

ESTIMASI PARAMETER ITEM DAN *LATENT CLASS* DENGAN MODEL DINA UNTUK DIAGNOSIS KESULITAN BELAJAR

Kusaeri¹, Suryanto², Kumaidi³

¹IAIN Sunan Ampel, Jalan Jenderal Ahmad Yani 117 Surabaya

²Universitas Negeri Yogyakarta, Kampus Karangmalang, Yogyakarta

³Universitas Muhammadiyah Surakarta, Jl. A. Yani Tromol Pos 1, Pabelan, Surakarta
e-mail: erinda_1@yahoo.co.id

Abstract: Estimation of Item Parameter and Latent Class with DINA Model to Diagnose Learning Difficulties. This study aims to estimate item parameter of diagnostic test developed with DINA model and identify attribute profiles of each test participant. The instrument of this study was diagnostic test using multiple choice format with 4 options. The data were analyzed using Mplus software, R program and ITEMAN. The results show that out of 8 items measuring social arithmetic and comparison, there was only one item that had low guessing and slip parameter. The study also found that basic operation and concept in arithmetic and verbal questions were problematic for most students.

Keywords: DINA model, diagnostic test, latent class, attribute profile

Abstrak: Estimasi Parameter Item dan Latent Class dengan Model DINA untuk Diagnosis Kesulitan Belajar. Penelitian ini bertujuan untuk mengestimasi parameter item dari tes diagnostik yang dikembangkan dengan model DINA dan mengidentifikasi profil atribut setiap peserta tes. Instrumen penelitian ini berupa tes diagnostik berbentuk pilihan ganda dengan 4 pilihan jawaban. Data dianalisis dengan menggunakan *software* Mplus, program R dan ITEMAN. Hasil penelitian menunjukkan bahwa dari 8 item yang mengukur materi aritmetika sosial dan perbandingan, hanya ada satu item dengan parameter *guessing* dan *slip* rendah. Temuan lain operasi dan konsep dasar dalam aritmetika serta soal bentuk verbal masih menjadi masalah bagi sebagian besar siswa.

Kata kunci: model DINA, tes diagnostik, *latent class*, profil atribut

Kajian-kajian mutakhir mengelompokkan berbagai pandangan tentang belajar dalam suatu kontinum, mulai dari teori empiris yang berlandaskan psikologi behavioristik hingga teori rasionalis yang berdasarkan psikologi kognitif. Empirisme menekankan pengaruh eksternal yaitu pengalaman anak, dan bukan faktor-faktor internal, sebagai faktor utama pendorong belajar. Empirisme juga lebih mendasarkan diri pada perilaku yang dapat diobservasi, seperti guru memberi pujian, teguran, senyum, dan marah jika anak mengganggu teman lainnya (Santrock, 2004). Teori ini berasumsi bahwa suatu perilaku dapat dijelaskan dari jenis stimulus yang diberikan (*stimulus response bond*). Oleh karena itu, untuk memunculkan suatu perilaku yang diinginkan secara konsisten yang disebut kebiasaan (*habit formation*), diperlukan adanya suatu proses pengondisian atau *conditioning* (Marhaeni, 2005).

Sebagai reaksi terhadap teori behavioris, muncul teori rasionalis yang didasarkan atas psikologi kognitif. Teori rasionalis berasumsi bahwa manusia memiliki kapasitas atau kemampuan internal (*innate capacity*). Menurut teori ini, pikiran, perasaan dan motivasi merupakan hal penting dalam proses pembelajaran, meskipun kesemuanya tidak dapat diamati. Proses mental seperti pemahaman anak tentang suatu konsep, perasaan senang terhadap guru dan motivasi anak akan mengontrol perilaku dan hasil belajar anak (Santrock, 2004). Penekanan kognitif seperti ini akhirnya menjadi basis bagi banyak pendekatan dalam pembelajaran dewasa ini.

