

MENENTUKAN UKURAN MATRIKS Q PADA MODEL DINA UNTUK DIJADIKAN DASAR MENYUSUN ITEM TES DIAGNOSTIK

Kusaeri¹ & Kumaidi²

¹UIN Sunan Ampel Surabaya, Jl. Jenderal Ahmad Yani 117 Surabaya

²Universitas Muhammadiyah Surakarta, Jl. A. Yani Tromol Pos 1, Pabelan Surakarta
e-mail: kusaeri@uinsby.ac.id

Abstract: Determining Q-Matrix Value of DINA Model for Diagnostic-Test Item Construction. This study measures an efficient Q-matrix value using Monte Carlo simulation. The simulated Q-matrix, obtained from the simulation conducted based on the real data set of the responses of the seventh graders of middle schools, has 3, 4, and 5 attributes. The middle schools include SMPN 1 Panjatan Kulon Progo, SMPN 1 Sanden Bantul, and SMPN 1 Yogyakarta. The information criteria value used in determining the efficient Q-matrix are the Akaike value (AEC) and the Bayesian (BIC). The simulation results in $Q_{4 \times 3}$ matrix for the 3 attributes, $Q_{6 \times 4}$ matrix for the 4 attributes, and $Q_{6 \times 5}$ for the 5 attributes. Those matrixes are recommended in developing diagnostic test items within DINA model framework. In other words, in the development of diagnostic-test items, the Q-matrix should have more attributes than the number of the test items.

Keywords: DINA, Q-matrix, and Monte Carlo

Abstrak: Menentukan Ukuran Matriks Q pada Model Dina untuk Dijadikan Dasar Menyusun Item Tes Diagnostik. Penelitian ini bertujuan mencari ukuran matriks Q yang efisien melalui simulasi Monte Carlo. Matriks Q yang disimulasikan memiliki 3, 4 dan 5 atribut. Simulasi dilakukan berdasarkan pada data riil yang diperoleh dari jawaban siswa kelas VIII SMP, yaitu SMPN 1 Panjatan Kulon Progo, SMPN 1 Sanden Bantul, dan SMPN 1 Yogyakarta. Untuk memilih matriks Q yang efisien digunakan nilai *information criteria* berupa nilai Akaike (AIC) dan Bayesian (BIC). Hasil simulasi menunjukkan: matriks $Q_{4 \times 3}$ (untuk 3 atribut), $Q_{6 \times 4}$ (untuk 4 atribut), dan $Q_{6 \times 5}$ (untuk 5 atribut) merupakan ukuran matriks yang direkomendasikan ketika menyusun item tes diagnostik pada model DINA. Artinya, dalam pengembangan item tes diagnostik bila menggunakan acuan matriks Q maka banyaknya atribut harus lebih dari itemnya.

Kata Kunci: DINA, matriks Q, Monte Carlo

Kegiatan diagnosis kesulitan belajar pada prinsipnya merupakan bagian dari suatu kegiatan pembelajaran (Gierl, *et al.*, 2007). Tujuannya adalah mengidentifikasi permasalahan pembelajaran dan membantu mengatasinya. Untuk itulah diperlukan tes diagnostik yang terintegrasi dengan pembelajaran (Wahyuning-sih, *et al.*, 2013). Tes ini dikembangkan guna membantu guru memahami cara siswa berpikir dan menyelesaikan masalah selama pembelajaran (Hadi, *et al.*, 2015).

Tes diagnostik berbeda dengan tes prestasi (*achievement test*). *Pertama*, tes diagnostik tidak hanya memberikan informasi berupa angka sebagai indikator kemampuan siswa, namun mendeskripsi-

kan penguasaan siswa pada subkemampuan tertentu (Kusaeri, 2012). Jadi, tes diagnostik juga harus mampu membedakan antara kemampuan yang telah dikuasai seorang siswa dan kemampuan yang masih harus dipelajari (Hadi *et al.*, 2015). *Kedua*, tes diagnostik harus mampu memberikan informasi spesifik dari jawaban yang didapatkan dari siswa sehingga dapat diidentifikasi kelemahan atau ketidakkonsistenan pola pikirnya (Close, 2012). Dengan demikian, dapat memberikan potret yang utuh tentang kemampuan siswa (Gorin, 2007).

