THE USE AND MISUSE OF PSYCHOMETRIC MODELS

Jonathan Templin

UNIVERSITY OF KANSAS

Laine Bradshaw

THE UNIVERSITY OF GEORGIA

The purpose of our paper entitled *Hierarchical Diagnostic Classification Models: A Family of Models for Estimating and Testing Attribute Hierarchies* (Templin & Bradshaw, 2014) was two-fold: to create a psychometric model and framework that would enable attribute hierarchies to be parameterized as dependent binary latent traits, and to formulate an empirically driven hypothesis test for the purpose of falsifying proposed attribute hierarchies. The methodological contributions of this paper were motivated by a curious result in the analysis of a real data set using the log-linear cognitive diagnosis model, or LCDM (Henson, Templin, & Willse, 2009). In the analysis of the Examination for Certification of Proficiency in English (ECPE; Templin & Hoffman, 2013), results indicated that few, if any, examinees were classified into four of the possible eight attribute profiles that are hypothesized in the LCDM for a test of three binary latent attributes. Further, when considering the four profiles lacking examinees, it appeared that some attributes must be mastered before others, suggesting what is commonly called an attribute hierarchy (e.g., Leighton, Gierl, & Hunka, 2004). Although the data analysis alerted us to the notion that such a data structure might be present, we lacked the methodological tools to falsify the presence of such an attribute hierarchy. As such, we developed the Hierarchical Diagnostic Classification Model, or HDCM, in an attempt to fill the need for such tools.

We note that the driving force behind the HDCM is one of seeking a simpler, or more parsimonious, solution when model data misfit is either evident from LCDM results or implied by the hypothesized theories underlying the assessed constructs. As a consequence of the ECPE data results, we worked to develop a more broadly defined set of models that would allow for empirical evaluation of hypothesized attribute hierarchies. We felt our work was timely, as a number of methods, both new and old, are now using implied attribute hierarchies to assess examinees in many large scale analyses—from so-called intelligent tutoring systems (e.g., Cen, Koedinger, & Junker, 2006) to large scale state assessment systems for alternative assessments using instructionally imbedded items (e.g. the Dynamic Learning Maps Alternate Assessment System Consortium Grant, 2010–2015). Moreover, such large scale analyses are based on tremendously large data sets, many of which simply cannot fit with the types of (mainly unidimensional) models often used in current large scale testing situations. Furthermore, newly developed standards in education have incorporated ideas of learning progressions which indirectly imply the existence of hierarchically structured attributes (e.g., Progressions for the Common Core State Standards in Mathematics, Common Core State Standards Writing Team, 2012). In short, the current and

---

future needs for educational assessments are faced with inadequate psychometric and statistical methodology to make valid inferences about such complex multidimensional theories of learning and knowledge acquisition.

Once we developed the HDCM, we tested for the attribute hierarchy suggested by the LCDM analysis of the ECPE data. The analysis was meant to be an illustration of the HDCM as an extension of the LCDM. We concluded for this data set that a model with an attribute hierarchy was a better fitting model in comparison to the LCDM, although not across all tested models. We encourage readers to avoid making conclusions about the HDCM as a methodological tool based on idiosyncrasies of this specific analysis, and, instead, to focus on the results of the simulation study that provided strong evidence for the HDCM as a viable method for estimating an examinee's attribute pattern among a reduced set of attribute patterns imposed by the attribute hierarchy structure.

Overall, we view the commentary by von Davier and Haberman (2014) as having two significant and overlapping themes related to our paper and an additional theme on conjunctive models that, while not applying to our paper, we would like to address. First, and most significantly, does any data conform to multidimensional psychometric models, including those that are equipped to feature ordered categorical latent variables? The authors present their skepticism, stating "On a regular basis, in our experience, either less complex models or alternative specifications of models fit data as well or even better..." (see Introduction section). We view this issue as perhaps the most critical contemporary issue in our part of the psychometrics community and seek to broaden the discussion into two parts: (1) whether or not such multidimensional data has existed in the past or can exist, and (2) whether or not our current psychometric methods are sensitive enough to capture the multidimensional information in such data. In many respects, part (1) is analogous to the discussions in the early 20th century between Spearman and Thurston as to whether multiple factor analysis was necessary. Improving upon part (2) was the motivation for the development of the HDCM. Second, we seek to address the comments regarding conjunctive models. Third, and last, we will address the authors' beliefs regarding naming conventions of psychometric models.