Uraian pada paragraf terakhir menunjukkan bahwa ada dua variabel yang dapat diidentifikasi ketika berada dalam *domain* psikologi kognitif, yakni variabel manifes dan variabel laten. Dilihat dari namanya,

manifest, berarti variabel ini keberadaannya dapat diketahui secara kasat mata (tampak), dan besaran kuantitatifnya dapat diketahui secara langsung, seperti hasil belajar anak. Variabel kedua adalah variabel laten, yakni variabel yang nilai kuantitatifnya tidak dapat dilihat secara kasat mata (Umar, 2011), seperti pemahaman anak atau motivasinya.

Lazarsfeld pada tahun 1950 mengenalkan suatu model yang mengaitkan variabel manifes dan variabel laten (Bolt, 1999). Model yang memiliki karakteristik demikian dinamakan kelas laten (*latent class*). *Latent class* merupakan hasil pengelompokan atau pemartisan suatu populasi yang heterogen ke dalam sejumlah subpopulasi yang homogen. Menurut Croon (2002), masing-masing anak memiliki satu dari *latent class*, dan anak yang memiliki *latent class* yang sama memiliki karakteristik yang mirip dalam merespon variabel manifes.

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*) suatu materi yang diujikan. Misalnya, untuk kelompok anak yang menguasai materi diberi kode 1, dan yang tidak menguasai diberi kode 0. Hal itu, menurut Templin (2011), lebih mudah dilakukan dibandingkan dengan menempatkan peserta tes pada suatu skala berdasarkan skor yang diperolehnya. Penempatan peserta tes pada suatu skala juga sangat rawan muncul kesalahan sehingga memengaruhi hasil yang didapat. Sumber kesalahan umumnya terletak pada penetapan skor batas (*cut score*), karena hanya didasarkan pada estimasi.

Saat ini telah berkembang berbagai model *latent class* untuk kegiatan diagnosis (Bolt, 1999; von Schrader, 2006; de la Torre, 2008). Salah satu di antaranya yang berkembang pesat saat ini adalah model DINA. Hal ini ditandai dengan banyaknya kajian terhadap model ini berupa artikel konseptual dan hasil penelitian yang dipublikasikan di berbagai jurnal, penelitian untuk tugas akhir mahasiswa S-3 berupa disertasi, dan berbagai monograf (Anozie & Junker, 2007; von Davier, 2011).

DINA berasal dari kata *deterministic input*, *noisy*, dan *AND gate*. Komponen *deterministic input* menggambarkan kemampuan seorang peserta tes dalam menjawab suatu item (benar atau salah) yang didasarkan pada penguasaan mereka terhadap atribut yang diukur (Rupp dkk., 2010). Atribut merupakan kemampuan atau kompetensi yang harus dimiliki anak agar mampu menyelesaikan suatu item (Kusaeri, 2012). Penguasaan terhadap atribut tersebut direpresentasikan dalam suatu matriks Q. Matriks Q merupakan suatu matriks dengan M baris dan N kolom yang unsur-unsur di dalamnya

terdiri atas bilangan 0 dan 1 (bilangan biner). Komponen item menempati baris, dan komponen atribut menempati kolom pada matriks Q. Unsur pada matriks Q akan bernilai 1 apabila seorang peserta tes menguasai semua atribut yang diperlukan untuk menyelesaikan item tertentu. Sebaliknya, unsur pada matriks Q bernilai 0 bila peserta tes tidak menguasai salah satu atribut yang dipersyaratkan (Liu dkk., 2009; de la Torre dkk., 2010).

Sebagai contoh, berikut disusun matriks Q dengan 5 item dan 4 atribut (Gambar 1). Dengan demikian, pada matriks Q yang terbentuk terdapat 5 baris dan 4 kolom. Misalnya, untuk menyelesaikan item 1 diperlukan penguasaan terhadap atribut 1; item 2 memerlukan penguasaan terhadap atribut 1 dan atribut 2; item 3 memerlukan penguasaan terhadap atribut 1 dan atribut 3; item 4 memerlukan penguasaan atribut 2 dan atribut 4; serta item 5 memerlukan penguasaan semua atribut. Jadi, matriks Q yang diturunkan dari 5 item dan 4 atribut tersebut adalah seperti Gambar 1.

