

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 menunjukkan 1) tahapan pengembangan tes meliputi: identifikasi kompetensi dasar dan merumuskan indikator, menyusun *learning continuum*, menyusun hierarki materi, merumuskan atribut, mengonstruksi 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 mengkonstruksi 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), Suwanto (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 direpresentasikan 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 probabilitas 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 diagnostik 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, mengonstruksi 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 dijangkau 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 pelaksanaannya, 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 menunjukkan 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 mempertimbangkan 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, campuran dan desimal; (31) mengurangi 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.

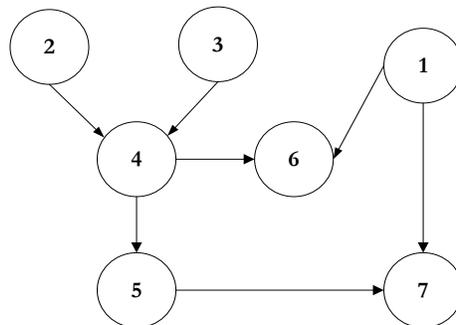
c. Hasil Penyusunan Hierarki Materi

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 lainnya. 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.

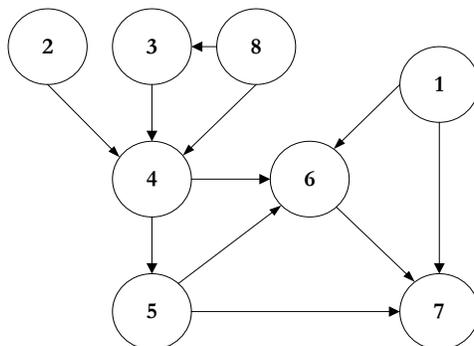


Gambar 1. Diagram Hirarki

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:

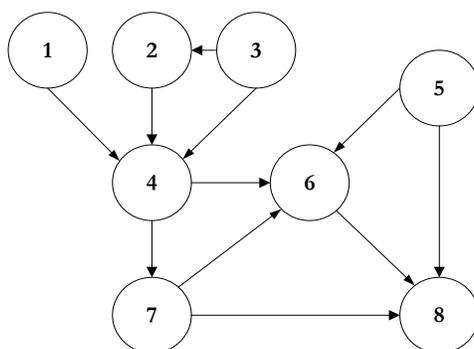
Hasil perbaikan berdasarkan saran praktisi, selanjutnya dikonsultasikan dengan promotor. Menurut promotor, ada yang kurang tepat dalam penempatan nomor urut pada diagram. Seharusnya, nomor urut di puncak diagram diawali dari yang terkecil, berlanjut ke nomor urut yang lebih besar. Artinya, materi atau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



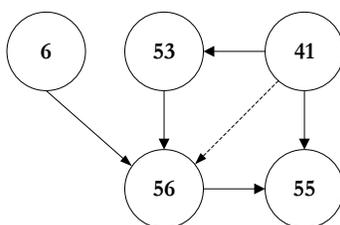
Gambar 2.

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 dikonstruksi, 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 $i = 1, 2, 3, \dots 62$.

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 dikembangkan, 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 dikonstruksi. 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) substansi isi.

Teknik penulisan mencakup: konsistensi dalam menggunakan istilah tertentu, misalnya dalam penulisan lambang atau simbol matematika seperti tanda kurang ($-$) atau kali (\times). 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. Substansi isi mencakup ketepatan suatu soal dalam mengukur indikator tertentu.

g. Hasil Uji Empirik

Hasil uji coba empirik yang dilakukan di SMPN 16 Yogyakarta, menunjukkan 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. Karakteristik 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 *entropy* 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 *entropy*, 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 kontribusi terhadap rendahnya nilai *entropy* 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 *entropy* kelompok item ini.

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

Kelompok Item	No. Item	Nilai <i>Entropy</i>	Kelompok Item	No. Item	Nilai <i>Entropy</i>
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 variabel 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 dikonstruksi 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, pengelompokan 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	
		<i>Guessing</i>	<i>Slip</i>
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 menyelesaikannya. 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, mengonstruksi 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 penelitian 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 penelitiannya.

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 untuk 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.