Alderson (2005) menyebutkan bahwa tes diagnostik seharusnya memiliki enam sifat berikut: (a) soal-soal yang ada dalam tes diagnostik dapat mengukur

pencapaian kompetensi siswa secara mendalam; (b) dapat menampilkan kompetensi yang telah dan belum dikuasai siswa; (c) kompetensi yang belum dikuasai siswa ditunjukkan dengan jelas pada hasil tes; (d) hasil tes mampu mengarahkan siswa untuk mempelajari kompetensi yang perlu dipelajari kembali; (e) hasil tes langsung dapat diketahui siswa setelah siswa selesai melaksanakan tes, dan (f) hasil tes dapat langsung ditindaklanjuti siswa untuk memperbaiki pencapaian kompetensi mereka.

Hal penting lainnya yang perlu dibahas adalah bagaimana bentuk tes diagnostik dikembangkan agar efektif diimplementasikan. Menurut Kato (2009), bentuk pilihan ganda tepat digunakan untuk tes diagnostik bila digunakan untuk skala yang luas. Bentuk item pilihan ganda juga memiliki potensi membuka proses berpikir tingkat tinggi dan dapat memberikan informasi diagnostik bila item itu dikonstruksi secara hati-hati (Osterlind, 1998). Oleh karena itu, perlu dibuat pembeda antar pengecoh pada item pilihan ganda. Pengecoh yang dibuat harus dapat memberikan informasi dan kesimpulan tentang apa yang dikuasai dan belum dikuasai anak (Ciofalo & Wylie, 2006).

Guna mengembangkan tes diagnostik bentuk pilihan ganda, dapat menggunakan acuan matriks Q sebagaimana yang dianjurkan oleh Tatsuoka (2009). Matriks Q memainkan peranan penting dalam pengembangan tes sebagai kisi-kisi (Leighton, *et al.*, 2004). Matriks Q adalah sebuah matriks dengan m baris dan n kolom yang unsur-unsurnya di dalamnya terdiri atas bilangan 0 dan 1. Baris pada matriks Q tersusun atas item dan kolom pada matriks Q terdiri atas sejumlah atribut yang diukur oleh item. Matriks Q akan bernilai 1 jika atribut tertentu diperlukan untuk menjawab sebuah item. Sebaliknya akan bernilai 0.

Menurut Torre (2008), matriks Q perlu disusun secara hati-hati dan perlu mempertimbangkan ukuran atau dimensinya agar diperoleh informasi struktur *latent class* yang maksimum. Penelitian Kusaeri, *et al.* (2012) menunjukkan bahwa matriks Q yang disusun tanpa memperhatikan kaidah yang tepat dapat menyebabkan tidak munculnya beberapa informasi struktur *latent class* yang dihasilkan. Padahal, informasi tersebut sangat diperlukan untuk kegiatan diagnosis berisi penguasaan peserta tes yang dipakai dalam menjawab setiap item.

Berdasarkan kenyataan di atas, perlu dilakukan penelitian guna mencari ukuran matriks Q yang efisien. Matriks Q efisien adalah matriks Q yang mampu memberikan informasi struktur *latent class* yang maksimum. Cara yang tepat untuk menemukannya adalah melalui studi simulasi. Simulasi ini juga sebagai

cara yang lebih efisien dibandingkan dengan penelitian eksperimen sesungguhnya. Model simulasi yang dipilih adalah simulasi Monte Carlo yaitu simulasi yang didesain seperti penelitian eksperimen dengan data dibangkitkan melalui bantuan *software* komputer.

METODE

Penelitian ini merupakan penelitian simulasi, dengan cara menyimulasikan berbagai kondisi ukuran matriks Q dan banyaknya peserta tes. Jenis simulasi yang digunakan adalah simulasi Monte Carlo. Jenis simulasi Monte Carlo dipilih dengan pertimbangan kemampuannya dalam mengestimasi parameter populasi berdasarkan data riil yang diperoleh dari sampel terbatas (Muthen & Muthen, 2010). Data riil diperoleh dari jawaban siswa SMP pada tes diagnostik yang melibatkan siswa SMPN 1 Panjatan Kulon Progo, SMPN 1 Sanden Bantul, dan SMPN 1 Yogyakarta.

