of concern, examinees with similar abilities are most likely to share information about the contents and items of a test. Consequently, it requires exposure control conditional on ability. They presented a method to control exposure rate conditional upon examinee ability level (i.e., Stocking-Lewis [SL]). This method is applicable for unidimensionalCAT procedures. Later, for multidimensional cases, Finkelman et al. (2009) introduced a generalized version of SL method known as generalized Stocking-Lewis (GSL) method.
The SL method sets the exposure control boundary along a set of U discrete θ-levels, $\theta_{1,}...,\theta_{U,}$ which approximately satisfies the boundary for all ability levels. Notice that the constraint is not implemented over all ability levels, it rather is implemented over U meaningfully selected values of θ. Therefore, this method establishes the relation below for all j and all $u = 1,...,U,$

$$\pi_{j,\boldsymbol{\theta}_U}(A) \leq r \tag{16}$$

Inequality16 is used as surrogate for the desired relation of

$$\pi_{j,\boldsymbol{\theta}}(A) \leq r. \tag{17}$$

The computed proportion of selection and administration for each θ_u are then denoted as $P_{j,\theta u}(S)$ and $P_{j,\theta u}(A)$, respectively. When item j is selected as a candidate item, the item exposure control parameter $P_{j,\theta u}(A|S)$ that corresponds to θ_u is used as it is closest to the current theta estimate.

In multidimensional cases, direct extension of SL method would require *inequality 16* for all j and all u over an D-dimensional grid where ability is represented by a vector $\boldsymbol{\theta}$ $(\theta_1,..., \theta_D)$. Finkelman et al. (2009) discussed that because the number of $\boldsymbol{\theta}$ values requiring *inequality 16* exponentially increases as the number of dimensions increases, complete crossing of discrete values in the grid is intractable even when D is moderate. The GSL method proposed by Finkelman et al. (2009), ideally, maintains a good number of quadrature points regardless of D. Instead of using *inequality16*, this method performs an exposure control conditional on θ^*, where θ^* is considered to be a function of $\boldsymbol{\theta}$. However θ^* is a scalar such that $\theta^* = \lambda' \boldsymbol{\theta}$ where λ is a set of weights. Therefore, item exposure control is conditional on a linear combination of D dimensions of $\boldsymbol{\theta}$. Then the established relationship between probability of administering item j at θ^* and desired exposure rate is

$$\pi_{j,\boldsymbol{\theta}^*}(A) \leq r, \tag{18}$$

for all j and all values θ^*. All operational steps are same for SL and GSL methods, the only difference is that the conditioning variable in GSL becomes θ^* rather than θ.

For item exposure control in the context of cognitive diagnosis, Wang, Chang, and Huebner (2011) proposed a restrictive stochastic item selection method, which is a modified version of the *progressive method* (PM; Revuelta, 1995) for traditional CAT procedures. The PM method weights items based on an information component with a random part, where as the test progresses the relative impact of information increases. Wang et al. (2011) modified the PM by adding a stochastic component such that it would not always select the item producing the highest information at the current stage. The restrictive progressive (RP) method that employs PWKL information is defined as

$$RP-PWKL_j = \left(1-\frac{exp_j}{r}\right)\left[\left(1-\frac{m-1}{n}\right)R_j + \frac{PWKL_j\beta_{(m-1)}}{n}\right], \tag{19}$$

where exp_j is the preliminary exposure rate of item j, r is the pre-defined exposure control rate, $m-1$ is the number of administered items, n is the test length, R_j is a random value that is drawn from a uniform distribution between 0 and $PWKL_j$ for the items in the pool, and β is an arbitrary number to control the balance between the test security and estimation accuracy. Smaller β tends to produce more secure test, however, tests with small β yield less accurate estimation.

Kaplan (2016) has recently incorporated the RP method for exposure control with the GDI and MPWKL item selection rules for CD-CAT application. The notation below is the RP method representation where Δ_j indicates the information on item j (e.g., $MPWKL_j$ and ζ_j)

$$RP-\Delta_j = \left(1-\frac{exp_j}{r}\right)\left[\left(1-\frac{m-1}{n}\right)R_j + \frac{\Delta_j\beta_{(m-1)}}{n}\right], \tag{20}$$

Notice that the current form of RP method is applicable to fixed length (i.e., n) tests. In his study, Kaplan (2016) modified it such that minimum of the maximum of the posterior distribution is used as the test termination rule. Then, the modified item selection index incorporating the RP method is

$$RP-\Delta_j=\left(1-\frac{exp_j}{r}\right)\left[f(x)R_j+\Delta_j\beta\frac{\pi(\hat{\boldsymbol{\alpha}}_l\mid \boldsymbol{X}_j)}{P}\right], \quad (21)$$

Where $f(x)=\min\left(0,1-\frac{\pi(\hat{\boldsymbol{\alpha}}_l\mid \boldsymbol{X}_j)}{P}\right)$ and P is the predetermined minimax value.

Stopping Rule

A rule determining when to stop administering items adaptively is referred to as *stopping rule*. Stopping rules can be set for fixed- and variable-length tests (Reckase, 2009; Frey & Seitz, 2009). A stopping rule for a fixed-length test sets a predetermined number for item administration. Once the test reaches this pre-specified length, it is terminated and final ability estimate is computed. A variable-length test is terminated based on a pre-specified standard error of the ability estimate. In other words, stopping rule in a variable-length test becomes a statistical criterion on the measurement precision. In variable-length tests, the number of items to be administered depends on the location of the examinee in the latent ability-space, consistency of examinee responses, and the information provided by the item pool relative to the ability level (Reckase, 2009) or his/her attribute profile in CDM cases.

