

TESTLET SEBAGAI SALAH SATU MODEL INSTRUMEN PENGUKURAN KOMPETENSI MATEMATIKA DI SMA

I Wayan Widana

Dosen Jurusan Pendidikan Matematika FPMIPA IKIP PGRI Bali

iwyn_widana@yahoo.co.id

ABSTRACT

Testlet as An Instrument Model Measurement of Competence In Math High School

This study aimed to describe one of the alternative forms of measurement instruments in high school math competency. One of the assumptions in item response theory (IRT) is the responses given by the examinees on test items are free of locally independence. However, the previous studies have shown that many test current local grains that contain dependence. Although local dependence is not expected in the IRT, but there is an important reason for entering the items each dependence on an assessment. Many real problems require settlement related to the issue or problem solving in a single step by step manner. Usage-based scoring testlet can minimize the negative consequences of dependence this item.

Key word: Testlet, Instrument Model, Measurement.

PENDAHULUAN

Kompetensi matematika pada jenjang pendidikan dasar dan menengah dijabarkan dalam bentuk Kompetensi Dasar (KD), yang tersebar pada masing-masing kelas pada setiap jenjang pendidikan. Pada ujian nasional (UN), KD yang sedemikian banyak akhirnya diringkas menjadi beberapa kompetensi yang diukur menggunakan beberapa indikator sebagaimana dinyatakan dalam kisi-kisi UN. Indikator-indikator tersebut selanjutnya dijabarkan menjadi beberapa butir soal pilihan ganda dengan 5 pilihan jawaban. Model penskoran yang digunakan adalah model dikotomis, apabila peserta didik menjawab benar maka diberi skor 1, dan bila menjawab salah atau tidak menjawab diberikan skor 0. Model penskoran tersebut diduga mempengaruhi nilai fungsi informasi tes. Hal itu dapat terjadi dalam penskoran *graded response model (GRM)*, karena sebuah kompetensi yang diukur menggunakan banyak butir soal, sehingga apabila pada butir soal tertentu peserta didik tidak bisa menjawab soal dengan benar seolah-olah peserta didik belum menguasai kompetensi tersebut, walaupun butir soal lain yang mengukur kompetensi yang sama bisa dijawab dengan benar.

Sebelum model *testlet* dipopulerkan, model *Item Response Theory/IRT* politomus adalah metode utama untuk mengestimasi

parameter butir soal. Saat ini, model *testlet* tersebut secara luas telah digunakan untuk mengembangkan dan mengkonstruksi tes. Dengan demikian, alasan teoretis untuk memilih model *testlet* (Wainer & Wang, 2000) atas model *IRT* politomus dalam analisis *testlet* diarahkan pada peningkatan mutu instrumen pengujian. Namun, beberapa kelemahan terhadap model *testlet* juga harus dipertimbangkan. Pertama, model *testlet* lebih kompleks dari model *IRT* standar dan model *IRT* politomus karena menambah parameter *testlet*. Kedua, ketika salah satu parameter *testlet* ditambahkan dalam model, suatu sifat laten tambahan juga ditambahkan dalam model sehingga multidimensionalitas terjadi dan hasil peningkatan analisis kompleksitas. Oleh karena itu, kebermanfaatan dalam menggunakan model *testlet* (Wainer & Kiely, 1987) harus dipertimbangkan terhadap kompleksitas ditambahkan dalam analisis data.

Meskipun Wainer dan Wang (2000) menyatakan bahwa keuntungan dari model *testlet* pada model politomus, penting untuk dibandingkan dalam berbagai kondisi, Pitt, Kim, dan Myung (2003) menunjukkan bahwa tujuan dari pemilihan model tidak hanya untuk menemukan model yang menyediakan *fit* maksimal untuk satu set data yang diberikan, tapi untuk mengidentifikasi model dari satu set model yang paling sesuai dengan karakteristik atau

yang mendasari proses kognitif. Secara singkat, model terbaik adalah model yang sesuai dengan tujuan penelitian dan bisa menjelaskan semua *fitur* penting dari data yang sebenarnya tanpa menambah kompleksitas yang tidak perlu.

PEMBAHASAN

Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) No. 59 Tahun 2014 tentang Kurikulum 2013 Sekolah Menengah Atas/Madrasah Aliyah, pada lampiran III dinyatakan bahwa kompetensi merupakan seperangkat sikap, pengetahuan, dan keterampilan yang harus dimiliki, dihayati, dan dikuasai setelah mempelajari suatu muatan pembelajaran, menamatkan suatu program, atau menyelesaikan satuan pendidikan tertentu. Berdasarkan analisis kebutuhan, potensi, dan karakteristik sosial, ekonomi, dan budaya daerah, maka pemerintah perlu merumuskan dan menetapkan standar kompetensi lulusan (SKL) sebagai kriteria mengenai kualifikasi kemampuan lulusan yang mencakup sikap, pengetahuan, dan keterampilan.

