anak. Kondisi ini mendorong dilakukannya perubahan terhadap logika berpikir yang digunakan pada pengecoh itu. Dengan demikian, logika berpikir pada pengecoh yang digunakan dalam penelitian ini, diharapkan mampu mencerminkan salah konsepsi yang umumnya dilakukan anak.

Sementara itu, item baru yang digunakan untuk mengganti item sebelumnya, belum ada kerangka pikir terhadap pengecoh yang digunakan. Ada 3 item baru yang termasuk dalam kelompok ini. Oleh karena itu, dari uji coba ini diperoleh kerangka pikir sebagai dasar menyusun analisis pengecoh.

Hasil uji coba empirik yang dilakukan di SMPN 16 Yogyakarta dan SMPN 1 Panjatan Kulon Progo, diperoleh 37 item tes diagnostik. Ke-37 item itu selanjutnya diujikan pada siswa kelas VIII SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul, dan SMPN 1 Panjatan Kulon Progo.

2. Karakterisktik Item

Beberapa hal yang menjadi perhatian pada karakteristik item tes diagnostik di antaranya aspek item yang *fit*, tingkat kesukaran item dan indeks daya beda item. Data yang dijadikan dasar untuk menganalisis ketiga karakteristik tersebut berasal dari data hasil uji coba yang melibatkan 180 anak berasal dari SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul, dan SMPN 1 Panjatan Kulon Progo.

a. Uji Kecocokan (*Fit*)

Tabel 1 menyajikan nilai *entropy* untuk setiap kelompok item. Kelompok item 1 dan 2 menguji materi bentuk aljabar; kelompok item 3 sampai 5 menguji materi persamaan dan pertidaksamaan linier satu variabel; kelompok item 6 menguji materi aritmetika sosial; kelompok item 7 menguji materi perbandingan; dan kelompok item 8 sampai 10 menguji materi himpunan. Dengan demikian ada 10 kelompok item yang memuat 37 item tes diagnostik.

Sejauh ini belum ada referensi yang secara khusus membahas tentang garis batas (tinggi-rendah) nilai *entropy*. Untuk itu, dalam tulisan ini digunakan garis batas 0,75 sebagai batas tinggi-rendahnya *entropy*. Artinya, bila sekelompok item memiliki nilai *entropy* lebih dari sama dengan 0,75 maka dikategorikan tinggi, sebaliknya bila sekelompok item memiliki nilai *entropy* kurang dari 0,75 maka dikategorikan rendah. Bila garis batas itu digunakan, maka dari Tabel 1 diperoleh kelompok item 1, 2, 3, 4, dan 9 memiliki nilai *entropy* rendah.

Pada kelompok item yang memiliki nilai *entropy* rendah, selanjutnya ditelusuri nilai *bivariate model fit information.* Hasil penelusuran pada kelompok item 1 ditemukan bahwa terdapat sepasang item yang diduga memberikan sumbangan terhadap rendahnya nilai *entropy*. Pasangan item itu adalah item nomor 2 dan 3, karena terdapat selisih antara nilai pada H_1 (data hasil observasi atau *observed data*) dengan H_0 (data yang diprediksi model atau *model-predicted data*).

Pada kelompok item 2, hanya memuat dua item yakni item nomor 4 dan 5. Rendahnya nilai *entrophy* pada kelompok ini, tentu saja dipengaruhi oleh pasangan item ini. Pada kelompok item 3 memuat item nomor 6

sampai 8, dan penelusuran nilai *bivariate model fit information* tidak ditemukan pasangan item yang memiliki selisih antara H_1 dengan H_0 berbeda. Dengan demikian, pasangan item mana yang memiliki kontribusi terhadap rendahnya nilai *entrophy,* tidak dapat dideteksi.

Pada kelompok item 4, terdapat sepasang item yakni item nomor 9 dan 11 yang memiliki nilai H_1 dan H_0 berbeda. Sepasang item ini diduga memberikan konstribusi terhadap rendahnya nilai *entrophy* kelompok item ini. Pada kelompok item 9, yang memuat item nomor 28 hingga 30, juga ditemukan sepasang item yang memiliki nilai H_1 dan H_0 berbeda. Pasangan item itu adalah item bernomor 28 dan 30. Dengan demikian, pasangan item ini diduga memiliki sumbangan terhadap rendahnya nilai *entrophy* kelompok item ini.

Tabel 1. Rangkuman Nilai *Entropy* Masing-masing Kelompok Item

Kelompok Item	**No. Item**	**Nilai *Entropy***	**Kelompok Item**	**No. Item**	**Nilai *Entropy***
1.	1-3	0,685	6.	16-19	0,837
2.	4-5	0,518	7.	20–23	0,718
3.	6-8	0,594	8.	24-27	0,880
4.	9-12	0,615	9.	28-30	0,561
5.	13-15	0,772	10.	31-37	0,894

Terkait rendahnya nilai *entropy* pada kelima kelompok item ini, menurut de la Torre (2008:360) disebabkan karena kurang tepatnya ukuran matriks Q. Ketepatan yang dimaksud dilihat dari dua aspek, yaitu ukuran matriks Q dan kesesuaian matriks Q dengan struktur kelas laten. Ketepatan matriks Q merupakan bagian penting dari model DINA, yang sering diabaikan dalam analisis fit model.

Dari aspek ukuran, de la Torre menyarankan agar banyaknya item di dalam matriks Q lebih dari banyaknya atribut. Hasil simulasi menunjukkan

bila banyaknya item kurang dari atau sama dengan banyaknya atribut, maka Mplus tidak memberikan informasi apapun tentang struktur kelas laten yang dicari. Mplus hanya memberikan out put berupa pernyataan : "INPUT READING TERMINATED NORMALLY; Monte-Carlo Simulation for DINA Model." Pernyataan ini mengisyaratkan bahwa hasil analisis akan tidak fit bila dipaksakan dengan ukuran matriks seperti ini. Rupp & Templin (2008:83) mengungkapkan bahwa setiap penambahan sejumlah atribut pada matriks Q, seharusnya panjang tes juga ditambah agar dapat memberikan informasi yang reliabel tentang semua varibel laten.

Dari aspek kesesuaian, idealnya dalam mendesain matriks Q harus berkorespondensi dengan struktur kelas laten (de la Torre et al., 2010:233). Hal ini dapat dilakukan bila atribut pada matriks Q dijadikan sebagai kisi-kisi selama proses penyusunan item. Artinya, atribut pada matriks Q dikonstruk dan ditetapkan terlebih dahulu dan item tes dibangun yang mengacu pada sejumlah atribut. Padahal dalam penelitian ini, matriks Q dikembangkan atas dasar item dan hierarki materi yang telah dibangun sebelumnya. Item tes tidak dibangun berdasarkan matriks Q. Oleh karena itu, wajar bila kesesuaian antara matriks Q dengan struktur kelas laten kurang, sehingga kemungkinan rendahnya nilai *entropy* kelompok item ini juga dipengaruhi oleh aspek ini.

Penelusuran terhadap ke-15 item yang memiliki nilai *entropy* rendah, juga dilakukan pada parameter *guessing* dan *slip*-nya. Langkah ini dilakukan mengacu pendapat Rupp & Templin (2008:81). Menurut Rupp & Templin, bila model dapat menggambarkan data dengan baik atau fit, maka kedua parameter item pada model DINA (*guessing* dan *slip*) harus kecil. Sebaliknya, jika salah satu dari kedua parameter itu tinggi, menunjukkan item yang bersangkutan dikatakan tidak fit.

Data penelitian menunjukkan bahwa dari 15 item yang memiliki nilai *entropy* rendah, ditemukan 10 item yang memiliki parameter *guessing* tinggi yakni item nomor 1, 2, 3, 4, 6, 8, 9, 10, 11 dan 30. Hasil ini melengkapi *out put* dari Mplus berupa *bivariate model fit information.* Berdasarkan tiga bukti di atas, yakni rendahnya nilai *entropy,* tingginya parameter *guessing,* dan *bivariate model fit information* maka kesepuluh item (nomor 1, 2, 3, 4, 6, 8, 9, 10, 11 dan 30) yang digunakan dalam tes diagnostik ini tidak fit.

b. Tingkat Kesukaran

Level parameter *slip* dan *guessing* selanjutnya digunakan untuk menentukan tingkat kesukaran item. Oleh karena itu, pengelompokkan item yang mengacu pada parameter slip dan *guessing* perlu dibuat lebih rinci. Tabel 2 menyajikan distribusi item berdasarkan level parameter *slip* dan guessing yang dimilikinya.

Tabel 2. Distribusi Item Berdasarkan Parameter *Guessing* dan *Slip*

No	Kategori /Level	Nomor Item	
		Guessing	*Slip*
1.	Rendah	12, 16, 31, 32, 36, 37	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 16, 25, 27, 30, 33, 36, 37
2.	Sedang	5, 13, 19, 20, 23, 27, 28, 29	5, 19, 21, 22, 28, 29
3.	Tinggi	9, 17, 21, 24, 25, 26, 34, 35	10, 14, 18, 20, 24, 34, 35
4.	Sangat Tinggi	1, 2, 3, 4, 6, 7, 8, 10, 11, 14, 15, 18, 22, 30, 33	13, 15, 17, 23, 26, 31, 32

Selanjutnya, hasil pengklasifikasian item berdasarkan tingkat kesukarannya sangat membantu dalam memetakan profil item pada tingkat materi/topik yang diujikan. Dengan demikian, dapat dipotret materi-materi tertentu yang memiliki tingkat kesukaran mudah, sedang dan sulit. Tabel 3 merangkum tingkat kesukaran item yang dikorespondensikan dengan materi/topik yang diuji oleh item tersebut.

Berdasarkan Tabel 3, terlihat bahwa kelompok item yang menguji materi bentuk aljabar, sebagian besar memiliki tingkat kesukaran item mudah. Sebaliknya, pada materi aritmetika sosial memiliki tingkat kesukaran item sedang dan sulit. Item-item yang menguji bentuk aljabar dirancang untuk menguji kemampuan prosedural anak, sedangkan pada aritmetika sosial lebih menekankan pada pengujian aspek konsep.

Pesan yang dapat ditangkap dari kasus ini, umumnya anak akan mengalami kesulitan pada item yang menggali aspek konsep. Mengapa demikian? Item nomor 16 sampai 19 memaksa anak harus memahami konsepnya terlebih dahulu, seperti mengidentifikasi variabel yang terlibat pada soal dan memodelkannya sebelum dapat menyelesaikan. Hal itu berbeda dengan item 1 sampai dengan 5 yang hanya menguji kemampuan prosedural anak dan hanya menuntut aspek keterampilan rutin yang dapat dilatihkan.

Tabel 3. Distribusi Tingkat Kesukaran Item pada Setiap Materi

No	Topik/Materi	Tingkat Kesukaran		
		Mudah	Sedang	Sulit
1.	Bentuk aljabar	1, 2, 3, 4	5	-
2.	Persamaan dan pertidaksamaan linier satu variabel	6, 7, 8, 9	14	10, 11, 12, 13, 15
3.	Aritmetika sosial	-	16 dan 18	17 dan 19
4.	Perbandingan	21 dan 22	-	20 dan 23
5.	Himpunan	25, 30, 33	27, 28, 29, 36, 37	24, 26, 31, 32, 34, 35

c. Indeks Daya Beda

Hasil estimasi parameter *slip* dan *guessing* digunakan pula untuk menentukan indeks daya beda item. Tabel 4 menyajikan rangkuman distribusi item berdasarkan indeks daya beda item. Dari Tabel 4, ada enam item yang memiliki indeks daya beda jelek. Item yang jelek memiliki arti bahwa item yang bersangkutan tidak mampu membedakan anak yang telah menguasai seluruh atribut ($\eta = 1$) dan anak yang belum menguasai atribut ($\eta = 0$). Item yang memiliki sifat demikian, kemungkinan disebabkan oleh:

(a) materi yang ditanyakan terlalu sulit, sehingga banyak anak yang menebak; (b) pengecoh kurang berfungsi; dan (c) sebagian besar anak walaupun telah memahami materi, mereka beranggapan ada informasi yang salah pada item itu (Kusaeri & Suprananto, 2012:176).

Tabel 4. Distribusi Item Berdasarkan Kualitas Item

No.	Kualitas Item	Nomor Item
1.	Jelek	10, 15, 17, 23, 26, 33
2.	Kurang	2, 4, 7, 13, 14, 18, 22, 30, 32
3.	Cukup	1, 31, 35
4.	Baik	3, 5, 6, 8, 9, 11, 12, 16, 19, 20, 21, 24, 25, 27, 28, 29, 34, 36, 37

Dari kelima item yang memiliki indeks daya beda jelek, empat di antaranya memiliki tingkat kesukaran item sulit, yakni item nomor 10, 15, 17, 23 dan 26, sedangkan item nomor 33 memiliki tingkat kesukaran mudah. Untuk lima item yang disebutkan pertama, hasil penelusuran respon 180 anak melalui program ITEMAN ditemukan kelimanya memiliki nilai biserial yang rendah. Item yang memiliki nilai biserial seperti ini menurut Crocker & Algina (1986:315), termasuk dalam kategori item yang jelek.

Selanjutnya, rendahnya indeks daya beda item nomor 33 disebabkan karena kurang berfungsinya pengecoh. Hasil analisis yang dilakukan melalui program ITEMAN diperoleh sebagian besar (84% atau 151 dari 180 anak) memilih kunci jawaban. Selebihnya terdistribusi secara proporsional pada pengecoh yang disediakan, yakni 8,3% memilih pengecoh A, hanya 2,8% memilih pengecoh C, dan 3,9% memilih pengecoh D. Deskripsi ini menggambarkan bahwa ada dua pengecoh yang kurang diminati oleh anak, karena dipilih kurang dari 5%. Dengan kata lain, dua pengecoh tersebut kurang berfungsi.

Dengan demikian keenam item yang memiliki indeks daya beda jelek harus dihilangkan dari paket tes. Hal itu disebabkan karena indeks daya beda berkaitan dengan kualitas item. Bila indeks daya bedanya jelek maka kualitas item yang bersangkutan juga jelek.

Simpulan

Penelitian ini telah menghasilkan instrumen tes diagnostik yang dikembangkan dengan menggunakan model DINA. Selanjutnya, berdasarkan hasil penelitian dan pembahasan dapat dikemukakan simpulan penelitian sebagai berikut: (a) Tahapan pengembangan tes yang digunakan dalam penelitian ini meliputi: identifikasi kompetensi dasar (KD) dan merumuskan indikator, menyusun *learning continuum*, menyusun hierarki materi, merumuskan atribut, mengonstruk soal, validasi ahli dan uji empirik. Setelah melalui tujuh tahapan tersebut, berhasil dikembangkan 37 item tes diagnostik; dan (b) Dari 37 item tes diagnostik yang dikembangkan, 15 item di antaranya harus dihilangkan/dihapus dari paket tes. Item-item yang dihilangkan/dihapus disebabkan karena tidak fit dan memiliki kualitas item jelek (indeks daya beda kurang dari 0,2).

Namun, penelitian ini memiliki beberapa keterbatasan karena pengembangan tes diagnostik berbasis pada *latent class* (model DINA) merupakan hal baru dalam dunia pengujian pendidikan di Indonesia. Beberapa kelemahan dalam penelitan ini di antaranya: (i) Penyusunan item tes dilakukan dengan mengacu pada kisi-kisi tes seperti lazimnya pada proses pengembangan tes prestasi. Hal itu berbeda dengan langkah yang seharusnya digunakan pada model DINA, yaitu item tes dibangun dan diturunkan berdasarkan sejumlah atribut; (ii) Cakupan materi pada tes diagnostik ini terlalu luas, yakni semua materi aljabar yang diajarkan di kelas VII SMP (ada 5 materi). Luasnya cakupan materi tes berdampak pada kompleksnya atribut yang mendasari setiap item. Akibatnya, matriks Q yang terbentuk memiliki atribut lebih banyak, dibandingkan jumlah item; dan (iii) Pelaksanaan uji coba penelitian ini baru dikenakan pada sejumlah siswa kelas VIII SMP negeri, dan dilakukan di akhir semester gasal. Kondisi ini tentunya akan berbeda bila uji coba juga melibatkan siswa SMP

swasta, dan dikenakan pada siswa kelas VIII yang sedang berada di awal semester gasal atau siswa kelas VII di akhir semester genap. Dengan demikian, dari aspek validitas internal masih mengandung beberapa keterbatasan, baik menyangkut karakteristik subjek penelitian, waktu dan tempat penelitianya.

Daftar Pustaka

Awal Isgiyanto. 2011. *Analisis untuk menemukan informasi diagnostik data ujian nasional matematika.* Disertasi doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Chi, M.T.H. 2008. Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. Dalam S. Vosniadou (Eds), *Handbook of research on conceptual change* (pp. 61-82). New Jersey: Lawrence Erlbaum Associates.

Crocker, L. & Algina, J. 1986. *Introduction to classical and modern test theory.* Forth Worth: Holt, Rinehart and Winston.

Depdiknas 2006. *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 22 tahun 2006, tentang, Standar Isi untuk Satuan Pendidikan Dasar dan Menengah.*

de la Torre, J. 2008. DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics,* 39 (1): 115-130.

de la Torre, J. & Karelitz, T.M. 2009. Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement.* 46 (4): 450-469.

de la Torre, J., Yuan Hong & Weiling Deng. 2010. Factors affecting the item parameter estimation and classification accuracy of the DINA Model. *Journal of Educational Measurement,* 47 (2): 227-249.

Fauzan. 2010. *Pengembangan tes diagnostik kesulitan belajar matematika di SMA.* Disertasi doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Kusaeri. 2012. *Pengembangan tes diagnostik dengan menggunakan model DINA unuk mendapatkan informasi salah konsepsi dalam aljabar.* Disertasi doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Kusaeri & Suprananto 2012. *Pengukuran dan penilaian pendidikan.* Yogyakarta: Graha Ilmu.

Muthen, L.K. & Muthen, B. O. 2010. *Mplus user's guide (6th ed*). Los Angeles, CA: Muthen & Muthen.

Rakes, C.R. 2010. *Misconception in rational numbers, probability, algebra, and geometry*. Disertasi doktor, tidak diterbitkan. The University of Louisville.

Rupp, A.A., Templin, J. & Henson, R.A. 2010. *Diagnostic measurement: Theory, methods and applications.* New York: The Guilford Press.

Russel, M. & O'dwyer, L.M. 2009. Diagnosing students'misconceptions in algebra: Results from an experimental pilot study. *Behavior Research Methods,* 41, 414-424.

Sugiharto. 2003. *Diagnosis kesulitan siswa SMU dalam menyelesaikan soal matematika.* Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Suwarto. 2011. *Pengembangan tes diagnostik untuk mengungkap kesulitan siswa dalam memahami reproduksi sel.* Disertasi doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Templin, J. 2008. *Cognitive diagnosis modeling with Mplus (User guide).* Diambil pada tanggal 19 Desember 2011 dari http://jtemplin.coe.uga.edu/teaching/ dcm08 ncme/cdmuserguide.html

Templin, J. 2011. *Diagnostic measurement: Theory, methods and application.* Diambil pada tanggal 8 Desember 2011 dari http://jtemplin.coe.uga.edu/workshops/dcm/uga_dcm1.html.

Venables, W. R. & Smith, D.M. 2011. *An introduction to R: A programming environment for data analysis and graphics version 2.14.1.* Diambil pada tanggal 17 Januari 2012 dari http://cran.r-project.org/mirrors.html.

Waskito 2010. *Pengembangan model analisis prestasi belajar untuk mengidentifikasi kesulitan belajar peserta didik.* Disertasi doktor, tidak diterbitkan. Universitas Negeri Padang, Padang.

Xiaobao Li 2006. *Cognitive analysis of student's errors and misconceptions in variables, equations, and functions*. Disertasi doktor, tidak diterbitkan, A & M University, Texas.

Ying Liu, Douglas, J.A., & Henson, R.A. 2009. Testing person fit in cognitive diagnosis. *Applied Psychological Measurement,* 33 (8): 579-598.

Zhang, W. 2006. *Detecting differential item functioning using the DINA model.* Disertasi doktor, tidak diterbitkan. The University of North Carolina, Greensboro.

IMPLEMENTASI DAN PERMASALAHAN DALAM IMPLEMENTASI PERMENDIKNAS NOMOR 2 TAHUN 2008 TENTANG BUKU

Lina Miftahul Jannah, Muh. Azis Muslim, Eko Sakapurnama
Universitas Indonesia
miftahul@ui.ac.id; muhazis@ui.ac.id; eko.sakapurnama@ui.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui implementasi Permendiknas Nomor 2 Tahun 2002 tantang buku dan mendeskripsikan masalah-masalah yang ditimbulkannya. Penelitian dilaksanakan dengan metode campuran kuantitatif dan kualitatif di Kota Depok, Kabupaten Cianjur, Kota Makassar, dan Kabupaten Maros dengan menggunakan angket, wawancara mendalam dan diskusi kelompok terfokus untuk mengumpulkan data. Hasil penelitian menunjukkan bahwa implementasi Permendiknas yang bertujuan untuk menyediakan buku murah dan berkualitas bagi masyarakat belum sepenuhnya tercapai. Terdapat beragam masalah dalam pelaksanaannya antara lain minimnya sosialisasi, tidak memadainya infrastruktur teknologi informasi, ketidaksesuaian materi dengan kurikulum, rendahnya kualitas cetakan, dan Permendiknas dianggap sebagai kebijakan elitis yang mematikan iklim persaingan usaha.

Kata kunci: *kebijakan, pelaksanaan, buku cetak, buku sekolah elektronik*

IMPLEMENTATION AND ISSUES IN IMPLEMENTATION PERMENDIKNAS NUMBER 2 OF 2008 ON THE BOOK

Lina Miftahul Jannah, Muh. Azis Muslim, Eko Sakapurnama
Universitas Indonesia
miftahul@ui.ac.id; muhazis@ui.ac.id; eko.sakapurnama@ui.ac.id

Abstract

This research aims to figure out and elaborate the implementation of Cultural and Educational Ministry's regulation No 2, 2002 as well as the problems occurring during the implementation. The research was conducted by applying mixed method -qualitative and quantitative- held in Depok City, Cianjur Regency, Makassar City, and Maros Regency using questionnaires, interviews and focus group discussion to collect data. The result shows that the implementation of regulation, which originally aimed to provide affordable and quality book for the people, fails to achieve its purpose. There are various problems found in the implementation, e.g. lack of socialization, lack of information technology infrastructure, unfitness between the materials with the curriculum, lack of printing quality, and the prevailing opinion that the regulation is an elitist policy that kills business competition.

Keywords: *implementation, policy, education, text book*

Pendahuluan

Perkembangan dan kemajuan suatu bangsa sangat dipengaruhi oleh pendidikan. Melalui pendidikan diharapkan dapat menciptakan sumber daya manusia yang berkualitas sehingga mampu menghadapi perubahan yang terjadi dalam masyarakat. Dalam Pembukaan Undang-Undang Dasar 1945 juga ditegaskan bahwa salah satu tujuan negara Indonesia adalah mencerdaskan kehidupan bangsa dan juga disebutkan bahwa pendidikan merupakan tanggung jawab negara.

Sejalan dengan sistem pemerintahan Indonesia yang mengalami perubahan pasca Undang-Undang Nomor 22 Tahun 1999 tentang Pemerintahan Daerah yang kemudian direvisi kedalam Undang-Undang Nomor 32 Tahun 2004 dari sentralisasi menjadi desentralisasi, mengakibatkan perubahan pula dalam pembagian kewenangan, termasuk dalam bidang pendidikan. Pemerintah pusat selain mengelola Pendidikan Tinggi juga membuat kebijakan nasional, sedangkan Pemerintah Daerah (Kabupaten/Kota) berwenang mengelola Pendidikan Dasar dan Menengah. Pemerintah, baik pusat maupun daerah, wajib memberikan layanan dan kemudahan, serta menjamin terselenggaranya pendidikan yang bermutu bagi setiap warga negara tanpa diskriminasi. Pemerintah harus menyediakan segala kebutuhan pendidikan, baik kurikulum, sarana dan prasarana, guru, dan buku pelajaran. Terkait dengan buku pelajaran atau yang dikenal dengan istilah buku teks, penyediaannya ternyata mengalami banyak masalah, mulai dari keluhan orangtua siswa mengenai tingginya harga buku teks pelajaran, kurangnya persediaan buku, kecenderungan pemakaian buku teks yang telah ditentukan sebelumnya, juga keluhan mengenai penerbit yang cenderung mengambil posisi dominan dalam pengadaan buku

Sejak 2005, pemerintah juga telah mengatur tentang buku teks pelajaran melalui Peraturan Menteri Pendidikan Nasional (Permendiknas) Nomor 11 Tahun 2005. Pemerintah juga telah mengeluarkan program Bantuan Operasional Sekolah Buku. Kebijakan ini semakin dipertegas dalam Permendiknas Nomor 2 Tahun 2008 tentang Buku. Permendiknas ini mengatur bagaimana buku harus dibuat, bagaimana buku ini dapat

diedarkan, siapa yang dapat menjual, dan berapa buku ini harus dijual. Langkah yang telah dilakukan pemerintah adalah membeli hak cipta buku pelajaran dari penulis.

Kebijakan ini mempermudah distribusi buku dan mempermurah biaya mendapatkan buku. Salah satunya melalui jaringan internet dalam bentuk PDF *(e-book)* sehingga masyarakat (siswa, guru, orangtua, dan siapapun) dapat mengaksesnya secara langsung. Dalam kenyataannya, masih banyak guru, orangtua, dan tentu juga siswa yang belum dapat merasakan manfaat dari program Buku Sekolah Elektronik (BSE) ini karena terbentur beberapa masalah, seperti terbatasnya ketersediaan perangkat komputer, belum adanya fasilitas jaringan internet untuk dapat mengunduh BSE secara online, dan masih banyak kendala-kendala lainnya sehingga sulit untuk memanfaatkan program ini secara maksimal. Untuk itu menarik untuk mengetahui bagaimana implementasi Permendiknas Nomor 2 Tahun 2008 tentang Buku dan masalah-masalah apa yang terjadi dalam implementasi Permendiknas Nomor 2 Tahun 2008 tentang Buku.

Hipotesis yang diajukan dalam penelitian ini adalah Permendiknas Nomor 2 Tahun 2008 belum terlaksana dengan baik dan mengalami permasalahan dalam implementasinya. Penelitian dilaksanakan dengan *mixed method* kuantitatif dan kualitatif. Dengan menggunakan metode campuran ini, diharapkan peneliti dapat memperoleh gambaran menyeluruh tentang implementasi dan masalah-masalah dalam implementasi Permendiknas Nomor 2 Tahun 2008 tentang Buku. Metode kuantitatif yang digunakan adalah survei (melalui penyebaran kuesioner untuk siswa Sekolah Menengah Atas/SMA). Jumlah responden di wilayah kota Makassar, Kabupaten Maros, dan Kabupaten Cianjur masing-masing sebanyak 25 responden, sedangkan untuk Kota Depok sebanyak 27 reseponden. Total responden pada penelitian ini adalah 102 responden. Pemilihan siswa sebagai responden dari setiap wilayah dilakukan secara *accidental sampling*. Terkait dengan metode kualitatif, teknik pengumpulan data yang digunakan adalah wawancara mendalam kepada sekolah yang diwakili oleh kepala sekolah, lembaga swadaya masyarakat yang bergerak di bidang pendidikan, dan Badan Standar Nasional Pendidikan (BSNP), dan

Dinas Pendidikan. dan diskusi kelompok terarah dilakukan kepada guru, orangtua, dan penerbit/pedagang buku.

Penelitian ini mengambil lokasi di Provinsi Jawa Barat yaitu Kota Depok dan Kabupaten Cianjur, dan Provinsi Sulawesi Selatan yaitu Kota Makassar, Kabupaten Maros. Asumsi pemilihan lokasi kota yaitu mengindikasikan tingkat penggunaan internet yang tinggi dan kualitas pendidikan lebih baik, sedangkan kabupaten mengindikasikan tingkat penggunaan internet yang rendah, kualitas pendidikan lebih rendah.

Ada dua teori utama yang digunakan dalam penelitian ini, yaitu impelementasi kebijakan dan buku teks. Implementasi kebijakan merupakan proses atau tahapan yang penting dalam sebuah tahapan kebijakan. Bagaimanapun, sebuah kebijakan yang telah dihasilkan apabila tidak dilaksanakan maka akan sia-sia atau tidak dapat mengatasi suatu permasalahan. Implementasi juga penting karena menentukan berhasil atau tidaknya suatu kebijakan dibuat guna memecahkan suatu masalah. Crabbé and Leroy (2008, p.15) menjelaskan bahwa implementasi kebijakan merupakan sebuah proses di mana berbagai keputusan perlu untuk terus-menerus dibuat, dan seluruh kekuasaan, penerimaan, informasi, kapasitas dan elemen lain memainkan sebuah peranan yang penting. Sementara itu O'Toole (2000) mendefinisikan implementasi kebijakan sebagai apa yang berkembang antara pembentukan sebuah maksud yang jelas pada bagian pemerintah untuk melakukan sesuatu, atau untuk berhenti melakukan sesuatu, dan dampak akhir dalam serangkaian tindakan.

