Graduate
Management
Admission
Council®

Kullback-Leibler Information in Multidimensional Adaptive Testing: Theory and Application

Chun Wang and Hua-Hua Chang

University of Illinois at Urbana-Champaign

Presented at the Multidimensional CAT Paper Session, June 3, 2009

Abstract

Built on multidimensional item response theory (MIRT), multidimensional adaptive testing (MAT) can, in principle, provide a promising choice to ensuring efficient estimation of each ability dimension in a multidimensional vector. Currently, two item selection procedures have been developed for MAT, one based on Fisher information embedded within a Bayesian framework, and the other powered by Kullback-Leibler (KL) information. It is well-known that in unidimensional IRT that the second derivative of KL information (also termed "global information") is Fisher information evaluated at θ_0. This paper first generalizes the relationship between these two types of information in two ways—the analytical result is given as well as the graphical representation, to enhance interpretation and understanding. Second, a KL information index is constructed for MAT, which represents the integration of KL nformation over all of the ability dimensions. This paper further discusses how this index correlates with the item discrimination parameters. The analytical results would lay foundation for future development of item selection methods in MAT which can help equalize the item exposure rate. Finally, a simulation study is conducted to verify the above results. The connection between the item parameters, item KL information, and item exposure rate is demonstrated for empirical MAT delivered by an item bank calibrated under two-dimensional IRT.

Acknowledgment

Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Copyright © 2009 by the Authors

Citation

Wang, C. and H. Chang (2009). Kullback-Leibler information in multidimensional adaptive testing: theory and application. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

Chun Wang, University of Illinois, Department of Psychology, 603 E. Daniel St., Champaign, IL, 61820, U.S.A., Email: cwang49@illinois.edu; or Hua-Hua Chang, University of Illinois, Department of Psychology, 603 E. Daniel St., Champaign, IL, 61820, U.S.A., Email: hhchang@illinois.edu

Kullback-Leibler Information in Multidimensional Adaptive Testing: Theory and Application

In some standardized achievement tests, it is believed that the test items differentiate levels of multiple traits. For example, abilities concerning both problem solving and computation might be needed to solve an applied mathematical problem. When a test measures a number of attributes, a multidimensional approach can be used to obtain diagnostic information at the subscale level, e.g., the multidimensional compensatory item response theory modeling approach (MIRT). Denote $\vec{\theta}_i = (\theta_{i1}, \ldots\ldots \theta_{iM})^T$ as the ability vector for an examinee where M is the number of subscales or the number of attributes in cognitive diagnosis. In addition to providing one summative score, this approach will provide a fine breakdown of the domain score for each dimension. The advantage of this approach is that a continuous estimate of each dimension, is obtained, as an alternative to using only dichotomous master/non-master results, thereby gaining more information on each subscale for each examinee.

Several methods have been proposed to deliver a multidimensional adaptive test (MCAT). Segall (1996) formulated a Bayesian procedure for θ estimation and adaptive item selection. His method is based on the Fisher information matrix. van der Linden (1999) derived an algorithm that minimizes the asymptotic error variance, when the linear combination of different ability dimensions is of interest. Veldkamp & van der Linden (2002) incorporated various constraints in MCAT through a shadow test approach and used Kullback-Leibler (KL) information as their objective criterion.

Segall's Determinant of Fisher Information

Under the MIRT framework, Fisher information is no longer a scalar but a matrix. Motivated by the expression for the volume of the multivariate normal ellipsoid (Anderson,1984), Segall constructed the following item selection criterion:

Maximizing the determinant of the matrix

$$[I(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_j) + I(\underset{\sim}{\theta}, u_{k'}) + \phi^{-1}], \tag{1}$$