$$Q_{5 \times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Gambar 1. Matriks Q dari Lima Item dan Empat Atribut

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

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

Kelebihan model DINA apabila digunakan dalam konteks diagnosis adalah kemampuannya memberikan informasi diagnosis tanpa menuntut adanya analisis pengecoh. Artinya, dalam proses pengembangan tes diagnosis tidak diperlukan analisis pengecoh seperti lazimnya dalam pengembangan tes diagnostik konvensional. Dengan demikian, dari aspek waktu dan

konstruksi tes lebih mudah. Keunggulan lain adalah informasi diagnostik yang diperoleh dari model ini lebih kaya dibandingkan tes diagnostik konvensional. Menurut Templin (2011) reliabilitas hasil diagnostik yang dimiliki model ini lebih tinggi dibandingkan dengan tes diagnostik konvensional.

Mengacu kepada karakteristik yang dimiliki Model DINA tersebut, tujuan penelitian ini adalah untuk mengestimasi parameter item (*guessing* dan *slip*) dari tes diagnostik yang dikembangkan dengan model DINA, dan mengidentifikasi profil atribut berupa *latent class* dari masing-masing peserta tes. Materi yang menjadi fokus kajian penelitian ini adalah aljabar kelas VII SMP khususnya aritmetika sosial dan perbandingan. Materi ini dipilih dengan pertimbangan keduanya merupakan materi yang sering kali dianggap sulit oleh sebagian besar anak.

METODE

Penelitian ini merupakan penelitian deskriptif eksploratori. Informasi yang dicari di antaranya karakteristik item dan profil atribut peserta tes. Karakteristik item yang dimaksud mencakup parameter *guessing* dan *slip* masing-masing item. Sementara itu, penelitian deskriptif digunakan untuk mendeskripsikan dan menganalisis jawaban peserta tes. Deskripsi seperti ini dimaksudkan agar diperoleh kesimpulan yang kuat dan saling melengkapi profil atribut peserta tes tentang miskonsepsi anak.

Instrumen penelitian ini berupa tes bentuk pilihan ganda dengan 4 pilihan jawab (*option*), yang terdiri dari 8 item. Dari 8 item itu, 4 item digunakan untuk menguji materi aritmetika sosial dan 4 item sisanya menguji materi perbandingan. Instrumen tersebut diambil dari paket tes diagnostik yang dikembangkan untuk menguji materi aljabar kelas VII SMP yang meliputi materi bentuk aljabar, persamaan linier satu variabel, pertidaksamaan linier satu variabel, aritmetika sosial, perbandingan dan himpunan (sebanyak 37 item). Instrumen tersebut diujikan pada siswa kelas VIII SMP di 3 sekolah, yakni SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul dan SMPN 1 Panjatan Kulon Progo. Kegiatan uji coba ini diikuti oleh 180 siswa, sehingga dapat dijarah 180 data berupa respon siswa.

Data respon siswa selanjutnya dianalisis dengan menggunakan 3 *software* komputer, yakni program Mplus, program R dan ITEMAN. Program Mplus digunakan untuk menganalisis dan memberikan informasi tentang profil atribut peserta tes, dan digunakan program Mplus versi 6 dari Muthen & Muthen (2010). Dalam pelaksanaannya, program ini didukung oleh program *Cognitive Diagnostic Model* (CDM) yang di-

kembangkan oleh Templin (2008) untuk membangkitkan *syntax* Mplus. Program R versi 2.14.1 dari Venables & Smith (2011) digunakan untuk mengestimasi parameter item (berupa *guessing* dan *slip*). Sementara itu, program ITEMAN digunakan untuk menganalisis pola respon siswa.

HASIL DAN PEMBAHASAN

Hasil Estimasi Parameter Item

Tabel 1 menyajikan hasil estimasi parameter *guessing* dan *slip* pada masing-masing item. Pada Tabel 1 juga dilengkapi informasi tentang *standard error* (SE) setiap parameter. Hal ini penting untuk memberikan informasi bahwa ada suatu kuantitas yang perlu diperhatikan dan nilainya tak tentu selama proses estimasi parameter item dilakukan. *Standard error* cenderung dan bahkan akhirnya mendekati 0 ketika estimasi parameter yang didapatkan hampir sama dengan parameter populasi.