Lester dan Stewart (1996 : 97-98) menyatakan bahwa implementasi kebijakan dapat dilihat dari proses, *output*, dan *outcomes*. Dalam hal proses, implementasi merupakan serangkaian keputusan dan tindakan yang diarahkan agar peraturan yang ada dapat menghasilkan suatu efek. Dalam hal *output*, implementasi dapat dipahami sebagai sejauh mana tujuan suatu program didukung, seperti tingkatan pengeluaran yang dikomitmenkan untuk sebuah program atau jumlah pelanggaran yang diterbitkan agar ada kepatuhan terhadap petunjuk pelaksanaan suatu progam. Dalam hal *outcomes*, dapat dipahami bahwa hasil implementasi menyiratkan bahwa telah ada beberapa perubahan yang terukur pada masalah yang lebih besar yang telah ditangani oleh program tersebut, hukum publik, atau keputusan

pengadilan. Definisi van Meter dan van Horn (Lester, J. P and Stewart, J, JR. 2000). Memperlihatkan bahwa implementasi kebijakan merupakan tindakan dari aktor pemerintah atau swasta yang diarahkan pada pencapaian suatu tujuan tertentu. Dengan demikian, implementasi lebih melihat pada tindakan daripada melihat dampak dari kebijakan tersebut.

Dalam proses implementasi atau pelaksanaan kebijakan, kemungkinan akan selalu terbuka terjadinya perbedaan antara apa yang diharapkan atau yang sudah direncanakan oleh para pembuat kebijakan dengan apa yang senyatanya dicapai (sebagai hasil atau prestasi dari pelaksanaan kebijakan tadi). Hal inilah yang kemudian disebut dengan *implementation gap* atau kesenjangan. Pada batas tertentu kesenjangan ini masih dapat dibiarkan, sekalipun dalam monitoring harus diidentifikasi untuk segera diperbaiki apalagi jika kesenjangan yang terjadi lebih besar dari batas toleransi maka harus segera diperbaiki.

Implementation gap dapat terjadi antara lain karena, pertama tidak melaksanakan atau dilaksanakan tidak sebagaimana mestinya (*non implementation*). Kedua, karena tidak berhasil atau mengalami kegagalan dalam proses (*unsuccesful implementation*). Ketiga, pelaksanaan dilakukan sesuai dengan ketentuan yang ada, tetapi dalam prosesnya terjadi hambatan yang tidak dapat diatasi.

Terdapat beraga jenis buku yang biasa digunakan di sekolah. Salah satunya adalah buku teks. Buku teks menurut A.J. Loveridge merupakan buku sekolah yang memuat bahan yang telah diseleksi mengenai bidang studi tertentu, dalam bentuk tertulis yang memenuhi syarat tertentu dalam kegiatan belajar mengajar, disusun secara sistematis untuk diasimilasikan. Chambliss dan Calfee (1998) juga menjelaskannya bahwa buku teks merupakan alat bantu siswa untuk memahami dan belajar dari hal-hal yang dibaca dan untuk memahami dunia (di luar dirinya).

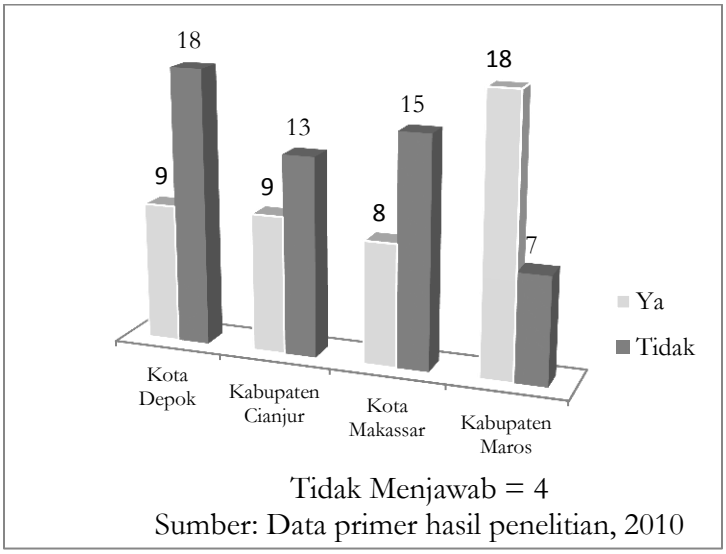
Dalam Permendiknas Nomor 2 Tahun 2008 tentang Buku disebutkan bahwa buku teks adalah buku acuan wajib untuk digunakan di satuan pendidikan dasar dan menengah atau perguruan tinggi yang memuat materi pembelajaran dalam rangka peningkatan keimanan, ketakwaan, akhlak mulia, dan kepribadian, penguasaan ilmu pengetahuan dan teknologi, peningkatan kepekaan dan kemampuan estetis, peningkatan

kemampuan kinestetis dan kesehatan yang disusun berdasarkan standar nasional pendidikan. Buku teks berguna bukan hanya bagi guru dan siswa tetapi juga bagi orang tua. Bagi guru dan siswa buku teks berguna sebagai acuan wajib karena dapat digunakan sebagai pedoman materi pembelajaran.

Hasil Penelitian dan Pembahasan

Hasil penelitian menunjukkan bahwa Permendiknas Nomor 2 Tahun 2008 sudah dilaksanakan, namun demikian belum berjalan dengan baik di semua pemangku kepentingan. Pasal 6 dan 7 Permendiknas tentang Buku telah menyebutkan bahwa buku teks sebagai acuan wajib oleh pendidik dan peserta didik dalam proses pembelajaran. Pendidik dapat menggunakan buku panduan pendidik, buku pengayaan, dan buku referensi dalam proses pembelajaran. Pendidik dapat menganjurkan kepada peserta didik yang mampu untuk memiliki buku. Anjuran tersebut seharunya bersifat tidak memaksa atau tidak mewajibkan. Dalam kenyataannya, peneliti menemukan bahwa di Kabupaten Maros, sebagian besar responden menyatakan bahwa guru mewajibkan siswa memiliki buku teks.

Di wilayah Kota Makasar, Kabupaten Cianjur, dan Kota Depok, responden tidak terlalu mengalami kesulitan untuk memperoleh buku teks, namun tidak demikian di Kabupaten Maros. Penyediaan buku diatur dalam Pada Pasal 7 ayat (4) Permendiknas tentang Buku menyebutkan "Satuan pendidikan wajib menyediakan buku teks di perpustakaan dan pendidik menganjurkan kepada semua peserta didik untuk meminjam buku teks pelajaran di perpustakaan satuan pendidikan atau memilikinya." Sebagian besar responden menyatakan bahwa mereka memperoleh buku teks dengan cara membeli dari koperasi sekolah (38,2%), membeli di toko buku, baik toko buku besar maupun toko buku kecil (24,5%), dan meminjam di perpustakaan (12,7%). Sisanya memperoleh buku teks dengan meminjam kepada kerabat atau teman, membeli pada guru, membeli di toko loak, dan menggandakan sendiri.



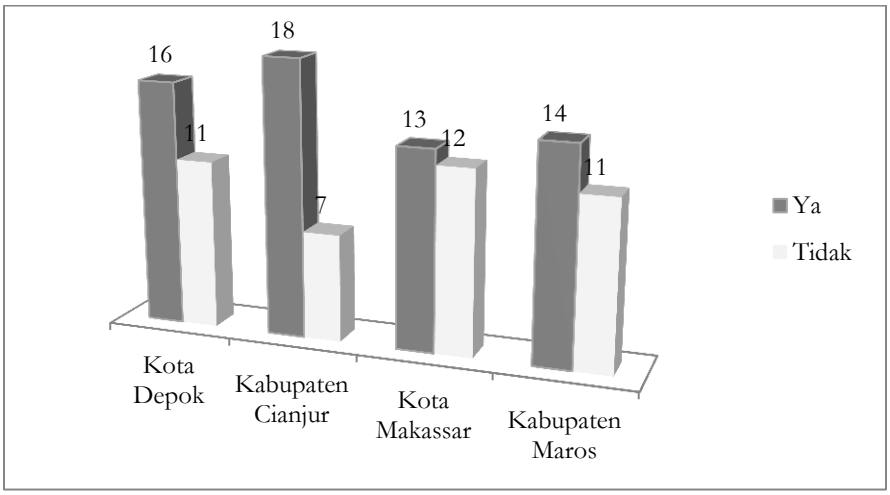
Tidak Menjawab = 4
Sumber: Data primer hasil penelitian, 2010

Gambar 1. Grafik Guru Mewajibkan Memiliki Buku Teks Berdasarkan Wilayah

Terkait dengan kebijakan ini, Pasal 11 Permendiknas ini menyebutkan bahwa pendidik, tenaga kependidikan, anggota komite sekolah/madrasah, dinas pendidikan pemerintah daerah, pegawai dinas pendidikan pemerintah daerah, dan/atau koperasiyang beranggotakan pendidik dan/atau tenaga kependidikan satuan pendidikan, baik secara langsung maupun bekerjasama dengan pihak lain, dilarang bertindak menjadi distributor atau pengecer buku kepada peserta didik di satuan pendidikan yang bersangkutan atau kepada satuan pendidikan yang bersangkutan, kecuali untuk buku-buku yang hak ciptanya sudah dibeli oleh Departemen, departemen yang menangani urusan agama, dan/atau Pemerintah daerah sebagaimana dimaksud dalam Pasal 3 ayat (4) dan dinyatakan dapat diperdagangkan sebagaimana dimaksud dalam Pasal 8 ayat (1). Kondisi ini jelas melanggar ketentuan, karena guru atau koperasi di sekolah ternyata menjual buku teks di luar yang dibeli hak ciptanya oleh pemerintah. Alasan yang diberikan, baik oleh koperasi dan guru, adalah

untuk mempermudah siswa daripada harus mencari ke toko buku, terutama di Kabupaten Maros dan Cianjur. Padahal dari hasil wawancara dengan penerbit, didapatkan informasi bahwa guru memperoleh komisi dari buku teks yang dijualnya antara 30-40%. Menurut para guru, komisi ini tidak diterimanya semua karena dibagi juga dengan kepala sekolah dan komite sekolah.

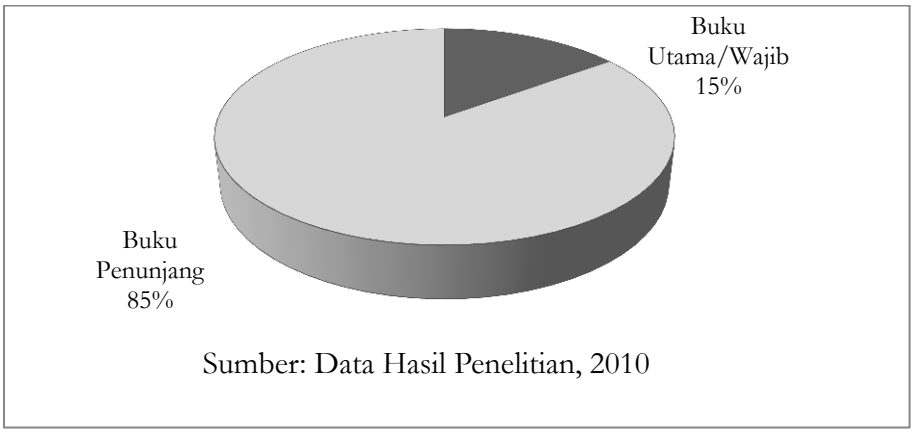
Kebijakan BSE sudah dikeluarkan sejak tahun 2008. Sesuatu yang mencengangkan jika melihat bahwa lebih dari 2 tahun, masih ada pihak yang terkait dengan pendidikan, dalam hal ini siswa, tidak mengetahui tentang BSE. Ternyata, bukan hanya siswa saja yang tidak tahu. Diakui oleh beberapa guru di Depok, bahwa mereka sendiri tidak tahu tentang BSE. Menurut mereka, tidak pernah ada sosialisasi dari dinas pendidikan. Informasi ini ditentang oleh Dinas Pendidikan yang menyatakan sudah menyampaikan informasi kepada kepala sekolah.



Gambar 2. Pengetahuan Responden tentang Buku Sekolah Elektronik Berdasarkan Wilayah

Pendapat resonden tentang BSE berdasarkan wilayah disajikan pada gambar 2. Berdasarkan hasil tersebut, diperoleh bahwa responden memperoleh informasi tentang BSE lebih lebih banyak melalui media cetak dan media internet. Internet yang diharapkan dapat juga memberikan informasi memiliki kendala dalam jangkauan terutama di wilayah yang terpencil.

Berdasarkan situs www.bse.kemdiknas.go.id diperoleh data bahwa hingga saat terdapat 958 judul BSE Buku-buku ini telah dinilai kelayakan pakainya oleh Badan Standar Nasional Pendidikan (BSNP) dan telah ditetapkan sebagai buku teks pelajaran yang memenuhi syarat kelayakan untuk digunakan dalam pembelajaran melalui Peraturan Menteri Pendidikan Nasional (Permendiknas) Nomor 46 Tahun 2007, Nomor 22 Tahun 2007, Nomor 27 Tahun 2007, Nomor 12 Tahun 2008, Nomor 34 Tahun 2008, Nomor 41 Tahun 2008, Nomor 69 Tahun 2008, Nomor 80 Tahun 2008, Nomor 81 Tahun 2008, dan Nomor 9 Tahun 2009. Untuk tingkat SD, buku teks mata pelajaran yang dibeli hak ciptanya oleh pemerintah adalah Bahasa Indonesia, Matematika, IPA, IPS, dan PKn (Pendidikan Kewarganegaraan), sedangkan untuk tingkat SMP yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, IPS, dan PKn.



Sumber: Data Hasil Penelitian, 2010

Gambar 3. Grafik Kegunaan Buku Sekolah Elektronik

Informasi yang diperoleh dari guru menunjukkan bahwa biasanya BSE memang hanya dipakai sebagai buku penunjang saja, karena biasanya lebih tidak menarik dan tidak lengkap dibandingkan dengan buku yang diterbitkan oleh penerbit. Diakui oleh penerbit, mungkin karena harus melalui internet dengan kapasitas terbatas, BSE dibuat lebih sederhana.

Tabel 1. Penilaian Responden tentang Buku Cetak Biasa dan Buku Sekolah Elektronik

Indikator	Buku Cetak Biasa	Buku Sekolah Elektronik
Harga	Mahal (62,1%)	Murah (48,9%)
Kualitas kertas	Bagus (54,9%)	Bagus (52,1%)
Kualitas cetakan	Baik (71,0%)	Baik (58,7%)
Isi materi	Lengkap (39,6%)	Lengkap (51,1%)
Penjelasan materi	Jelas (47,9%)	Jelas (55,3%)
Warna	Menarik (42,7%)	Menarik (51,1%)

Walaupun hasil penelitian menunjukkan tidak ada perbedaan yang terlalu berarti antara BSE dengan buku cetak biasa, namun hasil pengujian statistik menunjukkan kebalikannya. Nilai signifikansi adalah 0,000. Jika dibandingkan dengan nilai alpha 0,05, maka nilai signifikansi berarti lebih besar dibandingkan dengan nilai alpha. Iniberarti bahwa ada perbedaan penilaian yang signifikan antara buku cetak biasa dengan BSE.

Ini terbukti dari jawaban responden ketika ditanyakan mana yang lebih dipilih oleh responden antara buku cetak biasa dengan buku sekolah elektronik, mayoritas responden, yakni sebanyak 61,86%, lebih memilih buku cetak biasa dibandingkan buku sekolah elektronik.

Ada beberapa masalah terkait dengan pelaksanakan Permendiknas tentang Buku ini. *Pertama,* minimnya sosialisasi program BSE. Salah satu masalah utama yang dihadapi adalah minimnya informasi yang diterima berbagai pihak terkait penyelenggaraan BSE.

"Tidak pernah ada sosialisasi tentang BSE. Dengar sudah pernah…ngapain BSE, sementara buku tersebut tidak ada di BSE" (FGD Depok)

Tabel 2. Hasil Pengujian Statistik menggunakan Uji Perbedaan Wilcoxon

		N	Mean Rank	Sum of Ranks
BSE – BUKU CETAK	Negative Ranks	8(a)	14,50	116,00
	Positive Ranks	29(b)	20,24	587,00
	Ties	7(c)		
	Total	44		

	BSE – BUKU CETAK
Z	-3,586
Asymp. Sig. (2-tailed)	0,000

Masalah informasi ini bukan hanya pada saat buku telah selesai, tetapi juga saat penyeleksian dan penilaian buku yang diajukan. Tidak jelas kapan waktunya. Sebenarnya, terkait dengan informasi yang tidak jelas ini, dari situs http://bsnp-indonesia.org dapat ditemukan informasi yang lengkap tentang instrumen penilaian buku teks dan juga pengumuman penilaian buku teks pelajaran. Jadi, mungkin informasi ini yang tidak diakses atau diterima oleh para penulis buku.

Masalah *kedua* yaitu infrastruktur teknologi informasi tidak memadai. Tidak setiap siswa dan sekolah memiliki jaringan internet.

"Kalaupun ada jaringan internet sekolah harus bayar sendiri. Biasanya kerjasama dengan telkom. (FGD Makassar, Maros, dan Cianjur)"

BSE yang diharapkan lebih murah karena dapat diakses di mana saja, malah menjadi mahal karena akses yang lambat dan belum tersedianya perangkat di sekolah. Di Depok saja, ketika peneliti mencoba mencari toko buku yang menjual BSE, jumlahnya sangat sedikit. Mata harus jeli melihat

di sepanjang jalan yang memang memasang pengumuman menjual BSE. Pemerintah sebenarnya dapat memfasilitasi BSE tidak seperti model cetak seperti saat ini. Penggunaan cakram padat atau cakram digital menjadi salah satu pilihan. Biaya penggantian dua keping cakram digital berisi total 800an judul buku hanya dikenai biaya Rp 95.000,-(sudah termasuk ongkos kirim kemana saja di Indonesia dengan Pos Khusus). Tapi, tentu saja. untuk menggunakan cakram digital ini diperlukan perangkat pemutar, dan itu berarti biaya. Bagi yang memiliki komputer atau laptop dengan pemutar cakram, mungkin menjadi suatu pilihan lebih murah.

Masalah *ketiga* adalah ketidaksesuaian materi dengan kurikulum yang berlaku. BSNP adalah lembaga yang salah satu kewenangannya untuk menilai kelayakan isi, bahasa, penyajian, dan kegrafikaan buku teks pelajaran. Buku-buku yang dibeli hak ciptanya oleh pemerintah harus melalui tahap penilaian oleh lembaga ini. Dengan melalui uji kelayakan ini, seharusnya materi yang ada di buku teks atau BSE dapat dipastikan sesuai dengan kurikulum yang berlaku.

Namun demikian, tidak semuanya setuju dengan pernyataan bahwa kualitas BSE sama dengan buku cetak lain yang dikeluarkan oleh penerbit besar.

> *"Materi muatan lokalnya kurang. Jadi kurang lengkap dibandingkan dengan buku cetak." (FGD Maros)*

Masalah *keempat* dalam pelaksanaan Permendiknas Nomor 2 Tahun 2008 adalah ada indikasi kerjasama guru/sekolah dengan penerbit dalam menjual buku cetak. Memang tidak ada bukti bahwa guru atau sekolah melakukan kerjasama dalam menjual buku cetak. Namun, dari hasil FGD diperoleh informasi bahwa ada komisi yang diberikan penerbit kepada sekolah atau guru atas buku-buku yang dijual yang nilai besaran komisi minimal 10% dan bisa mencapai hingga 30 sampai 40 persen. Jadi penerbit biasanya akan mendatangi sekolah atau guru untuk menawarkan buku terbitannya.

> *"Sejak ada kebijakan BSE itu malah penerbit semakin massif." (Wawancara dengan LSM)*

Distribusi tidak merata menjadi masalah *kelima*. Penetapan harga eceran tertinggi (HET) oleh pemerintah terhadap BSE versi cetak yang dinilai terlalu rendah, banyak pengganda enggan mendistribusikan ke daerah terpencil karena biaya operasional tidak sesuai dengan margin karena HET terlalu rendah. Informasi yang diperoleh dari IKAPI Jakarta menyebutkan bahwa harga kertas di Aceh jauh lebih mahal dibandingkan dengan Jakarta atau di Pulau Jawa. Jadi, kalau siswa, orang tua, atau guru mau mencetak sendiri, biayanya lebih mahal dibandingkan dengan membeli buku di toko buku. Di Depok, misalnya, kertas ukuran A4 70 gram setiap rimnya hanya sekitar Rp. 25.000,00 saja, tetapi di wilayah lain bisa mencapai Rp. 35.000,00.

Kualitas cetakan rendah menjadi masalah yang *keenam*. Menurut Pedagang BSE di Depok, satu buku dihargai minimal Rp. 10.000,00, sedangkan menurut penerbit di Cianjur harga BSE di bawah 25.000,00. Harga ini memang di bawah harga buku-buku terbitan penerbit ternama yang mencapai Rp. 40.000,00. Bahkan akibat HET yang telah ditetapkan ini, kualitas buku menjadi rendah. Menurut siswa yang diwawancara, terbukti dengan ditemukannya buku-buku yang cetakannya tidak jelas terbaca. Selain itu, banyak buku BSE yang dicetak dengan kertas yang tidak sesuai dengan ketentuan. Menurut pedagang buku di Depok, hal ini harus dilakukan untuk menekan biaya produksi.

Ketujuh, Permendiknas ini juga dianggap **m**ematikan iklim persaingan usaha. Sejak diberlakukan BSE, Pemutusan Hubungan Kerja besar-besaran di industri penerbitan tidak dapat dihindari. Banyak karyawan penerbitan yang harus kehilangan pekerjaan karena pekerjaan mereka telah digantikan oleh peran pemerintah. Sebelum Permendiknas 2008 dikeluarkan, penerbit besar membuka kantor cabang di beberapa kota besar di Indonesia. Setelah 2008, beberapa penerbit menutup kantornya ini, seperti yang terjadi di Kabupaten Cianjur.

Masalah *kedelapan* terkait dengan kualitas sumber daya manusia pendidik dan siswa yang belum merata. Tidak semua wilayah memiliki sumber daya pendidik dan siswa dengan kualitas yang sama. Di daerah perkotaan, melek informasi dan teknologi menjadi suatu hal yang lumrah. Menurut data Kementerian Komunikasi dan Informatika

(Kominfo), jumlah pengguna internet dari komputer dan ponsel hingga Juni 2010 sudah mencapai 45 juta. Angka ini meningkat tajam jika dibandingkan dengan kondisi tahun 1999 yang hanya satu juta pengguna. Dalam kenyataannya internet hanya digunakan sekedar mengirim surat (*mailing*), mengobrol *(chatting)* melalui jejaring sosial, atau melihat *(browsing)*. Responden dalam penelitian ini misalnya, sebagian besarnya menggunakan internet terkait dengan jaringan pertemanan. Selain telepon genggam, responden juga biasa menggunakan warnet untuk mengakses internet (35,3%). Hanya 25,5% saja yang menggunakan komputer atau komputer jinjing pribadi untuk mengakses internet. Jarang sekali yang menyatakan menggunakan internet untuk kepentingan belajar. Kecepatan dan kemampuan para siswa dalam mengadopsi teknologi informasi tidak dapat diikuti oleh para pendidik. Jika para siswa adalah generasi *native speaker technology* maka sebagian besar guru lahir sebelum era 80-an, atau lebih dikenal generasi *digital immigrant* yang terkadang gagap teknologi. Terkait dengan penggunaan BSE, hasil penelitian di Cianjur menunjukkan bahwa hanya 25% saja guru yang menggunakan BSE.

Terlambatnya distribusi BSE yang telah dicetak oleh pemerintah akibat keterlambatan pencairan anggaran menjadi masalah *kesembilan*. Untuk tingkat SD dan SMP, pada tahun 2010 dan sebelumnya, pemerintah menganggarkan pembelian buku melalui dana BOS Buku. BOS (Bantuan Operasional Sekolah) adalah program pemerintah untuk penyediaan pendanaan biaya operasi non personalia bagi satuan pendidikan dasar sebagai pelaksana program wajib belajar. Salah satu jenis BOS adalah BOS Buku yang hanya diperuntukkan bagi penyediaan buku yang dibutuhkan dalam proses belajar mengajar. Mulai Januari 2009, besaran BOS Buku per siswa per tahun untuk SD di kota dan di kabupaten masing-masing sebesar Rp. 400.000,00 dan Rp. 397.000,00, sedangkan untuk SMP di kota dan di kabupaten masing-masing sebesar Rp. 575.000,00 dan Rp. 570.000,00 Melalui BOS Buku, pihak sekolah mengajukan buku dan jumlah yang akan dibeli. Dengan catatan, buku yang dibeli harus yang tercatat dalam daftar buku yang memang telah diuji kelayakannya. Hanya saja, tidak semua buku cetak yang dibutuhkan dapat dibeli setiap tahunnya. Misalnya tahun ajaran 2008 hanya bisa membeli buku matematika dan bahasa Indonesia. Tahun

berikutnya membeli buku dengan judul yang lain. Jumlahnya pun terbatas, tidak semua siswa dapat meminjam buku tersebut. Siswa yang tidak kebagian buku, akhirnya harus mengkopi dari buku siswa lain atau membelinya. Pada akhir tahun ajaran, buku tersebut harus dikembalikan lagi ke sekolah. Masalah yang muncul selain kurangnya jumlah dan jenis buku adalah buku tersebut tidak bisa diperoleh di awal tahun ajaran. Ini sangat tergantung dari pencairan anggaran. Dengan kondisi seperti ini, akhirnya muncul situasi pada awal tahun ajaran, siswa tetap harus membeli buku karena proses belajar mengajar tidak boleh ditunda karena ketiadaaan buku.

Selain itu, masalah yang lainnya adalah saat pengembalian buku di akhir tahun ajaran. Buku yang dikembalikan juga sebagian tidak dapat digunakan lagi karena hilang atau rusak. Untuk yang rusak, orangtua harus mengembalikan dengan buku yang sama, sedangkan untuk yang rusak, tidak ada sanksi apapun. Pada akhirnya, buku ini tidak dapat digunakan lagi untuk masa tahun ajaran berikutnya.

Masalah terakhir adalah *Permendiknas tentang Buku dianggap kebijakan yang elitis*. Kebijakan ini didefinisikan bebas sebagai kebijakan yang hanya mementingkan kelompok tertentu. Jadi hanya melihat kebutuhan sekelompok orang-orang yang sudah memiliki akses lebih, padahal kan hal-hal seperti ini dirancangnya untuk memenuhi kebutuhun terbesar mayoritas rakyat Indonesia yang ada di pedesaan.

Simpulan dan Saran

Hasil penelitian menunjukkan bahwa Permendiknas Nomor 2 Tahun 2008 sudah dilaksanakan, namun demikian belum berjalan dengan baik di semua pemangku kepentingan. Masalah-masalah yang dihadapi dalam pelaksanaan Permendiknas ini adalah minimnya sosialisasi tentang Buku Sekolah Elektronik (BSE), tidak memadainya infrastruktur teknologi informasi yang dibutuhkan untuk mengakses BSE, ketidaksesuaian materi BSE dengan kurikulum yang berlaku, ada indikasi kerjasama guru/sekolah dengan penerbit dalam menjual buku cetak, tidak meratanya distribusi buku pelajaran, rendahnya kualitas cetakan BSE, Permendiknas tentang Buku

dianggap mematikan iklim persaingan usaha, **k**ualitas sumber daya manusia pendidik dan siswa yang belum merata, terlambatnya distribusi BSE yang telah dicetak oleh pemerintah akibat keterlambatan pencairan anggaran, dan Permendiknas tentang Buku dianggap sebagai kebijakan yang elitis.

Dengan munculnya masalah-masalah yang cukup mendasar ini, ada baiknya pemerintah dalam membuat kebijakan yang akan diberlakukan secara nasional harus juga mempertimbangkan kondisi nasional tanpa menghilangkan latar belakang dan karakteristik masyarakat yang ada. Sosialisasi program secara terpadu dapat membantu berjalannya sebuah program secara efektif dan efisien. Berbagai media dan jaringan dapat dipergunakan untuk mendapat hasil maksimal dalam proses sosialisasi. Bagi pihak swasta disarankan untuk tidak hanya berpikir profit semata, tanggung jawab membangun bangsa dengan mencerdaskan masyarakat merupakan cita-cita bersama yang harus ditopang oleh semua pihak. Kepentingan sesaat yang terbangun dalam proses penyampaian buku (bahan ajar) tidak seharusnya mengorbankan pihak lain dalam hal ini masyarakat (murid dan wali murid). Bagi guru, sebagai aktor sentral dalam kegiatan belajar mengajar, bukan saatnya lagi menjadi pihak yang hanya menerima buku apalagi terlibat dalam *rent seeking*. Sudah saatnya bagi guru untuk berpikir proaktif dan kreatif dalam mengembangkan kegiatan pengajaran dengan melakukan pengayaan keilmuan melalui berbagai sumber. Bagi masyarakat (murid/wali murid dan komite sekolah) yang selama ini seringkali merasa menjadi korban kebijakan, sebaiknya dapat melakukan fungsi kontrol secara lebih baik dan berkesinambungan dalam kegiatan belajar mengajar yang berlangsung.

Daftar Pustaka

Ann, Crabbé and Leroy, Pieter. 2008. *The Handbook of Environmental Policy Evaluation*, London and Sterling, VA: Earthscan.

Chambliss, Marilyn J and Calfee, Robert C. 1998.. *Texbook for Learning: Nurturing Children's Minds, Wiley-Blackwell.*

Lester, J. P and Stewart, J, JR. 2000, *Public Policy: An Evolutionary Approach*, 2nd edition. Wadsworth, Belmont.

Loveridge, AJ 1972, *terj.* Hasan Amin*, Persiapan Naskah Buku Peladjaran Pembimbing bagi Pengarangn di Negara-Negara Berkembang,* Balai Pustaka.