Penelitian simulasi ini dilakukan dengan beberapa tahapan. *Pertama*, memilih desain simulasi. Pemilihan desain simulasi dimaksudkan untuk mempermudah pembangkitan data, analisis, dan pemaknaan. Pada tahapan ini, dipilih variabel bebas, variabel terikat, serta banyaknya replikasi. Sebagai variabel bebas adalah ukuran matriks Q, yang dirancang sembilan variasi: $Q_{4 \times 3}$, $Q_{6 \times 3}$, $Q_{8 \times 3}$, $Q_{4 \times 4}$, $Q_{6 \times 4}$, $Q_{8 \times 4}$, $Q_{4 \times 5}$, $Q_{6 \times 5}$, dan $Q_{8 \times 5}$. Diperhatikan pula besarnya sampel, dibuat 250, 5000 dan 1000.

Variabel terikat berupa nilai *information criteria* yakni *Akaike's Information Criteria* (AIC) dan *Bayesian Information Criteria* (BIC). *Akaike's Information Criteria* merupakan suatu statistik yang digunakan untuk menentukan kecocokan (*fit*) relatif dari model statistik yang berbeda. *Bayesian Information Criteria* merupakan suatu statistik yang digunakan untuk menilai kecocokan (*fit*) relatif dari model diagnostik berbeda dalam kerangka estimasi Bayesian. Banyaknya replikasi dilakukan 5 hingga 10 kali untuk setiap kasus, untuk memperkuat hasil estimasi.

Kedua, memilih program komputer. Program komputer ini diperlukan untuk membangun *syntax*, membangkitkan data, dan menganalisisnya. Untuk membangun *syntax* digunakan program *Cognitive Diagnostic Model* (CDM) yang dibuat oleh Templin (2009). Untuk membangkitkan data dan menganalisis digunakan paket software Mplus dari Muthen & Muthen (2010).

Ketiga, memilih uji *fit* guna menentukan ukuran matriks Q yang efisien berupa *information criteria*. Pemilihan ini didasarkan pada pendapat Templin

(2009). Mplus melaporkan nilai *information criteria* berupa nilai *Akaike Information Criteria* (AIC) dan *Bayesian Information Criteria* (BIC). Kedua informasi nilai tersebut, semuanya digunakan untuk memilih matriks Q yang efisien. Nilai paling rendah pada kedua *information criteria* menunjukkan bahwa matriks Q itu paling efisien karena mampu memberikan informasi struktur *latent class* secara maksimum. Sebenarnya menurut Nylund et al. (2009) penggunaan nilai BIC dalam konteks seperti ini lebih tepat dibandingkan dengan AIC. Alasannya (a) hasil AIC sering lebih tinggi (*overestimate*) dibandingkan nilai sebenarnya dan (b) nilai BIC lebih konsisten dibandingkan AIC. Kedua nilai digunakan agar kelemahan *information criteria* yang satu bisa dilengkapi lainnya. Dengan demikian, keputusan yang dibuat berdasarkan informasi tersebut dapat lebih komprehensif.

HASIL DAN PEMBAHASAN

Matriks Q dengan 3 Atribut

Hasil simulasi pada matriks Q dengan 3 atribut, nilai AIC dan BIC semakin besar seiring dengan bertambahnya panjang tes (4, 6 dan 8). Ini berarti, nilai AIC dan BIC terendah terjadi pada tes dengan panjang 4. Tes dengan panjang 4 dapat dikatakan lebih disarankan dibandingkan dengan panjang 6 dan 8 item. Oleh karena itu, bila ingin mengonstruksi item tes diagnostik dengan model *latent class* yang mengacu pada matriks Q dengan 3 atribut, maka tes dengan panjang 4 adalah pilihan yang tepat dibandingkan tes dengan panjang 6 dan 8 item. Hasil selengkapnya pengaruh panjang tes terhadap nilai *information criteria* disajikan pada Tabel 1.