$I(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_j)$ is the Fisher information matrix based on the items administered previously. The $\{r^{th}, s^{th}\}$ element of this matrix is given by $I(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}_j) = -E\left[\frac{\partial^2 InL}{\partial\theta_r \partial\theta_s}\right]$, where $I(\underset{\sim}{\theta}, u_{k'})$ is the item information matrix for the k^{th} item, with diagonal elements

$$I_{rr}(\underset{\sim}{\theta}, u_{k'}) = \frac{\left[\frac{\partial P_i(\underset{\sim}{\theta})}{\partial\theta_r}\right]^2}{P_i(\underset{\sim}{\theta}) Q_i(\underset{\sim}{\theta})}, \tag{2}$$

and off-diagonal elements

$$I_{rs}(\underset{\sim}{\theta}, u_{k'}) = \frac{\dfrac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_r} \times \dfrac{\partial P_i(\underset{\sim}{\theta})}{\partial \theta_s}}{P_i(\underset{\sim}{\theta}) Q_i(\underset{\sim}{\theta})}. \tag{3}$$

If the prior distribution of $\underset{\sim}{\theta}$ is taken into consideration, ϕ^{-1} is the inverse of the covariance matrix of the prior distributions.

This criterion is an analogue to the maximum Fisher information criterion in unidimensional CAT; by minimizing the determinant, we can narrow down the confidence ellipsoid in a high dimensional θ space, and therefore, estimation accuracy can be enhanced. However, the drawback of this method is that the contribution of each item cannot be independently determined. Despite this disadvantage, the method has been reported to be efficient. It is also of theoretical interest to include this method in our study, and compare this Fisher information based method to the KL information based methods.

The KL Information Index

The application of KL information in the context of CAT was first introduced by Chang and Ying (1996). Their study indicated that KL not only outperformed Fisher information at the beginning of a CAT, but also it preserved all the good features of a CAT. Moreover, KL information can be obtained directly from item response functions, whether continuous or not, whereas Fisher information can be obtained only from continuous response functions whose second derivative can be taken.

Veldkamp and van der Linden (2002) proposed a KL information measure for MCAT, where the KL information index is presented as

$$K_i(\hat{\underset{\sim}{\theta}}^{\hat{k}-1}) = \int_{\underset{\sim}{\theta}} \ldots\ldots \int K_i(\underset{\sim}{\theta}, \underset{\sim}{\theta}^{k-1}) \partial \underset{\sim}{\theta} = \int_{\theta_1 - \frac{3}{\sqrt{n}}}^{\theta_1 + \frac{3}{\sqrt{n}}} \ldots\ldots\ldots \int_{\theta_m - \frac{3}{\sqrt{n}}}^{\theta_m + \frac{3}{\sqrt{n}}} K_i(\underset{\sim}{\theta}, \underset{\sim}{\theta}^{k-1}) \partial \underset{\sim}{\theta}. \tag{4}$$

The integrand $K_i(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}^{k-1})$ is the KL item information, measuring the distance between two probabilities over the same parameter space (Lehmann & Casella, 1998), and it is defined as

$$K_i(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}^{k-1}) = p(\hat{\underset{\sim}{\theta}}^{k-1}) \log\left(\frac{p(\hat{\underset{\sim}{\theta}}^{k-1})}{p(\underset{\sim}{\theta})}\right) + \left[1 - p(\hat{\underset{\sim}{\theta}}^{k-1})\right] \log\left[\frac{1 - p(\hat{\underset{\sim}{\theta}}^{k-1})}{1 - p(\underset{\sim}{\theta})}\right]. \tag{5}$$

$\hat{\underset{\sim}{\theta}}^{k-1}$ is the current update of ability after administering *(k* $-$ 1) items. Following the assumption of local independence for a test of *n* items, the KL test information is equal to the summation of item information,

$$K_n(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}^{\hat{k}-1}) = \sum_{i=1}^{n} K_i(\underset{\sim}{\theta}, \underset{\sim}{\theta}^{k-1}). \tag{6}$$

Therefore, KL information maintains the additivity in the multidimensional case, and the information carried by each item is independently identified, which is another advantage over Fisher information.

Analytical Derivations

Relationship Between Fisher Information and KL Information

Since both Fisher information (FI) and Kullback-Leibler information (KLI) are useful information measures, it is often of theoretical interest to explore the connections between them. We limit our exploration to the two-dimensional case with the 2PL model to facilitate derivations, but the results can be generalized to higher dimensions.

