Graduate
Management
Admission
Council®

The Modified Maximum Global Discrimination Index Method for Cognitive Diagnostic Computerized Adaptive Testing

Ying Cheng and Hua-Hua Chang

Department of Psychology, University of Illinois at Urbana-Champaign

Presented at the CAT and Cognitive Structure Paper Session, June 7, 2007

Abstract

This paper proposes a new item selection method, namely the modified maximum global discrimination index (MMGDI) method for cognitive diagnostic computerized adaptive testing (CD-CAT). The modified index captures two aspects of the appeal of an item: (1) how much contribution it can make toward adequate coverage of every attribute and (2) how much contribution it can make toward recovering the latent cognitive profile. The simulation study demonstrated that the method is capable of ensuring adequate coverage of every attribute measured by the test. Furthermore, compared to the original global discrimination index (GDI) method, it improved the recovery rate of each attribute and of the entire cognitive profile, especially the latter.

Acknowledgment

Presentation of this paper at the 2007 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Citation

Cheng, Y. and Chang, H.-H. (2007). The modified maximum global discrimination index method for cognitive diagnostic computerized adaptive testing. In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from
www.psych.umn.edu/psylabs/CATCentral/

Author Contact

Ying Cheng, Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign IL 61820, U.S.A. ycheng6@uiuc.edu

The Modified Maximum Global Discrimination Index Method for Cognitive Diagnostic Computerized Adaptive Testing

Cognitive diagnosis has received much attention recently, especially since the No Child Left Behind Act (2001) mandated that diagnostic feedback should be provided to students, teachers and parents. Instead of a summative score or several summative subscale scores, a cognitive diagnostic test offers a profile for each examinee, specifying which concepts and skills (often called “attributes” in the cognitive diagnosis literature) the students have mastered and on which areas remedial instruction is needed. A cognitive diagnostic test, therefore, not only serves evaluative purposes, but also offers valuable information regarding each individual examinee's educational needs.

Another flourishing research area in psychological and educational measurement is computerized adaptive testing (CAT). The major feature of CAT is that items are selected sequentially based on examinees' performance on the previous items, and thus each test is tailored to their latent trait levels (the latent trait can be, for example, variables such as general intelligence, math ability, or English proficiency). To be more specific, in CAT after an examinee responds to an item, his or her ability estimate is updated and the next item will be selected to closely match his or her latest ability estimate. Therefore, CAT can provide more efficient estimate of the latent trait of interest (Weiss, 1982).

Researchers have also tried to combine the two above-mentioned research problems and developed cognitive diagnostic computerized adaptive test (CD-CAT) item selection algorithms (e.g., Xu, Chang, & Douglas, 2003; McGlohen, 2004). However, the current literature does not address how to balance the attribute coverage in CD-CAT. To be more specific, it is critical to ensure that each cognitive attribute is measured by an adequate number of items such that accurate diagnostic information can be gathered from the test. This paper proposes an item selection method for CD-CAT, namely the modified maximum global discrimination index (MMGDI) method, which can ensure balanced coverage of every attribute.

The DINA Model

Many cognitive diagnostic models have been proposed over the last three decades, including the rule space model (Tatsuoka, 1983), the binary skills model (Haertel, 1984; Haertel & Wiley, 1993), the Bayesian inference network (Mislevy, Almond, Yan, & Steinberg, 1999), and conjunctive latent class models such as the NIDA model (Maris, 1999), the Fusion model (Hartz, 2002; Hartz, Roussos & Stout, 2002) and the model used in this study, the “Deterministic Input; Noisy ‘And' Gate" (DINA) model (Doignon & Falmagne, 1999; Haertel, 1989; Junker & Sijstma, 2001; Macready & Dayton, 1977).