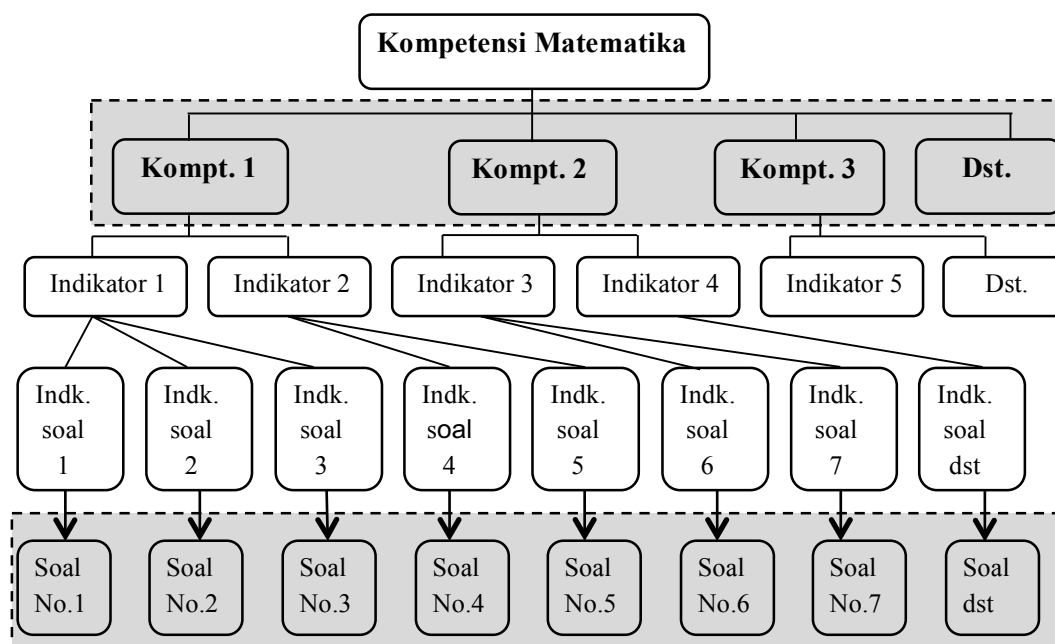
Lebih lanjut Permendikbud No. 59 Tahun 2014 lampiran III, mengamanatkan bahwa mata pelajaran matematika diberikan bertujuan agar peserta didik memiliki kompetensi matematika antara lain sebagai berikut: 1) memahami konsep matematika, merupakan kompetensi dalam menjelaskan keterkaitan antarkonsep dan menggunakan konsep maupun algoritma, secara luwes, akurat, efisien, dan tepat, dalam pemecahan masalah; 2) menggunakan pola sebagai dugaan dalam penyelesaian masalah, dan mampu membuat generalisasi berdasarkan fenomena atau data yang ada; 3) menggunakan penalaran pada sifat, melakukan manipulasi matematika baik dalam penyederhanaan, maupun menganalisa komponen yang ada dalam pemecahan masalah dalam konteks matematika maupun di luar matematika (kehidupan nyata, ilmu, dan teknologi) yang meliputi kemampuan memahami masalah, membangun model matematika, menyelesaikan model dan menafsirkan solusi yang diperoleh termasuk dalam rangka memecahkan masalah dalam kehidupan sehari-hari (dunia nyata); 4) mengomunikasikan gagasan, penalaran serta mampu menyusun bukti matematika dengan menggunakan kalimat lengkap, simbol, tabel, diagram, atau media lain untuk memperjelas

keadaan atau masalah; 5) memiliki sikap dan perilaku yang sesuai dengan nilai-nilai dalam matematika dan pembelajarannya, seperti taat azas, konsisten, menjunjung tinggi kesepakatan, toleran, menghargai pendapat orang lain, santun, demokrasi, ulet, tangguh, kreatif, menghargai kesemestaan (konteks, lingkungan), kerjasama, adil, jujur, teliti, dan cermat; dan 6) melakukan kegiatan-kegiatan motorik yang menggunakan pengetahuan matematika.

Anas Sudijono (1996) mengemukakan bahwa salah satu alat ukur (instrumen) yang sering digunakan untuk mengukur hasil belajar peserta didik di sekolah adalah tes. Tes adalah himpunan pertanyaan yang harus dijawab, pernyataan-pernyataan yang harus dipilih/ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang dites dengan tujuan untuk mengukur suatu aspek (perilaku) tertentu dari orang yang dites. Secara umum tes memiliki 2 (dua) fungsi utama, yaitu: a) sebagai alat ukur keberhasilan peserta didik, dalam hal ini tes berfungsi mengukur tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah menempuh pembelajaran dalam kurun waktu tertentu; dan b) sebagai alat ukur keberhasilan program pembelajaran, artinya melalui seperangkat tes tertentu dapat diketahui keberhasilan program pembelajaran yang telah disusun sebelumnya.