O'Toole, L.J, 2000. "Research on Policy Implementation," *Assessment and Prospects. Journal of Public Administration Research and Theory*, 10(2).

Republik Indonesia, *Peraturan Menteri Pendidikan Nasional Nomor 11 Tahun 2005 tentang Buku Teks Pelajaran.*

_____, *Peraturan Menteri Pendidikan Nasional Nomor 2 Tahun 2008 tentang Buku*.

_____, *Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional*, Lembaran Negara Tahun 2003 Nomor 78.

www.maroskab.go.id

PENGEMBANGAN ALAT UKUR KECEMASAN OLAHRAGA

Nyak Amir
FKIP Universitas Syah Kuala Banda Aceh
info@infoaxe.net

Abstrak

Tujuan penelitian ini adalah mengembangkan alat ukur yang memiliki validitas dan reliabilitas yang baik dan dapat diterapkan sesuai kondisi di Indonesia. Subjek penelitian (N=406) adalah seluruh atlet sepakbola pemula Kota BandaAceh. Metode pengembangan alat ukur ini dilakukan dengan dua kegiatan, yaitu: adaptasi instrumen, dan pengumpulan butir baru melalui item pool dan *screening of item pool* (Q-sort). Selanjutnya alat ukur ini diujicobakan melalui dua tahap, yakni uji coba tahap pertama dilakukan pada 406 atlet klub sepakbola pemula, dan uji coba tahap kedua dilakukan pada 1000 atlet klub sepakbola Provinsi Aceh. Data dianalisis melalui pembuktian validitas dan estimasi reliabilitas. Hasil penelitian menunjukkan bahwa validitas berada pada kategori baik, dan reliabilitas skala kecemasan pada kategori cukup, dan skala kecemasan olahraga yang terdiri atas 4 faktor dan 35 butir pernyataan ini dapat dipakai untuk mengukur kecemasan olahraga.

Kata kunci: *pengembangan, alat ukur, kecemasan olahraga*

SPORTS DEVELOPMENT TOOLS MEASURING ANXIETY

Nyak Amir
FKIP Universitas Syah Kuala Banda Aceh
info@infoaxe.net

Abstract

The aim of this research is to develop standardized instrument with a good validity and reliability to measure sport anxiety which is suitable to local conditions in Indonesia. Subjects of the research were (N =406) beginner soccer players from Banda Aceh Municipality. The development of this measurement tool was conducted through two activities, namely instrument adaptation, and collection of new items through item pooling and screening of the item pool (Q-sort). Then the instrument was validated throught two phases, firstly on 406 beginner soccer players, and secondly, on 1000 soccer players from Aceh Province. Data were analysed through validity and reliability testing and factor analysis. Results show either the validity or the reliability level of anxiety scale was high enough, and that the sport anxiety scale comprising four factors and 35 items could be used to measure sport anxiety.

Keywords: *development, measurement tool, sport anxiety*

Pendahuluan

Permasalahan mendasar yang dihadapi oleh dunia persepakbolaan di Indonesia dan juga klub sepakbola pemula di Provinsi Aceh adalah hingga dewasa ini belum ada penerapan program pembinaan mental secara khusus dalam sistem pembinaan prestasi olahraga. Selain program pembinaan mental, belum juga ditemui adanya alat ukur baku yang dapat digunakan untuk mengukur tingkat kecemasan olahraga para atlet. Simpulan tersebut diperoleh melalui hasil pengamatan pada beberapa klub sepakbola pemula, serta didukung oleh hasil wawancara langsung dengan 18 atlet dan 6 pelatih klub sepakbola pemula di Kota Banda Aceh. Tidak adanya program pembinaan mental dan instrumen pengukuran kecemasan baku menyebabkan tidak terdeteksinya tingkat gangguan kecemasan para atlet, sekalipun beberapa atlet mengaku telah beberapa kali mengalami kecemasan saat menghadapi pertandingan.

Sampai saat ini tampaknya belum ada alat ukur kecemasan olahraga yang baku dengan tingkat validitas dan reliabilitas tinggi, yang dapat diterapkan di Indonesia sebagai alat untuk mengukur tingkat kecemasan olahraga. Sementara ini, khususnya di negara Barat telah terdapat sekurang-kurangnya sepuluh instrumen kecemasan yang dapat digunakan sebagai alat mengukur kecemasan olahraga secara valid dan reliabel. Sayangnya, kesepuluh instrumen tersebut dikembangkan dalam konteks yang berbeda dengan latar belakang budaya Indonesia. Dalam kondisi seperti itu, instrumen tersebut tidak serta merta dapat dipergunakan untuk para atlet olahraga Indonesia. Oleh karena itu, diperlukan upaya-upaya mengadaptasi, memodifikasi, dan mengembangkan instrumen kecemasan olahraga yang dapat dipergunakan secara valid dan reliabel agar dapat diterapkan sesuai kondisi Indonesia. Terkait dengan hal tersebut, penelitian ini bertujuan mengembangkan alat ukur yang memiliki validitas dan reliabilitas yang baik dan dapat diterapkan sesuai kondisi di Indonesia

Kecemasan olahraga

Secara umum, kecemasan dapat dibagi dalam dua kategori, yakni *state anxiety* dan *trait anxiety*. Ketakutan yang tidak proporsional terhadap satu situasi tertentu disebut dengan *state anxiety*. Jenis kecemasan ini merupakan

kondisi emosi yang bersifat sementara dan berlangsung untuk suatu situasi tertentu saja. Jenis kecemasan berikutnya adalah *trait anxiety*, jenis kecemasan yang lebih menetap dan menyebar ke berbagai aspek kehidupan individu. Individu merasa cemas, kapan dan sehingga timbul rasa khawatir dan tegang.

Terkait dengan olahraga, kecemasan yang timbul saat pertandingan merupakan reaksi emosional negatif atlet ketika harga dirinya dirasa terancam. Hal ini terjadi apabila atlet menganggap pertandingan sebagai tantangan berat untuk berhasil, mengingat kemampuan penampilannya. Kecemasan ini biasanya dipicu pula oleh karena atlet banyak memikirkan akibat dari kekalahannya. Kecemasan akan selalu terjadi pada diri individu apabila sesuatu yang diharapkan mendapat rintangan sehingga kemungkinan tidak tercapainya harapan menghantui pikirannya. Kecemasan olahraga adalah perasaan khawatir, gelisah, dan tidak tenang dengan menganggap pertandingan sebagai sesuatu yang membahayakan (Martens, Vealey, & Burton, 1990).

Unsur yang paling dominan menyebabkan kecemasan adalah unsur kognitif yakni kekhawatiran dan pikiran negatif bahwa proses dan hasil pertandingan dapat mengancam posisi atlet (Smith & Sarason, 1993). Anshel menjelaskan bahwa kecemasan olahraga menggambarkan perasaan atlet bahwa sesuatu yang tidak dikehendaki akan terjadi (1997). Hal yang tidak dikehendaki misalnya atlet tampil buruk, lawannya dipandang demikian superior, atlet akan mengalami kekalahan, kekalahan menyebabkan dirinya dicemooh oleh teman-teman dan seterusnya membentuk kecemasan berantai. Kondisi ini memberikan dampak yang sangat tidak menguntungkan pada atlet apalagi jika rasa percaya diri atlet kurang tinggi. Atlet cenderung tampil kaku, bingung, dan gerakan-gerakannya menjadi kurang terkontrol dengan baik. Spielbelger (1972) rnenerjemahkan kecemasan sebagai takut mengalami kegagalan (*fear offailure*) atau takut menderita kekalahan. Spielbelger juga mendefinnisikan pikiran negatif berhubungan dengan anggapan mengenai bahaya yang akan menimpa diri. Sroufe (1996) mengemukakan bahwa remaja yang berada pada masa menuju kematangan mempunyai kemungkinan yang besar untuk mengalami kecemasan. Pada masa ini, remaja digambarkan aktif

menjelajahi berbagai pilihan untuk menentukan identitas diri. Mereka masih kebingungan untuk menentukan identitas yang sesuai dengan dirinya sehingga emosi mereka sangat labil. sebagai akibatnya mereka sering keliru menanggapi suatu situasi.

Hasil penelitian Cratty (1973) mengenai fluktuasi *anxiety* pada umur-umur tertentu memperoleh simpulan bahwa: (a) *anxiety* akan memuncak pada usia sekitar dua puluh tahun (*late adolescent years*), (b) pada usia sekitar tiga puluh tahun, *anxiety* cenderung untuk menurun, dan (c) setelah usia enam puluh tahun, *anxiety* biasanya mulai naik lagi. *Anxiety* memuncak pada umur dua puluhan karena mendekati puncak-puncak potensi fisik (*physical potentials*) yaitu tahun paling produktif dalam karier seorang atlet.

Tidak dapat disangkal situasi pertandingan memberikan pengaruh yang menekan pada atlet. Reaksi tersebut sangat bergantung pada atlet yang bersangkutan. Pada atlet yang sensitif (peka), situasi ini dapat menimbulkan kecemasan. Sifat kecemasan olahraga juga berubah sesuai situasi pertandingan, yaitu sebelum, selama, dan mendekati akhir pertandingan. Hal ini sesuai dengan yang dikemukakan Cratty (1973) sebagai berikut. (a) Biasanya sebelum pertandingan, *anxiety* naik disebabkan oleh bayangan akan beratnya tugas atau pertandingan yang akan datang, (b) selama pertandingan, tingkat *anxiety* biasanya menurun karena atlet telah beradaptasi dengan situasi pertandingan, dan (c) mendekati akhir pertandingan, tingkat *anxiety* biasanya mulai naik kembali, terutama apabila skor pertandingan sama atau hanya berbeda sedikit saja. Dalam beberapa cabang olahraga yang durasinya berlangsung untuk waktu yang lama seperti menembak, panahan, dan beberapa nomor atletik, fluktuasi kecemasan atlet biasanya semakin tinggi.

Strategi mengatasi kecemasan olahraga

Kecemasan olahraga menjadi suatu masalah yang memerlukan jalan keluar. Hardy, Jones & Gould (1991) mengemukakan bahwa pertandingan olahraga berlangsung dalam suatu lingkungan yang sangat menekan sehingga kemampuan mengawasi tekanan sangat penting. Para atlet boleh menggunakan berbagai strategi dan keterampilan untuk mengatasi situasi-situas tersebut. Gunarsa (2004) menyatakan bahwa untuk membantu atlet

yang mengalami kecemasan dapat digunakan pendekatan konseling yang sesuai dengan karakteristik atlet tersebut ataupun yang sesuai dengan permasalahannya. Setyobroto (1989) juga menyatakan bahwa untuk membantu memecahkan masalah kecemasan yang dihadapi atlet, dapat digunakan pendekatan psikologik dengan terapi konseling. Suinn (1990) menyarankan agar atlet diberi program konseling yang mengajarkan strategi mengatasi kecemasan. Kecemasan olahraga merupakan gangguan emosi yang dihadapi oleh atlet sehingga konseling sangat berperan dalam membantu atlet yang mengalami gangguan emosi dengan memberikan konseling berdasarkan pendekatan yang dapat dilaksanakan. Salah satu bentuk konseling adalah *rational emotive behavior therapy*.

Perangkat instrumen kecemasan olahraga

Beberapa instrument kecemasan olahraga yang telah dikembangkan untuk mengukur kecemasan atlet di antaranya adalah *Taylor Manifest Anxiety Scale* (TMAS), *Autonomic Perception Questionnaire* (APQ), *Affective Adjective Checklist* (AACL), *Activation Deactivation Checklist* (AD-ACL), *Trait Anxiety Inventory* (TAI), *Sport Competition Anxiety Test* (SCAT), *State Anxiety Inventory* (SAI), *Competitive State Anxiety Inventory* I (CSAI I), *Competitive State Anxiety Inventory* II (CSAI 11), dan *Sport Anxiety Scale* (SAS). *Taylor Manifest Anxiety Scale* (TMAS) dikembangkan tahun 1951 (sitat dalam, 1989). Tes ini terdiri atas 50 butir dengan alternatif jawaban ya/tidak. Instrumen ini juga dikembangkan untuk anak-anak yaitu *Children Manifest Anxiety Scale* (CMAS). *Autonomic Perception Questionnaire* (APQ) dikembangkan oleh Mendler dan Urviller (1958, sitat dalam Gunarsa, 1989) yang terdiri atas tiga bagian. *Affective Adjective Checklist* (AACL) dikembangkan oleh Zuckerman (1960, sitat dalam Gunarsa, 1989). Instrumen ini terdiri dari 21 butir (sebelas butir digolongkan sebagai kecemasan yang positif dan sepuluh butir digolongkan sebagai kecemasan yang negatif). *Activation Deactivation Checklist* (AD-ACL) dikembangkan Thayer (1967, sitat dalam Gunarsa, 1989) untuk mengukur aktivitas melalui empat dimensi yang independent. *Trait Anxiety Inventory* (TAI) dikembangkan oleh Spielbelger, Gorsuch, &Lushene(1970, sitat dalam Gunarsa, 1989) untuk mengukur general *non-transitory anxiety*. Tes ini terdiri atas 20 butir. *Sport Competition*

Anxiety Test (SCAT), ini dikenal untuk anak-anak dan untuk orang dewasa. Tes ini merupakan modifikasi dari TAI yang dikembangkan oleh Martens (sitat dalam Gunarsa, 1989) dan terdiri atas 15 butir. Jawaban tes dalam bentuk tiga skala Likert. *State Anxiety Inventory* (SAI) dikembangkan oleh Spielbelger (1970, sitat dalarn Gunarsa, 1989). SAI dan TAI untuk, trait-A dianggap sebagai tes pasangan. Tes ini terdiri atas 21 pertanyaan yang dinilai dengan kata: "Perasaan saya...." Tes ini menggunakan empat tipe skala Likert. *Competitive State Anxiety Inventory* I (CSAI I), dibuat berdasarkan penelitian Spielbelger (1970, sitat dalam Gunarsa, 1989). Lima butir dari dua puluh butir SAI dipakai sebagai sub-skala. CSAI yang terdiri atas sepuluh butir yang berasal dari tes induk lalu dimodifikasi oleh Martens. *Competitive State Anxiety Inventory* II (CSAI II) dikembangkan oleh Martens, Burton, Vealey, Smith, & Bump (1981, sitat dalam Gunarsa, 1989). Instrumen ini mengukur aspek state-A yang multidimensional pada kompetisi (somatik, kognitif, dan rasa percaya diri). Pendekatan multidimensional memberi penjelasan yang lebih banyak tentang respon seorang atlet terhadap situasi kompetisi. Tes ini terdiri atas 27 butir dan dijawab dalam bentuk empat skala Likert. Terakhir, mengingat kebutuhan yang jelas bagi suatu ukuran multidimensi dari kecemasan kompetisi olahraga yang membedakan aspek kognitif dari somatik (badan) berdasarkan konsep tersebut, Smith, Smoll, & Schutz (1990) mengembangkan skala kecemasan olahraga (*Sport Anxiety Scale*, SAS).

Di antara kesepuluh contoh instrument yang ada, SAS dinilai sebagai yang paling sesuai dengan penelitian kali ini, karena terkait langsung dengan pengukuran kecemasan subjek. Penyusunan SAS berawal dengan serangkaian kajian analisis faktor. Smith etal. (1990) memberikan suatu versi tiga puluh butir dari skala itu pada atlet SMU pria dan wanita (N=451) dan pada suatu sampel lepas para pemain football antar-kampus (N=123). Hasil analisis komponen-prinsip (*principal-components*, PC) penjajakan terpisah atas kedua himpunan data menemukan tiga faktor yang serupa. Meskipun analisis-analisis penyelidikan ini menghasilkan faktor-faktor yang bisa ditafsirkan, delapan soal yang bermasalah dihapus dari skala tersebut. Smith etal. (1990) kemudian memberikan ulang versi tiga puluh soal dari SAS tersebut pada 384 atlet dari sampel SMU semula dan

memberikan suatu versi 22 soal terevisi dari skala tersebut kepada suatu sampel terpisah yang terdiri etas 490 atlet SMU. Dengan menggunakan serangkaian analisis faktor penegasan (*confirmatory factor analyses*/CFA) kemungkinan maksimum, Smith etal. (1990) menemukan bahwa solusi yang paling cocok diberikan oleh suatu versi 21 soal yang terdiri atas tiga faktor dari skala tersebut. Versi akhir SAS terdiri dari suatu sub-skala kecemasan somatic sembilan soal, sub-skala kecemasan sembilan soal, dan suatu sub-skala gangguan konsentrasi tiga soal. Tingkat-tingkat konsistensi internal yang dapat diterima menurut Cronbach (1951, sitat dalam Dunn, Wilson, & Syrotuik, 2000) bagi semua sub-skala dilaporkan oleh Smith etal. (1990) dari analisis-analisis terpisah atas kedua kumpulan data CFA: kecemasan somatik 0,880 dan O.920; khawatir, 0,820 dan 0,860; serta gangguan konsentrasi, 0,074 dan 0,810.

Meskipun SAS diakui sebagai instrumen *competitive trait anxiety* (CTA) pilihan di kalangan banyak psikolog olahraga tetapi para ahli psikometrik telah berulang-ulang memperingatkan para peneliti untuk hati-hati dalam menyimpulkan bahwa struktur faktor dan kompos isi faktor suatu instrumen itu konsisten antar-sampel, sebelum pengulangan atas usaha-usaha analisis faktor aslinya dilakukan (Gorsuch, 1983; Messick 1989; Pedhazur & Schmelkin, 1991). Gauvin dan Russel (1993) juga menganjurkan para peneliti psikologi olahraga mengulang bukti validitas asli mengenai alat pengukuran sebelum berlanjut menggunakan alat tersebut dalam situasi-situasi riset, sekalipun tidak ada perbedaan budaya atau bahasa antara sampel-sampel validitas aslinya dan populasi sasaran yang dipilih. Mengingat keterbatasan ketersediaan bukti-bukti empiris yang terbatas yang mendukung komposisi dan struktur faktor SAS, tampaknya masuk akal untuk mengisyaratkan bahwa dibutuhkan lebih banyak bukti validitas mengenai struktur laten dari instrumen tersebut. Dunn et al, (2000) telah melakukan studi untuk menyelidiki dan memastikan komposisi dan struktur faktor dari SAS menggunakan tiga kumpulan data yang diperoleh dalam proyek-proyek riset mandiri. Sekalipun hasil yang diperoleh cukup memuaskan, tetapi instrumen tersebut belum tentu dapat mengukur kecemasan olahraga di Indonesia secara valid dan reliabel, karena belum tentu sesuai dengan karakteristik kepribadian atlet Indonesia.

Oleh karena itu, penulis ingin mencoba mengembangkan instrumen kecemasan olahraga yang valid dan reliabel untuk atlet Indonesia.

Metode Penelitian

Variabel

Variabel penelitian adalah kecemasan olahraga. Kecemasan olahraga didefinisikan sebagai keadaan cemas, gelisah, dan tidak tenang dengan menganggap pertandingan sebagai sesuatu yang membahayakan. Dalam penelitian ini, kecemasan olahraga diwakili dengan skor angket kecemasan olahraga hasil adaptasi dari tes SAS. Instrumen ini meliputi empat aspek kecemasan olahraga, yaitu: motorik, afektif, somatik, dan kognitif.

Instrumen

Instrumen ini merupakan sejumlah butir pernyataan yang menggambarkan gejala dan gangguan kognitif, afektif 'somatik, dan motorik yang dialami atlet saat menghadapi pertandingan; gejala dan gangguan ini merupakan indikasi kecemasan, serta skala kecemasan olahraga ini dirancang dalam bentuk *self report* (laporan diri) (Stodolsky, 1985). Tujuannya ialah agar subjek mengungkapkan pikiran dan perasaannya sesegera mungkin setelah menghadapi pertandingan.

Populasi dan Sampel

Penelitian ini melibatkan para atlet dari enam klub sepakbola pemula Kota Banda Aceh. Jumlah total subjek penelitian adalah 406 atlet dan 6 orang pelatih. Rincian subjek penelitian adalah sebagai berikut. Tahap wawancara sebanyak 18 atlet dan 6 orang pelatih, tahap grup nominal sebanyak 84 atlet, tahap Q-sort sebanyak 8 orang ahli, dan tahap uji coba satu sebanyak 406 atlet klub sepakbola pemula Kota Banda Aceh. Uji coba tahap kedua dilakukan pada atlet klub sepakbola pemula Provinsi Aceh, dengan populasi yang terdiri atas 32 klub dan sejumlah 1860 atlet. Subjek uji coba tahap kedua hanya 15 klub dengan 1000 atlet. Pemilihan subjek penelitian dilakukan, dengan teknik pengambilan subjek berumpun *(clustered sampling)* dengan teknik *purposive sampling*, yaitu berdasarkan prestasi klub

yang baik, tingginya frekuensi mengikuti liga, dan manaiemen klub yang baik.

Prosedur Pengembangan

Dalam penelitian ilmiah ada tiga jenis instrumen yang paling sering dipakai, yaitu angket, tes dan skala nilai (Hadi, 1991:1). Lebih lanjut Hadi menjelaskan: "Angket digunakan untuk menyelidiki pendapat subjek mengenai hal atau untuk mengungkapkan keadaan pribadi responden. Tes digunakan untuk mengungkapkan karakteristik individu, khususnya kemampuan, bakat, minat, sikap dan kepribadian." Skala nilai digunakan untuk menilai keadaan pribadi orang lain atau mengenai sesuatu hal tertentu, Berdasarkan penjelasan tersebut, ketiga bentuk instrument baik tes, angket maupun skalai nilai memiliki kesamaan, terutama dari tujuan penelitian. Oleh sebab itu, instrumen kecemasan olahraga sepakbola pemula Kota Banda Aceh berisi pernyataan dengan skala penilaian berkisar 1 (satu) sampai 4 (empat) sesuai model yang dikembangkan oleh Likert. Penggunaan skala nilai 1 sampai 4 diharapkan dapat memamahi salah satu persyaratan panting yang harus dimiliki oleh suatu instrumen penelitian yaitu ketelitian, di samping kesahihan dan keterandalan (Hadi, 1991).

Atlet diminta untuk merespon butir pernyataan itu sesuai dengan yang dialaminya dengan memilih butir yang paling sesuai dengan dirinya pada saat menghadapi pertandingan. Alternatif jawaban responden telah ditentukan dengan menggunakan skala Likert yakni; "Sangat Sesuai (SS), Sesuai (S), Agak Sesuai (AS), dan Tidak Sesuai (TS)." Pemberian skor untuk skala kecemasan olahraga disesuaikan dengan jawaban butir pertanyaan, yakni: SS = 4, S = 3, AS = 2, TS = 1. Penentuan tingkat kecemasan tes didasarkan pada skor yang tercantum pada Tabel 1.

Tabel 1. Klasifikasi Tingkat Kecemasan Berdasarkan Skor Setiap Skala

Jenis Instrumen		Tingkat Kecemasan		
	Rendah	Agak Rendah	Agak Tinggi	Tinggi
Skala Kecemasan Olahraga	1-22	23-44	45-66	67-88

Prosedur pengembangan instrumen pada penelitian ini mengikuti Costin (1989) yang mengemukakan, bahwa kecemasan mempengaruhi aspek kepribadian individu dan bersifat: cognitive *(cognitively)*, affective *(affectively)*, somatic *(somatically),* dan motoric *(motorically)*. Upaya ini terbagi menjadi dua tahap, yaitu: (1) mengadaptasi instrumen *Sport Anxiety Scale* (SAS) yang dikembangkan oleh Smith, Smoll, dan Schutz (1990), dan (2) proses pengumpulan butir baru.

Adaptasi instrumen

Proses adaptasi terhadap instrumen SAS yang dikembangkan oleh Smith, Smoll, dan Schutz (1990) melalui dua tahap, yaitu: menerjemahkan instrumen tersebut ke dalam bahasa Indonesia dan meminta bantuan teman sejawat untuk memeriksa terjemahan tersebut, selanjutnya mengonsultasikan hasil terjemahan tersebut kepada ahli.

Pengumpulan butir-butir baru (item pool)

Pengumpulan butir baru dilakukan sebagaimana telah disarankan oleh Mutohir (1986, 1987, 1994). Pengumpulan butir-butir baru meliputi empat tahap, yaitu: (a) pengumpulan bakal butir, (b) pemilihan butir butir, (c) penyusunan skala, dan (d) penguji cobaan instrumen.

Pengumpulan bakal butir. Bakal bulir dikumpulkan melalui dua cara, yaitu wawancara dan proses grup nominal. Teknik wawancara meliputi studi pendahuluan terhadap 18 orang atlet dan 6 orang pelatih klub sepakbola pemula Kota Banda Aceh. Untuk mempermudah teknik wawancara, peneliti membuat panduan wawancara. Wawancara tersebut bertujuan mengenali gejala dan gangguan yang dialami atlet saat menghadapi pertandingan. Hasil wawancara dicatat dan digunakan untuk melengkapi teknik proses grup nominal.

Pengumpulan bakal butir kedua dilakukan dengan teknik proses grup nominal. Teknik ini dikembangkan oleh Delbecq dan Vande Ven sejak l971. Teknik ini memberi kesempatan kepada setiap peserta diskusi untuk berpartisipasi aktif secara bergantian sesuai giliran. Setiap peserta diminta menuliskan pendapat mereka pada secarik kertas. Pendapat ini akan dinilai oleh setiap anggota kelompok secara anonim untuk menjamin kebebasan berpendapat (Sample, 1984).

Teknik grup nominal dilakukan pada atlet sepakbola pemula Kota Banda Aceh yang berjumlah 84 atlet atau 20 persen dari seluruh jumlah atlet klub sepakbola pemula Kota Banda Aceh. Langkah-langkah teknik grup nominal dalam mengumpulkan bakal butir telah disederhanakan Mutohir (1987) menjadi dua tahap. Tahap pertama, para atlet dikumpulkan dalam satu ruangan dan mereka masing-masing diminta untuk menulis pada kertas yang disediakan tentang gejala serta gangguan yang dialami atlet saat menghadapi pertandingan. Tahap kedua, hasil adaptasi dan hasil wawancara dengan atlet dan pelatih digunakan dalam proses diskusi kelompok. Hasil wawancara selanjutnya diklasifikasi secara bersama antara peneliti dengan anggota grup Q-sort menurut empat sub-dimensi yang telah ditentukan sebelumnya.

Pemilihan butir-butir (screening of item pool) dengan teknik Q-sort. Sesuai dengan pendapat Mutohir (1986, 1987, 1994), proses pemilihan butir (*screening process of item pool*) dilakukan untuk mereduksi butir-butir yang mencerminkan gejala dan gangguan kecemasan olahraga. Untuk seleksi butir dilakukan dengan kegiatan "Q-sort" dan "analisis faktor." Kegiatan Q-sort dilakukan melalui pengumpulan setiap butir dan ditulis dalam kertas ukuran 5 x 5 cm. Prosedur kegiatan Q-sort adalah: (1) menentukan anggota kelompok Q-sort (peneliti dibantu oleh 8 orang dosen FKIP Unsyiah, 5 orang dosen jurusan pendidikan olahraga dan 3 orang dosen jurusan pendidikan bimbingan konseling), (2) pemberian penjelasan tentang pengertian dan tujuan Q-sort kepada para anggota, dan (3) penyaringan butir-butir oleh anggota kelompok untuk setiap dimensi menjadi tiga kategori menurut kepentingannya, yaitu: "amat penting", "cukup penting", dan "tidak penting." Kriteria penyaringan adalah kejelasan dimensi yang diwakili dan penilaian derajat kepentingan butir oleh mayoritas anggota grup Q-sort (>60%). Melalui tahap ini, diperoleh dan disepakati 54 gejala dan gangguan kecemasan olahraga yang dinilai paling penting.

Penyusunan skala (construction of scales). Instrumen yang dikembangkan pada penelitian ini diharapkan dapat berfungsi sebagai *diagnostic feedback* sehingga sekalipun memiliki cakupan yang luas, instrumen tersebut tetap harus memuat butir-butir spesifik untuk dapat mengukur gejala dan gangguan yang dialami atlet sewaktu bertanding secara reliabel dan valid.

Oleh sebab itu instrumen disusun melalui prosedur-prosedur tertentu sehingga dapat dipertanggungjawabkan secara ilmiah, baik adaptasi, pengumpulan butir, seleksi butir, uji coba dan penyusunan skala penilaian.

Penguji-cobaan skala. Pada tahap awal, dilakukan penetapan dimensionalitas instrumen melalui *factorial validity*. Tahap ini bertujuan mengenali faktor-faktor utama yang merupakan gejala dan gangguan kecemasan olahraga menurut atlet. Penyusunan skala meliputi: (1) analisis butir, (2) reliabilitas instrumen, (3) analisis factor, dan (4) penyusunan skala penilaian. Semua tahapan tersebut bertujuan untuk menghasilkan instrumen handal untuk mengungkap tingkat kecemasan olahraga.

Analisis data

Data yang telah dikumpulkan selanjutnya dikategorikan dan dianalisis baik secara kualitatif maupun kuantitatif. Data kualitatit digunakan agar dapat lebih menjelaskan permasalahan yang dibahas secara naratil. Sedangkan data kuantitatif dianalisis dengan menggunakan teknik-teknik statistik.