The main purpose of these models is to relate item responses to a set of latent attributes. An attribute is a task, subtask, cognitive process, or skill involved in answering an item. The purpose of cognitive diagnosis is to identify which attributes are mastered by an examinee and which ones are not. For each examinee, the mastery profile translates into a vector:

$$\alpha_i = \left(\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik}\right)'$$

where $\alpha_{ik} = 1$ indicates that the *i*th examinee masters attribute k and $\alpha_{ik} = 0$ otherwise. *K* is the total number of attributes measured by the test.

Items, on the other hand, are related to attributes by a *Q*-matrix (Tatsuoka, 1995). *Q* is a $J \times K$ matrix, with the entry $q_{jk} = 1$, meaning that a correct response to item *j* requires mastery of attribute *k* and $q_{jk} = 0$ otherwise. The *Q* matrix is usually constructed by content experts and psychometricians.

Let $\underset{\sim}{Y}_i$ denote a vector of dichotomous item responses for the *i*th examinee. It is reasonable to believe that $\underset{\sim}{\alpha}_i$ should be able to account for the pattern of $\underset{\sim}{Y}_i$ to a large extent. However, it is also necessary to consider "slipping" and "guessing" behaviors. Let $\eta_{ij} = 1$ denotes that the *i*th examinee masters all the attributes required by item *j* and $\eta_{ij} = 0$, otherwise:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} \tag{1}$$

When $\eta_{ij} = 1$, the *i*th examinee should be able to answer item *j* correctly, unless he or she "slips." Similarly, when $\eta_{ij} = 0$, the *i*th examinee should not be able to answer item *j* correctly, unless he or she is a lucky guesser.

The DINA model treats slipping and guessing at the item level. The parameter s_j indicates the probability of slipping on the *j*th item when an examinee has mastered all the attributes, and the parameter g_j denotes the probability of correctly answering the *j*th item when an examinee does not master all the required attributes.

The item response function therefore can be written as:

$$P\left(Y_{ij} = 1 \mid \underset{\sim}{\alpha}_i\right) = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}} \tag{2}$$

With the assumption of local independence and independence among all examinees, the joint likelihood function of the DINA model can be written as:

$$L(s, g; \alpha) = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[\left(1 - s_j\right)^{y_{ij}} s_j^{1-y_{ij}}\right]^{\eta_{ij}} \left[g_j^{y_{ij}} \left(1 - g_j\right)^{1-y_{ij}}\right]^{1-\eta_{ij}}$$

(3)

The DINA model requires only two easily interpretable parameters for each item, i.e. s_j and g_j. Both item parameters and examinees' cognitive profiles can be estimated conveniently using MLE. Therefore the DINA model is a very computationally efficient model and a good candidate for building a real-time CAT program.

The MMGDI Method for Item Selection in CD-CAT

The critical component of CAT is the item selection algorithm. Currently many CAT programs use the maximum Fisher information method (Wainer et. al., 2000). Fisher information is the amount of information that an observable random variable *X* carries about an unknown parameter λ upon which the likelihood function of *X*, $L(\lambda) = f(X;\lambda)$, depends. In the context of CAT, *X* is the item response vector of an examinee and the unknown parameter of interest, λ, is his or her ability level θ. Under the three-parameter logistic (3PL) model, the Fisher information of the *j*th item evaluated at the latest θ estimate is given by (Hambleton, Swaminathan, & Rogers, 1991):

$$I_j(\hat{\theta})=\frac{(1-c_j)a_j^{2}e^{a_j(\hat{\theta}-b_j)}}{\left[1+e^{a_j(\hat{\theta}-b_j)}\right]^2\left\{(1-c_j)+c_j\left[1+e^{a_j(\hat{\theta}-b_j)}\right]\right\}} \tag{4}$$

where a_j, b_j and c_j are the discrimination, difficulty and pseudo-guessing parameter for the *j*th item respectively. The item with the largest Fisher information evaluated at $\hat{\theta}$ will be selected as the next item on the test.