Untuk mengukur ketercapaian kompetensi matematika dalam ranah pengetahuan tersebut di atas, maka dalam penilaian kompetensi matematika dikelompokkan menjadi sejumlah sub kompetensi, sebagaimana dituangkan dalam kisi-kisi UN. Tercapainya sub kompetensi matematika dalam kisi-kisi UN ditandai oleh tercapainya sejumlah indikator pencapaian kompetensi dalam sub kompetensi tertentu. Indikator pencapaian kompetensi dalam kisi-kisi UN tersebut dapat dijabarkan lagi menjadi beberapa indikator soal yang lebih spesifik, sebagai acuan untuk mengembangkan alat ukur (butir soal). Untuk menjelaskan pengelompokan kompetensi matematika ranah kognitif dan pengukuran ketercapaian kompetensi tersebut, dapat dilihat dalam bagan 1 di bawah ini.

Bagan 1. Kompetensi Matematika dan Pengukurannya



Pada bagan 1. di atas, kompetensi-1, kompetensi-2, dan seterusnya merupakan sub kompetensi-sub kompetensi matematika yang wajib dicapai oleh peserta didik. Sebagai contoh, tercapainya kompetensi-1 ditandai oleh tercapainya indikator-1 dan indikator-2, dimana indikator-1 dapat dijabarkan menjadi beberapa indikator soal yaitu indikator soal-1, indikator soal-2, dan indikator soal-3. Demikian pula indikator-2 dapat dijabarkan menjadi indikator soal-4, dan indikator soal-5. Selanjutnya masing-masing indikator soal dibuatkan butir soalnya. Dalam pelaksanaan UN, untuk mencapai kompetensi-1 diukur menggunakan 5 butir soal pilihan ganda. Dengan demikian, untuk mengukur kompetensi-1 yang selama ini dilakukan dalam UN menggunakan 5 butir soal pilihan ganda, dapat dilakukan dengan mengukur kompetensi-1 menggunakan sebuah butir tes bentuk *testlet* yang terdiri atas 5 langkah penyelesaian. Dimana kelima langkah penyelesaian tersebut terdiri atas 5 butir soal pilihan ganda yang diskor dengan model

dikotomus. Sedangkan bila kompetensi-1 diukur menggunakan sebuah butir soal bentuk *testlet* dengan 5 langkah penyelesaian, maka model penskoran yang digunakan pada butir soal bentuk *testlet* adalah model politomus. Kumpulan 5 butir soal pilihan ganda yang digunakan untuk mengukur kompetensi-1 yang selanjutnya diskor menggunakan model politomus disebut *testlet-1*. Demikian seterusnya, kumpulan beberapa butir soal pilihan ganda yang digunakan untuk mengukur kompetensi-2, disebut *testlet-2*, dan seterusnya. Sehingga dari 40 butir soal pilihan ganda dalam UN Matematika SMA, setelah dikelompokkan menurut kompetensi tertentu akan menghasilkan beberapa butir soal *testlet*.

Wainer, H. & Kiely, G. L. (1989) mengemukakan bahwa *testlet* merupakan sekumpulan butir soal atau pertanyaan yang berkaitan dengan sebuah materi (*content*) tertentu yang dikembangkan menjadi sebuah unit, berisi sejumlah langkah yang telah ditetapkan sehingga dapat diikuti seorang

peserta tes. Model *testlet* pada mulanya digunakan untuk mengkonstruksi dan menganalisis model tes adaptif (*Computerized Adaptive Testing/CAT*), dengan harapan bahwa *testlet* dapat memudahkan observasi dan pemilihan taraf sukar butir terkait dengan algoritma pada tes yang dibuat.

Lebih lanjut Cees A. W. Glas (2012) mengemukakan bahwa *testlet* merupakan satu set pertanyaan yang dikelompokkan di sekitar stimulus umum disebut sebuah *testlet*. Materi yang dapat diujikan dalam bentuk *testlet* adalah soal-soal pemahaman membaca dan penalaran analitis yang disajikan dalam format *testlet*. Penelitian terbaru telah merekomendasikan bahwa dua model potensial untuk mengevaluasi apakah perilaku respon *testlet* dapat diprediksi berdasarkan berbagai fitur *testlet*.

Embretson, S. E., & Reise, S. P., (2000) menyatakan bahwa secara garis besar, terdapat dua pendekatan dalam pengukuran, yaitu pendekatan teori tes klasik (*Classical Test Theory/CTT*) dan pendekatan teori tes modern (*Item Response Theory/IRT*). Teori tes klasik telah berkembang secara luas pada bidang psikologi dalam kurun waktu yang cukup lama. Namun dalam perjalanan tersebut dijumpai beberapa kelemahan dalam teori tes klasik, antara lain nilai statistik butir yang sering digunakan dalam pengukuran seperti tingkat kesukaran (*difficulty*) dan daya pembeda (*discrimination*) tergantung pada karakteristik peserta tes (Hambleton, R. K., dan Swaminathan H., 1985). Bila kemampuan peserta tes tinggi maka butir soal menjadi mudah dan sebaliknya bila kemampuan peserta tes rendah maka butir soal memiliki tingkat kesukaran yang tinggi. Statistik lain yang juga tergantung pada kemampuan peserta tes adalah koefisien validitas dan reliabilitas. Model ketergantungan tersebut sering disebut sebagai *group dependent* dan *item dependent*.