Butir-butir yang dikumpulkan sebagai indikator gejala dan gangguan kecemasan yang telah diperoleh melalui adaptasi, wawancara, proses grup, seleksi dan kategori dengan menggunakan Q-sort akan menjadi butir-butir yang digunakan dalam proses uji coba. Selanjutnya hasil uji coba instrumen dianalisis dengan teknik statistik berikut ini (a) Analisis reliabilitas dengan menggunakan Alpha Cronbach*,* (b) Analisisi faktor analisis dengan menggunakan, "teknik Principal *Axis Factoring dan Rotation Method Oblimin with Kaiser Normalization.*" Seluruh analisis dilakukan dengan menggunakan bantuan komputer melalui program *Statistical Package for Social Sciences* (SPSS) (Nie, 1975).

Prosedur pelaksanaan

Penelitian pengembangkan alat ukur kecemasan olahraga dilaksanakan pada atlet dan pelatih sepakbola pemula di Provinsi Aceh. Waktu pelaksanaannya mulai bulan Mei sampai Oktober 2011, dengan rincian waktu pelaksanaan, yakni tahap observasi dilaksanakan pada bulan Mei 2011, tahap wawancara, teknik grup nominal, Q-sort, dan uji coba dilaksanakan pada bulan September sampai Oktober 2011.

Adapun prosedur pelaksanaan penelitian, peneliti menghubungi pelatih semua klub yang menjadi subjek penelitian untuk memperoleh izin untuk mendekati pemain mereka untuk ikut dalam penelitian ini. Keikutsertaan ini bersifat sukarela dan persetujuan tertulis diperoleh dari tiap atlet sebelum pengumpulan data. Peneliti melakukan pertemuan dengan atlet sesuai dengan jadwal yang ditentukan oleh masing-masing pelatih. Selanjutnya peneliti melakukan tahap-tahap, yakni wawancara, grup nomimal, Q-sort, dan pernbagian skala kecemasan olahraga kepada subjek pada tahap uji coba.

Hasil uji ooba instrumen

Berdasarkan hasil pengolahan data, dari 54 butir pernyataan ternyata ada delapan butir pernyataan yang tidak memiliki konsistensi internal. Dengan demikian delapan butir tersebut tidak diikutsertakan ke dalam instrumen penelitian, sehingga instrumen penelitian ini hanya terdiri atas 46 butir pernyataan. Beberapa butir yang tidak diikutsertakan dalam alat ukur penelitian ini adalah butir 5 (sering mencabut-cabut rambut), 8 (sering mengatup geraham), 19 (merasa kehilangan energi), 24 (memikirkan tentang tampilan buruk), 42 (menjadi pendiam), 43 (banyak berbicara), 50 (sering mengigit bibir), dan 54 (merasa tidak nyaman). Estimasi reliabilitas dengan menggunakan formula Alpha Cronbach menunjukkan, bahwa keempat faktor memiliki koefisien reliabilitas antara 0,531 sampai dengan 0,856, yang dikategorikan cukup baik sampai baik.

Setelah diadakan uji coba, diperoleh 46 butir pernyataan dalam satu instrumen. Selanjutnya ke-46 butir pernyataan tersebut dijadikan instrumen untuk mengukur tingkat kecemasan atlet klub sepakbola pemula di Provinsi Aceh. Hasil uji coba tahap kedua instrumen kecemasan olahraga dikenakan pada subjek 1000 atlet pada atlet klub sepakbola Provinsi Aceh, selanjutnya dianalisis kembali. Hal ini untuk menjawab apakah instrumen kecemasan olahraga yang dikembangkan pada atlet klub sepakbola pemula Kota Banda Aceh dapat digunakan oleh atlet secara reliabel? Data hasil uji coba tahap kedua dianalisis dengan menggunakan estimasi reliabilitas. Hasil analisis tersebut sebagai berikut.

Tabel 2. Muatan faktor keempat puluh satu butir instrumen

Keterangan	Muatan factor > 0,300 & hanya muatan muncul pada satu faktor	Muatan Faktor > 0,300 tetapi muncul pada beberapa faktor	Muatan faktor > 0,300
Nomor Butir	52, 18, 13, 35, 39, 32, 36, 28, 29, 25, 21, 48, 1, 26, 22, 31, 41, 23, 49, 20, 47, & 9	30, 11, 6, 40, 51, 38, 10, 27, 33, 2, 12, 45, 53, & 46	34, 17, 15, 16, & 7

Berdasarkan hasil pengolahan data, ternyata 46 butir pernyataan seluruhnya memiliki konsistensi inetrnal yang baik. Keempat puluh enam butir tersebut diikutsertakan dalam instrumen penelitian. Oleh karena, nilai probabilitas ke 46 butir tersebut lebih kecil dari 0,05. Adapun estimasi reliabilitas dengan menggunakan formula Alpha Cronbach menunjukkan, bahwa keempat faktor memiliki koefisien reliabilitas dengan alpha antara antara 0,631 sampai dengan 0,823. Oleh sebab itu, instrumen tersebut dinilai dapat memberikan hasil yang dapat dipercaya (andal), untuk selanjutnya melalui proses analisis faktor.

Uji KMO and Bartlett's Test. Uji KMO and Bartlett's test dilakukan untuk mengetahui apakah variabel dan sampel yang ada dapat dianalisis lebih lanjut atau tidak. Adapun hasil uji KMO and Bartlett's test adalah 0,734 dengan signifikansi $p<0,001$. Oleh karena angka tersebut sudah di atas 0,50 dan signifikansi jauh di bawah 0,05, variabel dan sampel yang ada dapat dianalisis lebih lanjut.

Uji anti-image matrices. Uji *anti-image matrices* atau *anti-image correlation* dilakukan untuk mengetahui butir yang masuk dalam faktor dan butir yang tidak masuk dalam faktor. Adapun hasil uji *anti-image correlation* ternyata dari empat puluh enam butir yang ada hanya empat puluh satu butir yang masuk dalam faktor dan lima butir tidak masuk dalam faktor, antara lain adalah butir-butir 3 (mudah jengkel), 4 (kurang bergairah), 14 (pesimis), 37

(khawatir tentang pencapaian tujuan), dan 44 (merasa malu). Hal ini disebabkan hasil *measure of sampling adequacy* yang diperoleh lebih kecil dari pada 0,50.

Tabel 3. Hasil Uji Pattern Matrix

Butir		Faktor			
		1 Motorik	2 Afektif	3 Somatik	4 Kognitif
25.	Cepat putus asa		0,666		
21.	Sembrono.		0,671		
48.	Memiliki keraguan diri		0,694		
1.	Jantung berdebar-debar keras			0,343	
26.	Selalu ingin buang air kecil			0,631	
22.	Mengalami ketegangan			0,733	
31.	Pernafasan tidak teratur			0,738	
41.	Sering minum air			0,881	
23.	Berkeringat dingin			0,711	
49.	Suka tidur			0,799	
20.	Memikirkan tidak bisa berkonsentrasi				0,567
47.	Berpikir tentang hal tidak berhubungan.				0,496
9.	Pikiran negatif mengganggu konsentrasi				0,364
52.	Raut muka dan dahi berkerut.	0,695			
18.	Gemetar	0,933			
13.	Kaki terasa berat	0,497			
35.	Sering menggaruk-garuk kepala	0,676			
39.	Otot-otot sakit	0,794			
32.	Sering jalan mondar-mandir.	0,632			
36.	Badan lesu	0,339			
28.	Tubuh terasa kaku	0,722			
29.	Mengalami ketegangan otot (kram)	0,562			

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin with Kaiser Normalization

Berdasarkan hasil analisis faktor dari 41 yang tersebar menjadi empat faktor, ternyata dari beberapa kali melakukan analisis faktor, hanya 22 butir yang memiliki muatan faktor lebih besar dari 0,30 pada *pattern matrix* dan sekaligus hanya muncul pada satu faktor. Terdapat enam butir memiliki muatan faktor lebih kecil dari 0,30 pada *pattern matrix*. Selain itu, terdapat pula 13 butir yang memiliki muatan faktor lebih besar dari 0,30 tetapi muncul pada beberapa faktor. Butir-butir yang memiliki muatan faktor lebih kecil dari 0,30 dan muncul pada beberapa faktor digugurkan. Hasil akhir dapat dilihat pada Tabel 3 yang telah tersaji di atas.

Factor correlation matrix. Nilai-nilai yang diperoleh dari korelasi berdasarkan nilai analisis faktor, terlihat dengan jelas bahwa muatan faktor dari 22 butir yang terdistribusi pada 4 faktor dan masing-masing butir bermuatan secara signifikan pada faktor yang ditargetkan untuk diukur. Tampak faktor-faktor dalam Skala kecemasan olahraga saling berkorelasi satu dengan yang lain, walaupun secara analisis faktor masing-masing faktor tampak jelas mengukur dimensi kecemasan yang harus diukur. Adapun *factor correlation matrix* hasil analisis faktor dapat dilihat pada Tabel 4.

Tabel 4. Hasil *factor correlation matrix*

Faktor		1 Motorik	2 Afektif	3 Somatic	4 Kognitif
1	Motorik	1,000	0,481	0,417	0,508
2	Afektif	0,481	1,000	0,424	0,403
3	Somatik	0,417	0,424	1,000	0,412
4	Kognitif	0,508	0,403	0,412	1,000

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin with Kaiser Normalization

Hasil Penelitian

Secara singkat, pengembangan instrumen kecemasan olahraga ini melalui dua kegiatan, yakni kegiatan adaptasi dan kegiatan pengumpulan butir-butir baru dengan langkah-langkah sebagai berikut: (1) wawancara, (2) grup nominal, dan (3) grup Q-sort. Selanjutnya instrumen tersebut diuji cobakan sebanyak dua tahap, yakni uji coba satu dan uji coba dua. Hasil uji coba selanjutnya dianalisis dengan menggunakan pengujian validitas, reliabilitas, dan analisis faktor.

Berdasarkan hasil analisis, dapat dilihat faktor dan butir yang mencerminkan faktor-faktor gejala dan gangguan kecemasan olahraga yang diikutsertakan dalam skala kecemasan olahraga adalah sebagai berikut.

Faktor motorik. Berdasarkan hasil pengujian validitas, pengujian reliabilitas, dan analisis faktor hanya sembilan gejala dan gangguan dari faktor motorik yang mencerminkan gejala dan gangguan kecemasan olahraga yang diikutsertakan dalam skala kecemasan olahraga. Adapun gejala dan gangguan kecemasan olahraga tampak pada diri atlet melalui keadaan raut muka dan dahi berkerut, gemetar, kaki terasa berat, sering menggaruk-garuk kepala, otot-otot sakit, sering jalan mondar-mandir, badan lesu, tubuh terasa kaku, dan mengalami ketegangan otot. Faktor motorik adalah faktor pertama dalam kecemasan olahraga dengan korelasi antara skor butir dengan skor faktor terletak pada rentang 0,946 sampai 0,358.

Faktor afektif. Berdasarkan hasil pengujian, validitas, pengujian reliabilitas, dan analisis faktor hanya tiga gejala dan gangguan dari faktor afektif yang mencerminkan gejala dan gangguan kecemasan olahraga yang diikutsertakan dalam skala kecemasan olahraga. Adapun gejala dan gangguan kecemasan olahraga tampak pada diri atlet melalui pengakuan atlet seperti merasa cepat putus asa, sembrono, dan memiliki keraguan diri. Faktor afektif adalah faktor kedua dalam kecemasan olahraga dengan korelasi antara, butir dengan skor faktor terletak pada rentang 0,773 sampai 0,654.

Faktor somatik. Berdasarkan hasil pengujian validitas, pengujian reliabilitas, dan analisis faktor hanya tujuh gejala dan gangguan dari faktor

somatik yang mencerminkan gejala dan gangguan kecemasan olahraga yang diikutsertakan dalam skala kecemasan olahraga. Adapun gejala dan gangguan kecemasan olahraga tampak pada diri atlet dalam keadaan jantung berdebar-debar keras, ingin buang air kecil, mengalami ketegangan, pemafasan tidak teratur, sering minum air, berkeringat dingin, dan sukar tidur. Faktor somatik adalah faktor ketiga dalam kecemasan olahraga dengan korelasi antara skor butir dengan skor faktor terletak pada rentang 0,861 sampai 0,392.

Faktor kognitif. Berdasarkan hasil pengujian validitas, pengujian reliabilitas, dan analisis faktor hanya tiga gejala dan gangguan dari faktor kognitif yang mencerminkan gejala dan gangguan kecemasan olahraga yang diikutsertakan dalam skala kecemasan olahraga. Adapun gejala dan gangguan kecemasan olahraga tampak pada diri atlet dalam wujud tidak bisa berkonsentrasi, berpikir tentang hal-hal yang tidak berhubungan, dan pikiran negatif yang mengganggu konsentrasi. Faktor kognitif termasuk faktor keempat dalam kecemasan olahraga dengan korelasi antara skor butir dengan skor faktor terletak pada rentang 0,501 sampai 0,334.

Berdasarkan hasil reduksi, pengujian validitas, pengujian reliabilitas, dan analisis faktor, dapat disimpulkan dari sekian banyak butir yang dirancang untuk skala kecemasan olahraga setelah melalui reduksi dan analisis statistik hanya 22 butir pernyataan yang terdiri atas 4 faktor yang dapat dipakai untuk skala kecemasan olahraga yang memiliki tingkat kesahihan yang sedang serta memiliki tingkat keterandalan yang tinggi.

Di samping itu, mengingat kebutuhan yang jelas terhadap aspek multidimensi suatu instrumen yang bisa membedakan aspek-aspek kecemasan olahraga, skala ini disusun atas empat dimensi kecemasan olahraga, yakni dimensi motorik, dimensi afektif, dimensi somatik, dan dimensi kognitif. Adapun skala kecemasan olahraga (SKO) tersebut di bawah ini.

Tabel 5. Skala kecemasan olahraga

No.	Butir Pernyataan	SS	S	AS	TS
1.	Jantung saya berdebar-debar keras saat menghadapi pertandingan.				
2.	Tubuh saya kaku saat menghadapi pertandingan.				
3.	Pikiran-pikiran negatif mengganggu konsentrasi saya saat pertandingan.				
4.	Saya sukar tidur saat menghadapi pertandingan				
5.	Kaki saya berat saat menghadapi pertandingan.				
6.	Saya gemetar saat menghadapi pertandingan.				
7.	Saya memikirkan tidak akan mampu berkonsentrasi saat pertandingan.				
8.	Saya sembrono saat pertandingan.				
9.	Saya mengalami ketegangan saat menghadapi pertandingan.				
10.	Saya berkeringat dingin saat menghadapi pertandingan.				
11.	Saya cepat putus asa saat pertandingan, apabila berada dalam keadaan tertekan.				
12.	Saya selalu ingin buang air kecil saat menghadapi pertandingan.				
13.	Saya mengalami ketegangan otot (krem) saat pertandingan.				
14.	Saya memiliki keraguan diri saat pertandingan.				
15.	Pernafasan saya tidak teratur saat menghadapi pertandingan.				
16.	Saya sering jalan mondar-mandir saat menghadapi pertandingan.				
17.	Saya sering menggaruk-garuk kepala saat menghadapi pertandingan.				
18.	Badan saya lesu saat menghadapi pertandingan.				
19.	Otot-otot saya sakit saat menghadapi pertandingan.				
20.	Saya sering minum air saat menghadapi pertandingan.				
21.	Saya menemukan diri saya berpikir tentang hal yang tidak berhubungan saat pertandingan.				
22.	Raut muka dan dahi saya berkerut saat menghadapi pertandingan.				

Simpulan

Hasil penelitian kali ini menunjukkan bahwa skala kecemasan olahraga yang terdiri atas empat faktor dan 22 butir pernyataan sudah dapat dinilai valid dan reliabel untuk mengukur kecemasan olahraga. Instrumen kecemasan olahraga yang dikembangkan di Provinsi Aceh sudah memperoleh replikasi kelayakan dan keandalan. Bagaimanapun juga instrumen kecemasan olahraga tersebut masih perlu dikembangkan dan diujicobakan kepada subjek yang berbeda sehingga benar-benar memperoleh implikasi kelayakan dan keandalan untuk digunakan pada situasi penelitian yang lebih luas.

Penelitian pengembangan instrumen kecemasan olahraga ini masih memiliki beberapa kelemahan, antara lain adalah: (1) analisis faktor yang masih terbatas pada tahap penggunaan analisis faktor penyelidikan (exploratory factor analyses/ EFA), belum sampai pada tahap analisis faktor penegasan (confirmatory factor analyses/CFA), dan (2) pelaksanaan uji coba tahap satu dan dua seharusnya dilakukan setelah menghadapi pertandingan (subjek harus bertanding sebelum menjawab angket), tetapi dalam pelaksanaannya subjek curna dikondisikan untuk mereview kembali tentang gejala dan gangguan yang subjek rasakan saat menghadapi lawan yang paling tangguh.

Daftar Pustaka

Anshel, M.H. 1997. *Sport psychology: From theory to practice*. Scottsdale, AZ:Gorsuch Scarisbrick.

Cratty, B.J. 1973. *Psychology in contemporary, sport.* New Jersey: Prentice Hall, Inc.

Costin, F. & Draguns, J. G. 1989. Abnormal psychology: Patterns, issues, and inventions. New York: John Wiley & Sons.

Dunn, J. G. H., Wilson, P., & Syroruik, D. G. 2000. Reexamining the factorial composition and factor structure of the sport anxiety scale. *Journal of Sport and Exercise Pscyhology*, 22,183-193.

Gorsuch. R. L. 1983. *Factor analyses* (2"led.). Hillsdale. NJ: Erlbaum

Gunarsa, S.D. 1989. *Psikologi olahraga. Jakarta*: PT. Gunung Mulia.

Gunarsa, S.D. 2004, 8 Mei. *Latihan mental terlupakan* [on-line]. Diambil 8 Juli 2004, dari http://www.kompas.com/kompas%2Dcetak/0405/08/or/1013124.htm

Hadi, S. 1991. Analisis butir untuk instrumen angket tes dan skala nilai dengan Basic. Yogyakarta: Andi Offset.

Hardy, L., Jones, G., & Gould, D. 1996. *Understanding psychology preparation for sport*: Theory and practice of elite performers. Chichester: Wiley.

Martens, R., Vealey, R. S. & Burton, D. 1990. *Competitive Anxiety in Sport.* Champaign, Illinois: Human Kinetics.

Messick, S. 1989. Validity. In R.L. Linn (Ed.). *Education measurement* (3rd ed., pp. 1317). New York: Amereican Council on Education.

Mutohir, T. C. 1986. *The Development and examination of student evaluation of teaching a effectiveness in an Indonesian higher education setting.* Thesis, unpublished, Australia Macquarie University, Sydney.

Mutohir, T. C. 1987. *Laporan penelitian pengembangan instrumen evaluasi efektifitas pengajaran di perguruan tinggi (suatu rintisan).* Surabaya, Pusat Penelitian IKIP Surabaya, Depdikbud.

Mutohir, T. C. 1994. *Evaluasi keefektivan pengajaran studi kasus di IKIP Surabaya.*

Nie, N_ H., Hull, C. H., Jenkins. J.G., Steinbrenner, ' K., & Bent, D. H. 1975. *Statistical package for social sciences.* New York: McGraw-Hill.

Pedhazur, E.J. & Schmelkin, L.P. 1991. *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Sample, J.A. (1984). Nominal group technique: An alternative to brainstorming. J*ournal of Extension,* 22 (2). Retrieved July 8, 2064, from: http:/'/'www.joe.org/ foe 1984 march/ iw2. html.

Setyobroto, S. 1989. *Psikologi olahraga.* Jakarta: Anem Kosong Anem.

Sroufe, L. Alan, Cooper, Robert G., and Ganie B. DeHart. *Child developinent, Its nature and course.* New York: McGraw. Hill

Smith, R.E. & Sarason, I.G. 1993. *Psychology thefrontiers of behavior.* New York: Harper & Row Publisher.

Smith, R. E., Smoll, F. L, & Schutz. R. W. 1990. *Measurement and correlates ofsport specific cognitive and somatic trait anxiety:* The sport anxiety scale. *Anxiety Research,* 2, 263280.

Spielberger, C.S. 1972. Theory and research on anxiety: *Anxiety behaviour.* Academic press.

Stodolsky, S. 1985. Telling math: Origin of math aversion and anxiety. *Educational Psyhologist*, 3, 125-133.

EVALUASI PROGRAM PENINGKATAN KULAIFIKASI GURU MADRASAH DI SUMATERA SELATAN

Nyayu Khodijah
Fakultas Tarbiyah IAIN Raden Fatah Palembang
nyayu-dd@yahoo.com

Abstrak

Penelitian ini bertujuan untuk mengevaluasi program peningkatan kualifikasi guru, khususnya yang dilaksanakan oleh Kementerian Agama terhadap guru madrasah, dan untuk menilai apakah program yang berjalan telah memberikan dampak yang diharapkan. Penelitian dilakukan menggunakan metode deskriptif evaluatif menggunakan model evaluasi konteks, input, proses dan produk (context, input, process, product, CIPP), dan difokuskan pada evaluasi proses dan produk. Penelitian dilaksanakan di provinsi Sumatera Selatan. Responden penelitian ini adalah guru madrasah yang menjadi peserta program kualifikasi angkatan pertama. Pengumpulan data dilakukan menggunakan teknik angket dan wawancara, sedang analisis data menggunakan teknik analisis kuantitatif. Hasil penelitian menunjukkan bahwa pelaksanaan program peningkatan kualifikasi guru madrasah sebagian besar telah berjalan sesuai dengan yang direncanakan. Namun, bila dilihat satu per satu, khusus pada proses pelaksanaannya masih ada beberapa hal yang belum sepenuhnya sesuai dengan yang direncanakan. Dampak yang dirasakan oleh guru dengan diterapkannya program ini terlihat pada peningkatan kompetensi mengajar, peningkatan karir, serta peningkatan kesejahteraan, tapi dampaknya pada kemungkinan penerapan materi perkuliahan yang diperoleh masih belum optimal.

Kata kunci: *guru madrasah, program peningkatan kualifikasi guru*

EVALUATION OF MADRASAH TEACHERS' QUALIFICATIONS IMPROVEMENT PROGRAM IN SOUTH SUMATRA

Nyayu Khodijah
Fakultas Tarbiyah IAIN Raden Fatah Palembang
nyayu-dd@yahoo.com

Abstract

This study aimed to evaluate the program of improving teacher qualifications, especially those implemented by the Ministry of Religious Affairs of the madrasah teachers, and to assess whether the program has given the expected impact. The study was conducted by using descriptive evaluative method. The research was conducted in South Sumatra province. The respondents were all madrasah teachers who became the first group participants in that program. Data collection techniques used were questionnaire and interview, data analysis techniques used are of quantitative analysis. The results showed that the program for improving teacher qualifications had been running largely as planned. However, when viewed one by one, specifically on the implementation process there are still many things that are not completely as planned. Moderate impact felt by the teachers can be seen at improving teaching competencies, improving careers, as well as improving the welfare, but still less on the possibility of applying materials obtained from lecturing.

Keywords: *madrasah teachers, the program for improving teacher qualifications*

Pendahuluan

Madrasah merupakan lembaga pendidikan yang dalam beberapa hal memiliki banyak kekurangan. Di antara kelemahan terpenting yang dimiliki madrasah adalah rendahnya kualitas lulusan madrasah. Rendahnya kualitas lulusan madrasah tampak dari masih rendahnya prosentase siswa yang dapat melanjutkan ke jenjang pendidikan yang lebih tinggi. Hanya 88,7% lulusan MI dapat melanjutkan ke MTs, SMP atau Pesantren; 63,9% lulusan MTs melanjutkan ke MA, SMA, SMK atau Pesantren, dan hanya 35,6% lulusan MA dapat melanjutkan ke Perguruan Tinggi (Depag, 2007: 3).

Ada banyak hal yang menyebabkan kompetensi lulusan madrasah itu rendah, diantaranya adalah karena kondisi guru madrasah yang kurang mendukung. Diantara kondisi yang cukup memprihatinkan adalah kualifikasi sebagian besar guru madrasah yang masih berada di bawah S1/D4. Berdasarkan data Direktorat Pendidikan pada Madrasah Dirjen Pendidikan Islam Depag RI tahun 2007, dari 584.095 orang guru madrasah, hanya 224.886 (38,5%) saja yang memiliki kualifikasi S1. Bila dilihat untuk masing-masing jenjang, hanya 6,9% guru RA yang memiliki kualifikasi S1, 16,8% untuk guru MI, 54,6% untuk guru MTs, dan 72% untuk guru MA. Dari angka-angka tersebut jelas terlihat masih belum memadainya kualifikasi guru madrasah.

Menyadari pentingnya peningkatan kualifikasi guru dalam rangka mengejar ketertinggalan kualitas lulusan madrasah, Kementerian Agama memberikan prioritas tersendiri pada upaya peningkatan kualifikasi guru madrasah. Dengan meningkatnya kualifikasi guru diharapkan mampu meningkatkan kualitas guru. Peningkatan kualitas guru ini menjadi bagian dari arah kebijakan strategis pengembangan kualitas madrasah. Sejalan dengan arah kebijakan tersebut, salah satu program peningkatan kualitas guru madrasah yang dilaksanakan Kementerian Agama adalah program peningkatan kualifikasi guru ke jenjang S1/D-4 melalui kerjasama dengan LPTK. Program ini ditujukan untuk membantu guru madrasah dalam meningkatkan kualifikasi akademiknya sesuai dengan amanah Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan.

Di Propinsi Sumatera Selatan sendiri, program peningkatan kualifikasi guru madrasah diselenggarakan oleh Fakultas Tarbiyah IAIN Raden Fatah. Program ini sudah berjalan selama ± 4 tahun. Akan tetapi, sejauh ini belum pernah dilakukan evaluasi eksternal apakah program yang dilaksanakan betul-betul dilaksanakan sesuai harapan dan juga mencapai hasil yang diharapkan. Dalam pengamatan sementara yang dilakukan, ditemukan ada beberapa permasalahan dalam proses pelaksanaan program itu sendiri, baik dari segi persiapan maupun pelaksanaannya. Dalam konteks itulah penelitian ini dipandang sangat signifikan, sebagai upaya mengevaluasi secara komprehensif terhadap pelaksanaan program peningkatan kualifikasi guru madrasah yang sudah berjalan.

Penelitian ini bertujuan untuk mengevaluasi program peningkatan kualifikasi guru, khususnya yang dilaksanakan oleh Kementerian Agama terhadap guru-guru madrasah di propinsi Sumatera Selatan. Kriterianya adalah apakah proses pelaksanaan program peningkatan kualifikasi tersebut telah sesuai dengan perencanaan. Selain itu, penelitian ini juga ditujukan untuk menilai apakah program peningkatan kualifikasi guru madrasah yang berjalan telah memberikan dampak yang diharapkan berdasarkan penilaian guru.

Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode deskriptif evaluatif guna mendeskripsikan dan mengevaluasi pelaksanaan program peningkatan kualifikasi guru madrasah berdasarkan standar kriteria perencanaan yang disusun dan tujuan yang hendak dicapai. Model evaluasi yang digunakan adalah model evaluasi CIPP *(Context, Input, Process, Product)* dari Stufflebeam (1985: 169-179). Mengingat penelitian ini berfokus pada evaluasi pelaksanaan program yang sudah berjalan implementasinya, maka evaluasi dibatasi pada evaluasi proses dan evaluasi produk. Evaluasi proses bertujuan untuk menilai sejauh mana perencanaan yang disusun telah dapat terlaksana dan apa yang harus direvisi. Standar yang digunakan adalah perencanaan yang telah disusun oleh Kemenag. Sedang evaluasi produk bertujuan untuk mengetahui dampak dirasakan oleh guru dengan

penerapan program tersebut. Standar yang digunakan adalah tujuan penyelenggaraan program sertifikasi guru agama yang telah digariskan oleh Kemenag.

Penelitian dilaksanakan di provinsi Sumatera Selatan. LPTK (PTAI) penyelenggara program peningkatan kualifikasi guru madrasah di provinsi ini adalah Fakultas Tarbiyah IAIN Raden Fatah Palembang. Responden penelitian ini adalah guru madrasah yang menjadi peserta program kualifikasi angkatan pertama yang berjumlah 250 orang, sedang sampelnya adalah 30 orang.

Pengumpulan data menggunakan teknik angket dan wawancara. Teknik angket digunakan untuk menggali data proses pelaksanaan program peningkatan kualifikasi guru madrasah yang sudah berjalan selama ini, sedang wawancara digunakan untuk memperoleh data pelengkap tentang proses pelaksanaan program peningkatan kualifikasi yang belum tergali melalui angket dan untuk menggali data tentang dampak yang dirasakan oleh guru dari pelaksanaan program peningkatan kualifikasi berdasarkan penilaian (judgement) guru. Teknik analisis data yang digunakan adalah teknik analisis kuantitatif dengan menggunakan statistik persentase. Selain itu juga dilakukan analisis kualitatif terhadap data kualitatif yang terkumpul melalui wawancara.

Hasil Penelitian

Proses Pelaksanaan Program

Evaluasi terhadap proses pelaksanaan program peningkatan kualifikasi guru madrasah dalam hal ini meliputi evaluasi terhadap persiapan program dan evaluasi terhadap proses pembelajaran.

Dalam hal persiapan program, secara umum hasil evaluasi menunjukkan bahwa persiapan program Peningkatan Kualifikasi Guru Madrasah sebagian besar sudah berjalan sesuai dengan yang direncanakan. Akan tetapi jika dilihat dari masing-masing aspek, beberapa belum sepenuhnya sesuai. Hasil analisis data selengkapnya sebagai berikut:

1. Prosedur Penyaluran Bantuan

Evaluasi terhadap prosedur penyaluran bantuan dalam hal ini meliputi waktu penerimaan bantuan, kesesuaian jumlah bantuan yang diterima, sistem penyaluran dana, efektifitas mekanisme penyaluran dana, kemudahan dalam menerima bantuan biaya pendidikan, serta potongan biaya baik oleh lembaga penyalur (bank) maupun LPTK penyelenggara.

