Diagnostic Classification Models: Thoughts and Future Directions

Robert A. Henson
The University of North Carolina at Greensboro

In general, I feel that diagnostic classification modeling (DCMs; also commonly referred to as cognitive diagnosis models) holds great potential in low stake situations because of the promise of providing more detailed information related to those attributes (sometimes called skills, abilities, or traits) that an individual should improve. Thus, diagnostic modeling provides a tool that can aid in the development of tailored lesson plans (for a student or a class), which could save students and teachers time. However, this true potential of diagnostic modeling has yet to be seen. Most papers in the literature seem to have focused on theoretical issues where the application is used mostly as an example of the methodology. In addition, most of these *example applications* do not provide a typical low stakes situation, but instead represent nearly ideal situations with large sample sizes that provide for the demonstration of a new methodology. Probably the most common example of such a data set is in math (e.g., mixed fraction subtraction which has been used by Tatsuoka, 1990; de la Torre and Douglas, 2004; Templin, Henson, & Douglas, 2008; Henson, Templin, & Willse, 2008), although other examples that do not use a math assessment exist.

I believe that this direction of the literature is due to an initial demand for DCMs to address many technical questions related to issues that have already been addressed by alterative *competing* theories of test assessment. That is, new methodologies tend to be evaluated based on those characteristics that are most familiar and so much of the DCM literature was attempting to catch up to what had already been developed in methodologies, such as factor analysis and item response theory. Although this direction was necessary as a first step toward establishing a very basic set of statistical principles, the growing emphasis of the methodology for DCMs is now on providing evidence that these models, *in application*, can provide the information that has been promised. Specifically, the field must now address the question, "Can a DCM be applied in an interdisciplinary setting, such as a typical teaching setting, where alternative models have already proven useful?" For such a question to be answered, DCMs must be accessible to a broader audience that is not familiar with the typical jargon used to describe these models. Therefore, a set of papers must be written describing the models that are available, how they differ, and how they can be estimated, among other issues, in a way that is accessible to this audience.

The paper, "Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State-of-the-Art" by Drs. Rupp and Templin (2008) addresses this concern by providing an ambitious summary of many of these basic concepts. Specifically, the authors have provided: (i) a definition of DCMs, including when they are appropriate; (ii) a summary of a basic set of core models; (iii) the methods used to estimate these models and the available software;

Correspondence should be addressed to Robert A. Henson, Ed Research Methodology, The University of North Carolina at Greensboro, 207 Curry Building, 1000 Spring Garden Street, Greensboro, NC 27403. E-mail: rahenson@uncg.edu

(iv) general models used to describe the attribute space (i.e., the proportion of people with each attribute profile); and (v) general approaches to assess model fit. In doing so, many of the concepts that must be considered from a *start-to-finish* application of DCMs have been discussed.

In addition, the authors have provided a very general definition of what is meant by diagnostic modeling and have provided nine characteristics that are used to compare traditional factor analysis and multidimensional item response theory to DCMs. Probably most importantly, the authors have provided a general discussion related to the common terminology of *cognitive diagnosis models*. In my past experience, this terminology is misleading. I agree with the alternative naming convention of diagnostic classification models in that, although these models may be used to study cognition, the broader application is to classify examinees in a way that would be useful for remediation. In their discussion of the original naming convention, the authors provide an insightful discussion as to how these models may be used to explore cognition based on model selection, although this may not be their primary purpose. This confirmatory approach is based on the fact that models differ on how the mastered attributes relate to the probability of a correct (or positive) response.

However, I do not agree with all the comparisons made based on these nine characteristics. Probably, the point of most contention is the statement that DCMs are criteria-referenced. DCMs are norm-referenced in the sense that classification into *groups* depends on the sample. That is, individuals are always placed into a high or low group, and thus no clear reference to a criterion is implied by the model. In classifying individuals, one cannot be certain that an individual classified as a master will perform as a master based on some criterion. Instead, one can only assume that a master would perform better than a nonmaster. Having said this, standard setting in any latent variable modeling approach is a good way to obtain a criterion-referenced interpretation. Standard setting approaches specific to DCMs have been discussed by Henson and Templin (2008) when a benchmark Algebra II assessment was used to determine what class objectives should be reviewed by each examinee. Specifically, teachers were asked how well a *master* should perform on the benchmark. In this instance, teachers were told that a student who has mastered an ability is defined by a student that would not need to review (or study) that ability prior to taking the end-of-course exam, thus setting the criterion.