In unidimensional IRT, Chang & Ying (1996) showed that FI at θ_0 equals the second derivative of KLI evaluated at the same true value, θ_0, which is expressed as

$$\frac{\partial^2}{\partial\theta^2} KL(\theta\|\theta_0)\Big|_{\theta=\theta_0} = I(\theta_0)\,. \tag{7}$$

For any given θ, KLI represents the ease or difficulty of distinguishing θ from θ_0. In particular, for θ varying around θ_0, KLI reduces to FI. Due to the reason that FI can be fully recovered from KLI, Chang & Ying (1996) termed KLI as "global information" and FI as "local information". Geometrically speaking, if KLI is viewed as a curve on the plane, FI becomes the curvature of the curve at $\theta = \theta_0$.

In two dimensions, FI extends to a matrix instead of a scalar. The $\{r^{th}, s^{th}\}$ element of this matrix is given by $I(\underset{\sim}{\theta}, u_k) = -E\left[\frac{\partial^2 InL}{\partial\theta_r \partial\theta_s}\right]$, where $I(\underset{\sim}{\theta}, u_k)$ is the item information matrix for the k^{th} item, with each element denoted by either Equation 2 or Equation 3. KLI, on the other hand, stays the same and it is expressed as:

$$KL_j(\underset{\sim}{\theta}\|\underset{\sim}{\theta}_0) = \sum_{y=0}^{1} \log\left[\frac{P(u_k|\underset{\sim}{\theta})}{P(u_k|\underset{\sim}{\theta}_0)}\right] P(u_k|\underset{\sim}{\theta}_0)\,. \tag{8}$$

It can be shown that each entry of FI matrix $I(\underset{\sim}{\theta}, u_k)$ can be obtained by taking second derivatives of $KL_j(\underset{\sim}{\theta}\|\underset{\sim}{\theta}_0)$. In particular, the $\{r^{th}, s^{th}\}$ element of $I(\underset{\sim}{\theta}_0)$ is obtained through $\frac{\partial^2 KL_j(\underset{\sim}{\theta}\|\underset{\sim}{\theta}_0)}{\partial\theta_r \theta_s}\Big|_{\theta=\theta_0} = I_{rs}(\underset{\sim}{\theta}_0)$.

Therefore, whenever KLI is available, the Fisher information matrix is determined. But KLI cannot be recovered from FI, so the terms global and local information maintain their meaning in two-dimensions.

The KLI Index and Item Discrimination

The multidimensional KLI (KI) proposed by Veldkamp and van der Linden (2002) is reduced to the following form in the two dimensional case:

$$KI(\hat{\underset{\sim}{\theta}}^{\hat{k-1}}) = \iint_D KL_i(\underset{\sim}{\theta}, \underset{\sim}{\theta}^{k-1}) d\theta_1 d\theta_2 = \int_{\theta_{10}-\frac{3}{\sqrt{n}}}^{\theta_{10}+\frac{3}{\sqrt{n}}} \int_{\theta_{20}-\frac{3}{\sqrt{n}}}^{\theta_{20}+\frac{3}{\sqrt{n}}} KL_i(\underset{\sim}{\theta}, \underset{\sim}{\theta}^{k-1}) d\theta_1 d\theta_2 \tag{9}$$

Clearly, they assumed that the integration domain is a square centered at $(\theta_{10}, \theta_{20})$ with side length of $6/\sqrt{n}$, and the two dimensions take equal priority in item selection. However, this integration domain can be adjusted based upon the test requirement, which reflects the potential flexibility of this method. Denote the integration domain as *D*, which is central symmetric with center $(\theta_{10}, \theta_{20})$ (in a CAT, this is the updated point estimate), and we can consider two different cases:

(a) Circular domain $D = \left\{ (\theta_1, \theta_2) \middle| \theta_1^2 + \theta_2^2 \le r^2 \right\}$,

and

(b) Elliptic domain $D = \left\{ (\theta_1, \theta_2) \middle| \frac{\theta_1^2}{r_1^2} + \frac{\theta_2^2}{r_2^2} \le 1 \right\}$.