Dilihat dari waktu penerimaan bantuan, diketahui bahwa dari 30 orang sampel guru, 9 orang (30%) menyatakan menerima bantuan pada bulan Maret-April, 6 orang (20%) menyatakan menerimanya di bulan Mei-Juni, 5 orang (16,66%) menyatakan menerimanya di bulan Januari-Februari, 5 orang (16,66%) menyatakan menerimanya di bulan September-Oktober, 4 orang (13,33%) menyatakan menerimanya di bulan Nopember-Desember, dan hanya 1 orang (3,33%) menyatakan menerimanya di bulan Juli-Agustus. Dilihat dari kesesuaian jumlah bantuan yang diterima, diketahui bahwa 18 orang (60%) menyatakan bahwa jumlah bantuan yang diterima tidak sesuai dengan yang seharusnya dan 12 orang (40%) menyatakan sesuai. Dilihat dari sistem penyaluran dana, diketahui bahwa semua responden (100%) menyatakan bahwa dana bantuan yang diterima tidak melalui prosedur pengiriman langsung ke rekening pribadi.

Pada efektifitas mekanisme penyaluran dana, diketahui bahwa 15 orang (50%) menyatakan bahwa mekanisme penyaluran dana sudah efektif dan 15 orang (50%) lainnya menyatakan tidak efektif. Dilihat dari kemudahan dalam menerima bantuan, diketahui bahwa 22 orang (73,33%) merasakan kemudahan dalam menerima bantuan yang diberikan dan 8 orang (26,67%) merasakan tidak mudah. Dilihat dari potongan biaya oleh lembaga penyalur (bank), diketahui bahwa 18 orang (60%) menyatakan adanya potongan oleh bank. Dilihat dari potongan biaya oleh LPTK penyelenggara program, diketahui bahwa 17 orang (56,67%) menyatakan adanya potongan oleh LPTK dan 13 orang (43,33%) menyatakan tidak ada potongan.

2. Prosedur Registrasi

Evaluasi terhadap prosedur registrasi dalam hal ini meliputi sistem registrasi, sistem seleksi, dan kesulitan dalam memenuhi persyaratan administrasi.

Dilihat dari sistem registrasi untuk menjadi calon penerima bantuan, diketahui bahwa dari 30 orang sampel guru, 28 orang (93,33%) menyatakan sistem registrasinya mudah dan hanya 2 orang (6,67%) menyatakan sulit/berbelit-belit. Dilihat dari sistem seleksi untuk menjadi calon penerima bantuan, diketahui bahwa 19 orang (63,33%) menyatakan sistem seleksinya sudah dilakukan sesuai prosedur dan 11 orang (36,67%) menyatakan tidak sesuai prosedur. Dilihat dari kesulitan dalam memenuhi persyaratan administrasi untuk memperoleh bantuan, diketahui bahwa 20 orang (66,67%) menyatakan tidak mengalami kesulitan dan 10 orang (33,33%) menyatakan mengalami kesulitan.

Dalam hal proses pembelajaran, **s**ecara umum hasil evaluasi menunjukkan bahwa proses pembelajaran pada program Peningkatan Kualifikasi Guru Madrasah sebagian besar sudah berjalan sesuai dengan yang direncanakan. Akan tetapi, jika dilihat dari masing-masing aspek, baik dilihat dari kesiapan dosen dalam mengajar, kualitas pengelolaan perkuliahan, metode pembelajaran, ruang kuliah, fasilitas penunjang proses pembelajaran, sarana dan prasarana, lingkungan kampus, serta pelayanan akademik dan administrasi, beberapa belum sepenuhnya sesuai. Hasil analisis data selengkapnya sebagai berikut:

a. Kesiapan Dosen Dalam Mengajar

Evaluasi terhadap kesiapan dosen dalam mengajar dalam hal ini meliputi kesesuaian bidang keahlian dosen dengan mata kuliah yang diajarkan, tingkat kualifikasi pendidikan dosen, rata-rata pemahaman dosen terhadap kurikulum, rata-rata kesiapan dosen dalam kelengkapan dokumen perangkat pembelajaran, rata-rata kesiapan dosen dalam merancang tugas-tugas, waktu luang untuk memberikan bimbingan akademik, ketepatan waktu mengajar sesuai jadwal, serta ketepatan waktu memberikan nilai hasil belajar mahasiswa sesuai jadwal.

Dilihat dari kesesuaian bidang keahlian dosen dengan mata kuliah yang diajarkan, diketahui bahwa dari 30 orang sampel guru, 15 orang (50%) menyatakan bahwa kesesuaian bidang keahlian dosen dengan mata kuliah yang diajarkan berkisar antara 75% - 99%, 9 orang (30%) menyatakan kesesuaiannya berkisar antara 50% - 74%, 3 orang (10%) menyatakan kesesuaiannya sudah 100%, dan 3 orang (10%) menyatakan kesesuaiannya < 50%. Mengkaji tingkat kualifikasi pendidikan dosen, diketahui bahwa 12 orang (40%) menyatakan bahwa 75% - 99% berkualifikasi pendidikan S2 ke atas, 9 orang (30%) menyatakan kurang dari 50% dosen yang berkualifikasi pendidikan S2 ke atas, 6 orang (20%) menyatakan 100% dosen sudah berkualifikasi pendidikan S2 ke atas, dan 3 orang (10%) menyatakan bahwa dosen yang berkualifikasi pendidikan S2 ke atas sebanyak 50% - 74%.

Dilihat dari tingkat pemahaman dosen terhadap kurikulum, diketahui bahwa 16 orang (53,33%) menyatakan tingkat pemahaman dosen terhadap kurikulum berada dalam kategori baik, 12 orang (40%) menyatakan cukup, 2 orang (6,66%) menyatakan sangat kurang, dan yang menyatakan kurang tidak ada sama sekali. Kesiapan dosen terhadap dokumen perangkat pembelajaran, diketahui bahwa 20 orang (66,66%) menyatakan kesiapan dosen mencapai 75 – 90%, 7 orang (23,33%) menyatakan kesiapan dosen lebih dari 90%, 3 orang (10%) menyatakan kesiapan dosen antara 50 – 74%, dan yang menyatakan kesiapan dosen terhadap dokumen perangkat pembelajaran kurang dari 50% tidak ada sama sekali.

Dilihat dari kesiapan dosen dalam merancang tugas-tugas, diketahui bahwa 18 orang (60%) menyatakan kesiapan dosen mencapai 75 – 90%, 6 orang (20%) menyatakan kesiapan dosen lebih dari 90%, 6 orang (20%) menyatakan kesiapan dosen antara 50 – 74%, dan yang menyatakan kesiapan dosen dalam merancang tugas-tugas kurang dari 50% tidak ada sama sekali. Dilihat dari waktu luang dosen untuk memberikan bimbingan akademik di luar jam kuliah, diketahui bahwa 14 orang (46,67%) menyatakan hanya sebagian kecil dosen yang mempunyai waktu untuk memberikan bimbingan akademik di luar jam kuliah, 10 orang (33,33%) menyatakan tidak ada dosen yang mempunyai waktu untuk memberikan

bimbingan akademik di luar jam kuliah, 4 orang (13,33%) menyatakan sebagian besar dosen mempunyai waktu untuk memberikan bimbingan akademik di luar jam kuliah, dan 2 orang (6,67%) menyatakan rata-rata dosen mempunyai waktu untuk memberikan bimbingan akademik di luar jam kuliah.

Dilihat dari ketepatan waktu dosen dalam mengajar, diketahui bahwa 16 orang (53,33%) menyatakan sebagian besar dosen mengajar sesuai dengan jadwal, 6 orang (20%) menyatakan rata-rata dosen mengajar sesuai dengan jadwal, 5 orang (16,67%) menyatakan hanya sebagian kecil dosen yang mengajar sesuai dengan jadwal, dan 3 orang (10%) menyatakan tidak ada dosen yang mengajar sesuai jadwal. Dilihat dari ketepatan waktu dosen dalam memberikan nilai, diketahui bahwa 14 orang (46,67%) menyatakan rata-rata dosen memberikan nilai tepat waktu, 11 orang (36,67%) menyatakan sebagian besar dosen memberikan nilai tepat waktu, 4 orang (13,33%) menyatakan tidak ada dosen yang memberikan nilai tepat waktu, dan 1 orang (3,33%) menyatakan hanya sebagian kecil dosen yang memberikan nilai tepat waktu.

b. Kualitas Pengelolaan Perkuliahan

Evaluasi terhadap kualitas pengelolaan perkuliahan dalam hal ini meliputi pola pengelolaan perkuliahan yang digunakan dosen, penerapan program remedial, pengayaan, dan percepatan dalam implementasi prinsip-prinsip pembelajaran tuntas, respon dan antusiasme mahasiswa program peningkatsn kualifikasi guru dalam perkuliahan, dan interaksi dosen dan mahasiswa dalam proses perkulihan di dalam kelas.

Dilihat dari pola pengelolaan perkuliahan yang digunakan dosen, diketahui bahwa dari 30 orang sampel guru, 13 orang (43,34%) menyatakan dosen menggunakan gabungan pola individual dan kelompok, 10 orang (33,33%) menyatakan menggunakan pola kelompok belajar, 7 orang (23,33%) menggunakan pola klasikal, dan yang menyatakan dosen menggunakan pola individual tidak ada sama sekali. Dilihat dari penerapan program remedial, pengayaan, dan percepatan dalam implementasi prinsip-prinsip pembelajaran tuntas, diketahui bahwa 17 orang (56,67%) menyatakan bahwa penerapan program tersebut dalam proses

pembelajaran oleh dosen berada dalam kategori cukup, 7 orang (23,33%) menyatakan kurang, 6 orang (20%) menyatakan sudah baik, dan yang menyatakan tidak adanya penerapan program tersebut tidak ada sama sekali.

Dilihat dari respon dan antusiasme mahasiswa, diketahui bahwa 21 orang (70%) menyatakan bahwa respon dan antusiasme mahasiswa dalam proses pembelajaran berada dalam kategori baik, 5 orang (16,67%) menyatakan cukup, 4 orang (13,33%) menyatakan kurang, dan yang menyatakan respon dan antusiasme mahasiswa sangat kurang tidak ada sama sekali. Dilihat dari interaksi dosen dan mahasiswa dalam proses perkuliahan di kelas, diketahui bahwa 24 orang (80%) menyatakan bahwa interaksi dosen dan mahasiswa dalam proses perkuliahan di kelas berada dalam kategori baik, 4 orang (13,33%) menyatakan cukup, 2 orang (6,67%) menyatakan kurang, dan yang menyatakan interaksi dosen dan mahasiswa sangat kurang tidak ada sama sekali.

c. Metode Pembelajaran

Evaluasi terhadap metode pembelajaran dalam hal ini meliputi frekuensi penggunaan multimedia/ICT dalam perkuliahan serta penggunaan metode pembelajaran yang bervariasi.

Dilihat dari frekuensi penggunaan multimedia, diketahui bahwa 24 orang (80%) menyatakan bahwa frekuensi penggunaan multimedia dalam proses pembelajaran sering dilakukan, 4 orang (13,33%) menyatakan kadang-kadang saja, 2 orang (6,67%) menyatakan jarang, dan yang menyatakan tidak pernah tidak ada sama sekali. Dilihat dari penggunaan penggunaan metode pembelajaran yang bervariasi, diketahui bahwa 13 orang (43,34%) menyatakan bahwa hanya sebagian kecil dosen yang mengajar dengan metode bervariasi, 10 orang (33,33%) menyatakan rata-rata dosen mengajar dengan metode bervariasi, 7 orang (23,33%) menyatakan sebagian besar dosen mengajar dengan metode bervariasi, dan yang menyatakan metode mengajar dosen tidak bervariasi tidak ada sama sekali.

d. Ruang Kuliah

Evaluasi terhadap ruang kuliah dalam hal ini meliputi kondisi bangunan ruang kuliah, penerangan, pencahayaan, dan sirkulasi udara di ruang kuliah, kondisi kebersihan ruang kuliah, serta kondisi umum ruang perkuliahan.

Dilihat dari kondisi bangunan ruang kuliah, diketahui bahwa dari 30 orang sampel guru, 21 orang (70%) menyatakan kondisi bangunan ruang kuliah cukup baik, 5 orang (16,67%) menyatakan kurang baik, 4 orang (13,33%) menyatakan baik, dan yang menyatakan sangat jelek tidak ada sama sekali. Dilihat dari penerangan, pencahayaan, dan sirkulasi udara di ruang kuliah, diketahui bahwa 22 orang (73,34%) menyatakan kondisi penerangan, pencahayaan, dan sirkulasi udara di ruang kuliah cukup baik, 4 orang (13,33%) menyatakan baik, 4 orang (13,33%) menyatakan kurang baik, dan yang menyatakan sangat jelek tidak ada sama sekali.

Dilihat dari kondisi kebersihan ruang kuliah, diketahui bahwa 20 orang (66,67%) menyatakan kondisi kebersihan ruang kuliah cukup bersih, 6 orang (20%) menyatakan bersih, 3 orang (10%) menyatakan kotor, dan yang menyatakan sangat kotor 1 orang (3,33%). Dilihat dari kondisi umum ruang perkuliahan, diketahui bahwa 18 orang (60%) menyatakan kondisi umum ruang perkuliahan cukup kondusif, 8 orang (26,67%) menyatakan kondusif, 4 orang (13,33%) menyatakan kurang kondusif, dan yang menyatakan sangat kurang kondusif tidak ada sama sekali.

e. Fasilitas Penunjang Proses Pembelajaran

Evaluasi terhadap fasilitas penunjang proses pembelajaran dalam hal ini meliputi kondisi fasilitas perkulihan, keberadaan ICT/Multimedia di ruang perkuliahan, dan penggunaan ICT/Multimedia yang ada di ruang kuliah.

Dilihat dari kondisi fasilitas perkulihan (*whiteboard*, spidol, OHP, dan lain-lain), diketahui bahwa dari 30 orang sampel guru, 10 orang (33,33%) menyatakan kondisi fasilitas perkuliahan cukup memadai, 10 orang (33,33%) menyatakan kurang memadai, 9 orang (30%) menyatakan memadai, dan yang menyatakan sangat kurang memadai hanya 1 orang (3,34%). Dilihat dari kondisi keberadaan ICT/Multimedia (TV, Komputer,

VCD, OHP, dll) di ruang perkuliahan, diketahui bahwa 16 orang (53,33%) menyatakan bahwa ICT/Multimedia tidak ada sama sekali, 8 orang (26,67%) menyatakan keberadaan ICT/Multimedia masih kurang, 4 orang (13,33%) menyatakan cukup, dan yang menyatakan lengkap hanya 2 orang (6,67%).

Dilihat dari penggunaan ICT/Multimedia (TV, Komputer, VCD, OHP, dll) yang ada di ruang kuliah, diketahui bahwa 15 orang (50%) menyatakan bahwa ICT/Multimedia tidak difungsikan karena tidak ada, 6 orang (20%) menyatakan sering difungsikan, 5 orang (16,67%) menyatakan jarang difungsikan, dan 4 orang (13,33%) menyatakan hanya difungsikan kadang-kadang.

f. Pelayanan Akademik dan Administrasi

Evaluasi terhadap pelayanan akademik dan administrasi dalam hal ini meliputi kualitas pelayanan akademik yang diberikan, cara pengelola program merespon kesulitan akademik yang dihadapi mahasiswa, perhatian pengelola program terhadap kebutuhan dan minat mahasiswa, perhatian personal yang didapatkan mahasiswa, kualitas pelayanan administrasi yang diberikan, kesediaan pegawai administrasi dalam memberikan pelayanan administrasi kepada mahasiswa, serta kemudahan mahasiswa untuk memperoleh informasi tentang nilai/kelulusan mata kuliah yang diikuti.

Dilihat dari kualitas pelayanan akademik yang diberikan, diketahui bahwa dari 30 orang sampel guru, 24 orang (80%) menyatakan bahwa kualitas pelayanan akademik yang diberikan sudah memuaskan dan 6 orang (20%) menyatakan tidak memuaskan. Dilihat dari cara pengelola program dalam merespon kesulitan akademik yang dihadapi mahasiswa, diketahui bahwa 15 orang (50%) menyatakan bahwa pihak pengelola program hanya membantu seperlunya, 9 orang (30%) menyatakan membantu dengan simpatik, dan 2 orang (6,67%) menyatakan pengelola program acuh tak acuh.

Dilihat dari perhatian pengelola program terhadap kebutuhan dan minat mahasiswa, diketahui bahwa 13 orang (43,33%) menyatakan bahwa perhatian pengelola program terhadap kebutuhan dan minat mahasiswa sudah baik, 8 orang (26,66%) menyatakan biasa-biasa saja, 5 orang

(16,66%) menyatakan sangat baik. Dilihat dari perhatian personal yang didapatkan mahasiswa dari pengelola program, diketahui bahwa 17 orang (56,67%) menyatakan bahwa perhatian personal yang didapatkan mahasiswa dari pengelola program sudah baik, 8 orang (26,67%) menyatakan sangat baik, 5 orang (16,66%) menyatakan biasa-biasa saja, dan yang menyatakan kurang perhatian tidak ada sama sekali.

Dilihat dari kualitas pelayanan administrasi yang diberikan, diketahui bahwa 19 orang (63,33%) menyatakan bahwa kualitas pelayanan administrasi yang diberikan sudah cukup memuaskan, 8 orang (26,67%) menyatakan memuaskan, 2 orang (6,67%) menyatakan kurang memuaskan, dan 1 orang (3,33%) menyatakan tidak memuaskan. Dilihat dari kesediaan pegawai administrasi dalam memberikan pelayanan administrasi kepada mahasiswa, diketahui bahwa 14 orang (46,67%) menyatakan bahwa kesediaan pegawai administrasi dalam memberikan pelayanan administrasi kepada mahasiswa sudah baik, 10 orang (33,33%) menyatakan cukup, 6 orang (20%) menyatakan sangat baik, dan yang menyatakan kurang baik tidak ada sama sekali.

Dilihat dari kemudahan mahasiswa untuk memperoleh informasi tentang nilai/kelulusan mata kuliah yang diikuti, diketahui bahwa 23 orang (76,67%) menyatakan mengalami kemudahan dalam memperoleh informasi tentang nilai/kelulusan mata kuliah yang diikuti dan 7 orang (23,33%) menyatakan mengalami kesukaran.

Dampak yang Dirasakan oleh Guru dari Pelaksanaan Program

Secara umum, hasil evaluasi terhadap dampak program Peningkatan Kualifikasi Guru Madrasah baik terhadap peningkatan kompetensi mengajar guru, peningkatan karir guru, peningkatan kesejahteraan guru, maupun terhadap kemungkinan penerapan materi perkuliahan yang diperoleh sebagian besar sudah sesuai dengan yang diharapkan.

Dilihat dari dampak program Peningkatan Kualifikasi Guru Madrasah terhadap peningkatan kompetensi mengajar guru, diketahui bahwa dari 30 orang sampel guru, 26 orang (86,67%) menyatakan adanya dampak program terhadap peningkatan kompetensi mengajar guru dan 4

orang (13,33%) menyatakan tidak ada dampaknya. Dilihat dari dampak program Peningkatan Kualifikasi Guru Madrasah terhadap peningkatan karir guru, diketahui bahwa 29 orang (96,67%) menyatakan adanya dampak program ini terhadap peningkatan karir guru dan 1 orang (3,33%) menyatakan tidak ada dampaknya.

Dilihat dari dampak program Peningkatan Kualifikasi Guru Madrasah terhadap peningkatan kesejahteraan guru, diketahui bahwa 27 orang (90%) menyatakan adanya dampak program terhadap peningkatan kesejahteraan guru dan 3 orang (10%) menyatakan tidak ada dampaknya. Dilihat dari dampak program Peningkatan Kualifikasi Guru Madrasah terhadap kemungkinan penerapan materi perkuliahan yang diperoleh, diketahui bahwa 12 orang (40%) menyatakan bahwa materi perkuliahan yang diperoleh dapat diterapkan, 11 orang (36,67%) menyatakan sebagian besar dapat diterapkan, 5 orang (16,67%) menyatakan sebagian kecil dapat diterapkan, dan 2 orang (6,66%) menyatakan tiadk dapat diterapkan.

Pembahasan

Berdasarkan hasil evaluasi terhadap pelaksanaan program peningkatan kualifikasi guru madrasah menunjukkan bahwa secara umum sebagian besar sudah berjalan sesuai dengan yang direncanakan. Akan tetapi, jika dilihat satu persatu, khusus pada proses pelaksanaannya masih ada beberapa hal yang belum sepenuhnya sesuai dengan yang direncanakan.

Berkaitan dengan proses pelaksanaan program peningkatan kualifikasi guru madrasah, baik pada tahap persiapan program dan proses pembelajaran yang dilaksanakan beberapa diantaranya masih belum sepenuhnya sesuai harapan. Pada tahap persiapan program, khususnya pada prosedur penyaluran dana, ada keragaman yang cukup tinggi dalam waktu penerimaan bantuan, sebagian besar responden merasakan ketidaksesuaian jumlah bantuan yang diterima, bantuan yang diterima tidak melalui prosedur pengiriman langsung ke rekening pribadi, serta mekanisme penyaluran dana masih belum sepenuhnya efektif, adanya potongan oleh bank maupun oleh LPTK.

Pada proses pembelajaran, baik dilihat dari kesiapan dosen dalam mengajar, kualitas pengelolaan perkuliahan, metode pembelajaran, ruang kuliah, fasilitas penunjang proses pembelajaran, sarana dan prasarana, lingkungan kampus, maupun pelayanan akademik dan administrasi, sebagian besar masih belum sesuai dengan rencana. Dilihat dari kesiapan dosen dalam mengajar, hanya 10% yang menyatakan kesesuaian bidang keahlian dosen dengan mata kuliah yang diajarkan sudah mencapai 100%, hanya 20% yang menyatakan 100% dosen sudah berkualifikasi pendidikan S2, hanya 23,33% yang menyatakan kesiapan dosen terhadap dokumen perangkat pembelajaran mencapai lebih dari 90%, hanya 20% yang menyatakan kesiapan dosen dalam merancang tugas-tugas mencapai lebih dari 90%, serta hanya sebagian kecil dosen yang mempunyai waktu untuk memberikan bimbingan akademik di luar jam kuliah. Dilihat dari kualitas pengelolaan perkuliahan, hanya 20% yang menyatakan penerapan program remedial, pengayaan, dan percepatan dalam implementasi prinsip-prinsip pembelajaran tuntas oleh dosen sudah baik. Dilihat dari penggunaan metode pembelajaran, menurut sebagian besar responden hanya sebagian kecil dosen yang mengajar dengan metode bervariasi. Dilihat dari ruang kuliah, hanya 13,33% yang menyatakan ruang kuliah dalam kondisi baik, hanya 13,33% yang menyatakan penerangan, pencahayaan, dan sirkulasi udara di ruang kuliah dalam kondisi baik, hanya 20% yang menyatakan ruang kuliah dalam kondisi bersih, serta hanya 26,67% yang menyatakan ruang perkuliahan kondusif. Dilihat dari fasilitas penunjang proses pembelajaran, hanya 30% yang menyatakan kondisi fasilitas perkulihan memadai, hanya 6,67% yang menyatakan kondisi keberadaan ICT/Multimedia lengkap, kalaupun ada sebanyak 50% responden menyatakan bahwa ICT/Multimedia tersebut tidak difungsikan. Dilihat dari pelayanan akademik dan administrasi, sebanyak 6,67% responden menyatakan pengelola program acuh tak acuh dalam merespon kesulitan akademik yang dihadapi mahasiswa.

Sangat berbeda dengan hasil evaluasi terhadap proses pelaksanaannya, dampak yang dirasakan oleh guru dari program peningkatan kualifikasi guru madrasah ini sebagian besar sudah sesuai dengan yang diharapkan, baik dilihat dari dampaknya terhadap

peningkatan kompetensi mengajar guru, peningkatan karir guru, dan peningkatan kesejahteraan guru. Hanya saja dilihat dari dampak program terhadap kemungkinan penerapan materi perkuliahan yang diperoleh, hanya 40% responden yang menyatakan bahwa materi perkuliahan yang diperoleh dapat diterapkan.

Simpulan

Berdasarkan hasil penelitian dan pembahasan di atas, dapat diambil kesimpulan sebagai berikut:

1. Proses pelaksanaan program peningkatan kualifikasi guru madrasah di Sumatera Selatan secara umum sebagian besar sudah berjalan sesuai dengan yang direncanakan. Akan tetapi, jika dilihat satu persatu, khusus pada proses pelaksanaannya masih ada beberapa hal yang belum sepenuhnya sesuai dengan yang direncanakan. Hal-hal yang masih belum sesuai dengan yang direncanakan nampak baik pada tahap persiapan program maupun pada proses pembelajaran.
2. Dampak yang dirasakan oleh guru madrasah dengan diterapkannya program peningkatan kualifikasi guru ini terlihat pada peningkatan kompetensi mengajar guru, peningkatan karir guru, serta peningkatan kesejahteraan guru. Hanya saja dilihat dari dampak program terhadap kemungkinan penerapan materi perkuliahan yang diperoleh, hanya 40% responden yang menyatakan bahwa materi perkuliahan yang diperoleh dapat diterapkan.

Daftar Pustaka

Daradjat, Z., et al. 1996. Pendidikan Islam. Jakarta: Bumi Aksara Bekerjasama dengan Direktorat Jenderal Pembinaan Kelembagaan Agama Islam, Departemen Agama RI.

Djamas, N. 2005. "Posisi Madrasah Di Tengah Perubahan Sistem Pendidikan Islam". Jurnal Edukasi. Volume 3, Nomor 1, Januari-Maret 2005. Hlm. 23-29.

Idi, A. 2006. "UU No. 14/2005 Tentang Guru/Dosen: Antara Cita dan Fakta", Intizar: Jurnal Kajian Agama Islam dan Masyarakat, Vol. 12/No.2/Desember 2006. Hlm. 133-148.

Kementerian Agama RI. 2007. Buku Statistik Pendidikan Islam TP 2006/2007. Jakarta: Direktorat Jenderal Pendidikan Islam

Kunandar. 2007. Guru Profesional: Implementasi Kurikulum Tingkat Satuan Pendidikan (KTSP) dan Persiapan Menghadapi Sertifikasi Guru, Jakarta: PT. Raja Grafindo Persada.

Stufflebeam, Daniel L. & Shinkfield, A. J. 1985. Systematic Evaluation. USA: Kluwer-Nijhoff Publishing.

PENGARUH METODE DAN UKURAN SAMPEL TERHADAP VARIANSI SKOR HASIL PENYETARAAN

Tri Rijanto
Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Surabaya
hari_tri2001@yahoo.com

Abstrak

Penelitian ini bertujuan untuk memperoleh informasi perbedaan variansi skor hasil penyetaraan (*equating*) metode linear dan metode ekipersentil untuk ukuran sampel 200, 400, dan 800 pada Ujian Akhir Sekolah Berstandar Nasional (UASBN). Metode yang digunakan adalah simulasi dengan variabel metode penyetaraan dan banyaknya responden. Data penelitian berupa respons peserta UASBN SD/MI tahun pelajaran 2008/2009 mata pelajaran IPA di Jakarta Timur, yang ditentukan menggunakan teknik penarikan sampel acak dengan pengembalian. Hipotesis diuji menggunakan uji kesamaan variansi. Hasil penelitian dengan $\alpha = 0{,}05$ menunjukkan: (1) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp200}) tidak berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin200}) untuk ukuran sampel 200, (2) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp400}) tidak berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin400}) untuk ukuran sampel 400, dan (3) variansi skor penyetaraan metode ekipersentil (σ^2_{ekp800}) berbeda dengan variansi skor penyetaraan metode linear (σ^2_{lin800}) untuk ukuran sampel 800.

Kata kunci: *variansi skor, equating, metode ekipersentil, metode linear*

THE IMPACT OF METHODS AND SAMPLE SIZE TO THE SCORE VARIANCE OF EQUATING RESULT

Tri Rijanto
Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Surabaya
hari_tri2001@yahoo.com

Abstract

This study was aimed to obtain information on the difference of score variance as a result of equating linear method and equipercentile method for the sample size of 200, 400, and 800 in the Final Examinition of National Standardized Schools. The method used was a simulation of variables equating method and the number of respondents. The population are examinees from the 2008/2009 elementary school final examination for science class in East Jakarta. Random sampling with replacement technique was used. The hypotheses were tested using similarity variance. The results with $\alpha = 0{,}05$ showed that: (1) the equated score variance from equipercentile method (σ^2_{ekp200}) was not different from the equated score variance from linear method (σ^2_{lin200}) for the sample size of 200, (2) the equated score variance from equipercentile method (σ^2_{ekp400}) was not different from the equated score variance from linear method (σ^2_{lin400}) for the sample size of 400, and (3) the equated score variance from equipercentile method (σ^2_{ekp800}) was different from the equated score variance from linear method (σ^2_{lin800}) for the sample size of 800.