Reckase (2009) argued that test length could be determined based on some practical considerations such as testing time, or by taking the average of the variable length tests. Wang, Chang, and Boughton (2012) argued that the literature on MAT focuses on the stopping rules for fixed-length tests, which provide less accurate ability estimates for examinees whose ability locations are substantially different than the average difficulty level of the item bank.

Despite the fact that fixed- and variable-length stopping rules were well explored in the unidimensional CAT, Wang et al. (2012) noted that because the precision of multiple ability dimensions should be considered simultaneously, these well-defined stopping rules cannot straightforwardly be generalized to multidimensional adaptive testing situations. They further discussed that in order to set a stopping criterion for MAT, firstly, an index such as generalized variance, total variance, or entropy should be set to quantify the estimation accuracy of θ-vector. Moreover, Kaplan and de la Torre (in press) is among the limited variable-length CD-CAT studies where they use the minimum of the maximum of the posterior distribution as the test termination rule.

DISCUSSION

This paper intended to compile and highlight the recent developments in CAT procedures by reviewing the generalization and/or modification of traditional CAT components for MAT and CD-CAT applications. This paper also intended to provide researchers and practitioners with pragmatic information such that they can use this information toward their own research and application purposes.

When traditional CAT is intended, unidimensional items need to be written and calibrated in accordance with unidimensional IRT models. Similarly, item development and item calibration need to be in accordance with the MIRT models when MAT is intended. Item pool development for CD-CAT can be much more challenging, as CDM applications require a Q-matrix specifying the item-by-attribute associations. Construction of a Q-matrix requires collaboration among measurement experts and content experts. The misspecification of the Q-matrix can reduce the credibility of CD-CAT applications.

Generalization of traditional CAT procedures for MAT is challenging because the new procedures needs to be tractable and computationally manageable as the number of dimensions increases. Further, due to the discrete nature of CDMs, not all conventional item selection rules and exposure control rates can be modified for CD-CAT implementations. For example, item selection algorithms based on Fisher information cannot be applied in the context of cognitive diagnosis (Xu, Chang, & Douglas, 2003). Alternatively, the MPWKL and GDI can be used as item selection algorithms in CD-CAT.

Another vital practical consideration in adaptive testing is item exposure rates. It should be noticed that constraints on item exposure comes at a price because item selection algorithms can no longer use the most informative items in every step. As discussed by Finkelman, Nering, and Roussos (2009), employment of item exposure control methods on the item selection algorithms results in reduction in estimation accuracy. In other words, there is a trade-off between item exposure control and measurement precision. There is not much research conducted for item exposure control in CD-CAT applications. Wang, Chang, and Huebner (2011) proposed a restrictive progressive item selection method, which is a modified version of PM (Revuelta, 1995; Revuelta & Ponsoda, 1998) for traditional CAT applications. Although the RP was proposed for and used with fixed-length tests, it was recently modified for variable-length tests.

Another practical consideration in adaptive testing is the starting point (i.e., with which item to start), for which there is not enough research for CD-CAT. Similarly, impact of type of attribute profile estimation (i.e., MLE, MAP, and EAP) may impact item selection and consequently exposure rates differently. Impact of ability estimation methods and the prior distribution on CD-CAT can be further research topics.

REFERENCES

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

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74,* 619-632.

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

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guilford Press.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34,* 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76,* 179-199.

de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicologia Educative, 20,* 89-97.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39(1),* 1-38.

Finkelman, M., Nering, M. L., & Roussos, L. A. (2009). A conditional exposure control method for multidimensional adaptive testing. *Journal of Educational Measurement*, *46*, 84-103.

Flaugher, R. (2000). Item pools. In H. Wainer (Ed.),*Computerized adaptive testing: A primer* (2nd ed., pp. 37-59). Mahwah, NJ: Lawrance Erlbaum Associates, Inc, Publishers.

Frey, A., & Seitz, N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, *35*, 89-94.

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.

Kaplan, M. (2016). *New item selection and test administration procedures for cognitive diagnosis computerized adaptive testing*. Unpublished doctoral dissertation, Rutgers, The State University of New Jersey, NJ.

Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied psychological measurement, 39,* 167-188.

Lee, Y., Ip, E. H., &Fuh, C. (2008). A strategy for item exposure in multidimensional computerized adaptive testing. *Educational and Psychological Measurement*, *68*, 215-232.

Meijer, R. R., &Nering, M. L. (1999). Computerized adaptive testing: overview and introduction. *Applied Psychological Measurement*, *23*, 187-194.

Reckase, M. D. (2009). *Multidimensional item response theory, statistics for social and behavioral sciences*. New York, NY: Springer Science Business Media, LLC.

Revuelta, J. (1995). *El control de la exposicion de los items en tests adaptativosinformatiozados [Item exposure control in computerized adaptive tests]*. Unpublished master's dissertation, Universidad Autonoma de Madrid, Spain.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure methods in computerized adaptive testing. *Journal of Educational Measurement*, *35*, 311-327.

Rupp, A., & Templin, J. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-art. *Measurement, 6,* 219-262.

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

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

Wainer, H., &Mislevy, R. J. (2000). Item response theory, item calibration, and proficiency estimation. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2nd ed., pp. 61-100). Mahwah, NJ: Lawrance Erlbaum Associates, Inc, Publishers.

Wang, C., Chang, H.-H., &Boughton, K. A. (2011). Kullback-Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika*, *76*, 13-39.

Wang, C., Chang, H.-H., &Boughton, K. A. (2012). Deriving stopping rules for multidimensional computerized adaptive testing. *Applied Psychological Measurement*, *37*, 99-122.

Xu, X., Chang, H.-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.

Yi, Q., Zhang, J., & Chang, H.-H. (2008). Severity of organized item theft in computerized adaptive testing: A simulation study. *Applied Psychological Measurement*, *32*, 543-558.