Dali S. Naga (2013) menyatakan bahwa teori tes modern *IRT* dikembangkan

untuk mengatasi berbagai keterbatasan dalam teori tes klasik. Hasil yang diperoleh melalui pendekatan teori klasik dirasakan kurang akurat. Teori responsi butir berusaha meningkatkan akurasi pengukuran melalui pemisahan taraf sukar butir dari kemampuan responden. Pemisahan ini berarti bahwa nilai taraf sukar butir terpisah dari kemampuan responden. Berapapun kemampuan responden, nilai taraf sukar tidak berubah. Karakteristik butir ditentukan oleh responsi para responden (baik kemampuan tinggi maupun kemampuan rendah) sehingga dikenal sebagai teori responsi butir (*Item Response Theory*).

Model matematis dalam *IRT* mempunyai makna bahwa probabilitas peserta tes menjawab benar sebuah butir tertentu tergantung pada kemampuan peserta tes dan parameter butir. Hal ini berarti bahwa peserta tes dengan kemampuan tinggi, memiliki probabilitas menjawab dengan benar sebuah butir lebih besar daripada peserta tes dengan kemampuan rendah. Agar butir pada *IRT* mengikuti karakteristik butir, maka diperlukan sejumlah persyaratan. Syarat-syarat tersebut diperlukan untuk menjamin bahwa bentuk teori responsi butir sesuai dengan karakteristik butir yang digunakan. Hambleton, R. K., dan Swaminathan H. (1985) menjelaskan bahwa terdapat 3 (tiga) asumsi dasar dalam *IRT* yaitu: (a) unidimensi berarti bahwa setiap butir tes hanya mengukur satu kemampuan, (b) invariansi parameter artinya karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes, dan parameter yang menjadi ciri peserta tes tidak tergantung pada ciri butir soal, dan (c) independensi lokal artinya parameter kemampuan peserta tes yang satu tidak tergantung pada kemampuan peserta tes yang lainnya, demikian pula parameter butir yang satu tidak tergantung pada parameter butir soal yang lainnya.

Abadyo (2009) dalam penelitiannya menyatakan bahwa walaupun ketergantungan

lokal tidak diharapkan dalam *IRT*, namun ada alasan penting untuk memasukkan butir-butir yang saling bergantung pada suatu asesmen. Banyak persoalan nyata membutuhkan penyelesaian yang terkait pada persoalan atau penyelesaian soal tunggal dalam cara tahap demi tahap. Termasuk butir-butir bergantung pada konteks di suatu tes dapat meningkatkan validitas konstruk. Contoh-contoh mengenai konstruk-konstruk yang bertautan, butir-butir yang bergantung termasuk butir-butir yang menuntut responden menyelesaikan persoalannya dan selanjutnya memberikan penjelasan bagaimana mereka memperoleh jawaban itu atau penggunaan butir berganda untuk mengukur secara komprehensif pada bagian bacaan, skenario, atau grafik. Oleh sebab itu, yang menjadi tantangan bagi pengembang tes adalah tidak mengeliminasi ketergantungan butir, tetapi bagaimana membuat model yang pantas sedemikian sehingga ketergantungan lokal tidak terjadi. Permasalahan tersebut dapat di atas dengan beberapa metode untuk mendeteksi ketergantungan lokal, dan pemodelan yang tepat untuk konstruk-konstruk yang bertautan dengan ketergantungan lokal dalam model *IRT*.

Jika diperoleh kebergantungan dalam data ketika digunakan himpunan-himpunan butir dalam konteks kebergantungan, maka salah satu metode penyekoran butir-butir itu adalah menggunakan model *IRT* politomus dan *testlet* (Thissen, et al., 1989; Thissen, Billeaud, McLeod, & Nelson, 1997; Yen, 1993, dalam Abadyo (2009)). Suatu *testlet* adalah unit penyekoran dalam suatu tes yang lebih kecil pada suatu tes, terdiri dari butir-butir yang mungkin atau mungkin tidak mempunyai kebergantungan lokal (Wainer & Kiely, 1987). Sebagai contoh, penggalan bacaan pada seksi *Reading Comprehension* dalam *TOEFL* dan butir-butir yang dikaitkannya dapat disusun menjadi satu *testlet*. Dalam penggunaan model *IRT* politomus untuk menyekor *testlet*, data itu

dapat dianalisis dan sementara itu kebebasan lokal di antara *testlet* yang berbeda tetap dapat dijaga.