Keywords: *score variance, equating, equipercentile method, linear method*

Pendahuluan

Peraturan Pemerintah Republik Indonesia Nomor 19 tahun 2005 tentang Standar Nasional Pendidikan, Pasal 63 Ayat (1) menyebutkan bahwa penilaian pendidikan pada jenjang pendidikan dasar dan menengah terdiri atas: (1) penilaian hasil belajar oleh pendidik, (2) penilaian hasil belajar oleh satuan pendidikan, dan (3) penilaian hasil belajar oleh pemerintah. Penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan teknologi dan dilakukan dalam bentuk ujian nasional (UN).

Ujian nasional untuk sekolah dasar/madrasah ibtidaiyah/sekolah dasar luar biasa (SD/MI/SDLB) dilakukan pertama kali pada 2008. Ujian tersebut bernama Ujian Akhir Sekolah Berstandar Nasional (UASBN) sesuai dengan Peraturan Menteri Pendidikan Nasional RI Nomor 39 tahun 2007. Ujian ini bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam (IPA) serta mendorong tercapainya target wajib belajar pendidikan dasar yang bermutu (Depdiknas, 2007:6).

Ujian Akhir Sekolah Berstandar Nasional SD/MI/SDLB dilaksanakan secara terintegrasi dengan ujian sekolah/madrasah. Artinya setiap paket soal UASBN terdiri dari 25% soal yang ditetapkan oleh Badan Standar Nasional Pendidikan (BSNP) dan berlaku secara nasional, serta 75% soal yang ditetapkan oleh penyelenggara UASBN tingkat provinsi berdasarkan spesifikasi yang ditetapkan oleh BSNP. Dengan kata lain dalam paket soal terdiri dari 25% butir *anchor items* dan sisanya butir soal yang dibuat oleh masing-masing provinsi.

Instrumen penilaian yang digunakan oleh pemerintah dalam bentuk ujian nasional (UN) menurut Peraturan Menteri Pendidikan Nasional RI Nomor 20 tahun 2007 tentang Standar Penilaian Pendidikan memenuhi persyaratan substansi, konstruksi, bahasa, dan memiliki bukti validitas empirik serta menghasilkan skor yang dapat diperbandingkan antarsekolah, antardaerah, dan antartahun (Depdiknas, 2007:9). Keterbandingan skor antarsekolah, kabupaten/kota, provinsi, dan antartahun dapat diperoleh

jika semua peserta tes mengerjakan soal-soal (paket tes) yang sama. Perbedaan skor antarmereka menunjukkan perbedaan tingkat kemampuannya. Dalam praktiknya, pengadministrasian soal-soal sama antartahun, merugikan peserta tes yang mengerjakan pada tahun-tahun pertama dan menguntungkan mereka yang ikut tes pada tahun-tahun terakhir. Juga, pengadministrasian soal-soal yang sama di setiap sekolah sangat beresiko terhadap kebocoran.

Sebagai jalan keluar agar keadilan *(fairness)* dan kerahasiaan *(test security)* soal-soal ujian terjaga mengharuskan pengadministrasian paket-paket tes berbeda antartahun, daerah, dan tempat tes. Tetapi masalah lain muncul, dengan mengadministrasian paket-paket yang berbeda, perbedaan skor antarpeserta tes tidak dapat langsung disimpulkan adanya perbedaan kemampuan antarmereka, karena tingkat kesukaran paket yang digunakan mempengaruhi perbedaan tersebut.

Untuk menanggulangi ketidakadilan tersebut dilakukan penyamaan atau penyetaraan matriks skor *(equating)*. Penyetaraan matriks skor merupakan cara untuk memperoleh suatu konversi nilai dari skor mentah suatu paket, ke skor mentah paket yang lain. Jadi, melalui penyetaraan matriks skor dimungkinkan siswa menjawab benar 32 soal di Paket A, misalnya, mendapat nilai sama dengan siswa menjawab benar 30 soal di Paket B, karena Paket B lebih sukar dua soal dari Paket A.

Akibat lebih serius dengan tidak adanya penyetaraan *(equating)* adalah ketidakadilan pada kelulusan. Batas lulus 5,5 misalnya, pada paket-paket sukar akan merugikan siswa-siswa yang mengerjakannya dan akan menguntungkan mereka yang mendapatkan paket-paket mudah. Oleh karena itu kemungkinan besar ada siswa yang seharusnya lulus tetapi karena mendapatkan paket sukar menjadi tidak lulus. Dengan demikian masalah penyetaraan menjadi penting untuk dikaji melalui penelitian yang komprehensif.

Untuk dapat membandingkan atau menyetarakan skor ujian nasional diperlukan desain penyetaraan yang tepat. Hal ini sejalan dengan pendapat Marco, *et al.,* (1983:148) bahwa karena penyetaraan merupakan prosedur empirik, maka dibutuhkan desain untuk mengumpulkan data dan suatu aturan untuk mentransformasikan skor dari suatu tes pada skor tes yang

lain. Menurut Hambleton dan Swaminathan (1985:198) ada tiga desain dasar penyetaraan, yaitu metode kelompok tunggal *(single group method)*, metode kelompok ekivalen *(equivalent group method)*, dan metode tes jangkar *(anchor test design)*.

Terdapat dua cara penyamaan skor pada teori klasik yaitu penyamaan metode linear dan metode ekipersentil. Menurut Hambleton, Swaminathan, dan Rogers (1991:124) asumsi penyetaraan metode linear adalah kedua skor tes distribusinya berbeda, distribusi tersebut terkait dengan rerata dan simpangan bakunya. Dengan demikian pada metode linear membutuhkan asumsi rerata dan simpangan baku atau variansinya berbeda.

Pada teori klasik, terdapat jumlah butir tertentu serta jumlah peserta tertentu. Dalam waktu yang bersamaan, semua peserta menjawab semua butir ujian. Jawaban setiap peserta terhadap setiap butir ujian dapat berupa jawaban benar atau berupa jawaban salah. Jawaban benar dan salah menjadi skor pada ujian tersebut. Ada kalanya jawaban benar dan salah tersebut dikaitkan dengan sukar dan tidak sukarnya butir ujian tersebut dirasakan oleh peserta ujian. Akan tetapi sukar dan tidak sukarnya butir ujian sangat bergantung kepada para peserta ujian. Butir yang dirasakan tidak sukar oleh satu peserta mungkin saja dirasakan sukar oleh peserta lain. Dengan demikian dapat saja terjadi peserta tertentu merasa semua butir ujian adalah sukar sementara ada pula peserta lainnya yang merasa semua butir ujian adalah tidak sukar.

Taraf sukar dan daya beda butir merupakan ciri butir pada ujian klasik. Taraf sukar butir pada tes yang menggunakan skor dikotomi, rerata skor butir yang diperoleh berkaitan dengan proporsi peserta tes yang menjawab benar butir tersebut. Proporsi ini untuk butir *ke-i* biasanya dinyatakan sebagai p_i, pada analisis butir klasik p_i disebut sebagai taraf sukar butir *(item difficulty)*. Nilai optimum p tergantung pada tujuan penyelenggaraan tes. Misalnya, untuk tes yang diadakan untuk seleksi beasiswa atau tes penempatan, nilai optimum p rendah, tetapi tes yang dirancang untuk menyeleksi siswa yang perlu remidial nilai optimum relatif tinggi. Untuk menyeleksi siswa diperlukan rentangan kemampuan yang cukup besar maka nilai optimum p mendekati 0,50 (Croker dan Algina, 1986:69). Jika kemampuan peserta lebih dari indeks taraf sukar butir, maka

probabilitas menjawab benar akan lebih dari probabilitas menjawab salah sehingga p>0,50. Sebaliknya, jika kemampuan peserta tes kurang dari indeks taraf sukar butir, maka probabilitas jawaban benar akan kurang dari probabilitas jawaban salah sehingga $p < 0{,}50$.

Ciri kedua dalam analisis butir klasik adalah daya beda butir untuk tes berbentuk dikotomi. Daya beda butir adalah kemampuan suatu butir untuk membedakan peserta tes dari kelompok kemampuan tinggi dengan peserta tes kelompok kemampuan rendah. Daya beda butir dinyatakan dalam bentuk indeks daya beda *(discrimination index)* yang biasanya diberi lambang *D*. Jika $D>0$ berarti peserta kelompok kemampuan tinggi menjawab benar lebih banyak dibandingkan peserta kelompok kemampuan rendah. Sebaliknya jika harga $D < 0$ berarti peserta kelompok kemampuan rendah menjawab lebih banyak dari kelompok kemampuan tinggi.

Daya beda butir sebenarnya diestimasi menggunakan metode korelasi butir total. Kaplan dan Saccuzzo (1982:147) menyebut metode ini sebagai *point biserial method* karena menggunakan korelasi biserial titik *(point biserial correlation)*. Harga koefisien beserial titik antara $-1 \leq r_{pbis} \leq 1$. Indeks daya beda minimum 0,2 menurut Nunnaly (1970), Croker dan Algina (1986), Mehren dan Lehman (1991), dan Aiken (1994), (dalam Naga, 2008:89) dikatakan memadai dan indeks daya beda < 0,2 diangap tidak memadai. Indeks taraf sukar butir dan daya beda butir pada teori klasik sangat tergantung pada kelompok peserta tes. Artinya, harga kedua indeks akan berubah jika peserta tes juga berubah yang dikenal dengan istilah tidak invarian.

Pengukuran dalam pendidikan mengenal dua macam kekeliruan, yaitu kekeliruan acak atau kekeliruan sampel *(sampling error)* dan kekeliruan sistematik *(systematic error)* (Naga, 1992:116). Kekeliruan sampel adalah perbedaan antara keadaan sebenarnya yang ada pada populasi *(true score)*. Hal ini disebabkan oleh karena hasil ukuran pada sampel tersebut hanya merupakan salah satu dari sekian banyak kemungkinan hasil pengukuran yang dapat dicuplik secara berulang-ulang dari suatu populasi. Kekeliruan sampel tetap saja muncul meskipun alat ukur yang dipakai, situasi dan kondisi pengukuran, maupun jenis kemampuan yang diukur tetap sama. Dengan demikian ukuran sampel berpengaruh terhadap hasil pengukuran.

Sampel menurut Steel dan Torrie (1980:3) adalah bagian dari populasi, kadang-kadang mencakup seluruh populasi dan umumnya informasi dari sampel digunakan untuk penarikan kesimpulan tentang populasi itu. Pada sisi lain, dikenal juga pengertian sampel butir. Lord dan Novick (1968:235-236), secara eksplisit membedakan antara sampel butir *(sample of items)* yang ditarik secara acak dari suatu populasi butir dan sampel peserta tes *(sampel of examinees)* yang ditarik secara acak dari suatu populasi peserta tes. Sedangkan Thordike *et al.* (1991:197) menyebutkan pengertian butir sampel *(sampel item)* dan bukan sampel butir, mereka mendefinisikan butir sampel sebagai contoh-contoh yang serupa dengan apa yang akan muncul pada tes, terkait dengan (1) atribut stimulus untuk membatasi kompleksitas dari stimulus yang dihadapi peserta tes dan (2) atribut respons untuk menggambarkan dengan cepat bagaimana konstruk itu akan diukur.

Naiman, Resenfeld, dan Zirkel (1993:156) menyatakan bahwa untuk sampel besar dengan ukuran sampel *(sample size)* >30 secara statistik akan menghasilkan suatu distribusi rerata cukup dekat pada distribusi normal, sehingga kalkulasi berdasarkan kurva normal masuk akal. Eid meneliti tentang pengaruh ukuran sampel *(sample size)* pada penyetaraan butir tes mengusulkan untuk menggunakan ukuran sampel 200, 400, dan 800. Setiadi (1997:7) dalam penelitiannya terhadap estimasi parameter butir menyatakan bahwa sampel yang relatif kecil berukutan 100 atau 200, sedangkan Livingston dan Feryok (1987:9-10) melakukan penelitian pada penyetaraan ekipersentil estimasi frekuensi dengan penghalusan pada sampel berukuran 100 sampai dengan 3000 dan akurasi penyetaraan terjadi pada sampel berukuran 300. Dengan demikian penelitian ini menggunakan ukuran sampel 200, 400, dan 800.

Tujuan penyetaraan skor adalah untuk mambandingkan skor yang diperoleh dari perangkat tes yang satu (X) dan perangkat tes lainnya (Y) yang dilakukan melalui proses penyetaraan skor pada kedua perangkat tersebut (Hambleton dan Swaminathan, 1990:199). Selanjutnya Croker dan Algina (1986:4) mengemukakan bahwa dua skor, hasil pengukuran yang menggunakan instrumen X dan instrumen Y dapat disetarakan skornya jika keduanya mengukur kemampuan atau trait yang sama. Kemudian Aiken (1997:82) juga mengemukakan bahwa skor hasil pengukuran dua perangkat

tes yang paralel dapat disetarakan untuk mengkonversi skor dari perangkat tes satu terhadap skor dari perangkat tes lainnya.

Menurut Wiersma dan Jurs (1990:148) penyetaraan skor adalah suatu prosedur empiris yang diperlukan untuk mentransformasikan skor suatu tes ke skor tes yang lain. Karena merupakan prosedur empiris maka penyetaraan skor didasarkan pada data skor tes. Pendapat lain dikemukakan oleh Brown (1976:216) yang mengatakan bahwa skor dua tes dapat disetarakan jika kedua tes tersebut diperoleh dari populasi yang sama. Senada dengan Brown, Kolen menyatakan bahwa penyetaraan skor dapat dilakukan jika kelompok peserta setara, karena ketidaksetaraan yang ekstrim akan berpengaruh dalam perhitungan (Keeves, 1997:733).

Penyetaraan skor di antara ujian yang berbeda dilakukan melalui rancangan tertentu. Menurut Hambleton (1990:199) ada tiga rancangan penyetaraan skor yang umum digunakan, yaitu rancangan kelompok tunggal *(single group design)*, rancangan kelompok acak *(random group design),* dan rancangan butir bersama *(anchor items design)*. Pada rancangan kelompok tunggal, secara acak, peserta dibagi ke dalam dua kelompok, yaitu kelompok K1 dan kelompok K2. Dengan keacakan itu dimaksudkan agar K1 dan K2 adalah setara, yakni mereka memiliki kemampuan yang sama dengan variansi yang sama pula. Kemudian kelompok K1 mengerjakan tes X dan kelompok K2 mengerjakan tes Y. Dari kedua hasil tes tersebut kemudian dicari patokan konversinya.

Rancangan kelompok acak, secara acak peserta dibagi ke dalam dua kelompok, yaitu kelompok K1 dan kelompok K2. Cara acak ini dimaksudkan agar kedua kelompok peserta adalah setara. Selanjutnya kelompok K1 mengerjakan tes X dan kemudian tes Y, serta kelompok K2 mengerjakan tes Y dan kemudian tes X. Dengan demikian setiap kelompok mengerjakan dua tes dalam urutan yang terbalik dan selanjutnya dari hasil tes tersebut disetarakan.

Pada rancangan butir bersama tidak perlu secara acak membagi dua kelompok peserta seperti halnya pada rancangan kelompok tunggal dan rancangan kelompok acak. Sudah cukup terdapat kelompok peserta K1 dan K2. Namun demikian, rancangan ini tidak menolak penggunaan kelompok yang setara, yaitu kelompok peserta seperti pada rancangan kelompok

tunggal dan rancangan kelompk acak. Selanjutnya, tes X ditambah dengan sejumlah butir sebagai tes Z serta tes Z yang sama ditambahkan juga ke tes Y. Jadi bersama tambahan tersebut terdapat perangkat tes (X+Z) dan perangkat tes (Y+Z). Butir ujian Z dikenal sebagai tes jangkar *(anchor test)*.

Selanjutnya kelompok K1 mengerjakan tes (X+Z) dan kelompok K2 mengerjakan tes (Y+Z). Dari kedua tes tersebut, khususnya tes Z yang dikerjakan oleh kedua kelompok kemudian dicari patokan konversinya. Dilihat dari penyiapan peserta tes, rancangan butir bersama adalah yang paling praktis karena tidak mensyaratkan kesetaraan di antara kelompok peserta. Oleh karena itu tulisan ini menggunakan rancangan butir bersama dan rancangan ini pula yang digunakan dalam UASBN. Panjang tes pada UASBN 2009 mata pelajaran IPA Paket 01 dan Paket 02 sebanyak 40 butir, yang terdiri dari 30 butir (75%) merupakan butir daerah dan 10 butir (25%) merupakan butir pusat atau butir bersama *(anchor items)*. Hal ini sejalan dengan hasil penelitian Yahya Umar (1987:105) bahwa dengan panjang tes 40 butir dibutuhkan minimal 10 butir (25%) *anchor items* agar diperoleh hasil yang baik. Demikian pula hasil penelitian Yetti Supriyati (2003) yang merekomendasikan proporsi minimal butir *anchor items* sebesar 20% dari butir total.

Pada rancangan butir bersama kelompok peserta K1 mengerjakan tes X dan Z sedangkan kelompok peserta K2 mengerjakan tes Y dan Z. Dengan demikian, tes Z dikerjakan oleh kedua kelompok. Untuk membedakan digunakan notasi berbeda pada hasil tes tersebut. Tes Z yang dikerjakan oleh kelompok K1 menghasilkan skor Z_1 dengan rerata skor μ_{Z1} dan simpangan baku σ_{Z1}. Tes Z yang dikerjakan oleh kelompok peserta K2 menghasilkan skor Z_2 dengan rerata skor μ_{Z2} dan simpangan baku σ_{Z2}. Hasil gabungan ujian Z yang dikerjakan oleh kelompok peserta K1 dan K2 adalah skor Z dengan rerata μ_Z dan simpangan baku σ_Z.

Tes X hanya dikerjakan oleh kelompok peserta K1 dan menghasilkan skor X dengan rerata μ_X dan simpangan baku σ_X, sedangkan tes Y hanya dikerjakan oleh kelompok K2 dan menghasilkan skor Y dengan rerata μ_Y dan simpangan baku σ_Y. Dari sini yang diperhatikan adalah regresi skor X terhadap Z yang dikerjakan oleh kelompok peserta K1, serta regresi skor Y

terhadap skor Z yang dikerjakan oleh kelompok peserta K2. Kedua regresi tersebut masing-masing menghasilkan koefisien arah b_{XZ1} dan b_{YZ2}.

Selisih nilai X yang disebabkan oleh selisih $\mu_{\tilde{z}}$ dan $\mu_{\tilde{z}1}$ adalah sebesar $b_{XZ1}(\mu_{\tilde{z}} - \mu_{\tilde{z}1})$ sehingga rerata μ_x bergeser dengan nilai itu. Demikian pula, dengan jalan yang sama, selisih nilai Y yang ditimbulkan oleh selisih $\mu_{\tilde{z}}$ dan $\mu_{\tilde{z}2}$ adalah sebesar $b_{XZ2}(\mu_{\tilde{z}} - \mu_{\tilde{z}2})$ sehingga rerata μ_Y juga bergeser dengan nilai itu. Hal ini juga terjadi pada simpangan baku.

Penyetaraan ekipersentil menurut Braun & Holland (1982) dan Lord (1950) didasarkan pada definisi bahwa skala skor untuk kedua tes sebanding dan distribusi skor kedua tes identik (Linn, 1987:247). Sementara itu menurut Croker dan Algina (1986:346), secara umum diterima bahwa skor pada dua tes dianggap ekivalen jika skor-skor tersebut memiliki tara persentil *(percentile rank)* yang sepadan.

Menurut Braun dan Holland (1982) dan Lord (1950) penyetaraan ekipersentil didasarkan pada definisi bahwa skala skor untuk kedua tes sebanding dan distribusi skor ke dua tes identik (Linn, 1987:247). Cook dan Petersen menyatakan penyetaraan ekipersentil tidak seperti metode penyetaraan lainnya, tidak membutuhkan asumsi-asumsi terhadap tes-tes yang akan disetarakan (Naga, 1992:226). Dengan demikian metode penyetaraan ekipersentil mengasumsikan bahwa skor pada tes X dan Y adalah identik atau ekivalen, bertujuan agar distribusi dari skor tes X yang diubah sama dengan distribusi skor tes Y, dan tidak membutuhkan asumsi-asumsi seperti pada metode linear. Dengan demikian penelitian ini bertujuan untuk mengetahui (1) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 200, (2) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 400, dan (3) perbedaan variansi skor hasil penyetaraan antara metode menyetaraan linear dan metode penyetaraan ekipersentil untuk ukuran sampel 800.

Metode Penelitian

Penelitian ini menggunakan simulasi dengan dua variabel yang diteliti mengggunakan pola jawaban peserta tes diperoleh melalui pelaksanaan UASBN SD/MI negeri dan swasta tahun pelajaran 2008/2009 mata pelajaran IPA. Variabel bebas dalam penelitian ini adalah metode penyetaraan dan jumlah responden. Metode penyetaraan pada penelitian ini adalah metode penyetaraan linear dan metode penyetaraan ekipersentil, sedangkan variabel jumlah responden berturut-turut adalah 200, 400, dan 800. Variabel terikatnya adalah variansi skor hasil penyetaraan.

Populasi peserta tes adalah peserta tes UASBN tahun pelajaran 2008/2009 mata pelajaran IPA se-Jakarta Timur. Populasi tersebut sebanyak 44.401 siswa yang terbagi menjadi dua bagian yaitu sebanyak 22.201 siswa yang mengerjakan Paket 01 dan sebanyak 22.200 siswa yang mengerjakan Paket 02. Pengambilan sampel menggunakan metode acak berulang dengan pengembalian *(random sampling with replacement),* yaitu cara pengambilan sampel secara acak dengan jumlah tertentu yang dikembalikan lagi menjadi populasi untuk mendapat peluang menjadi sampel berikutnya, sehingga sampel tersebut mempunyai peluang untuk dipilih kembali pada saat pengambilan sampel berikutnya.

Data diperoleh melalui Pusat Penilaian Pendidikan (Puspendik) Badan Penelitian dan Pengembangan (Balitbang) Kementerian Pendidikan Nasional. Teknik analisis data menggunakan uji kesamaan variansi dengan analisis Uji F. Analisis data menggunakan bantuan perangkat lunak MINITAB, *ITEMAN for Windows Version 3.50d*, SPSS, dan Excel. Hipotesis statistik dalam penelitian ini terdiri dari tiga hipotesis nol dan tiga hipotesis alternatif, yaitu (1) H_o: $\frac{\sigma^2_{eks200}}{\sigma^2_{lin200}} = 1$ H_1: $\frac{\sigma^2_{ekp200}}{\sigma^2_{lin200}} > 1$; (2) H_o: $\frac{\sigma^2_{ekp400}}{\sigma^2_{lin400}} = 1$ H_1: $\frac{\sigma^2_{ekp400}}{\sigma^2_{lin400}} > 1$; dan (3) H_o: $\frac{\sigma^2_{ekp800}}{\sigma^2_{lin800}} = 1$ H_1: $\frac{\sigma^2_{ekp800}}{\sigma^2_{lin800}} > 1$.

Hasil Penelitian dan Pembahasan

Analisis variansi skor hasil penyetaraan dilakukan terhadap masing-masing cuplikan dengan ukuran sampel 200, 400, dan 800 baik untuk Paket 01 maupun Paket 02, masing-masing sebanyak 41 kali cuplikan. Jadi, seluruhnya terdapat 246 analisis dan menghasilkan variansi skor hasil penyetaraan dengan metode linear dan metode ekipersentil seperti disajikan pada Tabel 1 dan Tabel 2.

Tabel 1. Variansi Skor Hasil Penyetaraan Metode Linear

Cuplikan Ke	Variansi Skor			Cuplikan Ke	Variansi Skor		
	s^2_{lin200}	s^2_{lin400}	s^2_{lin800}		s^2_{lin200}	s^2_{lin400}	s^2_{lin800}
1	25,064	28,625	25,983	22	24,495	25,035	27,130
2	25,237	26,277	24,257	23	29,022	23,641	25,383
3	21,936	27,998	27,538	24	25,459	23,062	23,032
4	24,242	24,417	25,349	25	24,874	25,579	23,420
5	20,485	25,177	23,256	26	27,721	26,519	25,128
6	23,112	28,163	24,750	27	27,420	22,658	25,284
7	26,317	23,150	23,665	28	23,673	22,169	24,004
8	24,373	27,485	26,838	29	25,249	25,525	22,833
9	25,285	24,518	25,575	30	24,257	24,460	25,386
10	24,644	28,006	25,055	31	22,896	23,741	25,969
11	27,669	24,281	23,592	32	23,406	26,971	25,015
12	25,250	24,867	24,373	33	20,842	22,833	25,364
13	28,172	24,663	25,480	34	26,415	21,903	24,816
14	22,533	22,732	25,685	35	28,342	28,106	24,988
15	26,620	29,690	26,312	36	28,256	25,483	26,331
16	27,942	23,645	23,761	37	24,354	21,502	25,544
17	24,316	23,972	23,763	38	24,883	23,815	26,394
18	21,786	22,707	23,286	39	28,987	24,498	25,297
19	22,232	24,167	24,185	40	28,083	23,828	24,764
20	24,573	25,782	25,393	41	21,455	24,977	23,709
21	24,322	22,560	25,268	Rerata	25,119	24,855	24,986
				Simpangan Baku	2,240	2,042	1,139

Tabel 2. Variansi Skor Hasil Penyetaraan Metode Ekipersentil

Cuplikan Ke	Variansi Skor			Cuplikan Ke	Variansi Skor		
	s^2_{ekp200}	s^2_{ekp400}	s^2_{ekp800}		s^2_{ekp200}	s^2_{ekp400}	s^2_{ekp800}
1	28,072	28,847	30,449	22	27,466	29,408	30,367
2	31,286	29,049	30,633	23	32,481	32,132	27,923
3	29,072	29,050	27,034	24	32,107	26,371	26,819
4	31,865	30,475	27,476	25	29,591	29,336	29,125
5	30,125	33,465	27,098	26	29,955	28,635	28,883
6	28,981	27,222	28,470	27	29,788	29,758	29,276
7	28,961	28,658	28,449	28	25,413	28,185	29,439
8	30,842	29,248	29,516	29	29,373	29,632	31,484
9	27,022	27,818	28,715	30	27,373	27,639	30,363
10	25,874	30,538	26,344	31	28,647	27,684	31,088
11	30,535	30,652	29,363	32	28,353	31,567	29,976
12	28,472	27,727	29,382	33	28,397	30,558	26,645
13	33,138	29,510	30,830	34	29,180	28,994	26,152
14	34,047	32,860	31,069	35	32,580	28,738	29,924
15	29,739	29,425	30,797	36	27,163	29,724	30,557
16	30,956	27,411	27,635	37	28,351	28,845	31,057
17	28,215	33,959	31,499	38	31,454	31,665	31,261
18	24,765	28,215	28,161	39	31,618	31,967	27,015
19	30,952	29,194	29,155	40	31,408	33,618	29,165
20	30,391	29,815	27,205	41	31,348	30,207	27,336
21	31,790	28,087	30,927	Rerata	29,645	29,642	29,168
				Simpangan Baku	2,122	1,832	1,601

Sebelum teknik analisis statistik uji kesamaan variansi digunakan, terlebih dahulu perlu diperiksa apakah data penelitian telah memenuhi persyaratan. Uji persyaratan analisis yang digunakan adalah uji normalitas Kolmogorov-Smirnov, rangkuman hasil uji normalitas dapat dilihat pada Tabel 3. Berdasarkan Tabel 3 dapat dinyatakan bahwa keenam data variansi skor penyetaraan berasal dari populasi yang berdistribusi normal dengan taraf signifikasi 5%.

Tabel 3. Hasil Uji Normalitas Varians Skor Penyetaraan

No	Varians	Uji Normalitas Kolmogorov-Smirnov			
		Statistik	db	Sgn.	Sebutan
1	s^2_{lin200}	0,114	41	0,200	Normal
2	s^2_{lin400}	0,089	41	0,200	Normal
3	s^2_{lin800}	0,073	41	0,200	Normal
4	s^2_{ekp200}	0,071	41	0,200	Normal
5	s^2_{ekp400}	0,108	41	0,199	Normal
6	s^2_{ekp800}	0,121	41	0,136	Normal

Ada tiga hipotesis yang diuji dalam penelitian ini, yaitu hipotesis yang membandingkan variansi skor penyetaraan s^2_{ekp200} dengan $s^2_{lin200,}$ variansi skor penyetaraan s^2_{ekp400} dengan s^2_{lin400}, dan variansi skor penyetaraan s^2_{ekp800} dengan s^2_{lin800}. Setelah dilakukan analisis terhadap data variansi skor penyetaraan diperoleh hasil yang selengkapnya dapat dilihat pada Tabel 4.

Tabel 4. Tabel 4. Statistik Variansi Skor Penyetaraan

No	Variansi	Statistik		
		Rerata	SD	Variansi
1	s^2_{ekp200}	29,645	2,122	4,460
2	s^2_{lin200}	25,119	2,240	5,218
3	s^2_{ekp400}	29,642	1,832	3,281
4	s^2_{lin400}	24,855	2,042	4,065
5	s^2_{ekp800}	29,168	1,601	2,582
6	s^2_{lin800}	24,986	1,139	1,305

Hipotesis pertama dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) lebih besar daripada metode

menyetaraan linear (s^2_{lin200}) untuk ukuran sampel 200. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) adalah 4,460 dan metode linear (s^2_{lin200}) sebesar 5,218. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h1} = s^2_{ekp200} / s^2_{lin200}$ = 4,460/5,218= 0,855.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1 - 1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h1} yang diperoleh lebih kecil dibandingkan dengan F tabel. Dengan demikian hipotesis nol diterima. Hal ini berarti bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp200}) tidak berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin200}) untuk ukuran sampel 200.