Tabel 1 Pengaruh Panjang Tes dan Besar Sampel terhadap *Information Criteria* (Matriks Q dengan 3 Atribut dan Replikasi Dilakukan 10 Kali)

<i>Information Criteria</i>	Panjang Tes		
	4	6	8
(N = 250)			
Akaike (AIC)	1007.45	1628.74	2208.69
Bayesian (BIC)	1060.27	1695.65	2289.68
(N = 500)			
Akaike (AIC)	1955.89	3249.34	4399.51
Bayesian (BIC)	2019.11	3329.42	4496.44
(N = 1000)			
Akaike (AIC)	3926.85	6514.20	8760.78
Bayesian (BIC)	4000.46	6607.45	8800.61

Dari Tabel 1, kecenderungan naiknya ketiga nilai *information criteria* juga tidak hanya terjadi pada besarnya N tertentu. Namun, berlaku untuk keseluruhan N yang disimulasikan. Sebagai contoh, pada N = 250 didapatkan nilai AIC 1007,45 (panjang tes 4), 1628,74 (panjang tes 6), menjadi 2208,69 (panjang tes 8). Hal yang sama juga terjadi pada nilai BIC.

Temuan lain dari penelitian ini adalah semakin besar sampel semakin bertambah besar pula ketiga nilai *information criteria*. Kondisi demikian terjadi pada tiga kondisi panjang tes yang disimulasikan. Hal ini berarti sampel yang berukuran 250 telah cukup menghasilkan informasi struktur *latent class* yang maksimum dibandingkan dengan sampel berukuran 500 atau 1000.

Hasil ini sejalan dengan dua penelitian sebelumnya, walaupun ada beberapa kondisi yang berbeda dengan penelitian ini. Misalkan, penelitian de la Torre et al. (2010) menunjukkan bahwa ukuran sampel yang relatif kecil (N = 1000) cukup untuk mengestimasi secara akurat parameter-parameter yang ada pada model DINA. Estimasi yang dihasilkan dari sampel berukuran besar (N = 2000 dan N = 4000) lebih bias. Atas dasar ini, mereka menetapkan untuk menggunakan sampel berukuran 1000 dalam mengestimasi parameter.

Sementara itu, Nylund et al. (2007) dalam studi simulasinya membandingkan informasi kriteria AIC dan BIC untuk menetapkan banyaknya kelas dalam analisis *latent class*. Studi tersebut menggunakan ukuran sampel 200, 500, dan 1000 dan menghasilkan temuan bahwa tingkat akurasi AIC menurun seiring dengan bertambahnya ukuran sampel.

Matriks Q dengan 4 Atribut

Ketika simulasi dilakukan pada matriks Q dengan 4 atribut dan cacah (banyaknya) item 4, Mplus tidak menyajikan *out put* apapun seperti lazimnya yang terjadi pada simulasi sebelumnya. Mplus hanya menyajikan kembali *syntax* yang sebelumnya dijadikan input. Artinya, *input* dan *out put* Mplus hampir sama dan hanya ada tambahan informasi pada bagian akhir *out put* Mplus berupa:

```
OUTPUT: Tech8; Tech9;
INPUT READING TERMINATED NORMALLY
Monte-Carlo Simulation for DINA
Model, 4 Attribute;
```

Hal tersebut menunjukkan bahwa untuk matriks Q yang memiliki 4 atribut, tidak disarankan menyusun item dengan cacah item yang bersesuaian dengan atribut tersebut sebanyak 4 item.

Kondisi itu berbeda ketika simulasi dilakukan pada tes dengan panjang 6 dan 8 item. Pada dua simulasi terakhir, selain dimunculkan kembali *syntax* yang menjadi input Mplus juga disajikan informasi lainnya tentang struktur *latent class* secara lengkap. Oleh karena itu, hasil penelitian Rupp & Templin (2008) sejalan dengan hasil simulasi ini. Berdasarkan penelitiannya, mereka membuat rekomendasi bahwa proporsi item yang mengukur atribut dan banyaknya atribut yang diukur oleh suatu item dapat mempengaruhi hasil estimasi. Oleh karena itu, ukuran matriks perlu menjadi perhatian.

Sementara itu, nilai ketiga *information criteria* pada tes dengan panjang 6 hasilnya lebih rendah dibandingkan tes dengan panjang 8. Artinya, matriks $Q_{6 \times 4}$ lebih ideal dibandingkan dengan matriks berukuran $Q_{8 \times 4}$. Pernyataan ini mengisyaratkan bahwa bila ingin mengonstruksi item tes diagnostik dengan model *latent class* yang didasarkan pada matriks Q dengan 4 atribut, maka tes dengan panjang 6 lebih baik hasilnya dibandingkan tes dengan panjang 4 dan 8 item. Hasil selengkapnya perubahan panjang tes terhadap nilai *information criteria* disajikan pada Tabel 2