1. On the Possible Existence of Multidimensional Data

von Davier and Haberman (2014) rely on results from our ECPE data analysis to question the added "analytic value" of the HDCM as a method (see last sentence of Introduction section). Their commentary does not attend to the simulation components of the paper that should be used to guide empirical claims about the model as a method. For example, von Davier and Haberman (2014) suggest that one should start with the simplest model, instead of using the top-down approach we suggested. Our suggestion was backed by our simulation results which, for example, showed that when the estimating measurement model was the simplest possible DCM (i.e., the DINA model), attribute hierarchies generated from the LCDM could not be identified because the conjunctive assumptions about attributes masked the attribute hierarchy. Our results indicated that one should start with the more general LCDM and then compare it to a given HDCM to statistically test the presence of an attribute hierarchy. Second, in the simulation study, we showed that if simulated data followed a linear attribute hierarchy, the HDCM *always* yielded better model-data fit when compared to a set of unidimensional models, including located latent class models (LLCMs) with varying numbers of classes and the 2-PL IRT model. This result was not surprising to us, but it was conducted to refute the claim in the paper's reviews *suggesting the opposite would occur*. This result also provides evidence to refute the claim that there is not a place for the new HDCM methodology as well as the claim that a model specifying a linear attribute hierarchy obstructs a more appropriate unidimensional model.

The ECPE analysis in our paper, like most real data analyses used for DCMs to this point, required use of data that originally were calibrated using a unidimensional model. As such, the result that a unidimensional model had the best model-data fit is unsurprising. Although the ECPE data did not fit with the HDCM, this result does not imply that the HDCM as a method is not of value. *It simply implies that the underlying construct of the ECPE is unidimensional*. Moreover, although we were not aware of the process of constructing the ECPE, common test construction processes often omit non-conforming or non-fitting items—such is recommended practice to make limited modifications to improve the fit of the data to the unidimensional model. The process of omitting such items thereby enforces a *subjective reality of unidimensionality* onto the test data and more broadly the construct, whether or not it is true. We view this as a potential breakdown of methodological tools that happens all too often in large scale psychometric assessment. Psychometric assessment should not be different from any other statistical method in that it should be held to the notion that constructs should be viewed as hypotheses that are to be empirically falsified. That said, we expect in the future, as more tests are designed from the ground-up to diagnose attributes with dependencies, that we will see real data analyses that do fit the HDCM better than a unidimensional model; however, for our case, the development of the methodology preceded the ground-up application.

Even though we feel the idiosyncratic ECPE results are independent of the methodological contributions of the HDCM, we appreciate that the real data analysis inspired comments which provided a forum for discussions. In the following sections, we hope to clarify the nature of the ECPE analyses and illuminate the value of modeling attribute hierarchies—even linear attribute hierarchies—in a DCM framework in order to further motivate ideation for and exploration of subsequent research in this area.

2. The Distinction of Guttman Scales and Linear Attribute Hierarchies

We primarily focus our discussion on linear attribute hierarchies in this section. Although von Davier and Haberman (2014) refer sometimes to attribute hierarchies generally, most of their claims are pronouncing disagreements under the specific hierarchy condition referred to as a linear attribute hierarchy (Leighton & Gierl, 2004). A linear attribute hierarchy is one where attributes are mastered in a specific, sequential order. For example, consider a test that measures three binary attributes, α_1, α_2, and α_3, such that Attribute 1 must be mastered before Attribute 2 and Attribute 2 must be mastered before Attribute 3. For this test, there exist only four possible patterns of attribute mastery: [000], [100], [110], and [111]. Suppose the possible patterns of a linear attribute hierarchy are elements of a set denoted $J_{(p)}$, where p denotes the permutation of attributes corresponding to the hierarchy in a set of attributes $A_{|A|} = \{1, 2, \ldots, A\}$, where $|A|$ is the cardinality of a set A. Then for our three attribute example, the attribute set is denoted by $A_3 = \{1, 2, 3\}$ and the linear attribute hierarchy patterns are elements of $J_{(123)}$.

A major premise of von Davier and Haberman's (2014) claims is that linear attribute hierarchies are isomorphic to Guttman scales; we disagree and will illustrate why there is not a one-to-one mapping. We seek to show that methodological options other than located latent class models (LLCM; e.g., Lindsay, Clogg, & Grego, 1991) or unidimensional DCMs, which *are* isomorphic to Guttman scales, are needed to model linear attribute hierarchies because of the possible states of empirical data. von Davier and Haberman (2014) showed that a Guttman scale given by set $G = \{0, 1, 2, 3\}$ was one-to-one with set $J_{123} = \{[000], [100], [110], [111]\}$ by

$$\begin{pmatrix} 0 \to [000] \\ 1 \to [100] \\ 2 \to [110] \\ 3 \to [111] \end{pmatrix}$$