Note that Fisher information is additive, so the test information *I* is the sum of the Fisher information of all the administered items. The rationale behind the maximum information method is that as $n \to \infty$, $\mathrm{SE}(\hat{\theta}) \to \sqrt{1/I}$ where $\mathrm{SE}(\hat{\theta})$ is the standard error of the latent trait estimate (Lord, 1980). Therefore the maximum information method yields the best measurement precision asymptotically.

However, Fisher information does not naturally lend itself to cognitive diagnosis because it requires the conditional distribution of *X* given the unknown parameter λ to be continuous and differentiable with respect to λ, whereas the latent structure underlying cognitive diagnosis involves discrete latent classes, which can also be viewed as discretized multidimensional traits. Nevertheless another information measure, the Kullback-Leibler information, is applicable.

The Kullback-Leibler information is a measure of “distance" or “divergence" between two probability distributions *f(x)* and *g(x)* (Cover & Thomas, 1991):

$$d[f,g]=E_f\left\{\log\left[\frac{f(x)}{g(x)}\right]\right\} \tag{5}$$

Note that the Kullback-Leibler information is not strictly a distance measure because it is not symmetric, i.e. *d*[*f,g*] ≠ *d*[*g,f*]. The reason why it is sometimes referred to as Kullback-Leibler distance is that the larger the *d*[*f,g*] is, the easier it is to statistically discriminate the two probability distributions *f(x)* and *g(x)* (Henson & Douglas, 2005).

In cognitive diagnostic assessments, we are interested in the conditional distribution of the *i*th examinee’s response to item *j*, Y_{ij} given $\underset{\sim}{\alpha}_i$. The Kullback-Leibler distance between the distribution of Y_{ij} conditioning on the current estimated latent state/class, i.e. $f(Y_{ij}\mid\hat{\underset{\sim}{\alpha}}_i)$ and the conditional distribution of Y_{ij} given one possible latent state $\underset{\sim}{\alpha}_c$, i.e. $f(Y_{ij}\mid\underset{\sim}{\alpha}_c)$, can be computed as

$$KL_j(\hat{\underset{\sim}{\alpha}}_i \,\|\, \underset{\sim}{\alpha}_c)=\sum_{y=0}^{1}\log\left[\frac{P(Y_{ij}=y\mid\hat{\underset{\sim}{\alpha}}_i)}{P(Y_{ij}=y\mid\underset{\sim}{\alpha}_c)}\right]P(Y_{ij}=y\mid\hat{\underset{\sim}{\alpha}}_i) \tag{6}$$

The quantity $KL_j(\hat{\underset{\sim}{\alpha}}_i \| \underset{\sim}{\alpha}_c)$ indicates how good the *j*th item is in distinguishing $\hat{\underset{\sim}{\alpha}}_i$ from $\underset{\sim}{\alpha}_c$. In another words, $KL_j(\hat{\underset{\sim}{\alpha}}_i \| \underset{\sim}{\alpha}_c)$ represents how discriminating the *j*th item is regarding $\hat{\underset{\sim}{\alpha}}_i$ and $\underset{\sim}{\alpha}_c$.

Note that when there are *K* attributes, there are 2^K possible latent cognitive states, and $\underset{\sim}{\alpha}_c$ is only one of them. Since the true latent state is unknown, Xu, Chang & Douglas (2003) proposed the following global discrimination index (GDI), which is basically the sum of the Kullback-Leibler distances between $f(Y_{ij} | \hat{\underset{\sim}{\alpha}}_i)$ and the conditional distribution of Y_{ij} given each of the 2^K possible latent cognitive states:

$$GDI_j(\hat{\alpha}_i) = \sum_{c=1}^{2^K} \left[\sum_{y=0}^{1} \log \left(\frac{P(Y_{ij} = y | \hat{\underset{\sim}{\alpha}}_i)}{P(Y_{ij} = y | \underset{\sim}{\alpha}_c)} \right) P(Y_{ij} = y | \hat{\underset{\sim}{\alpha}}_i) \right] \qquad (7)$$

An implicit assumption of this index is that all the 2^K states are equally likely to be the true state. Xu, Chang & Douglas (2003) proposed selecting the next item for the *i*th examinee that yields the largest global discrimination index, i.e. $KL_j(\hat{\alpha}_i)$ [There are also other indices proposed on the basis of the GDI; for details, see Henson & Douglas (2005)]. Their study showed that the GDI method recovered examinees' cognitive profiles fairly well. However, the method does not consider attribute coverage.