Ditinjau dari akurasi estimasi reliabilitas, estimasi yang paling akurat adalah pada butir-butir yang bebas lokal, karena butir-butir yang bergantung lokal cenderung menggelembungkan estimasi reliabilitas (Sireci et al., 1991, dalam Abadyo (2009)). Ketika nampak butir-butir berbeda terkait pada kebergantungan suatu penggalan, pengelompokan secara bersama butir-butir itu ke dalam suatu *testlet* merupakan model struktur tes yang lebih tepat. Dengan menggunakan strategi ini, kebebasan butir lokal dapat dipertahankan untuk semua *testlet*, karena *testlet* dimodelkan sebagai unit (sebagai butir politomus). Jadi, himpunan butir-butir yang bergantung lokal sebagai model *testlet* dari struktur tes berbasis *testlet* akan memenuhi asumsi independensi lokal dalam *IRT*.

Lebih lanjut Abadyo (2009) menyatakan bahwa salah satu kekhawatiran penggunaan model-model *IRT* politomus adalah hilangnya informasi penting yang termuat di masing-masing butir. Dengan menjumlahkan skor-skor butir dalam suatu *testlet* untuk menghitung skornya itu, informasi yang terkait dengan jawaban benar pada butir-butir tertentu dari para peserta tes berpeluang menjadi hilang. Sebagai contoh, jika ada 10 *testlet* yang masing-masing terdiri dari 5 butir disekor secara dikotomus menggunakan model *IRT* tiga-parameter, maka ada $3 \times 5 \times 10 = 150$ parameter butir yang akan diestimasi. Sebaliknya, jika tes itu dikalibrasi menggunakan model politomus bagi struktur *testlet* (misalnya dihitung dengan menggunakan model respon berjenjang, Samejima (1969)), maka hanya ada satu parameter pembeda dan 5 parameter *threshold* akan diestimasi untuk masing-masing *testlet* (total ada $2 \times 5 \times 10 = 100$ parameter). Jadi, beberapa informasi pengukuran mungkin akan hilang ketika butir-

butir itu diringkas ke dalam *testlet*, ketika kebergantungan butir tidak tampak, pembentukan *testlet* dan penyekoran secara politomus tidak akan memperbaiki estimasi parameter-parameter butir, tes, maupun kemampuan peserta tes. Oleh sebab itu, derajat keberadaan ketergantungan lokal pada suatu tes harus diketahui dengan pasti sebelum menentukan bagaimana sebaiknya model tes yang akan dipakai.

Korelasi bersyarat antar-butir juga telah diusulkan sebagai ukuran dari ketergantungan lokal (Ferrara, Huynh, & Baghi, 1997; Ferrara, Huynh, & Michaels, 1999; Huynh & Ferrara, 1994, dalam Abadyo (2009)). Dalam metode ini, peserta ujian dibagi ke dalam delapan sampai sepuluh kelompok berdasar pada skor tes total, dan korelasi antarbutir dihitung dalam masing-masing interval skor tes. Korelasi antarbutir dalam suatu *testlet* dapat dirata-ratakan sepanjang tiap-tiap tingkatan skor dan masing-masing butir untuk mendapatkan ukuran statistik dari ketergantungan lokal untuk masing-masing *testlet*. Ukuran dari ketergantungan lokal dalam *testlet* ini dapat dibandingkan dengan statistik yang dihitung pada seluruh *testlet*. Jika rata-rata korelasi dalam *testlet* lebih tinggi dari korelasi antar *testlet*, estimasi reliabilitas yang diturunkan dari penyekoran dikotomus butir-butir itu secara positif akan menjadi bias. Lee dan Frisbie (1999) juga menghitung rata-rata korelasi dalam dan antara-testlet dengan pendekatan teori perumuman untuk mengases reliabilitas tes yang tersusun atas testlet. Ketika penyekoran *testlet* digunakan pada himpunan butir-butir mereka, perbedaan antara hasil perhitungan koefisien reliabilitas dan perumuman adalah kecil. Berdasarkan keadaan itu dapat disimpulkan bahwa penyekoran testlet merupakan penyekoran yang tepat digunakan bila dibandingkan dengan penyekoran butir dikotomus.

Local item independence (LID) adalah salah satu asumsi untuk model *IRT*. Ini berarti

bahwa dalam responsi butir dalam kondisi yang *independent* terhadap kemampuan responden. Oleh karena itu, seharusnya tidak ada korelasi antara dua item setelah mengendalikan karakteristik kedua butir tersebut. Item hanya berkorelasi melalui sifat laten yang diukur dalam tes ini (Lord dan Novick, 1968). Namun, asumsi *LID* ini hampir selalu dilanggar dalam aplikasi nyata. Kadang-kadang korelasi yang signifikan antar butir tetap ada setelah mengontrol pengaruh parameter kemampuan responden. Korelasi yang signifikan ini, berarti bahwa item tergantung dengan yang lainnya merupakan dimensi dalam pengukuran yang tidak diperhitungkan oleh sifat dimensi menyeluruh. Independensi lokal selalu menjadi penyebab kehilangan informasi untuk model *IRT* (Chen & Thissen, 1997).