In addition, I would like to comment on the discussion of the log-linear cognitive diagnosis model (LCDM, Henson, Templin, & Willse, 2008) which is also related to Criterion 6. The LCDM has just recently been discussed as a general model that can be used to define a complete family of DCMs. In doing so, all core models (i.e., DINA, DINO, NIDA, NIDO, NC-RUM, C-RUM, and some applications of the GDM) can be fit, excluding the RSM and AHM. As a result, this model cannot be classified as only a compensatory model, but instead must also be a classified as a noncompensatory model. Using the log-linear model with latent variables (Haberman, 1974, 1979; Hagenaars, 1993), a diagnostic model was defined such that the probability of a correct response depends on attribute *main effects* (describing the change in the log-odds of the probability of a correct response when comparing masters to nonmasters), in addition to all *interaction* terms between attributes (describing the dependency of the main effects on mastery or nonmastery of any other attributes). The model has been fit by Henson, Templin, and Willse (2008) using an MCMC algorithm, but can also be fit using software such as Mplus (Templin, Henson, & Douglas, 2008). For a more thorough discussion see Henson, Templin, and Willse (2008).

In summary, the paper by Drs. Rupp and Templin provides a much needed step toward the general application of DCMs. The authors have provided a summary of many of the concepts that one must consider to properly apply a DCM (which ranges from model selection and estimation, to

assessing the appropriateness of the model using methods of model fit) in a way that is accessible to a broader audience. As a result, this paper will contribute to the interest of DCMs without requiring an individual to filter through a list of technical documents that have been distributed throughout several journals. It has also provided a basis for a number of discussions centering on the definition of DCMs and when these models are appropriate. However, I would like to suggest that another area of much needed research is test development. Specifically, it may be intuitive that items should not be written in the same way as they have been written when a traditional analysis is the focus, but how this procedure should differ needs to be addressed. By discussing test development, in addition to topics such as longitudinal modeling of DCMs, I feel that DCMs will continue to become a possible alternative to common methodologies of psychometrics.

REFERENCES

de la Torre, J., & Douglas, J. A. (2004). Higher order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.

Haberman, S. J. (1974). Loglinear models for frequency tables derived by indirect observation: Maximum likelihood equations. *Annals of Statistics*, *2*, 911–924.

Haberman, S. J. (1979). *Qualitative data analysis: Vol. 2. New developments*. New York: Academic Press.

Hagenaars, J. (1993). *Loglinear models with latent variables*. Thousand Oaks, CA: Sage Publications, Inc.

Henson, R., & Templin, J. (2008, March). Implementation of standards setting for a geometry end-of-course exam using cognitive diagnosis. American Educational Research Association Conference. New York.

Henson, R. A., Templin, J., & Willse, J. (2008). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika.* http://www.springerlink.com/content/x82844151581059/?p=ee8e0f33141d4f11b69e8d528f61666&pi=0

Rupp, A. A. & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*(4), 219–262.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ: Erlbaum.

Templin, J., Henson, R., & Douglas, J. (2009). General theory and estimation of cognitive diagnosis models as constrained latent class models. *Manuscript under review*.

Evidentiary Reasoning in Diagnostic Classification Models

Roy Levy

Arizona State University,

In *Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State-of-the-Art*, Rupp and Templin (2008 hereafter RT) undertake the ambitious task of

Correspondence should be addressed to Roy Levy, Arizona State University, Payne Hall, Room 302, PO Box 870611, Tempe, Arizona 85287-0611. E-mail: Roy.Levy@asu.edu