Untuk menentukan level parameter *guessing* dan *slip*, digunakan acuan yang dibuat oleh de la Torre dkk. (2010), yakni (a) rendah, bila terletak pada interval 0,05-0,15 dan (b) tinggi, bila terletak pada interval 0,20-0,30. Klasifikasi di atas selanjutnya dimodifikasi untuk menentukan level parameter *guessing* dan *slip*, sehingga menjadi (a) rendah, bila pada interval 0,00-0,15; (b) sedang, bila pada interval 0,16-0,25; (c) tinggi, bila pada interval 0,26-0,40; dan (d) sangat tinggi, bila pada interval 0,41-1,00.

Tabel 1. Hasil Estimasi Parameter *Guessing* dan *Slip* Masing-masing Item

No. Item	Parameter			
	<i>Guessing</i>	S.E	<i>Slip</i>	S.E
1.	0,112	0,011	0,034	0,021
2.	0,374	0,039	0,602	0,300
3.	0,451	0,041	0,313	0,131
4.	0,174	0,093	0,173	0,026
5.	0,202	0,031	0,332	0,138
6.	0,302	0,042	0,213	0,095
7.	0,534	0,060	0,227	0,064
8.	0,184	0,037	0,623	0,135

Berdasarkan Tabel 1 ditemukan bahwa ada 1 item yang memiliki parameter *slip* dan *guessing* rendah, 3 item dengan parameter *slip* dan *guessing* sedang, 2 item memiliki parameter *slip* dan *guessing* tinggi dan 2 item memiliki parameter *slip* dan *guessing* sangat tinggi. Item yang baik adalah item dengan parameter *slip* dan *guessing* yang rendah. Jadi, hanya ada 2 item

yang memenuhi sifat demikian, yakni item nomor 1 dan 4. *Standar error* dengan nilai lebih dari 10% juga ditemukan pada item nomor 2, 3 dan 5 ketika mengestimasi parameter *slip*. Artinya, hasil estimasi parameter *slip* pada ketiga item itu memiliki peluang yang tinggi berbeda dengan parameter populasi.

Selain hasil penelitian utama, terdapat beberapa hasil tambahan. Hasil tambahan pertama adalah level parameter *guessing* dan *slip* pada suatu item dapat digunakan untuk menentukan tingkat kesukaran item. Ketentuan ini mengacu kepada pendapat Zhang (2006) bahwa suatu item dikatakan mudah apabila item itu memiliki parameter *guessing* tinggi dan *slip* rendah; item yang sulit bila memiliki parameter *guessing* rendah dan *slip* tinggi; dan item sedang bila memiliki parameter *guessing* dan *slip* rendah. Mengacu kepada ketentuan ini, item nomor 1 dan 3 memiliki tingkat kesukaran sedang, item nomor 6 dan 7 memiliki tingkat kesukaran mudah, dan item nomor 2, 4, 5 dan 8 memiliki tingkat kesukaran item sulit.

Hasil tambahan lainnya berupa hasil estimasi parameter *guessing* dan *slip*, digunakan untuk menentukan indeks daya beda item (δ_j). Indeks ini dihitung melalui formula $\delta_j = P_{\alpha A} - P_{\alpha B} = (1 - s_i) - g_i$ (Rupp dkk., 2010). Indeks daya beda item pada kelas laten ternyata memiliki fungsi sama dengan daya pembeda butir pada teori tes klasik. Oleh karena itu, klasifikasi yang dibuat oleh Crocker & Algina (1986), digunakan dalam konteks ini. Suatu item apabila memiliki indeks daya beda (a) lebih dari atau sama dengan 0,40 dikatakan baik; (b) terletak di antara 0,30 sampai dengan 0,39 dikatakan cukup; (c) di antara 0,20 sampai dengan 0,29 dikatakan kurang; dan (d) kurang dari atau sama dengan 0,19 dikatakan jelek.