Hipotesis kedua dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) lebih besar daripada metode menyetaraan linear (s^2_{lin400}) untuk ukuran sampel 400. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) adalah 3,281 dan metode linear (s^2_{lin400}) sebesar 4,065. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h2} = s^2_{ekp400} / s^2_{lin400}$ = 3,281/4,065 = 0,807.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1 - 1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h2} yang diperoleh lebih kecil dibandingkan dengan F tabel. Dengan demikian hipotesis nol diterima. Artinya, variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp400}) tidak berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin400}) untuk ukuran sampel 400.

Hipotesis ketiga dinyatakan bahwa variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) lebih besar daripada metode menyetaraan linear (s^2_{lin800}) untuk ukuran sampel 800. Hasil analisis yang terdapat pada Tabel 4 menunjukkan bahwa rerata variansi skor hasil

penyetaraan yang dihasilkan dari penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) adalah 2,582 dan metode linear (s^2_{lin800}) sebesar 1,305. Untuk memastikan apakah kedua variansi skor ini berbeda secara signifikan atau tidak, maka dilakukan analisis uji-F sebagai berikut: $F_{h3} = s^2_{ekp800} / s^2_{lin800} = 2{,}582/1{,}305 = 1{,}979$.

Harga F tabel untuk taraf signifikansi (α) = 0,05 dengan pembilang (n_1-1) = 40 dan penyebut (n_2 - 1) = 40 sebesar 1,684. Angka F_{h3} yang diperoleh lebih besar dibandingkan dengan F tabel. Jadi, hipotesis nol ditolak. Artinya, variansi skor hasil penyetaraan menggunakan metode ekipersentil (s^2_{ekp800}) berbeda secara signifikan dengan variansi skor hasil penyetaraan menggunakan metode linear (s^2_{lin800}) untuk ukuran sampel 800.

Berdasarkan hasil pengujian hipotesis pertama dan kedua di atas dapat dinyatakan bahwa hipotesis nol diterima untuk ukuran sampel 200 dan 400. Artinya tidak ada perbedaan variansi antara metode linear dan ekipersentil dengan ukuran sampel 200 dan 400. Kenyataan data dalam menerima hipotesis nol dapat disebabkan oleh berbagai kemungkinan, antara lain adalah asumsi penggunaan penyetaraan metode linear tidak terpenuhi. Hal ini sesuai dengan pendapat Hambleton dan Rogers bahwa asumsi penyetaraan dengan cara linear adalah kedua skor tes distribusinya berbeda, distribusi tersebut terkait dengan rerata dan simpangan bakunya. Jika terdapat kasus yang demikian skor kedua kelompok akan sama. Sebaliknya bila asumsi tersebut benar, maka penyetaraan linear menjadi penyetaraan ekipersentil (Hambleton, Swaminathan, dan Rogers, 1991:124-125). Dengan demikian distribusi skor kedua kelompok pada penelitian mempunyai distribusi yang tidak berbeda, sehingga penyetaraan linear sama dengan penyetaraan ekipersentil.

Pengujian hipotesis ke tiga dengan ukuran sampel 800 menunjukkan hasil yang berbeda, yaitu menolak hipotesis nol. Artinya ada perbedaan variansi antara metode linear dan ekipersentil dengan ukuran sampel 800. Ini dipengaruhi oleh variansi metode linear yang makin kecil untuk ukuran sampel 800. Makin kecilnya variansi metode linear dengan ukuran sampel yang makin besar sesuai dengan penelitian yang dilakukan Hanson *et. al.* (1994) bahwa rerata kuadrat kesalahan penyetaraan *(mean-squared equating*

error) makin kecil untuk ukuran sampel yang makin besar (Kolen dan Brennan, 1995:102).

Simpulan

Sesuai dengan hasil penelitian dapat dikemukakan beberapa simpulan sebagai berikut: (1) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin200}) terbukti memiliki variansi yang sama dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp200}) dengan ukuran sampel 200. Artinya, tidak ada perbedaan variansi skor penyetaraan menggunakan metode linear maupun metode ekipersentil untuk ukuran sampel 200, (2) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin400}) terbukti memiliki variansi yang sama dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp400}) dengan ukuran sampel 400. Artinya, tidak ada perbedaan variansi skor penyetaraan menggunakan metode linear maupun metode ekipersentil untuk ukuran sampel 400, dan (3) variansi skor hasil penyetaraan menggunakan metode linear (σ^2_{lin800}) terbukti memiliki variansi yang berbeda dengan variansi skor hasil penyetaraan menggunakan metode ekipersentil (σ^2_{ekp800}) dengan ukuran sampel 800. Dengan demikian terdapat perbedaan variansi skor penyetaraan menggunakan metode linear dan metode ekipersentil untuk ukuran sampel 800.

Daftar Pustaka

Aiken, Lewis R. 1997. *Psychological Testing and Assemesment.* Boston: Allyn & Bacon.

Brown, Frederick G. 1976. *Principles of Educational and Psychologycal Testing.* New York: Holt, Rinehart and Wiston.

Croker, Linda dan James Algina. 1986. *Introduction to Classical & Modern Test Theory.* New York: Rinehart and Winston Inc.

Departemen Pendidikan Nasional. Peraturan Pemerintah Republik Indonesia Nomor 19 tahun 2005 tentang Standar Nasional Pendidikan. Jakarta: Badan Standar Nasional Pendidikan.

_____. Peraturan Pemerintah Republik Indonesia Nomor 20 tahun 2007 tentang Standar Penilaian Pendidikan. Jakarta: Badan Standar Nasional Pendidikan.

_____. Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 39 tahun 2007 tentang Ujian Akhir sekolah Berstandar Nasional (UASBN) untuk sekolah Dasar/Madrasah Ibtidaiyah/Sekolah Luar Biasa (SD/MI/SDLB) tahun Pelajaran 2007/2008. Jakarta: Badan Standar Nasional Pendidikan.

Eid, Ghada K. "*Effects of Sample Size in the Equating of Test Items*" http://findarticles.com/p/articles/mi_qa3673/is_200510/ai_n15641924 (diakses tanggal 12 Oktober 2008).

Hambleton, Ronald K. dan Hariharan Swaminathan. 1985. *Item Response Theory: Principles and Application.* Boston: Kluwer.

_____. 1990. *Item Respon Theory Principles and Applications.* Boston: Kluwer Nijhoff Publishing.

Hambleton, Ronald K., Hariharan Swaminathan, dan H. Jane Rogers. 1991. *Fundamentals of Item Respons Theory*. California: SAGE Publications, Inc.

Kaplan, Robert M. dan Dennis P. Saccuzzu. 1982. *Psychological Testing, Principle, Application and Issues.* California: Wadsworth, Inc.

Keeves, John (ed). 1997. *An International Hand Book of Measurement*. Oxford: Pergamon.

Kolen, Michael J. dan Robert L. Brennan. 1995. *Test Equating Methods and Practices.* New York: Springer.

Linn, Robert L. 1987. *Educational Measurement*. New York: Macmillan Publishing Company.

Livingston, S. A. dan N. J. Feryok. 1987. "Univariate Versus Bivariate Smooting in Frewuency Evuating" *Laporan Riset* No. 87-36. New Jersey: Educational Testing Service.

Lord, Frederick M. dan Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Massachusetts: Addison-Wesley Publishing Company, Inc.

Marco, Gory L. *et al.* 1983. *A Test of The Adequacy of Curvilinear Score Equating Models* (New Horizon in Testing). Weiss David J. New York: Academic Press.

Naga, Dali S. 1992. *Pengantar Teori Sekor.* Jakarta: Besbats.

_____. 2008. *Probabilitas dan Sekor pada Hipotesis Statistika.* Jakarta: Universitas Tarumanegara.

Naiman, Arnold, Robert Rosenfeld, dan Gene Zirkel. 1993. *Understanding Statistic.* Singapore: McGraw-Hill.

Setiadi, Hari. 1997. Small Sample IRT Item Parameter Estimates. *Disertasi* tidak dipublikasikan.

Steel, Robert G. D. dan James H. Torrie. 1980. *Principles and Procedur of Statistic.* Singapore: Mc.Graw-Hill.

Thordike, Robert M. *et al.* 1991. *Measurement and Evaluation in Psychology and Education.* New York: Macmillan Publishing Company.

Wiersma, William dan Stephen G. Jurs. 1990. *Educational Measurement and Testing.* Boston: Allyn and Bacon.

_____. 2005. *Research Methods in Education.* New York: Pearson Education Inc.

Yahya Umar. 1987. Robusness of the Simple Linking Procedure i Item Banking Using the Rasch Model, *Dissertation*, University of California, p. 105.

Yetti Supriati. 2003. Pengaruh Proporsi *Anchor Items* terhadap Kestabilan Koefisien Penyetaraan Sekor pada Evaluasi Belajar. *Disertasi* tidak dipublikasikan. Jakarta: Pascasarjana Universitas Negeri Jakarta.

PENGARUH PEMBOBOTAN DAN JENIS PENILAI TERHADAP FUNGSI INFORMASI TES PERFORMANSI

Wakhinuddin S
Fakultas Teknik Universitas Padang
wakhid_nuddin@yahoo.com

Abstrak

Tujuan penelitian adalah untuk mengetahui pengaruh skala penilaian dengan pembobotan (SPPP), tanpa pembobotan (SPTP), penilai internal dan eksternal terhadap fungsi informasi tes performansi. Penelitian dilakukan dengan pendekatan eksperimen pada tujuh SMK di Sumatera Barat; melibatkan sampel 864 subjek dan 46 performansi. Pengambilan sampel dengan metode random sampel sederhana, penempatan subjek dengan metode random matriks sampel. Data dikumpulkan dengan tes performansi dan dianalisis dengan metode Rasch. Analisis data penelitian menggunakan Anava (desain faktorial 2x2). Hasil penelitian menunjukkan bahwa: 1) Fungsi informasi tes SPPP lebih tinggi daripada fungsi informasi tes SPTP; 2) Fungsi informasi tes oleh penilai eksternal lebih tinggi daripada fungsi informasi tes oleh penilai internal; 3) Faktor interaksi antara ragam tes performansi dan kelompok penilai mempengaruhi fungsi informasi tes.

Kata kunci: *tes performansi, fungsi informasi tes, skala penilaian dengan pembobotan, skala penilaian tanpa pembobotan, penilai internal dan penilai eksternal*

THE INFLUENCE OF PERFORMANCE TESTS WITH WEIGHTING AND ASSESSOR TYPE ON INFORMATION FUNCTION OF PERFORMANCE TEST

Wakhinuddin S
Fakultas Teknik Universitas Padang
wakhid_nuddin@yahoo.com

Abstract

The objective of this research was to investigate the influence of weighted and unweighted rating scales, and internal and external assessors on information function of performance test. The study was conducted on the even semester of 2004 at seven vocational schools in West Sumatera, involving 864 students and 46 performances. The sample was drawn by using simple random sampling technique, and placement was done by means of random matrix method. Factorial 2x2 design was adopted in this study. Data were collected through a performance test, and analyzed by using Rasch and two-way analysis of variance. The results of the research are: (1) the information function of the weighted rating scale test was more effective than the unweighted rating scale test, (2) the information function of the student's test assessed by external assessors was higher than that assessed by internal assessors, (3) The interaction factor between types of performance test (weighted and unweighted) and assessor groups (internal and external) affected the information function of test, (4) the information function of unweighted rating scale used by internal assessors was higher than that used by external assessors, (5) the information function of unweighted rating scale used by internal assessors was higher than the information function of weighted rating scale used by internal assessors, (6) the information function of weighted rating scale used by external assessors was higher than that of weighted rating scale used by internal assessors, (7) the information function of weighted rating scale used by external assessors was higher than that of unweighted rating scale used by internal assessors.

Keywords: *Performance test, informations function, weighted rating scale, unweighted rating scale, internal assessor and eksternal assessor*

Pendahuluan

Undang-Undang Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 57, ayat (1) menyatakan evaluasi dilakukan dalam rangka pengendalian mutu pendidikan secara nasional sebagai bentuk akuntabilitas penyelenggaraan pendidikan kepada pihak-pihak yang berkepentingan. Sekolah menengah kejuruan (SMK) dengan Kurikulum edisi 1999 mencerminkan kurikulum berbasis kompetensi, di antaranya pada akhir program ada uji kompetensi dan uji produktif. Evaluasi kompetensi siswa tidak hanya dinilai pihak sekolah (internal), melibatkan pihak luar (eksternal), seperti pada ujian promosi kompetensi dan ujian produktif. Perpaduan kedua kelompok penilai diharapkan menghasilkan suatu hasil evaluasi yang bermutu. Banyak persoalan yang timbul dalam mengkombinasikan hasil penilai internal dan penilai eksternal, di antaranya persoalan perbedaan pandangan penilaian. Penilai eksternal cenderung fokus pada hasil pembelajaran (*outcomes*), analisis tugas, dan pembelajaran tuntas. Berbeda dari pandangan penilai eksternal, penilai internal fokus pada aktivitas kelas, holistik, dan perkembangan siswa. Penilai eksternal lebih pasti memberi nilai sebab dia memakai pendekatan tuntas, sedangkan penilai internal mengambang karena dia memakai pendekatan holistik, ketidakpastian nilai siswa akan lebih banyak pada penilai internal. Dikmenjur memadu penilai internal dan penilai eksternal dalam menilai keberhasilan siswa di akhir suatu program.

Pengukuran kompetensi siswa pada ujian kompetensi program keahlian otomotif SMK menggunakan tes performansi dengan skala penlaian (*rating scale*) sebagai alat ukur. Penyusunan skala penilaian berkaitan dengan penyekoran yang dirancang terlebih dulu. Penyekoran tidak lepas dari pembobotan, dan umumnya besar skor suatu butir (aktivitas) ditetapkan berdasarkan tingkat kesulitan, tingkat kepentingan dan lama kerja. Materi tes dalam hal ini adalah tugas otentik (*authentic task*) siswa, yaitu tugas yang didemonstrasikan siswa, setara dengan tugas pada bengkel mobil.

Penelitian eksperimen tes ini dilakukan dengan cara memanipulasi bobot, yaitu membelah butir (*split item*), mengempiskan kategori butir

(*collapse item categories*), mengerumunkan/mengelompokkan butir (*cluster item*), dan membuang butir (*drop item*). Dalam penelitian ini, perlakuan pembobotan dilakukan dengan mengempiskan atau menggelembungkan bobot butir. SPTP berarti bobot dikempiskan, sedangkan SPPP berarti bobot digelembungkan. Perubahan pembobotan pada tes jenis skala penilaian diduga dapat mengubah fungsi informasi.

Fungsi informasi kebalikan dari ketidakpastian, dapat diketahui dari besarnya variansi pengukuran walau dilakukan berulang-ulang. Terkait dengan penilai dan skala penilaian, ada kemungkinan penilai tertentu menilai bahwa seorang siswa memperoleh nilai optimal dengan menggunakan SPPP, tetapi tidak memperoleh nilai optimal bila memakai SPTP. Mungkin terjadi penilai tidak memberikan nilai optimal memakai SPPP, tetapi memberikan nilai optimal dengan SPTP. Kecocokan penilai dengan skala penilaian diduga melahirkan fungsi informasi maksimum.

Fungsi informasi dalam teori respons butir merupakan gambaran ketelitian, konsistensi atau stabilitas suatu alat ukur. Yang menjadi persoalan adalah skala penilaian mana dari kedua jenis penskalaan (SPTP dan SPPP) yang memiliki ketelitian yang tinggi? Penilai internal dan penilai eksternal mempunyai karakter tersendiri, bila mereka diberikan alat ukur dengan SPTP dan SPPP, dan dapat menghasilkan skor fungsi informasi tertentu. Berhubungan dengan itu, muncul pertanyaan: pasangan jenis penilai dan skala penilaian manakah yang memiliki kesesuaian sehingga menghasilkan fungsi informasi tes performansi yang maksimum?

Bila penilai (asesor) bekerja secara ideal dan profesional seharusnya menghasilkan penilaian yang tidak jauh berbeda, baik penilai internal maupun eksternal. Kenyataan pada saat uji kompetensi siswa SMK, pada umumnya ada perbedaan hasil dari penilai berbeda. Perbedaan ini menjadi penting diungkapkan agar dapat menjawab, bagaimana penilaian berbeda, bagaimana perbedaan ini dapat dihitung, dan bagaimana mengendalikannya dalam suatu model pengukuran.

Dari uraian di atas dirumuskan beberapa masalah, yaitu a) apakah ada perbedaan fungsi informasi tugas skala penilaian pakai pembobotan (SPPP) dengan fungsi informasi tugas skala penilaian tanpa pembobotan (SPTP)?; b) apakah ada perbedaan fungsi informasi tugas yang dinilai oleh penilai

eksternal dengan penilai internal? c) Apakah ada interaksi antara ragam tes performansi dan kelompok penilai dalam pengaruhnya terhadap fungsi informasi ujian? Secara praktis, penelitian ini dimaksudkan untuk: membuat tes performansi skala penilaian tanpa pembobotan (SPTP) dan skala penilaian pakai pembobotan (SPPP), memanfaatkan fungsi informasi tugas dari kedua jenis tes, mengetahui besar pengaruh tes performansi cara penilaian SPTP dan SPPP pada fungsi informasi ujian, mengetahui besar pengaruh kelompok penilai internal dan kelompok penilai eksternal pada fungsi informasi ujian. Manfaat praktis hasil penelitian adalah dapat digunakan sebagai dasar mengambil keputusan tentang kualitas keterandalan tes performansi. Secara teoretis memberikan gambaran pengembangan tes performansi, dan peranan penilai internal dan eksternal dalam menguji kompetensi siswa program mekanik otomotif SMK.

Teori Respons Butir (TRB) banyak dipakai dalam evaluasi pembelajaran terutama membahas kaitan butir dengan respons siswa, dan TRB dapat mengukur kombinasi kemampuan siswa dengan kesukaran aktivitas (butir) melalui parameter siswa dan parameter aktivitas. Parameter siswa adalah skor nilai siswa, cerminan kemampuan siswa, makin tinggi kemampuan siswa, maka semakin pula tinggi skor nilai siswa. Parameter kemampuan siswa (dituliskan dengan Θ). Kaitan kemampuan siswa dengan taraf sukar butir dapat dirumuskan sebagai berikut, $P_i(\Theta) = f(\Theta - b_i)$. Persamaan disebut model 1 (satu) Parameter. Dalam perkembangannya ternyata telah digunakan pula pada penskoran polikotomi (politomi). TRB dapat dipakai pada skala penilaian tes performansi. Pada penelitian dipakai TRB dengan pendekatan model perluasan Rasch, model ini interpretasinya sederhana karena melibatkan satu parameter, karakteristik butir dan responsden stabil dengan sampel kecil. Model Rasch hanya memakai taraf kesukaran butir (b_i) yang dimasukan ke dalam model. Model Rasch berkembang pada jenis tes bersifat multiaspek (*multifacets*), seperti interaksi responsden (siswa) dengan butir (tugas), misalnya pada proses penjurian, aspek berinteraksi adalah aspek penilai (juri). Untuk suatu situasi penjurian, dipakai alat ukur politomi, seperti skala penilaian (*rating scale*).

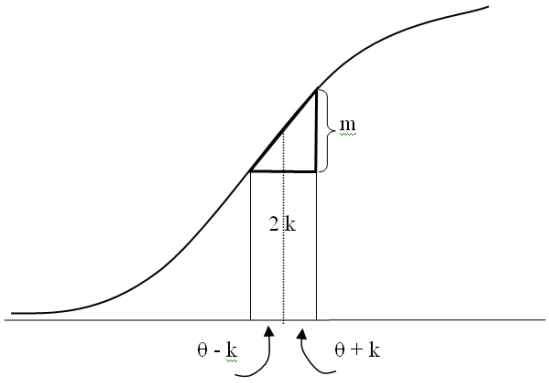
Rumus model perluasan Rasch untuk multi-aspek skala penilaian adalah: $\text{Log}(P_{nijk}/P_{nij(k-1)}) = B_n - D_i - C_j - F_k$ (Linacre., Wright., dan Lunz,

1990:2). Rumus ini digunakan untuk menilai tampilan kompetensi seseorang yang dinilai Penilai pada butir (aktivitas). Formula penilaian dengan butir sama namun penilai berbeda, menjadi : Log ($P_{nijk}/P_{nij(k-1)}$) = B_n - D_i - C_j - $F_{jk.}$ Model perluasan Rasch (*rasch extended*) dapat mengolah data politomi, seperti data tes kompetensi mekanik otomotif. Penggunaan model Rasch dapat meningkatkan kualitas pengukuran.

TRB dalam membahas ketepatan suatu alat ukur menggunakan konsep fungsi informasi butir (ujian). Definisi fungsi informasi butir adalah rasio kemiringan (*slope*) kuadrat dengan varian (Thorndike Robert L,1982: 81).

$$\text{Funsi informasi} = \frac{(slope\ (\theta))^2}{p_{(\theta)}\ \ q_{(\theta)}} = \frac{m}{2k}$$

Suatu fungsi informasi dapat dijelaskan sebagai berikut, misalkan beberapa siswa yang mempunyai tingkat kemampuan (θ) berbeda mengerjakan satu butir akan menghasilkan kurva karakteristik butir seperti pada Gambar 1.



Gambar 1. Slope kurva karakteristik butir

Pengukuran berulang-ulang tidak selalu sama hasilnya, ada varian. Ketidakpastian merupakan kejadian kombinasi kemampuan responsden dengan taraf kesukaran butir berkaitan dengan jawaban berpeluang betul

atau salah. Dalam ketidakpastian kebalikkan fungsi informasi berlaku semakin besar ketidakpastian, fungsi informasi semakin kecil, sebaliknya semakin kecil ketidakpastian semakin besar fungsi informasi. Fungsi informasi merupakan hubungan terbalik variansi, sedangkan variansi merupakan wujud lain daripada ketidakpastian (Dali S Naga,1992:309). Rumus fungsi informasi butir untuk butir ke-i ditulis di bawah ini.

$$I_j(\theta) = \frac{1}{KB_j(\theta)^2}$$

Dengan demikian, jika faktor kekeliruan baku diketahui, maka fungsi informasi dapat diketahui skornya. Fungsi informasi biasanya digambarkan sebagai sebuah grafik fungsi (I_o) terhadap kesukaran butir. Puncak grafik merupakan fungsi informasi maksimum, dalam penelitian ini yang dijadikan sebagai referensi nilai fungsi informasi adalah skor fungsi informasi maksimum. Fungsi informasi dipakai untuk: a) konstruksi perangkat ujian (tes); b) pemilihan butir; c) penilaian presisi pengukuran; d) komparasi sejumlah tes; e) penentuan bobot dalam penskoran; dan f) komparasi berbagai metode penskoran (Hambleton, Swaminathan, dan Rogers: 1991: 91-96). Penggunaan fungsi informasi pada TRB mempunyai tujuan yang sama dari analisis butir teori skor klasik. Pembahasan fungsi informasi ujian harus diawali dari pembahasan informasi butir. Rumus fungsi informasi butir satu parameter dituliskan di bawah ini.

$$I_j(\theta) = \frac{1}{P_j(\theta)Q_j(\theta)}\left[\frac{\partial P_j(\theta)}{\partial \theta}\right]^2 \qquad (1)$$

Fungsi informasi tes merupakan penjumlahan fungsi informasi butir aktivitas. Pada SPTP rumus penskoran ujian sebagai berikut.

$$X_{it} = \sum_{j=1}^{N} X_{ij} \qquad (2)$$

Rumus fungsi informsi ujian skala penilaian tanpa pembobotan (SPTP) seperti pada persamaan (1), yaitu sama dengan fungsi informasi

butir. Sifat fungsi informasi ujian diantaranya, keterikatan satu butir dengan butir lainnya tidak ada, efek butir hanya terjadi pada total nilai fungsi informasi ujian. Sifat lain pada fungsi informasi ujian berkaitan dengan penskoran, dan penskoran mempengaruhi fungsi informasi ujian (Dali S. Naga,1992:310). Bobot relatif butir ditentukan berdasarkan tingkat kepentingan dan kerumitan butir, secara teoritis dapat diketahui melalui analisis tugas dan tanggapan panelis instrumen. Instrumen pakai bobot disebut skala penilaian pakai pembobotan (SPPP), bobot diberikan ada tiga jenis, yaitu: 5, 10, dan 20. Skor akhir tugas yang memakai SPPP merupakan skor komposit (gabungan). Rumus penskoran pada SPPP memakai ragam kedua, yaitu penggabungan skor butir dengan pembobotan atau butir-butir dengan bobot yang berbeda-beda. Rumus penskoran ujian dengan pembobotan (Lord dan Norvick, 1968:96) adalah sebagai berikut.

$$X_{it} = \sum_{j=1}^{N} W_j X_{ij} \qquad (3)$$

Sejalan dengan rumus penskoran, maka rumus untuk fungsi informasi ujian dengan pembobotan (Dali S. Naga, 1992:332) adalah sebagai berikut.

$$I(\theta, X) = \frac{\sum_{j=1}^{N} w_j \left[\frac{\partial P_j(\theta)}{\partial \theta}\right]^2}{\sum_{j=1}^{N} w_j^2 P_j(\theta) Q_j(\theta)} \qquad (4)$$

Dari rumus terlihat besarnya peranan pembobotan (*w*) terhadap fungsi informasi tugas. Dengan demikian, dapat diduga fungsi informasi tugas yang memakai SPPP akan lebih tinggi daripada SPTP. Penskalaan (*scaling*) adalah suatu pengukuran kontinum pada suatu objek, person, atau peristiwa. Skala merupakan fasilitas yang sengaja dibuat untuk menghasilkan angka pada garis kontinum yang dijadikan skor yang mempunyai informasi.

Ragam pertama tes performansi adalah SPTP, keistimewaannya antara lain: mudah memakainya, mudah menghitungnya, mengurangi kelelahan, kebosanan penilai, dan mengurangi salah pengukuran. Ragam

kedua tes performansi adalah SPPP. Pembobotan pada butir tes bukan sesuatu yang baru pada program kejuruan. Keistimewaannya, setiap satu aspek penilaian memiliki beberapa kriteria (tidak selalu sama jumlah kriterianya), kriteria terbaik nilainya maksimum (ada skor 10 atau skor 20), sedangkan kriteria terjelek mendapat nilai kosong (skor 0).

Penilai dapat dibagi atas penilai internal dan penilai eksternal. Seorang penilai internal (*insider*) adalah individu atau kelompok yang merupakan kesatuan dari pelaksanaan program, sedangkan penilai eksternal individu atau kelompok di luar program yang dengan suatu alasan tertentu ikut memberi perhatian pada pelaksanaan program (Owen, 1993:33-41). Pada ujian praktek kompetensi di SMK penilai eksternal, adalah orang yang memiliki kepakaran di bidangnya, dapat berasal dari asosiasi profesi, majelis sekolah, DU/DI, BLK, PPPG kejuruan, perguruan tinggi, dan pihak lain yang relevan, sedangkan unsur internal adalah guru kejuruan yang memiliki bidang keahlian yang sesuai (Owen,1993:33-41). Untuk ujian praktek, konfigurasi adalah penilai eksternal untuk internal, yaitu penilai eksternal diminta menilai suatu program dengan siswa sekolah setempat. Keberadaan penilai eksternal pada sistem pendidikan tidak lepas dari peran dan fungsinya, yaitu untuk menjamin objekvitas materi ajar dan penilaian (Piper. 1994:28).

Posisi penilai eksternal dapat disamakan sebagai pengabsahan proses penilaian, penilai eksternal merupakan jembatan antara sekolah dengan DU/DI, secara umum berfungsi pengendali kualitas di SMK. Penilai eksternal hendaklah: a) memahami tujuan pembelajaran/kriteria unjuk kerja yang harus dikuasai siswa; b) memverifikasi topik-topik tugas yang akan diuji kepada siswa; c) memverifikasi kesediaan mesin dan peralatan; d) menguji kompetensi siswa; e) mengkonfirmasikan hasil penilaian kepada penilai internal yang satu tim; f) mengkomunikasikan hasil pengujian kepada tim penilai dan pihak sekolah. Tim penilai untuk satu paket kompetensi diuji oleh minimal tiga orang, satu tim penilai merupakan 2:1 atau 1:1 untuk tiap kelompok ujian dalam satu program keahlian. Artinya, 2 orang dari eksternal dan 1 orang dari internal. Aspek yang dinilai meliputi proses kerja, sikap dan waktu yang digunakan untuk menyelesaikan pekerjaan (tugas).

Ragam tes performansi adalah cara penilaian yang menuntut siswa melakukan tugas dalam bentuk perbuatan yang dapat diamati guru atau *penilai* lainnya dengan menggunakan skala penilaian. Kemampuan performansi dalam membedakan keadaan siswa menunjukkan bahwa performansi memiliki sifat mengukur dan menilai. Sifat tersebut membuatnya menjadi suatu instrumen penilaian yang disebut tes performansi, dan membuat tes performansi dikategori sebagai Tes acuan-kriteria.

Tes performansi mempunyai beberapa keistimewaan, antara lain: a) mengatasi beberapa hal (terutama keterampilan) yang tidak dapat dinilai dengan 'tes kertas-pensil'; b) lebih alami, langsung, lebih tuntas menilai keterampilan (*skill*); c) bermanfaat bagi siswa yang kurang ingin membaca; d) mendorong aplikasi pembelajaran kepada situasi kehidupan nyata. Namun, ada juga beberapa kelemahan penilaian performansi, yaitu: a) pemakaian waktu dan usaha yang banyak; b) penskoran dan penilaian cenderung subjektif, tidak adil, rendah reliabilitas; c) penilaian individual daripada kelompok (Gronlund, 1993: 115). Ada tiga elemen utama tujuan tes performansi, yaitu: 1) kondisi tugas yang akan ditampilkan; 2) deskripsi tugas; dan 3) kriteria kerja. Elemen ini menunjukkan, tes performansi tidak dapat dilepaskan dari suatu perbuatan (kerja). Beberapa penulis menggandeng kata performansi dengan kata tugas, dan dibaca tugas performansi (*performance task*).