Most of the traditional content balancing techniques in CAT do not apply to CD-CAT because they often require the content areas to be mutually exclusive. In other words, if one item belongs to content area 1, it cannot be in content area 2. With CD-CAT, however, an item can measure multiple attributes simultaneously. For example, an item can measure both addition and subtraction.

The Modified Maximum Global Discrimination Index

This paper proposes the modified maximum global discrimination index (MMGDI) method for item selection in CD-CAT. The basic idea is to compute an attribute-balancing index and multiply it by the global discrimination index. The attribute-balancing index measures how much contribution the item can make toward fulfilling the attribute-balancing requirement, while the global discrimination index concerns how attractive an item is in terms of psychometric property. The product therefore represents the general attractiveness of an item.

The attribute-balancing index is defined as follows:

$$\prod_{k=1}^{K} \left(\frac{B_k - b_k}{B_k} \right)^{q_{jk}} \qquad (8)$$

where B_k is the minimum number of items required that measures the *k*th attribute, b_k is the number of items measuring the *k*th attribute that are already selected. Then the modified global discrimination index (MGDI) becomes:

$$\mathrm{MGDI}_j(\hat{\alpha}_i) = \prod_{k=1}^{K} \left(\frac{B_k - b_k}{B_k} \right)^{q_{jk}} \times \mathrm{GDI}_j(\hat{\alpha}_i)$$

(9)

The next item to be selected will be the one in the bank that yields the largest MGDI, instead of the largest GDI.

Note that when $q_{jk} = 0$,

$$\left(\frac{B_k - b_k}{B_k}\right)^{q_{jk}} = 1 \qquad (10)$$

and consequently does not affect the MGDI. When $b_k = B_k$, meaning that one attribute is measured by a sufficient number of items, then items in the bank that measure the *k*th attribute will have an attribute-balancing index of 0 and the subsequent modified global discrimination index (MGDI) will be 0. Meanwhile, another item in the bank which does not measure attribute *k* but measures an attribute that is yet not adequately measured will have a positive MGDI. As a result, this item will be more attractive because its inclusion in the test will be contributing more to balancing the attribute coverage.

When all the b_ks reach the B_ks, the requirement of attribute coverage is met. Note that it is possible that fewer than

$$\sum_{k=1}^{K} B_k \qquad (11)$$

items are enough to meet the attribute coverage requirement because one item can measure more than one attribute. Meanwhile, reasonable B_ks should satisfy

$$\sum_{k=1}^{K} B_k \le L, \qquad (12)$$

where *L* is the test length. Therefore with a fixed-length CD-CAT it is not unlikely that the attribute coverage requirement is met before *L* items are chosen. The remaining items can be selected from the item bank using the original global discrimination index, instead of the modified global discrimination index.

In summary, item selection using the MMGDI can be implemented with the following algorithm:

If at least one $B_k > b_k$, meaning the attribute coverage requirement is not met yet, select the next item with the largest modified global discrimination index (MGDI);

Otherwise select the next item with the largest global discrimination index (GDI).

Simulation Design

A simulation study was implemented to examine the effectiveness of the MMGDI method. An item bank of 300 items was simulated, with slipping and guessing parameters generated from a Uniform(0.05, 0.25) distribution. Items measured up to six attributes. Thus, a 300×6 *Q*-matrix was also generated. Assuming independence among the items and independence among the attributes, the Q-matrix was generated entry by entry. One constraint imposed on the Q-matrix generation was that on average an item measured 20% of the attributes. Therefore, a random number was generated from a Uniform(0,1) distribution and compared to 0.2. If the random number was less than 0.2, the corresponding entry data value was 1, otherwise 0. Another constraint was that every item should measure at least one of the six attributes.