Purwo Susongko (2010) menyatakan bahwa banyak keuntungan yang didapat bila bentuk tes pilihan ganda disusun dalam bentuk tes *testlet*. Selain keunggulan-keunggulan bentuk tes objektif seperti penskorannya dapat dilakukan lebih objektif, hasil penskoran dapat dilakukan dengan cepat dan hasilnya segera dapat diumumkan, cocok untuk jumlah peserta tes yang banyak. Di samping itu tes dalam bentuk *testlet* juga mempunyai sistem penskoran yang bersifat politomus. Berdasarkan kajian teoretik dan empirik yang telah dilakukan ternyata dari segi penskoran, *testlet* lebih praktis dibanding bentuk uraian karena penskoran dapat dilakukan secara objektif dan bersifat politomus.

Lebih lanjut Purwo Susongko (2010) menyatakan bahwa selain mempunyai kelebihan-kelebihan *testlet* juga mempunyai kelemahan. Ada kelemahan penskoran *testlet* secara politomus menggunakan model *Graded Response Model (GRM)* yaitu menggunakan skor total sehingga kehilangan informasi yang berisi pola yang tepat dari respons penempuh tes. Dari segi konstruksi tes, bentuk *testlet* lebih membutuhkan

keterampilan yang jauh lebih kompleks dalam menyusunnya terutama berkaitan dengan pemilihan alternatif jawaban pada setiap item. Penentuan alternatif jawaban pada bentuk tes objektif menjadi sangat penting karena berkaitan dengan tingkat peluang penempuh tes menjawab benar dengan menebak. Meningkatnya peluang penempuh tes menjawab benar dengan menebak akan semakin menurunkan nilai fungsi informasi item sehingga memperbesar kesalahan pengukuran yang terjadi. Untuk memudahkan pemahaman terkait dengan contoh membuat instrumen dalam bentuk *testlet*, maka pada pembahasan ini akan digunakan contoh penyusunan *testlet* yang diadaptasi dari soal-soal UN matematika SMA Program IPA tahun pelajaran

2014/2015. Kisi-kisi UN mata pelajaran matematika SMA Program IPA tahun pelajaran 2014/2015 yang disusun oleh Badan Standar Nasional Pendidikan (BSNP) ditetapkan melalui Keputusan BSNP No. 0027/P/BSNP/IX/2014 tentang Kisi-Kisi UN untuk Satuan Pendidikan Dasar dan Menengah Tahun Pelajaran 2014/2015. Kisi-Kisi UN menyajikan pengelompokan kompetensi matematika peserta didik SMA Program IPA dan sejumlah indikator pencapaian kompetensi. Kompetensi-kompetensi tersebut disarikan dari sejumlah KD yang tersebar dari kelas X sampai kelas XII

Tabel 1. Kisi-Kisi UN Matematika SMA Program IPA

NO.	KOMPETENSI	INDIKATOR
1.	Menggunakan logika matematika dalam pemecahan masalah.	Menentukan penarikan kesimpulan dari beberapa premis.
		Menentukan ingkaran atau kesetaraan dari pernyataan majemuk atau pernyataan berkuantor.
2.	Menyelesaikan masalah yang berkaitan dengan aturan pangkat, akar dan logaritma, fungsi aljabar sederhana, fungsi kuadrat, fungsi eksponen dan grafiknya, fungsi komposisi dan fungsi invers, sistem persamaan linear, persamaan dan pertidaksamaan kuadrat, persamaan lingkaran dan garis singgungnya, suku banyak, algoritma sisa dan teorema pembagian, program linear, matriks dan determinan, vektor, transformasi geometri dan komposisinya, barisan dan deret, serta mampu menggunakannya dalam	Menggunakan aturan pangkat, akar, dan logaritma.
		Menggunakan rumus jumlah dan hasil kali akar-akar persamaan kuadrat.
		Menyelesaikan masalah persamaan atau fungsi kuadrat dengan menggunakan diskriminan.
		Menyelesaikan masalah sehari-hari yang berkaitan dengan sistem persamaan linear.
		Menentukan persamaan lingkaran atau garis singgung lingkaran.
		Menyelesaikan masalah yang berkaitan dengan teorema sisa atau teorema faktor.
		Menyelesaikan masalah yang berkaitan dengan komposisi dua fungsi atau fungsi invers.
		Menyelesaikan masalah program linear.
		Menyelesaikan operasi matriks.
		Menyelesaikan operasi aljabar beberapa vektor dengan syarat tertentu.
Menyelesaikan masalah yang berkaitan dengan besar sudut atau nilai perbandingan trigonometri sudut		