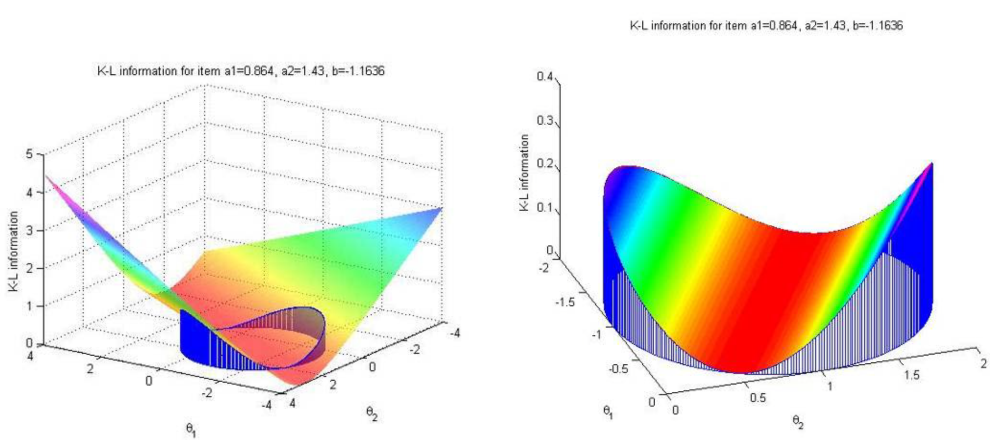
The first case presumes that the both dimensions are equally important in a test, while the second assumes that the two dimensions are weighted differently.

For ease of interpretation and derivation, we adopted the circular domain and transform the original rectangular coordinates to the Cylindrical coordinate, i.e.,

$$\begin{cases} \theta_1 = \theta_{10} + r\cos\alpha \\ \theta_2 = \theta_{20} + r\sin\alpha \end{cases}. \tag{10}$$

In terms of graphical representation, the KI is actually the volume under the information surface, bounded by a cylinder as shown in the following two figures.

Figure 1. Graphical Representation of KI as a Volume



In order to show what kinds of items are favored by KI, we explore the relationship between the magnitudes of item KI with its discrimination parameters. Denoting the item KI as $KI(\underset{\sim}{\theta}_0) = \iint_D KL_j(\underset{\sim}{\theta} \| \underset{\sim}{\theta}_0) \partial \underset{\sim}{\theta}$, where D is the central symmetric domain centered around $\underset{\sim}{\theta}_0$, we found that in the two-dimensional case, the size of $KI(\underset{\sim}{\theta}_0)$ is proportional to the function of item discrimination parameters $f(a_1, a_2)$ when the area of D approximates to zero. The form of the function depends on the shape of the integration domain. In particular, $f(a_1, a_2) = a_1^2 + a_2^2$ when D takes form (a) and $f(\underset{\sim}{a}) = (a_1 r_1)^2 + (a_2 r_2)^2$ when D takes form (b).

Reckase & McKinley (1991) first generalized item discrimination to the multidimensional case, and defined m*ultidimensional discrimination* as:

$$\text{MDISC}=\left(\sum_{k=1}^{m} a_{ik}^2\right)^{1/2}. \tag{11}$$

MDISC is an overall measure of the capability of an item to distinguish between individuals that are in different locations in the θ space. Intuitively, items with a high value of MDISC will provide a large amount of information somewhere in the θ space. Thus, it is reasonable that the item's KI relies on its MDISC, and MDISC can be regarded as an analogue of the unidimensional item discrimination parameter. This result will help us further develop some item selection rules to balance exposure rates.

Variants of KI

Sometimes, the different dimensions measured by a test are not equally important. For example, an applied math exam usually measure both math ability and reading ability; however the math ability is more important in this case than reading ability. van der Linden (1999) derived an algorithm that minimizes the asymptotic error variance, when the linear combination of different θ dimensions is of interest. In his paper, he considered the case when two dimensions are treated differently. In our study, we believed that KI is flexible enough to address this issue; the only modification needed is to change the shape of the integration domain D of the KI. For example, we can use form (b) as the integration domain.