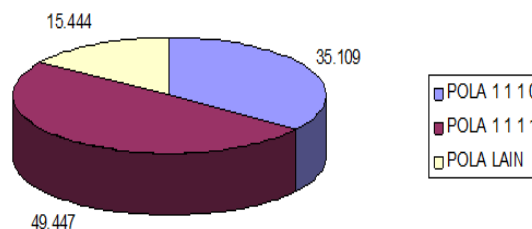
Indeks daya beda berkaitan dengan kualitas suatu item, dan hasilnya dirangkum sebagai berikut. Item nomor 1, 4, 5 dan 6 memiliki indeks daya beda lebih dari 0,4 sehingga dikategorikan **baik**. Item nomor 2 dan 8 memiliki indeks daya beda kurang dari 0,19 sehingga dikategorikan **jelek**. Item nomor 3 dan 7 dengan indeks daya beda di antara 0,2 dan 0,29 dikategorikan **kurang**.

Item yang memiliki indeks daya beda jelek, apabila ditelusuri lebih lanjut, memiliki tingkat kesukaran item sulit. Dengan demikian, peserta tes banyak yang tidak mampu mengerjakan 2 item ini. Kalaupun mereka ada yang mampu menjawabnya, hasil jawaban itu diperoleh dengan jalan menebak (*guessing*). Dengan kata lain, banyak peserta tes yang sebenarnya menguasai semua atribut pada kedua item ini, tetapi mereka menjawabnya secara salah karena *slip*. Item yang memiliki indeks daya beda jelek bukan persoalan krusial dalam tes diagnostik. Alasannya, tes ini tidak berfungsi untuk membedakan anak yang pandai dan kurang. Oleh

karena itu, kedua item itu tidak dihapus atau dihilangkan dari paket tes diagnostik.

Hasil Estimasi Profil Atribut Peserta Tes

Pada item nomor 1 sampai dengan 4, terdapat empat atribut yang dipersyaratkan. Empat atribut ini berkorelasi dengan 16 *laten class*. Keempat atribut itu adalah (a) memahami konsep harga jual dan harga beli, (b) menggunakan konsep aljabar untuk menghitung besar untung atau rugi suatu penjualan, (c) mendefinisikan pengertian rabat (diskon), bruto, tara dan neto, dan (d) mendefinisikan pengertian pajak (PPN dan PPh). Hasil *output* Mplus menunjukkan bahwa dari 16 kelas itu, pola jawaban peserta tes cenderung dominan di dua kelas, yakni kelas dengan pola 1 1 1 0 dan 1 1 1 1. Profil kedua kelas itu diperlihatkan oleh Gambar 2.



Gambar 2. Profil Atribut yang Dominan Muncul pada Item 1-4 (%)

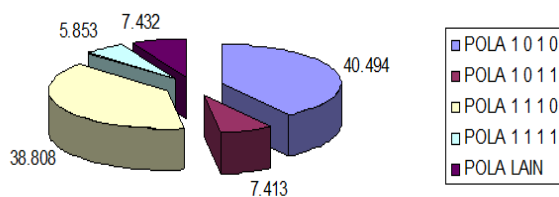
Dari Gambar 2, ada 49,45% peserta tes yang memiliki profil 1 1 1 1 atau menguasai seluruh atribut. Sebaliknya, sebagian peserta tes (35,10%) memiliki profil 1 1 1 0. Kelompok ini adalah peserta tes yang tidak memahami pengertian pajak (PPN dan PPh). Selanjutnya, penelusuran terhadap respon anak pada masing-masing item juga menguatkan hal di atas. Selain ditemukan banyaknya anak yang tidak memahami pengertian pajak (PPN dan PPh), juga ditemukan anak yang mengalami miskonsepsi dalam melakukan perkalian bilangan bulat dan pecahan, konsep untung dan rugi, konsep harga jual dan diskon, mengidentifikasi variabel yang terdapat pada soal, dan menyusun relasi bentuk aljabar yang terkait dengan untung dan rugi.