A 1000×6 *A* matrix was also generated, with the *i*th row vector, $\underset{\sim}{\alpha}_i$,representing the *i*th examinee's true cognitive state. Assuming independence among examinees and independence among attributes, the matrix was generated entry by entry. Another assumption made about the examinees was that on average one examinee mastered three out of the six attributes. Therefore, a random number was generated from a Uniform(0, 1) distribution and compared to 0.5. If the generated number was smaller than 0.5, the corresponding data value was 1, otherwise 0. Descriptive statistics of the *Q* and *A* matrices are reported in Table 1 and 2.

Table 1. Number of Items (Examinees) Measuring (Mastering) Each Attribute

	Attribute					
	1	2	3	4	5	6
Number of Items	87	74	82	84	77	84
Number of Examinees	530	502	490	473	476	512

Table 2. Number of Items (Examinees) Measuring (Mastering) a Certain Number of Attributes

	Attribute						
	0	1	2	3	4	5	6
Number of Items	0	159	100	35	6	0	0
Number of Examinees	13	106	247	287	234	92	21

Two CD-CATs of 24 items were simulated, one using the original GDI method and the second using MMGDI method. The required minimum number of items measuring each attribute was 4. Note that the initial cognitive state/profile estimate, $\hat{\underset{\sim}{\alpha}}^{(0)}$, was randomly generated with approximately half 0s and half 's. Every time an item was administered, the probability of answering this item correctly was obtained based on the DINA model. Item responses were simulated by comparing this probability with a randomly generated number from a Uniform(0, 1) distribution. Based on the item responses, maximum likelihood estimates (MLE) of the cognitive state/profile were updated. The latest MLE of $\underset{\sim}{\alpha}$ was then used to calculate the GDI or MGDI for the selection of the next item. After L items were administered, the last MLE, i.e. $\hat{\underset{\sim}{\alpha}}^{(L)}$ was the final estimate of the cognitive state. The results were compared in terms of the recovery rate of each attribute, the recovery rate of the entire cognitive state, and the attribute coverage in the resulting tests.

Results

Table 3 compares the recovery rate of each attribute and of the entire profile obtained from the two item selection methods. The recovery rate of each attribute and the entire profile of the MMGDI method was almost uniformly higher than that of the GDI method. While the difference in individual attribute recovery rate was small, the gain of the MMGDI method over the GDI method in terms recovering the entire pattern was substantial: 92.5% of the cognitive profiles were correctly recovered with the MMGDI method, and only 84.8% of them were correctly recovered with the GDI method. This is because recovering the entire profile requires correctly recovering every attribute. Therefore the gain at attribute-level accumulates and makes the overall gain large.

Table 3. Recovery Rate for Each Attribute and the Entire Cognitive State

Item Selection Method	Attribute						
	1	2	3	4	5	6	Entire Profile
GDI	0.975	0.975	0.989	0.985	0.969	0.948	0.848
MMGDI	0.985	0.983	0.988	0.990	0.988	0.988	0.925

The gain can be accredited to a more balanced test in terms of attribute coverage. Table 4 shows the percentage of tests that meet the attribute-coverage requirement, both at the attribute level and at the overall test level. For instance, the first entry in the table is 55.3, meaning 55.3% of the tests of the GDI method met the distributional requirement of the first attribute, i.e. 55.3% of the tests of the GDI method have four or more items measuring the first attribute. Note that the MMGDI method was very effective in balancing the attribute coverage: 100% of its tests met all the attribute coverage requirements, i.e. all its tests had four or more items measuring each of the six attributes. The difference at the overall test level was remarkably large: with the GDI method, only 0.2% of the tests—namely 2 out of the 1,000 tests—had adequate attribute coverage, whereas the MMGDI method ensured that every test had adequate coverage.