Simulation

The focus of the paper was to explore the relationship between FI and KI in the multidimensional case. Segall's method is a representative of FI while KI is based on KL information. Therefore, both of these methods were implemented in our simulation. A randomization method was also included as a baseline.

Examinee Generation

The true θ vectors were generated from a multivariate normal distribution with mean of zero, and the correlation between two dimensions was 0.5. The examinee sample size was set to 1,000.

Item Bank Construction

The item bank ws constructed following the two-parameter multidimensional IRT model. The two discrimination parameters were generated from a log-normal distribution, bounded within 0.5 to 3.0. The difficulty parameters were generated from standard normal distribution. The item bank size was set at 900.

Generation of Item Scores in CAT and Estimation of θ

During the CAT procedure, given each examinee's true θ vector, scores for each administered item were generated based on the two-dimensional IRT model. After obtaining the probability of a correct response $P(Y_{ij}=1|\underset{\sim}{\theta}_i)$, a random $U(0,1)$ variable u was generated, and the score Y_{ij} was defined as:

$$Y_{ij}=\begin{cases}1 & \text{If u} \le \text{P}(Y_{ij}=1|\underset{\sim}{\theta}) \\ 0 & \text{otherwise}\end{cases}. \tag{12}$$

Note that once an item was selected, the $\hat{\underset{\sim}{\theta}}$ was updated for choosing the next item. Bayesian expected a priori (EAP) estimation was used to estimate θ.

Results

Estimation Accuracy and Exposure Rate Balance

The item selection rule is a key part of a CAT. The evaluation of the various item selection rules is usually based upon two criteria: estimation accuracy and exposure rate balance.

Eestimation accuracy is usually quantified by mean squared error (MSE) and bias. MSE (Equation 13) captures the average squared discrepancy between estimated θ and true θ, thus it should be as small as possible. Bias (Equation 14), on the other hand, represents the average discrepancy between the estimated θ and true θ and it should also be as close to zero as possible. For the two distinct θ dimensions in this simulation study, both MSE and Bias were calculated at each dimension level.

$$MSE=\frac{1}{m}\sum_{i=1}^{m}(\theta_{ik}-\hat{\theta}_{ik})^2 , \tag{13}$$

and

$$. Bias=\frac{1}{m}\sum_{i=1}^{m}(\theta_{ik}-\hat{\theta}_{ik}) , \tag{14}$$

where m is the number of simulated examinees, θ_{ik} is the true ability for the i^{th} examinee and $\hat{\theta}_{ik}$ is the EAP estimator. K is the dimension, it was either 1 or 2.

Exposure rate balance is usually measured by a chi-square index defined as (Chang & Ying, 1999)

$$\chi^2=\sum_{j=1}^{N}(er_j-\overline{er_j})^2/\overline{er_j} \tag{15}$$

where er_j is the exposure rate of item j, and $\overline{er_j}=L/N$ is the desirable uniform rate for all items. This index represents the discrepancy between the observed and the ideal item exposure rates. The exposure balance not only indicates the efficiency of item bank usage, but also provides information concerning test security. The smaller the value, the more efficiently the item bank is used, and the more secure the test is. Usually, the objectives of exposure rate balance (or test security) and estimation accuracy cannot be maximized simultaneously, yielding an accuracy-security tradeoff. The results of the above three methods were as follows:

Table 1. Results of Different Methods

Method	MSE		Bias		Chi-square
	θ_1	θ_2	θ_1	θ_2	
KI (equal weight)	0.169	0.151	0.0013	0.0019	189.799
Segall's method	0.111	0.111	0.0009	0.0026	208.581
Random	0.262	0.271	-0.0061	-0.0080	0.3274

Table 1 shows that both Segall's method and KI performed much better than the random item selection methods, and both of them led to skewed item exposure rate distributions.

Item Exposure and Item Discriminations

We further checked the correlation between the item exposure rate and item discrimination parameters, and the results are shown in the Table 2.