Hasil penelitian ini menunjukkan bahwa operasi-operasi dasar ataupun konsep-konsep dasar dalam aritmetika masih memberikan andil terhadap penguasaan anak dalam aritmetika sosial. Ketidakkampuan menguasai konsep dasar dalam aritmetika menyebabkan anak tidak mampu menyelesaikan permasalahan dalam aritmetika sosial.

Kemampuan menyelesaikan soal verbal juga merupakan kunci penting ketika anak berada dalam konteks aritmetika sosial, karena permasalahan dalam arit-

metika sosial umumnya berbentuk soal verbal. Oleh karena itu, keberhasilan menyusun model matematika merupakan langkah awal dan penting. Agar anak mampu menyusun model matematika, diperlukan langkah-langkah sistematis seperti mengidentifikasi variabel yang terdapat soal dan menyusun relasi bentuk aljabar.

Pada item nomor 5 sampai 8, juga terdapat empat atribut atau 16 kelas yang bersesuaian. Keempat atribut itu adalah (a) memahami bahasa soal pada masalah yang berkaitan dengan persamaan atau pertidaksamaan, (b) mengkonversi satuan yang satu ke lainnya, (c) pengertian “perbandingan senilai” dan (d) pengertian “perbandingan berbalik nilai.” Dari empat atribut tersebut, sebagian besar peserta tes (79,3%) tidak menguasai atribut kedua dan keempat. Dengan demikian, peserta tes ini tidak mampu mengkonversi satuan yang satu ke lainnya dan tidak memahami perbandingan berbalik nilai. Hanya 5,85% peserta tes yang menguasai semua atribut, yakni peserta yang memiliki profil 1 1 1 1. Profil atribut yang dominan muncul pada kelompok item ini tampak pada Gambar 3.



Gambar 3. Profil Atribut yang Dominan Muncul pada Item 5-8 (%)

Hasil analisis terhadap pola respon siswa pada masing-masing item semakin melengkapi profil atribut tersebut di atas. Analisis ini juga menemukan bahwa miskonsepsi yang dominan terjadi pada saat anak mengkonversi satuan yang satu ke yang lainnya dan memahami perbandingan berbalik nilai serta membedakan perbandingan senilai dan berbalik nilai. Selain itu, menyederhanakan pecahan, memahami bahasa soal dan pengertian skala juga masih merupakan hal-hal mendasar yang belum dikuasai oleh anak.

Hasil penelitian ini semakin menguatkan bahwa pada soal berbentuk verbal masih dirasa sulit bagi anak seperti terjadi pada item nomor 5, 7 dan 8. Selain kesulitan dalam menyusun model matematika, item nomor 7 dan 8 juga menuntut kemampuan anak dalam membedakan perbandingan senilai dan berbalik nilai. Dengan demikian, pada item nomor 7 dan 8 paling tidak terdapat tiga dimensi atau aspek kemampuan yang terlibat di dalamnya. Dimensi itu adalah kemampuan membedakan perbandingan senilai dan berbalik nilai, menyusun model matematika, dan operasi hitung bilangan.

Temuan-temuan penelitian ini di kedua topik aljabar (aritmetika sosial dan perbandingan) menunjukkan bahwa operasi dasar dan konsep dasar aritmetika memberikan kontribusi penting terhadap keberhasilan anak dalam aljabar. Oleh karena, itu tepat apa yang dikatakan oleh Stacey (2009) bahwa anak tidak bisa melupakan sifat-sifat operasi dan hal-hal mendasar dalam aritmetika ketika mempelajari aljabar. Banyak anak mengalami kesulitan mempelajari aljabar karena mereka tidak yakin dengan sifat-sifat aritmatika yang digeneralisasikan ke aljabar. Hasil penelitian Blanco & Garrote (2007) juga menunjukkan bahwa materi aljabar akan dikuasai anak dengan baik apabila konsep-konsep dasar aritmetika dikuasai pula dengan baik.