Tabel 2 Pengaruh Panjang Tes dan Banyak Peserta Tes terhadap Information Criteria (Matriks Q dengan 4 Atribut dan Replikasi Dilakukan 10 Kali)

Information Criteria	Panjang Tes		
	4	6	8
(N = 250)			
Akaike (AIC)	-	1573.30	2128.86
Bayesian (BIC)	-	1668.38	2238.02
(N = 500)			
Akaike (AIC)	-	3113.88	4242.13
Bayesian (BIC)	-	3227.68	4372.78
(N = 1000)			
Akaike (AIC)	-	6220.24	8472.95
Bayesian (BIC)	-	6352.75	8625.10

Di sisi lain, untuk dua kondisi panjang tes yang disimulasikan, semakin banyak peserta tes semakin besar nilai-nilai AIC dan BIC. Artinya, ukuran sampel 250 mampu menghasilkan informasi struktur kelas laten lebih baik dibandingkan dengan sampel berukuran 500 atau 1000.

Temuan lain penelitian ini adalah pada simulasi yang melibatkan matriks Q dengan 4 atribut, waktu yang diperlukan untuk menganalisis (*elapsed time*) lebih lama dibandingkan pada matriks Q dengan 3 atribut. Bila pada simulasi untuk matriks Q dengan 3 atribut hanya membutuhkan waktu rata-rata 21 detik hingga 2 menit, maka pada simulasi

ini membutuhkan waktu lebih dari 1 jam. Semakin panjang tes yang disimulasikan, semakin lama waktu yang diperlukan. Sebagai contoh, pada simulasi menggunakan $N = 250$, waktu yang diperlukan untuk menganalisis 6 item adalah 1 jam 11 detik, sedangkan untuk 8 item diperlukan waktu 2 jam 8 menit 25 detik.

Kondisi seperti ini juga pernah dialami oleh Henson & Templin ketika melakukan kalibrasi data dengan menggunakan model DINA yang melibatkan 40 item, 7 atribut, dan $N = 3000$. Mereka membutuhkan waktu empat hari (Rupp & Templin, 2008). Dalam kegiatan simulasinya, mereka menggunakan laptop PC kapasitas 1,86 GHz dan RAM berukuran 1 GB. Software yang digunakan adalah Mplus yang didukung dengan program CDM.

Pesan yang dapat ditangkap dari pernyataan terakhir adalah analisis model DINA dengan menggunakan perangkat pendukung Mplus membutuhkan waktu lama bila menggunakan atribut yang banyak. Oleh karena itu, penggunaan atribut yang kecil merupakan pilihan yang disarankan.

Matriks Q dengan 5 Atribut

Ketika dilakukan simulasi pada matriks Q dengan 5 atribut dan cacah item 4, kondisi yang sama terjadi seperti pada simulasi pada matriks Q dengan 4 atribut dan cacah item 4. Artinya, Mplus tidak menyajikan *out put* apapun, dan Mplus hanya menyajikan kembali *syntax* yang sebelumnya dijadikan input. Rangkuman hasil simulasi untuk matriks Q yang memuat 5 atribut, disajikan pada Tabel 3.

Tabel 3 Pengaruh Panjang Tes dan Banyak Peserta Tes terhadap Information Criteria (Matriks Q dengan 5 Atribut dan Replikasi Dilakukan 5 Kali)

Information Criteria	Panjang Tes		
	4	6	8
(N = 250)			
Akaike (AIC)	-	1555.53	2071.37
Bayesian (BIC)	-	1706.95	2236.88
(N = 500)			
Akaike (AIC)	-	3013.68	4064.94
Bayesian (BIC)	-	3194.91	4263.03
(N = 1000)			
Akaike (AIC)	-	5978.05	8111.18
Bayesian (BIC)	-	6189.08	8341.85

Berdasarkan Tabel 3, nilai-nilai AIC dan BIC pada tes dengan panjang 6 lebih rendah dibandingkan tes dengan panjang 8. Kondisi itu terjadi pada tiga nilai N yang disimulasikan, yakni pada $N=250$,

$N=500$, dan $N=1000$. Pernyataan itu mengandung makna bila akan mengonstruksi item diagnostik berdasarkan model DINA dengan matriks Q yang memuat 5 atribut, maka tes dengan panjang 6 lebih baik dibandingkan tes dengan panjang 8 atau lebih. Tidak disarankan untuk menggunakan tes dengan panjang 4, sebab hasilnya akan kurang *fit* (cocok) dengan model.