NO.	KOMPETENSI	INDIKATOR
	pemecahan masalah.	antara dua vektor.
		Menyelesaikan masalah yang berkaitan dengan panjang proyeksi atau vektor proyeksi.
		Menentukan bayangan titik atau kurva karena dua transformasi atau lebih.
		Menentukan penyelesaian pertidaksamaan eksponen atau logaritma.
		Menyelesaikan masalah yang berkaitan dengan fungsi eksponen atau fungsi logaritma.
		Menyelesaikan masalah deret aritmetika.
		Menyelesaikan masalah deret geometri.
3.	Menentukan kedudukan, jarak dan besar sudut yang melibatkan titik, garis, dan bidang dalam ruang.	Menghitung jarak dan sudut antara dua objek (titik, garis dan bidang) di ruang dimensi tiga.
4.	Menggunakan perbandingan, fungsi, persamaan, identitas dan rumus trigonometri dalam pemecahan masalah.	Menyelesaikan masalah geometri dengan menggunakan aturan sinus atau kosinus.
		Menyelesaikan persamaan trigonometri.
		Menyelesaikan masalah yang berkaitan dengan nilai perbandingan trigonometri yang menggunakan rumus jumlah dan selisih sinus, kosinus dan tangen serta jumlah dan selisih dua sudut.
5.	Memahami konsep limit, turunan dan integral dari fungsi aljabar dan fungsi trigonometri, serta mampu menerapkannya dalam pemecahan masalah.	Menghitung nilai limit fungsi aljabar dan fungsi Trigonometri.
		Menyelesaikan soal aplikasi turunan fungsi.
		Menentukan integral tak tentu dan integral tentu fungsi aljabar dan fungsi trigonometri.
		Menghitung luas daerah dan volume benda putar dengan menggunakan integral.
6.	Mengolah, menyajikan dan menafsirkan data, serta mampu memahami kaidah pencacahan, permutasi, kombinasi, peluang kejadian dan mampu menerapkannya dalam pemecahan masalah.	Menghitung ukuran pemusatan atau ukuran letak dari data dalam bentuk tabel, diagram atau grafik.
		Menyelesaikan masalah sehari-hari dengan menggunakan kaidah pencacahan, permutasi atau kombinasi.
		Menyelesaikan masalah yang berkaitan dengan peluang suatu kejadian.

(Sumber: BSNP, 2014)

Pada tabel 1. di atas, terlihat bahwa kompetensi matematika SMA Program IPA dikelompokkan menjadi 6 sub kompetensi. Masing-masing sub kompetensi akan diukur menggunakan sejumlah butir soal *testlet* sesuai dengan banyaknya indikator pencapaian kompetensi. Misalnya sub

kompetensi-1 (*testlet-1*) “Menggunakan logika matematika dalam pemecahan masalah”, terdiri atas dua indikator pencapaian kompetensi yaitu: 1) menentukan penarikan kesimpulan dari beberapa premis, dan 2) menentukan ingkaran atau kesetaraan dari pernyataan majemuk atau pernyataan berkuantor. Pada indikator pencapaian

kompetensi nomor 1) dan 2), masing-masing indikator tersebut masih dapat dijabarkan menjadi beberapa indikator soal yang lebih spesifik, kemudian dibuatkan butir soalnya dalam bentuk pilihan ganda yang diskor menggunakan model dikotomus.

Langkah-langkah penyusunan instrumen model *testlet*, secara rinci dapat dijelaskan sebagai berikut.

1. Mengumpulkan data hasil UN serta salah satu paket naskah soal Matematika SMA Program IPA yang digunakan di salah satu provinsi atau kabupaten/kota. Agar hasil estimasi parameter butir soal (daya pembeda, tingkat kesukaran, dan tebakan semu) menggambarkan keadaan yang sesungguhnya, pilihlah salah satu provinsi atau kabupaten/kota yang dinyatakan bersih dalam penyelenggaraan UN (provinsi atau kabupaten/kota yang memiliki nilai integritas UN tinggi). Apabila data hasil UN sulit diperoleh, data respons peserta didik dapat pula diperoleh dengan cara melakukan pengujian secara langsung, serta hasilnya dicatat dan diperiksa dan disimpan dalam bentuk file *txt* (Notepad).
2. Menentukan nilai estimasi parameter butir soal (daya pembeda, tingkat kesukaran, dan tebakan semu). Estimasi terhadap parameter butir soal dapat dilakukan menggunakan pendekatan teori tes klasik atau teori tes modern (*item response theory/IRT*). Untuk memudahkan estimasi, dapat menggunakan bantuan aplikasi komputer yaitu: 1) pendekatan teori tes klasik dapat menggunakan program *iteman*, dan 2) pendekatan teori tes modern dapat menggunakan program *BILOG-MG*. Apabila tidak tersedia program aplikasi komputer, perhitungan estimasi nilai parameter butir soal dapat juga dilakukan dengan cara manual menggunakan rumus-rumus matematis.
3. Menyusun instrumen bentuk *testlet*. *Testlet* disusun dalam sebuah forum *focus*

group discussion (FGD). *FGD* beranggotakan guru mata pelajaran, pakar materi matematika dari perguruan tinggi, serta pakar pengukuran dan evaluasi. Dari salah satu paket soal UN Matematika SMA Program IPA yang diperoleh, selanjutnya dilakukan pengelompokan butir-butir soal berdasarkan kompetensi yang diukur dalam UN. Masing-masing kelompok butir soal tersebut disebut sebagai sebuah butir *testlet*. Menyusun urutan butir soal UN dalam *testlet* berdasarkan tingkat kesukarannya, dari yang paling mudah ke sukar.