Penelitian ini bertujuan untuk mengetahui pengaruh skala penilaian dengan pembobotan (SPPP), tanpa pembobotan (SPTP), penilai internal dan eksternal terhadap fungsi informasi tes performansi. Pada penelitian ini, ada dua ragam tes performansi yang berbeda konsep dan teknis pelaksanaan pengukuran kompetensi siswa, yaitu: 1) SPTP; skala penilaian mempunyai skor tetap 0 sampai 5, karena setiap jarak aktivitas mempunyai skor sama dan, 2) SPPP; pada SPPP skala penilai pada aktivitas mempunyai bobot berbeda, mulai dari skor 0-5, skor 0-10, dan skor 0-20.

Dalam banyak hal, prosedur dan produk adalah aspek penting performansi, umpamanya keterampilan memeriksa kerusakan dan memperbaiki mobil. Prosedur ditekankan di langkah awal dan produk belakangan, sesudah prosedur dikuasai. Dalam keterampilan khusus, seperti

kecekatan dievaluasi awal kegiatan, belakangan difokuskan pada kebersihan dan akurasi jenis bahan dan kecepatan prosedur.

Fungsi tes performansi adalah untuk mengetahui sejauh mana suatu program pendidikan berhasil diterapkan. Dalam pendidikan kejuruan performansi dikaitkan dengan tujuan pembelajaran (program), sehingga tujuan performansi diartikan sebagai tujuan belajar (Larson, 1972:126).

Pembuatan cara menilai performansi dengan: 1) memperhatikan tujuan pembelajaran yang adanya tampilan siswa; 2) menyeleksi topik yang akan dinilai, dan yang dinilai adalah proses kerja; 3) menyeleksi tingkat otentik tugas, pada langkah ini perhatikan relevansi antara topik pelajaran di kelas dengan tugas di lapangan kerja, 4) menyeleksi kepantasan melaksanakan tugas; 5) menetapkan kriteria tingkat keberhasilan tampilan siswa, seperti kualitas pekerjaan dan waktu kerja; 6) menetapkan metode observasi, di sini dipakai cara penilaian skala penilaian (*rating scale*) pakai bobot dan tanpa pembobotan. Pada SMK batas kelulusan ditetapkan berdasarkan normatif, Depdiknas menetapkan batas lulus untuk matapelajaran produktif/keahlian adalah tujuh (7).

Metodologi Penelitian

Penelitian ini merupakan penelitian eksperimen, pada butir tes, yaitu dengan perlakuan mengempiskan atau menggelembungkan bobot butir. Penelitian eksperimen pada butir telah pernah dilakukan seperti de Gruijter dan Van der Kamp (1991), penggunaan pada teori geneneralibitas telah dilakukan Cronbach (1971) (de Gruitjter, 2002: 50-55). Berdasarkan teori generalisabilitas dan teori pensekoran, maka desain penelitian yang cocok adalah Anava faktorial 2 x 2. Desain tersebut banyak dipakai pada psikometri, karena mampu membedakan perbedaan personal dan perbedaan butir. Perbedaan personal membicarakan kemampuan (*ability*) siswa dan perbedaan butir membahas kesukaran butir (aktivitas).

Pemilihan dan penetapan penelitian ini sebagai penelitian eksperimen didasari pandangan bahwa butir mempunyai karakter tersendiri, dengan demikian butir dapat dimanipulasi. Perlakuan dengan memanipulasi bobot butir, yaitu membelah butir (*split item*), mengempiskan

kategori butir (*collapse item categories*), mengelompokkan butir (*cluster item*), dan membuang butir (*drop item*).

Dalam melakukan eksperimen, ada beberapa tahapan yang akan dilalui: Proses pra-eksperimen, (1) melakukan *assesment* terhadap pelaksanaan program sesuai dengan kurikulum SMK edisi 1999; (2) memeriksa ketercapaian materi ajar di kelas responsden; (3) mencari beberapa tes format skala penilaian yang pernah dipakai dalam dua tahun belakangan sesuai materi kurikulum; (4) menetapkan topik tugas yang akan diuji ke siswa; (5) mengembangkan tes cara penilaian SPPP dan sekor SPTP sesuai topik tugas yang telah ditetapkan; (6) melakukan studi pendahuluan (*pilot studies*); (7) untuk verifikasi dicari koefisien reliabilitas dan validitas; (8) sesuai dengan studi pendahuluan, selanjutnya melakukan *upgrade* proposal penelitian, seandainya ditemukan kelemahan di lapangan; (9) membuat laporan.

Kegiatan pendahuluan (1) menentukan kriteria penilai internal dan eksternal; seperti: pengalaman 5 tahun; (2) *choacing,* seperti: cara menggunakan kedua jenis tes dan lainnya; (3) menentukan jadwal eskperimen; (4) penggandaan bahan tes; (5) mempersiapkan fasilitas bengkel; (6) menentukan dan mengundang responsden; (6) melakukan experiment sebanyak topik pengujian praktik selama ini ; (7) menjaga validitas internal dan eksternal penelitian. Pelaksanaan eskperimen mencakup: (1) menyediakan meja kerja penilai internal dan eksternal secara terpisah; (2) menyediakan ruang tunggu bagi siswa responsden; (3) menata letak mesin sehingga dapat terpantau kedua penilai; (4) set *stopwatch* sesuai lama pengujian; (5) penilai diperbolehkan menanya suatu hal kepada responsden berkaitan dengan perbuatan responsden; (6) sesama penilai tidak diperbolehkan berinteraksi selama ujian; (7) pengujian selesai tepat waktu; (8) Ujian bentuk sumatif, yaitu diakhir kelas II. Pengukuran dilakukan berulang-ulang, yaitu dalam bentuk beberapa tugas yang dipraktekkan siswa. Proses pasca-eksperimen: (1) lembaran tes diparaf penilai; (2) lembaran tes dikumpulkan segera; (3) lembaran tes yang cacat disortir; (4) data pada lembaran tes dipindahkan ke kerangka data mentah.

Ada beberapa metode yang dipakai memecahkan permasalahan penelitian adalah: 1) penyekoran : yaitu penyekoran tanpa pembobotan

(SPTP) dan penyekoran pakai pembobotan (SPPP); 2) acuan kriteria dijadikan dasar penilaian, yaitu skor nilai ≥ 7,0; dengan penetapan ini sebagai batas lulus maka ada pengkategorian siswa, yaitu: 1) tanpa kompetensi, 2) kompetensi kurang, 3) kompetensi minimal, 4) kompetensi bagus. Pengkategorian ini dijadikan dasar kode pengolahan data untuk dapat dipakai pada program *Facet* dari Linacre. Selanjutnya metode ke 3) yaitu penentuan unidimesi (ketunggalanmatra) digunakan metode analisis faktor; 4) penentuan independensi lokal digunakan statistik kai-kuadrat; 5) skor baku (z) digunakan untuk mentransformasi skala penilaian yang berbeda ke bentuk skala logit; dan 6) desain generalizabiliti, untuk menganalisis banyak faktor (*facet*), seperti: penilai, tugas, dan kemampuan siswa, dikenal dengan sebutan *multifacet*.

Sesuai dengan desain penelitian, ada dua sumber varian, yaitu tugas dan siswa. Populasi siswa adalah siswa kelas II SMK di Provinsi Sumatera Barat. Adapun jumlah keseluruhan SMK di Sumatera Barat, yang negeri sebanyak 14 dan Swasta sebanyak 39. Jumlah populasi siswa kelas II sekitar 16.552 orang, diambil dari sekolah SMK N I Pariaman, BLPT Padang (SMKN I dan V Padang), SMK N II Payakumbuh, SMK N II Solok, SMK Muhammadiyah I Padang, SMK Mitra Payakumbuh, SMK Muhammadiyah Solok.

Tabel 1. Desain penelitian anava faktorial dua jalur

Kelompok Penilai (B)	Ragam tes Performansi (A)	
	SPTP (A_1)	SPP (A_2)
Internal (B_1)	A_1B_1	A_2B_1
Eksternal (B_2)	A_1B_2	A_2B_2

Pengambilan sampel pada penelitian ini dilakukan secara bertahap: Tahap pertama, penentuan sekolah dilakukan teknik random sampling sederhana, yaitu dengan mengundi 7 dari 17 sekolah. Ada dua

pertimbangan utama dalam menentuan sekolah populasi, (1) legimitas pelaksanaan ujian, ke-17 sekolah telah mendapai izin dari Dinas Diknas Sumbar melaksanakan Ujian kompetensi secara mandiri; (2) kelengkapan fasilitas praktek memenuhi persyaratan pelaksanaan ujian kompetensi.

Tabel 2. Sebaran sampel penelitian dan asal sekolah

No.	Nama lembaga pendidikan	Status	Jumlah siswa
1	BLPT	Negeri	157
2	SMK N 2 Payakumbuh	Negeri	78
3	SMK N 2 Solok	Negeri	116
4	SMK N 1 Pariaman	Negeri	112
5	SMK Muhammadiyah 1 Padang	Swasta	147
6	SMK Muhammadiyah Solok	Swasta	112
7	SMK Mitra Payakumbuh	Swasta	142
Jumlah			864

Tahap ke dua pengambilan sampel, diawali pembuatan penomoran siswa, kemudian mengambil subyek dengan teknik sampel random sederhana. Dari proses ini didapat dua kerangka sampel sementara, yaitu sub kelompok dinilai dengan SPTP (A_1) sebanyak 432 siswa dan sub kelompok dinilai SPPP (A_2) sebanyak 432 siswa. Tahap ketiga penempatan siswa pada sub-sub matriks sampel.

Desain matriks sampel dibuat berdasarkan prinsip distribusi seimbang jumlah tugas (aproksimasi). Alasan pembuatan desain matriks sampel disebabkan banyak tugas yang akan diujikan kepada siswa, namun keterbatasan fisik siswa dan waktu pelaksanaan tidak memungkinkan ke-46 tugas dapat dilakukan satu siswa, maka setiap kompetensi dibagi tiga kelompok. Dengan demikian satu siswa mendapatkan tujuh tugas dan paling sedikit empat tugas.

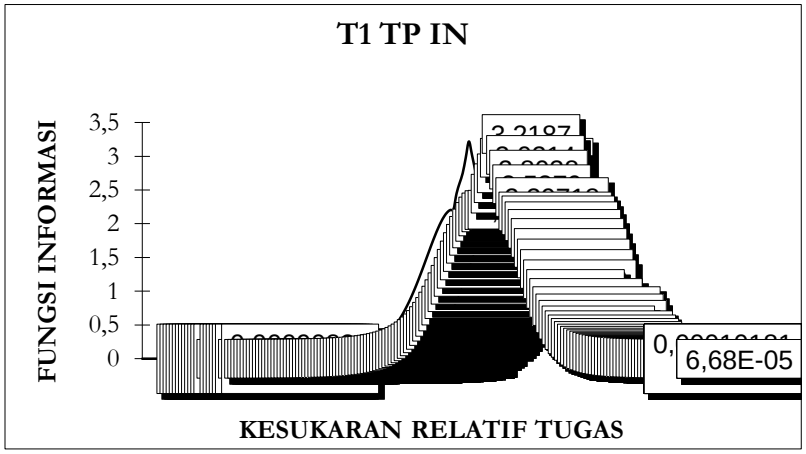
Teknik penempatan subyek pada sub-matrik juga menggunakan random acak sederhana. Tiga digit (angka) terakhir dijadikan referensi nomor defenitif subyek. Dari pemerosesan ini, matriks sampel terisi nama-nama subyek ke-1 sampai subyek ke-432. Setiap siswa diuji penilai internal dan penilai eksternal. Pengujian hipotesis penelitian digunakan teknik analisis varian (Anava) faktorial 2 x 2. Sebelum menggunakan Anava terlebih dahulu dilakukan uji persyaratan analisis, yaitu uji normalitas dan homogenitas data.

Hasil Penelitian dan Pembahasan

Suatu data yang akan dimasukkan ke dalam model Rasch harus memenuhi persyaratan analisis. Namun, ada dua hal pokok persyaratan suatu tugas (tes), yaitu independensi lokal dan unidimensi. Uji independensi menggunakan kai-kuadrat dan untuk menguji unidimensi tugas menggunakan analisis faktor. Ditemukan semua tugas memiliki indepedensi lokal, karena t hitung kurang dari t tabel ($\chi^2_{hit.tgs} < \chi^2_{(0,05)(5)}$). Dengan demikian dapat dikatakan, antara tugas satu dengan tugas lain tidak tergantung satu sama lainnya. Dari hasil pengolahan data juga ditemukan bahwa semua tugas adalah mengukur satu dimensi, ini teruji pada KMO = 0,859.

Hasil uji normalitas terhadap data fungsi informasi dari setiap perlakuan diketahui seluruh kelompok data normal. Pengujian homogenitas data sampel dilakukan uji Levene's, didapat nilai probabilitas uji Levene (0,792), di atas 0,05, sehingga H_o dapat diterima; varians sampel populasi adalah homogen.

Fungsi Informasi ujian Tugas 1 (T1) dari skala penilaian tanpa pembobotan (SPTP) oleh penilai internal.



Gambar 2. Fungsi informasi tugas 1 dinilai oleh penilai internal dengan skala penilaian tanpa pembobotan (SPTP).

Dari gambar di atas ditemukan fungsi informasi tugas sebesar 3,21 memakai skala penilaian tanpa pembobotan (SPTP) oleh penilai internal. Grafik tanpa meruncing ke atas, dan grafik melandai seimbang antara kiri dan kanan. Ini menunjukan bahwa tingkat kesukaran tugas satu cenderung berdistribusi seimbang, artinya semua siswa yang melaksanakan tugas satu relatif normal, dan keterandalan tugas satu baik karena mencapai 3,21 untuk skala lima.

Ada empat kategori dari setiap tugas dapat terjadi saat seorang penilai memberi nilai pada siswa. Pada gambar 2, dimulai dari hadap kiri, grafik paling kiri (1) adalah kategori kompetensi tak ada, grafik (2) adalah kategori kompetensi kurang, grafik (3) adalah kategori kompetensi minimal, dan grafik (4) adalah kategori kompetensi bagus. Dari hasil penilaian penilai, probabilitas siswa berada pada salah satu kategori dapat saja terjadi. Gambar 2 adalah tugas 38 dengan SPTP dengan penilai eksternal ditemukan probabilitas terbesar terjadi pada kategori (1) yaitu kompetensi tak ada, dan probabilitas terkecil terjadi pada kategori (3), yaitu kompetensi minimal. Ada empat kategori pada semua tugas (46) yang diujikan pada siswa. Statistik deskriptif fungsi informasi ujian disampaikan pada Tabel 3.

Tabel 3. Deskripsi Data Fungsi Informasi

Sumber Statistik		A_1	A_2	Σ
B_1	n	46	46	92
	$\bar{x}$	3,44	3,20	3,32
	s	0,74	0,57	0,62
B_2	n	46	46	92
	$\bar{x}$	2,68	4,53	3,60
	s	0,60	0,67	0,63
Σ	n	92	92	184
	$\bar{x}$	3,06	3,86	3,46
	s	0,67	0,62	1,25

Rata-rata terendah ada pada sel $A_1B_{1,}$ sebesar 2,68, dan rata-rata tertinggi ada pada sel $A_2B_{2,}$ sebesar 4,53. Ini menunjukan bahwa peningkatan fungsi informasi sebagai akibat perlakuan meningkat jauh pada kelompok penilai eksternal. Berbeda pada kelompok penilai internal terdapat penurunan rata-rata pada sel A_1B_1, yang semula sebesar 3,44 menjadi sebesar 3,20. Kesemua perubahan rata-rata fungsi informasi pada sel-sel disebabkan oleh perlakuan dari ragam tes SPTP ataupun SPPP.

Pengujian hipotesis dilakukan dengan menggunakan analisis varian (Anava) dua jalan yang dilanjutkan dengan uji Tukey. Anava dua jalan digunakan untuk menguji pengaruh utama (*main effect*) dan interaksi (*interaction effect*) variabel bebas ragam tes dan kelompok penilai terhadap variabel terikat fungsi informasi. Selanjutnya hasil analisis data dengan anava disajikan dalam bentuk Tabel 4.

Tabel 4. Tabel 4. Rangkuman Hasil Anava Data Fungsi Informasi

Sumber Varian	JK	dk	RJK	F_0	F_t	
					0,05	0,10
Antar Kolom	3,775	1	3,775	8,892**	3,84	6,63
Antar Baris	29,764	1	29,764	70,118**	3,84	6,63
Interaksi	50,085	1	50,085	117,989**	3,84	6,63
Dalam kelompok	76,408	180	0,424			
Jumlah	2366,712	183				

Pada perbedaan antarkolom F_0 =8,892 > F_t= 3,84 , berarti hipotesis nol (H_0) yang menyatakan bahwa tidak ada perbedaan fungsi informasi tugas yang dinilai dengan SPTP dengan fungsi informasi tugas SPPP ditolak, atau hipotesis alternatif (H_a) yang menyatakan fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP diterima. Dengan kata lain, bahwa fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP. Siswa yang dinilai SPTP awalnya lebih tinggi dari fungsi informasi yang dinilai dengan SPPP atau x_{A1} = 3,44 > x_{A2} = 2,68, namun karena adanya peran karakteristik skala penilaian dan karakteristik penilai maka pada saat tertentu fungsi informasi SPPP lebih tinggi daripada SPTP.

Skala dapat diklasifikasikan berdasarkan orientasi pengukuran yaitu: pendekatan sitimulus, pendekatan responss, dan pendekatan subjek. Penskalaan pada penelitian ini berorientasi pendekatan responss. Karakteristiknya, jawaban yang disusun bertingkat, dimulai dari butir jawaban rendah hingga ke jawaban tinggi. Prosedur yang demikian dikenal dengan penskalaan Guttman, bilamana siswa dapat mengerjakan aktivitas urutan kerja tertinggi, tentunya aktivitas yang lebih rendah dapat juga dikerjakannya. Skala penilaian tanpa pembobotan (SPTP) adalah satuan yang menilai kualitas tampilan siswa dengan jumlah skala konsisten,

kategori 0 sampai kategori 5. Pensekoran ini mencerminkan tingkat kualitas performansi. Kategori 5 berarti tingkat tertinggi, dinyatakan sebagai nilai kompetensi maksimum; dan angka kategori 0 berarti terendah, dinyatakan sebagai nilai tanpa kompetensi. Keistimewaan alat ukur SPTP: mudah memakainya, mudah menghitungnya, mengurangi kelelahan, kebosanan penilai, dan mengurangi salah pengukuran.

Pada skala penilaian pakai pembobotan (SPPP) mempunyai keistimewaan, setiap satu aspek penilaian memiliki beberapa kriteria (tidak selalu sama jumlah kriterianya), kriteria terbaik nilainya maksimum (ada sekor 10 dan 20), kriteria terjelek mendapat nilai kosong (sekor 0). Semakin besar pembobotan (w) semakin besar fungsi informasi, ini sesuai dengan rumusdi bawah ini.

$$I(\theta, X) = \frac{\sum_{j=1}^{N} w_j \left[\frac{\partial Pj(\theta)}{\partial \theta} \right]^2}{\sum_{j=1}^{N} w_j^2 P_j(\theta) Q_j(\theta)}$$

Dengan demikian temuan yang menyatakan fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP dapat diterima. Ini sesuai dengan temuan Lily Rochaya menemukan bahwa fungsi informasi butir soal bentuk pilhan ganda dengan lima pilihan jawaban lebih tinggi daripada nilai fungsi informasi butir soal bentuk pilihan ganda dengan tiga pilihan jawaban. Semakin besar bentangan penyebaran nilai, peluang kecocokan siswa dengan alat ukur akan lebih tinggi.

Pada perbedaan antarbaris harga F0 = 70,118 > Ft = 3,84, berarti hipotesis nol (H_0) yang menyatakan 'fungsi infromasi tugas yang dinilai penilai eksternal tidak berbeda daripada fungsi informasi tugas yang dinilai penilai internal'; ditolak, atau hipotesis alternatif (Ha) 'fungsi infromasi tugas yang dinilai penilai eksternal lebih tinggi daripada fungsi informasi tugas yang dinilai penilai internal'; dapat diterima. Dengan kata lain, dapat dinyatakan bahwa terdapat perbedaan fungsi informasi antara siswa yang dinilai oleh penilai internal dan siswa yang dinilai oleh penilai eksternal.

Fungsi informasi yang dinilai oleh penilai internal lebih rendah dari siswa yang dinilai oleh penilai eksternal (xB1=3,319 < xB2 =3,606).

Ini disebabkan, kebiasaan bekerja di industri (bengkel) yang mengacu pada disiplin, produktivitas tinggi, variatif dan teliti membuat seorang penilai eksternal cenderung banyak gagasan dan independen. Berbeda, dengan penilai internal lebih mementingkan proses pembelajaran sesuai keinginan kurikulum. Penilai internal menekankan adanya usaha siswa, dan secara rinci penilai internal cenderung lebih suka sasaran hasil segera, tidak variatif, terfokus, linear, sistematis, dan dependen.

Hipotesis nol (H_0) yang menyatakan bahwa tidak ada interaksi antara fungsi infromasi dan siswa yang dinilai oleh penilai internal tidak berbeda dengan siswa yang dinilai oleh penilai eksternal ditolak, atau hipotesis alternatif (H_a) yang menyatakan faktor interaksi antara ragam tes performansi dan kelompok penilai mempengaruhi fungsi informasi tugas diterima, dimana harga F_0= 117,989 > Ft= 3,84. Dengan demikian, pencapaian fungsi informasi dipengaruhi oleh interaksi antara ragam skala penilaian yang dipakai dalam ujian dengan kelompok penilai.

Ini dapat dijelaskan dari karakter pekerjaan yang setiap hari dilaksanakan penilai, cenderung membuat karakrakter berbeda dalam menilai siswa. Pada skala yang sama, penilai eksternal cenderung memberi nilai sangat variatif mulai dari nilai rendah sampai nilai tertinggi, berbeda daripada penilai internal lebih banyak penilai terfokus pada sekor nilai tertentu. Pada penilai eksternal keragaman hasil penilaiannya membuat kecocokan kemampuan siswa dengan tugas tinggi, berbeda dari pada hasil penilaian penilai internal cenderung terfokus membuat fungsi informasi tugas rendah. Dengan demikian dapat dikatakan temuan 'fungsi informasi tugas dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas dengan penilai internal' dapat diterima.

Dari uji hipotesis ketiga terbukti ada interaksi antara ragam skala penilaian dan kelompok penilai, ini menunjukkan bahwa efek utama ada perbedaan, sehingga perlu dilacak sel-sel mana saja yang berbeda. Teknik analisis yang dapat menjelaskan pengujian adalah uji Tukey. Analisis ini digunakan untuk menguji perbedaan nilai rerata absolut dari dua kelompok yang dipasang dengan cara membandingkan nilai tersebut dengan nilai

kritis HSD. Analisis ini merupakan analisis lanjut, yang akan mengungkapkan apakah ada perbebedaan antarsel. Hasil uji Tukey pada taraf signifikansi (α)=0,05. Pada kelompok A1B1 dan A2B1 harga q_0 = 11,53 > q_t 0,18, berarti hipotesis nol (H0) ditolak atau hipotesis alternatif keempat diterima, berarti fungsi informasi tugas SPTP dengan penilai internal lebih tinggi daripada fungsi informasi tugas (SPPP) dengan penilai internal.

Skala penilaian SPTP terbentuk dari pensekoran tetap, penggunaan pensekoran demikian melahirkan format lembaran observasi lebih mudah digunakan dalam menilai tampilan kerja siswa, format peniIaian yang tetap (monoton) dan tidak membutuhkan kriteria macam-macam sesuai dengan karakter penilai internal. Sedangkan SPPP memiliki pensekoran menggelembung dan banyak kriteria, dan skala penilaian lebar merupakan fasilitas bagi penilai eksternal memberi tanda nilai. Penilai sebagai subyek evaluasi mempunyai karakteristik masing-masing, karakter penilai internal cenderung dependen, sistematis dan tidak variatif. Sesuai dengan karakter ini, penilai internal lebih cocok dengan SPTP dibandingkan pada SPPP.

Pada kelompok A1B2 dan A2B2 harga q0 = 28,067 > q_t = 0,18, berarti hipotesis nol (H0) ditolak atau hipotesis alternatif ke lima diterima, hal ini berarti pada fungsi informasi tugas SPPP dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas SPTP dengan penilai eksternal. Bagi penilai eksternal yang memiliki kecenderungan jelimet dan rinci kriteria kerja, saat melaksanakan pengujian akan memberi responss berbagai pertimbangan untuk menentukan atau kemungkinan lainnya dalam menilai kompetensi siswa yang tampil. Karakter penilai eksternal demikian tersalurkan melalui SPPP, sehingga SPPP dengan kecenderungan karakter penilai eksternal mempunyai kecocokan tinggi.

Pada kelompok A1B1 dan A1B2 harga q0 = 0,282 > qt = 0,18, berarti hipotesis nol (H0) ditolak atau hipotesis alternatif keenam diterima, hal ini berarti pada fungsi informasi tugas SPTP dengan penilai internal lebih tinggi daripada fungsi informasi tugas SPTP dengan penilai eksternal. Berdasarkan karakter guru yang memiliki kecenderungan bertindak tidak variatif (monoton) namun sistematis, dapat dikatakan bahwa skala penilaian tanpa pembobotan (SPTP) lebih cocok untuk Penilai internal ketimbang

Penilai eksternal. Dengan demikian dapat dikatakan, ada kecenderungan sekor fungsi informasi yang dinilai oleh penilai internal menggunakan SPTP lebih tinggi daripada Penilai eksternal yang juga menggunakan SPTP.

Pada kelompok A2B1 dan A2B2 harga q0 = 27,91 > qt = 0,18, berarti hipotesis nol (H0) ditolak atau hipotesis alternatif ketujuh diterima, hal ini berarti pada fungsi informasi tugas SPPP dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas SPPP dengan penilai internal. Temuan ini dapat dijelaskan, bahwa penilai eksternal dalam ujian praktek berasal dari kalangan profesional, KADIN, dan perguruan tinggi. Pada penelitian ini, penilai eksternal yang dipakai pihak SMK adalah dari kalangan profesional, dari mekanik mobil. Mekanik dalam bekerja umumnya teliti, banyak alternatif metode kerja, dan kuantitatif (produktivitas) dijadikan ukuran dalam bekerja. Kebiasaan-kebiasaan ini terbawa dalam menilai kompetensi siswa, sehingga mekanik dalam menilai siswa mempunyai banyak kriteria dan cenderung presisi. Berbeda dari penilai internal, kebiasaan kerja logis, sistematis dan monoton membuat penilai internal bekerja kurang variasi. Karakter kerja mekanik (penilai eksternal) bekerja demikian tersalurkan melalui alat ukur SPPP.

Simpulan

Berdasarkan data dapat disimpulkan bahwa: a) Secara keseluruhan fungsi informasi ujian yang diukur dengan SPPP lebih tinggi daripada SPTP; b) Secara keseluruhan fungsi informasi tugas yang dinilai penilai eksternal lebih tinggi daripada fungsi informasi tugas yang dinilai penilai internal; c) Ada interaksi antara ragam tes performansi dan kelompok penilai dalam pengaruhnya terhadap pencapaian fungsi informasi tugas; bagi penilai internal, fungsi informasi tugas dinilai dengan SPTP lebih tinggi daripada menggunakan SPPP; bagi penilai eksternal, fungsi informasi tugas dengan SPPP lebih tinggi dari menggunakan SPTP.

Daftar Pustaka

Dali S Naga. 2002. Diktat perkuliahan '*Psikometri*'. Jakarta: Pascasarjana UNJ.

de Gruijter, Dato N.M. dan Van der Kamp, Leo J. Th. *Statistical Test Theory for Education and Psychology*.. May 2002. http://www.unt.edu/rss/class/rich/5840/test_theory_text.pdf.

Gronlund, N. E., 1993. *How to Make achievement Test and Assessment*. Boston: Ally and Bacon.

Hambleton, R.K., Swaminathan, H., Rogers, H.J. 1991.*Fundamentals of Item Responsse Theory*. Newbury Park: SAGE Publications.

Landy, F. J., Trumbo, D. A. 1980. *Psychology of Work Behavior*. Illionis: The Dorsey Press.

Larson, M. E. 1972. *Teaching Related Subjects in Trade and Industrial and Technical Education*. Ohio: Charles E, Merrill Publishing Co;

Lunz, M.E., Wright, B.D., Linacre, J.M. 1990. *Measuring The Impact of Judge Severity on Examination Scores*. Pp. 1-2. www.rasch.org/memo47.htm.

Rochaya. 2001. Fungsi Informasi Test Pilihan Ganda. Thesis. Jakarta: PPs UNJ.

Tay-Lim, B. S-H. Davis dan Tang. *The Impact Treatments on NAEP Reporting Scale Secore*, makalah pada National Council on Measurement in Education, 22-24 April 2003. Chicago:Educational Testing Service. www.ets.org/legal/copyright.

Undang-Undang Nomor 20 tahun 2003. Sistem Pendidikan Nasional.