Table 2. Correlation Between Exposure Rates and Item Parameters

Method	a_1	a_2	$(a_1^2 + a_2^2)$	b
K-L information	0.4061	0.4668	0.7139	-0.0425
Segall's method	0.2832	0.2550	0.5022	-0.0325
Random	-0.0181	-0.0265	-0.0461	0.0189

Clearly, KI relies heavily on the square of "multidimensional discrimination." This high correlation partly explains the skewness of the exposure rate distribution. As expected, there was no relationship between item exposure and the item difficulty parameter. Figure 1 shows the trend of the relationship between item exposure and MDISC.

Figure 1. Correlation Between Item Parameters and Item Exposure as Test Proceeds

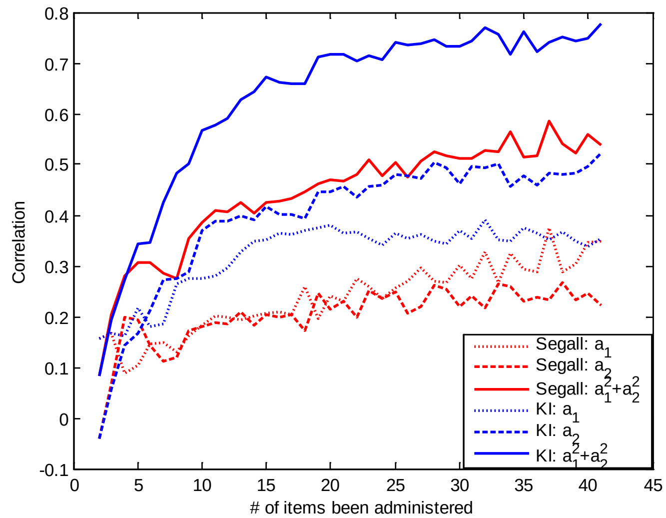


Figure 1 shows that item exposure obtained from both Segall's method and the KI method showed consistently higher correlations with MDISC than with the individual discrimination parameters. The figure also shows that the correlation increased when the number of items administered increases, but the correlation almost reached its highest point when the test length was 25 items. This is similar to the unidimensional case in which item exposure depends on the item discrimination parameter.

Variants of the KL information Index

When we changed the integration domain from a circle to a ellipse, the results changed somewhat, as shown in Table 3.

Table 3. The Performance of Each Integration Domain

Dimension Weights	MSE		Bias		Chi-square
	θ_1	θ_2	θ_1	θ_2	
r1 = r2 =3	0.169	0.151	0.00013	0.0019	189.7996
r1 = 4, r2 = 3	0.137	0.187	0.0009	0.0018	171.4478
r2 = 4, r1 = 3	0.213	0.121	0.0011	0.00023	175.1294

Table 3 shows that if one dimension is weighted higher than the other, the highly weighted dimension will lead to smaller MSE. Therefore, the practitioner can adjust the integration domain for their different needs. The correlation results are shown in Table 4.

Table 4. Correlation Between Exposure Rates and Item Parameters

Weights	a_1	a_2	$(a_1^2 + a_2^2)$	$(r_1^2 a_1^2 + r_2^2 a_2^2)$
w1 = w2 = 3	0.41886	0.46215	0.71786	0.71786
w1 = 4, w2 = 3	0.53729	0.20319	0.61048	0.6747
w2 = 4, w1 = 3	0.13819	0.5772	0.61854	0.67496

Similarly, if θ_1 is regarded as more important, the item selection will rely more on a_1. Also, the highest correlation was between exposure rate and the weighted square of the two item discriminations, not the unweighted MDISC.

Conclusions

This study explored the relationship between two information measure, Fisher information and Kullback-Leibler information, in the multidimensional case. Derivations showed that KL information maintains its global information feature since the whole FI matrix can be fully recovered from it. In addition, we investigated the characteristic of KI. Both analytical and simulation results showed that KI depends on the function of MDISC. Since MDISC functions similarly to the item discrimination parameter in unidimensional CAT, several methods for controlling exposure rate in unidimensional CAT can be easily generalized to multidimensional adaptive tests. This study further provided several variants of KI, which can be used in the situation where each dimension is treated differently. However, since the correlation between item exposure obtained from Segall's method and MDISC is not extremely high, it opens for further investigation the question of why the item exposure rate is skewed in Segall's method.