Selain itu, menyelesaikan soal bentuk verbal juga masih menjadi kendala bagi sebagian besar anak. Namun, kondisi seperti ini tidak hanya terjadi di Indonesia. Penelitian yang dilakukan oleh Fong & Lee (2009) di Singapura menunjukkan bahwa menyelesaikan soal bentuk verbal juga merupakan salah satu masalah besar dalam pembelajaran matematika di Singapura. Siswa menghadapi beragam kendala dalam menggunakan simbol aljabar untuk menggambarkan soal bentuk verbal, seperti memahami makna huruf yang digunakan sebagai simbol aljabar, menerjemahkan bahasa aslinya ke dalam bentuk persamaan, memahami struktur makna dari soal cerita, khususnya sifat dari hubungan antara kuantitas dan bagaimana mereka saling terkait, dan menggunakan makna yang didasarkan pada teks dalam pengkonstruksian model matematika. Di Amerika Serikat, para mahasiswa juga masih banyak mengalami kesulitan dan mengalami miskonsepsi saat menggunakan pernyataan aljabar untuk menyusun model matematika (Clement, 1982).

SIMPULAN

Berdasarkan uraian pada hasil penelitian dan pembahasan dapat dibuat kesimpulan sebagai berikut. Dari 8 item yang mengukur materi aritmetika sosial dan perbandingan, hanya ada satu item yang memiliki parameter *guessing* dan *slip* rendah; tujuh item lainnya memiliki parameter *guessing* dan *slip* sedang, tinggi dan sangat tinggi. Operasi-operasi dasar ataupun konsep-konsep dasar dalam aritmetika masih merupakan penyebab lemahnya penguasaan anak dalam aritmetika sosial dan perbandingan. Soal dalam bentuk verbal juga masih menjadi kendala bagi sebagian besar anak saat mereka dihadapkan pada permasalahan dalam aritmetika sosial dan perbandingan.

Penelitian ini telah dirancang dan dilakukan secara cermat, hati-hati dan sungguh-sungguh dengan melibatkan banyak pihak untuk menelaah hasil pengem-

bangun perangkat instrumen. Tahapan juga dilaksanakan secara ketat, sesuai dengan saran dari promotor/kopromotor, *reviewer* dan berbagai pihak agar dihasilkan perangkat tes diagnostik yang berkualitas. Namun, penelitian ini tidak terlepas dari beberapa keterbatasan, karena pengembangan tes diagnostik berbasis pada *latent class* (model DINA) merupakan hal baru dalam dunia pengujian pendidikan di Indonesia. Beberapa kelemahan dalam penelitian ini di antaranya adalah (a) Pelaksanaan uji coba penelitian ini baru dikenakan pada sejumlah siswa kelas VIII SMP negeri, dan dilakukan di akhir semester gasal. Kondisi ini akan berbeda apabila 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 ada beberapa keterbatasan, baik menyangkut karakteristik subjek penelitian, waktu dan tempat penelitiannya; dan (b) Penyusunan item tes dilakukan dengan mengacu kepada 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 pada matriks Q.