Table 4. Percent of Attribute-Balanced Tests

Item Selection Method	Attribute						
	1	2	3	4	5	6	Overall
GDI	55.3	50.3	67.1	74.3	82.5	53.5	0.2
MMGDI	100	100	100	100	100	100	100

Discussion and Conclusions

This paper proposed a new item selection method, namely the modified maximum global discrimination index (MMGDI) method for cognitive diagnostic computerized adaptive testing (CD-CAT). The simulation study showed that this method improved the recovery rate of each attribute and of the entire cognitive profile, especially the latter. Note that meeting content-balancing constraints in traditional CAT usually results in some loss in measurement precision of the underlying latent trait. An interesting question therefore is: why in CD-CAT, did meeting the attribute coverage requirement lead to measurement gain?

The reason is that CD-CAT is essentially multidimensional. The recovery rate of the entire cognitive profile summarizes the measurement precision along all dimensions: when one dimension was not adequately measured, the recovery rate of the entire cognitive profile suffered. On the other hand, traditional CATs are typically considered unidimensional, and the content-balancing constraints are imposed upon the test out of concern for test validity and defensibility (Hambleton, 2005) rather than of measurement precision. Therefore, the presence of the content balancing constraints only impedes the optimization of the psychometric property of a test. This study called our attention to the fact that a CD-CAT cannot be treated as a traditional CAT and balancing attribute coverage is not only important to its validity and defensibility, but also to the psychometric property of a CD-CAT program.

Although the current study demonstrated that the MMGDI method worked very successfully with the CD-CAT, it is limited in several aspects. First, the simulation study was based on an item bank with simulated guessing and slipping parameters and a simulated *Q*-matrix. The reason is that at present it is very difficult to find an item bank with a *Q*-matrix available and the process of identifying the *Q*-matrix for a large number of items can be very time-consuming. However, the results will certainly be more informative if a real item bank can be used. Another important issue is that in this study only fixed-length CD-CAT was considered. Additional future research is indicated on CD-CAT.

References

Cover, T. M. & Thomas. J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons, Inc.

Doignon, J. P., & Falmagne, J. C. (Eds.). (1999). *Knowledge Spaces*. New York: Springer Verlag.

Haertel, E. H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement, 8*, 333-346.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26,* 333-352.

Haertel, E. H., & Wiley, D. E. (1993). Presentations of ability structures: Implications for testing. In N. Frederiksen, R. J. Mislevey, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 359-384). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K. (2005). Preface to *Linear models for optimal test design.* In W. J. van der Linden, W. J. (2005). New York: Springer.

Hambleton, R. K, Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park CA: Sage.

Hartz, S. (2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practice. Unpublished doctoral thesis, University of Illinois at Urbana-Champaign.

Hartz, S., Roussos, L., & Stout, W. (2002). *Skill diagnosis: Theory and practice*. [Computer software user manual for Arpeggio software]. Princeton, NJ: ETS.

Henson, R. & Douglas J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29*, 262-277.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Mahwah, NJ: Erlbaum.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25,* 258-272.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics, 33,* 379-416.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64,* 187-212.

Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K .B. Laskey & H. Prade (Eds.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp. 437-446). San Francisco: Morgan Kaufmann.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20,* 345-354.

Tatsuoka, K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In R. D. Nichols, S .F. Chipman, & R. L. Brennan (Eds.), *Cognitive diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Erlbaum.

U.S. House of Representatives (2001) *Text of No Child Left Behind Act.*

Wainer, H. et. al. (2000). *Computerized adaptive testing: A primer* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 6*, 473-492.

Xu, X., Chang, H., & Douglas, J. (2003). *A simulation study to compare CAT strategies for cognitive diagnosis*. Paper presented at the annual meeting of National Council on Measurement in Education, Montreal, Canada.