References

Ackerman, A.T. (1991). Creating a test information profile for a two-dimensional latent space. *Applied Psychological Measurement*, *18,* 257-275

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement, 23,* 223-238.

Chang, H. (2007) Book review: Linear models for optimal test design. *Psychometrika*, *72*, 279-281.

Chang. H. Qian. J & Ying. Z (2001). *a*-stratified multistage computerized adaptive testing with *b* blocking. *Applied Psychological Measurement 25*, 333-341

Chang, H., & van der Linden, W. J. (2003). Optimal stratification of item pools in *a*-stratified adaptive testing. *Applied Psychological Measurement*, *27,* 262–274.

Chang, H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20, 213–229.

Chang, H., & Ying, Z. (1999). *a*-stratified multistage computerized adaptive testing. *Applied Psychological Measurement, 23,* 211–222.

Chang, H., & Ying, Z. (2008). To weight or not to weight? Balancing influence of initial items in adaptive testing. *Psychometrika, 73,* 441-450.

Cheng, Y., & Chang, H (2009). The constraints-weighted information method for severely constrained item selection in CAT. *British Journal of Mathematical and Statistical Psychology, 62,* 369-383.

Cheng, Y., & Chang, H. (2006, April). *Constraints-weighted information method for item selection of severely constrained computerized adaptive testing*. Paper presented at the 2006 annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Cheng, Y., Chang, H. & Budescu, D. (2007, April). *The dual information method for item selection in cognitive diagnosis CAT.* Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Hau, K. & Chang, H.-H. (2001). Item selection in computerized adaptive testing: Should more discriminating items be used first? *Journal of Educational Measurement, 38,* 249-266.

Reckase, M.D., & McKinely, R.L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, *15,* 361-373

Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation & Policy Analysis, 23*, 197-227.

Segall, D.O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 331-354.

McGlohen.M & Chang. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods 40,* 808-821.

van der Liden, W.J., & Chang. H. (2003). Implementing content constraints in alpha-stratified adaptive testing using a shadow test approach. *Applied Psychological Measurement*, *27,* 107-120

van der Liden, W.J., & Reese. L.M., (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement*, *22,* 259-270.

van der Linden, W. J. (2000). Constrained adaptive testing with shadow tests. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 27-52). Boston, MA: Kluwer.

van der Linden, W. J. (2007, June). *A universal model for adaptive testing: CAT with shadow tests.* 2007 GMAC Conference on Computerized Adaptive Testing, Minneapolis, MN (Keynote address)

van der Linden, W. J. & Veldkamp. B.P. (2003, April) . *Constraining item exposure in computerized adaptive testing with shadow tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

van der Linden, W. J., & Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, *29,* 273-291.

van der Linden, W. J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of Educational and Behavioral Statistics, 24,* 398-412.

Veldkamp. B.P. & van der Linden, W. J. (2002) Multidimensional adaptive testing with constraints on test content. *Psychometrika*, *67,* 575-588.

Van der Linden, W.J., Ariel, A. & Veldkamp, B.P. (2006) Assembling a computerized adaptive testing item pool as a set of linear tests. *Journal of Educational and Behavioral Statistics, 31,* 81-99

Wainer, H., Flaugher, R. Green, B. F., Mislevy, R, J, Steinberg, L., & Thissen,D. (1990). *Computerized adaptive testing: A primer.* Hillsdale, NJ, Lawrence Erlbaum Associates.

Xu, X., Chang, H., & Douglas, J. (2005). *Computerized adaptive testing strategies for cognitive diagnosis.* Paper presented at the annual meeting of National Council on Measurement in Education, Montreal, Canada.

Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory* [Computer software]. Monterey, CA: CTB/McGraw-Hill.

Yao, L. (2004, August). *Bayesian multivariate item response theory and BMIRT software.* Paper presented at the Joint Statistical Meetings, Toronto, Canada.