Hasil simulasi mempertegas bahwa untuk memotret tingkat penguasaan siswa terhadap suatu konsep harus digali melalui item tes diagnostik yang dimunculkan secara berulang-ulang. Tidak cukup, seorang siswa menjawab salah dari sebuah item langsung dihakimi yang bersangkutan tidak menguasai konsep tertentu. Hal ini sejalan dengan pendapat Xiabao Li (2006) yang menyatakan bahwa konsepsi siswa tentang materi tertentu bersifat stabil dan kuat. Oleh karena itu, cara melihatnya juga harus dilakukan secara berulang-ulang agar tidak salah dalam kita membuat kesimpulan diagnosis.

Untuk memperkuat uraian di atas, berikut disajikan sebuah contoh cara mengembangkan tes diagnostik dengan acuan matriks Q . Tes dikembangkan pada materi pecahan kelas V Sekolah Dasar. Kemampuan atau atribut yang diukur di antaranya:

- A_1 : mengubah pecahan campuran ke pecahan biasa,
- A_2 : menyamakan penyebut dan membuat pecahan ekuivalen,
- A_3 : menjumlahkan dua pembilang pada pecahan, dan
- A_4 : menyederhanakan/mengubah pecahan ke dalam bentuk paling sederhana.

Andaikan matriks Q yang dikembangkan dan dijadikan kisi-kisi penulisan item tes diagnostik adalah:

$$Q_{6 \times 4} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Dari matriks $Q_{6 \times 4}$ dapat dijelaskan bahwa untuk menyelesaikan item soal nomor 1, siswa membutuhkan penguasaan akan atribut 1, 3, dan 4. Item soal nomor 2 membutuhkan atribut 2, 3, dan 4. Hal yang sama untuk item soal nomor 3, 4, 5, dan 6.

DAFTAR RUJUKAN

Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.

Fakta ini menunjukkan bahwa untuk menghakimi seorang siswa yang tidak menguasai konsep tertentu (misalnya tidak mampu menjumlahkan dua pembilang pada pecahan atau A_3), maka diuji melalui 5 item soal, yakni soal nomor 1, 2, 3, 4, dan 5. Item tes yang dapat dikembangkan dengan mengacu matriks $Q_{6 \times 4}$ di atas adalah:

- (1) $2\frac{8}{6} + 3\frac{10}{6} = \dots$ (2) $\frac{2}{5} + \frac{12}{8} = \dots$ (3) $\frac{8}{5} + \frac{6}{5} = \dots$
- (4) $2\frac{1}{2} + 4\frac{2}{4} = \dots$ (5) $\frac{1}{2} + \frac{1}{7} = \dots$ (6) $\frac{1}{3} + \frac{1}{2} = \dots$
- (7) $\frac{3}{5} + \frac{1}{5} = \dots$ (8) $3\frac{1}{6} + 2\frac{3}{4} = \dots$ dan
- (9) $1\frac{3}{5} + \frac{2}{5} = \dots$

SIMPULAN

Berdasarkan tujuan penelitian dan paparan data sebelumnya, maka dapat diturunkan simpulan berikut. (1) Pada matriks Q dengan 3 atribut, tes dengan cacah item 4 lebih disarankan dibandingkan tes dengan cacah item 6 dan 8. (2) Pada matriks Q yang memiliki 4 atribut, tes dengan cacah item 6 lebih disarankan untuk dikonstruksi dibandingkan tes dengan cacah item 4 dan 8. (3) Pada matriks Q dengan 5 atribut, tes dengan cacah item 6 lebih disarankan dibandingkan tes dengan cacah item 4 dan 8.

Temuan sampingan lainnya dari penelitian ini adalah semakin banyak peserta tes semakin bertambah besar ketiga nilai *information criteria*. Hal tersebut mengandung makna bahwa sampel yang berukuran 250 telah cukup menghasilkan informasi struktur kelas laten yang maksimum dibandingkan dengan sampel berukuran 500 atau 1000.