4. Menghitung korelasi antar *testlet*, untuk memastikan bahwa antar *testlet* independen. Apabila antar *testlet* yang satu dengan yang lain mempunyai korelasi yang tinggi, berarti antar *testlet* saling bergantung sehingga model *IRT* tidak dapat digunakan untuk mengestimasi parameter kemampuan responden, dan sebaliknya.

PENUTUP

Testlet merupakan sekumpulan butir soal yang disusun menjadi sebuah unit sesuai dengan materi uji, dimana masing-masing soal tersebut memiliki sifat saling ketergantungan. Dalam praktiknya sangat sulit dihindarkan, karena untuk mencapai kompetensi tertentu dapat dilakukan melalui tahapan-tahapan secara hirarki seperti dalam mata pelajaran matematika. Konsep-konsep dalam matematika sifatnya hirarki, artinya untuk menguasai konsep berikutnya maka konsep sebelumnya harus dikuasai terlebih dahulu.

Jika dalam estimasi parameter butir ditemukan sifat kebergantungan antar butir, maka salah satu metode penyekoran butir-butir itu adalah menggunakan model *IRT* politomus dan model *testlet*. Sekumpulan butir soal yang dirangkum menjadi sebuah unit tes dapat disebut sebagai *testlet*, apabila antar butir tes pembentuk sebuah unit *testlet*

tersebut memiliki korelasi yang tinggi. Dengan pengelompokan demikian, antar *testlet* akan saling tergantung sehingga syarat independensi lokal dapat terpenuhi.

DAFTAR KEPUSTAKAAN

- Abadyo. Efek Kebergantungan Butir Lokal Pada Validitas Butir Maupun Tes Dalam Teori Respon Butir (Studi Pada Ujicoba Soal Olimpiade Matematika). *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA Fakultas MIPA*. Yogyakarta: Universitas Negeri Yogyakarta, 2009.
- Chen, W.H., Thissen, D. Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics* 22(3), 265-289, 1997.
- Cees A. W. Glas. *Fit to Testlet Models and Differential Testlet Functioning*. Enschede: LSA Council Research Report Publication, 2012.
- Dali S. Naga. *Teori Sekor pada Pengukuran Mental*. Jakarta: PT. Nagaran Citrayasa, 2013.
- Embretson, S. E., & Reise, S. P. *Item Response Theory for Psychologists*. NJ: Lawrence Erlbaum Associates Inc., 2000.
- Hambleton, R. K., dan Swaminathan H. *Item Response Theory*. Boston: Kluwer Nijhoff Publishing, 1985.
- Lee, G., & Frisbie, D. A. *Estimating reliability under a generalizability theory model for test scores composed of testlets*. *Applied Measurement in Education*, 12, 237-255, 1999.
- Lord, F. M., Novick, M. R. *Statistical Theories Of Mental Test Scores*. Reading Mass: Addison-Wesley, 1968.
- Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 59 Tahun 2014 tentang Kurikulum 2013 Sekolah Menengah Atas/Madrasah Aliyah, Lampiran III*.
- Peraturan BSNP No. 0027/P/BSNP/IX/2014 tentang Kisi-Kisi UN untuk Satuan Pendidikan Dasar dan Menengah Tahun Pelajaran 2014/2015*, hh. 27-28.
- Pitt, M.A., Kim, W., & Myung, I.J. Flexibility versus Generalizability in Model Selection. *Psychonomic Bulletin & Review*, 10, 29-44, 2003.
- Purwo Susongko. Perbandingan Keefektifan Bentuk Tes Uraian dan Testlet dengan Penerapan Graded Response Model (GRM). *Jurnal Penelitian dan Evaluasi Pendidikan Tahun 14, Nomor 2, 2010*, hh. 269-288.
- Samejima, F. Estimation of Latent Ability Using A Response Pattern of Graded Scores. *Psychometrika. Monograph Supplement, No. 17*, 1969.
- Sireci, S. G., Thissen, D., & Wainer, H. On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28(3), 237-247, 1991.
- Sudijono, A. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada, 1996.
- The Dynamic Learning Maps Alternate Assessment System. http://dynamiclearningmaps.org/sites/default/files/documents/Training/mod7_note_s.pdf (diakses tanggal 5 Mei 2015).
- Thissen, D., Steinberg, L., & Mooney, J. Trace Lines for Testlets: A Use of Multiplecategorical Response Models. *Journal of Educational Measurement*, 26, 247-260, 1989.
- Wainer, H. & Kiely, G. L. Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185-201, 1987.
- Wainer, H., & Wang, X. Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, 37, 203-220, 2000.