DAFTAR RUJUKAN

- Anozie, N.O. & Junker, B.W. 2007. *Investigating the Utility of a Conjunctive Model in Q-Matrix Assessment Using Monthly Student Records in an Online Tutoring System*. Makalah disajikan pada The Annual Meeting of the National Council on Measurement in Education (NCME), Illinois, Chicago, 10-12 April.
- Blanco, L.J. & Garrote, M. 2007. Difficulties in Learning Inequalities in Students of the First Year of Pre-university Education in Spain. *Eurasia Journal of Mathematics, Sains & Technology Education*, 3: 221-229.
- Bolt, D.M. 1999. *Psychometric Methods for Diagnostic Assessment and Dimensionality Representation*. Disertasi doktor tidak diterbitkan. Urbana: University of Illinois, Urbana.
- Clement, J. 1982. Algebra Word Problem Solutions: Thought Processes Underlying a Common Misconception. *Journal for Research in Mathematics Education*, 13: 16-30.
- Crocker, L. & Algina, J. 1986. *Introduction to Classical and Modern Test Theory*. Forth Worth: Holt, Rinehart and Winston.
- Croon, M. 2002. Ordering the Classes. Dalam J.A. Hagenaaers & A.L. McCutcheon, *Applied Latent Class Analysis* (hlm.137-162). New York: Cambridge University Press.
- 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. 2009. A Cognitive Diagnosis Model for Cognitively Based Multiple-choice Options. *Applied Psychological Measurement*, 33: 163-183.
- 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., Hong, Y., & Deng, W. 2010. Factors Affecting the Item Parameter Estimation and Classification Accuracy of the DINA Model. *Journal of Educational Measurement*, 47 (2): 227-249.
- Fong, S. Ng & Lee, K. 2009. The Model Method: Singapore Children's Tool for Representing and Solving Algebraic Word Problem. *Journal for Research in Mathematics Education*, 40 (3), 282-313.
- Henson, R.A. & Templin, J.L. 2007. *Q-Matrix Construction*. (Online), (http://jtemplin.coe.uga.edu/dcm/dcm07ncme/dcm07ncme_qmatrix.html), diakses 24 Desember 2011.
- Kusaeri. 2012. *Pengembangan Tes Diagnostik dengan Menggunakan Model DINA untuk Mendapatkan Informasi Salah Konsepsi dalam Aljabar*. Disertasi doktor tidak diterbitkan. Yogyakarta: Universitas Negeri Yogyakarta.
- Liu, Y., Douglas, J.A., & Henson, R.A. 2009. Testing Person Fit in Cognitive Diagnosis. *Applied Psychological Measurement*, 33 (8): 579-598.
- Marhaeni, A.A.I.N. 2005. *Pengaruh Asesmen Portofolio dan Motivasi Berprestasi dalam Belajar Bahasa Inggris terhadap Kemampuan Menulis dalam Bahasa Inggris*. Disertasi doktor tidak diterbitkan. Jakarta: Universitas Negeri Jakarta.
- Muthen, L.K. & Muthen, B.O. 2010. *Mplus User's Guide* (6th Ed). Los Angeles, CA: Muthen & Muthen.
- Rupp, A.A. & Templin, J. 2008. The Effects of Q-Matrix Misspecification on Parameter Estimation and Classification Accuracy in the DINA Model. *Educational Psychological Measurement*, 68: 78-96.
- Rupp, A.A., Templin, J., & Henson, R.A. 2010. *Diagnostic Measurement: Theory, Methods and Applications*. New York: The Guilford Press.
- Santrock, J.W. 2004. *Psikologi Pendidikan*. Terjemahan Tri Wibowo. 2008. New York: McGraw-Hill Company.
- Stacey, K. 2009. *The Transition from Arithmetic Thinking to Algebraic Thinking*. Australia: University of Melbourne.
- Templin, J. 2008. *Cognitive Diagnosis Modeling with Mplus (User Guide)*. (Online), (<http://jtemplin.coe.uga.edu/teaching/dcm08ncme/cdmuserguide.html>), diakses 19 Desember 2011.
- Templin, J. 2011. *Diagnostic Measurement: Theory, Methods and Application*. (Online), (<http://jtemplin.coe>).

- uga.edu/workshops/dcm/uga_dcm1.html), diakses 8 Desember 2011.
- Umar, J. 2011. *Penilaian dan Peningkatan Mutu Pendidikan di Indonesia: Kumpulan Tulisan Antara Tahun 1988-2008*. Jakarta: UIN Press.
- Venables, W.R. & Smith, D.M. 2011. *An Introduction to R: A Programming Environment for Data Analysis and Graphics Version 2.14.1*. (Online), (<http://cran.r-project.org/mirrors.html>), diakses 17 Januari 2012.
- von Davier, M. 2011. *Equivalency of the DINA Model and Constrained General Diagnostic Model*. Laporan penelitian tidak diterbitkan. New Jersey: Educational Testing Service (ETS).
- von Schrader, S. 2006. *On the Feasibility of Applying Skills Assessment Models to Achievement Test Data*. Disertasi doktor tidak diterbitkan. Iowa: The University of Iowa.
- Zhang, W. 2006. *Detecting Differential Item Functioning Using the DINA Model*. Disertasi doktor tidak diterbitkan. Greensboro: The University of North Carolina.