Berdasarkan temuan penelitian di atas, maka dapat dibuat saran berikut. (1) Matriks Q yang menjadi komponen penting dalam model DINA harus disusun secara hati-hati dan dijadikan dasar dalam proses penyusunan item tes diagnostik. Artinya, matriks Q harus benar-benar dijadikan acuan atau kisi-kisi dalam penyusunan item tes diagnostik. Terkait ukurannya, jumlah item harus lebih banyak dibandingkan dengan komponen atribut. (2) Hasil simulasi menunjukkan bahwa semakin banyak atribut yang digunakan, waktu yang dibutuhkan Mplus untuk analisis juga semakin lama. Oleh karena itu, penggunaan atribut yang kecil (misalnya 4-5 atribut) merupakan pilihan yang disarankan.

Ciofalo, J.F. & Wylie, E.C. 2006. Using Diagnostic Classroom Assessment: One Question at a Time. *Teachers College Record*, January 10, 2006.

- (<http://www.tcrecord.org/content.asp?contentid=12285>). Diakses 6 Januari 2014.
- Close, C.N. 2012. *An exploratory technique for finding the Q-matrix for the DINA Model in Cognitive Diagnostic Assessment: Combining theory with data*. Disertasi, tidak diterbitkan. Minneapolis: The University of The Minnesota.
- Gierl, M.J.; Leighton, J.P.; & Hunka, S.M. 2007. Using the Attribute Hierarchy Methods to Make Diagnostic Inferences about Examinees'. Dalam J.P. Leighton & M.J. Gierl (Eds). *Cognitive Diagnostic Assessment for Education* (hlm. 242-274). New York: Cambridge University Press.
- Gorin, J.S. 2007. Test Construction and Diagnostic Testing. Dalam J.P. Leighton & M.J. Gierl (Eds). *Cognitive Diagnostic Assessment for Education* (hlm. 361-389). New York: Cambridge University Press.
- Hadi, S.; Ismara, K.I.; & Tanumihardja, T. 2015. Pengembangan Sistem Tes Diagnostik Kesulitan Belajar Kompetensi Dasar Kejuruan Siswa SMK. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19 (1), 168-175.
- Kato, K. 2009. *Improving Efficiency of Cognitive Diagnosis by Using Diagnostic Item and Adaptive Testing*. Disertasi, tidak diterbitkan. St. Paul: The University of Minnesota.
- Kusaeri. 2012. Menggunakan model DINA dalam pengembangan tes diagnostik untuk mendeteksi salah konsepsi. *Jurnal Penelitian dan Evaluasi Pendidikan*, 16 (1): 281-306.
- Kusaeri; Suryanto; & Kumaidi. 2012. Estimasi parameter item dan *latent class* dengan Model DINA untuk diagnosis kesulitan belajar. *Jurnal Ilmu Pendidikan*, 18 (2): 187-193.
- Leighton, J. P.; Gierl, M. J.; & Hunka, S. 2004. The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41, 205-236.
- Muthen, L.K. & Muthen, B. O. 2010. *Mplus user's guide (6th ed)*. Los Angeles, CA: Muthen & Muthen.
- Nylund, K.L.; Asparouhov, T; & Muthen, B.O. 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535-569.
- Osterlind, S.J. 1998. *Constructing Test Item: Multiple-choice, Constructed-response, and Other Formats (2nd ed)*. Boston, MA: Kluwer Academic Publishers.
- 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.
- Tatsuoka, K.K. 2009. *Cognitive assessment: An introduction to the Rule Space Methods*. New York: Taylor & Francis Group.
- Templin, J. 2009. *Item parameter interpretation and model fit*. Slide presentasi disampaikan dalam National Council on Measurement in Education, Diagnostic modeling, tanggal 13 April 2009. (http://jonathantemplin.com/files/dcm/dcm09ncme/modelfit_ncme2009.pdf.) Diakses 21 Desember 2014.
- Torre, J. 2008. DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 39 (1): 115-130.
- 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.
- Wahyuningsih; Raharjo, T; & Masithoh, D.F. 2013. Pembuatan Instrumen Tes Diagnostik Fisika SMA Kelas XI. *Jurnal Pendidikan Fisika*, 1 (1), 111-117.
- Xiaobao Li. 2006. *Cognitive Analysis of Student's Error and Misconception in Variables, Equations, and Functions*. Disertasi, tidak diterbitkan. Texas: A & M University.