The implied isomorphic function $f : J \to G$ is $f(\boldsymbol{\alpha}_a) = \sum_{i=1}^{|A|} \alpha_{i,a}$ for $\boldsymbol{\alpha}_a \in J$, where $J = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{A+1}\}$ and $\boldsymbol{\alpha}_a = [\alpha_{1a} \ \alpha_{2a} \ \ldots \ \alpha_{Aa}]$. Based on this mapping, authors claim "in a linear hierarchy there is no information gained by knowing which attribute pattern was observed given that we know how many attributes are mastered ($=$ "1")" (see latter part of Hierarchies of Binary Variables section). Although the authors' reference to knowing *which* attribute pattern indicates an acknowledgement that other patterns exist, their claim is not supported by the isomorphic structure provided by f due to the existence of linear attribute patterns other than J_{123}. In fact, there are $|A|!$ permutations of the elements in set $A_{|A|}$, which correspond to $|A|!$ possible linear attribute hierarchies for a set of $|A|$ attributes, meaning $J_p \in H$ where $|H| = |A|!$. For our three attribute case, set J_{123} is a member of $H = \{J_{123}, J_{132}, J_{213}, J_{231}, J_{321}, J_{312}\}$. Thus, the following Guttman to linear attribute hierarchy mapping accurately depicts the scenario for linear attribute hierarchies, and the mapping is not one-to-one:

$$
\begin{pmatrix}
0 \to [000] \\
1 \to \begin{cases} 100 \text{ if } J_p \in \{J_{123}, J_{132}\} \\ 010 \text{ if } J_p \in \{J_{213}, J_{231}\} \\ 001 \text{ if } J_p \in \{J_{312}, J_{321}\} \end{cases} \\
2 \to \begin{cases} 110 \text{ if } J_p \in \{J_{123}, J_{213}\} \\ 101 \text{ if } J_p \in \{J_{132}, J_{312}\} \\ 011 \text{ if } J_p \in \{J_{231}, J_{321}\} \end{cases} \\
3 \to [111]
\end{pmatrix}
$$

This mapping illustrates the attribute-specific information that is lost if a Guttman scale replaces a linear attribute hierarchy. For example, in this scenario, reporting to an examinee that he or she has mastered two of the three attributes (using the Guttman scaling) loses a considerable amount of information in comparison to reporting to an examinee that he or she has mastered Attributes 1 and 3, but not Attribute 2 (using the HDCM). In an educational setting, knowing which two attributes have been mastered provides more information to guide the remediation or additional instruction a student may need. Thus, in direct opposition to von Davier and Haberman's claim that "the pattern of attributes is not informative, telling only how many attributes have been mastered" (see latter part of Hierarchies of Binary Variables section), the mapping above provides a straightforward illustration of the informative nature of knowing which linear attribute hierarchy exists.

The value in attribute hierarchies that depict an ordered, multidimensional space over unidimensional IRT models or ordered latent class models that depict an ordered, unidimensional space lies in this very non-isomorphic mapping to Guttman scales. Key to conceptualizing the difference among a linear attribute hierarchy and a single discrete or continuous trait is (a) the confirmatory nature of the HDCM analysis and (b) the multidimensional nature of the HDCM analysis.

Confirmatory Nature. Unlike continuous stages of progress or discrete stages of progress along a single continuum, attribute profiles in a linear attribute hierarchy are distinguished by a theoretical construct whose nature is hypothesized prior to the analysis. In a LLCM, quantifiable differences in the classes are not known beyond knowing that the classes have ordinally different ability levels with respect to answering the items on the test correctly. Beyond ordering, analysts are left to explore the meaning of and distinctions among the classes, as is typical in exploratory latent class analysis. This is true, too, for a multicategory, unidimensional DCM that was shown in Templin and Bradshaw (2014) to be a particular, constrained version of a LLCM.

In contrast to LLCMs, when linear attribute hierarchies are specified by the HDCM, mastery of specific sets of traits defines each class. The HDCM is a confirmatory analysis model because

the attribute classes, as well as the attribute-item alignment, are defined *a priori* to the analysis. In that sense, the HDCM can be used to parameterize, and then falsify, hypothesized attribute structures. To illustrate this difference, consider an elementary math test which measures four attributes thought to form a linear attribute hierarchy. For example Attributes 1–4 may be addition, subtraction, multiplication and division. An LLCM may order examinees along five categories of math ability, but there is no statistical structure on the model to suggest the five categories correspond to profiles that can be described by elements of J_{1234}. Without this statistical structure, inferences based on the results in this way could not be validated inferentially as they could be in the HDCM. Given the parameterization of the HDCM, an added benefit is that the *a priori* conjecture about the attribute hierarchy can be formulated as a testable hypothesis to provide empirical evidence for or against the theory behind the hierarchy's form. Conversely, if addition, subtraction, multiplication and division did not form a linear hierarchy, the LLCM could offer no statistical evidence against (or for) this conjecture. In this way, the HDCM can help build theories about the structure among attributes in ways that LLCMs, or IRT models, cannot.

Multidimensional Nature of HDCM. Attributes in a linear attribute hierarchy necessarily have a strict ordering dependency; however, this does not mean that the attributes are the same trait or that the attributes should be modeled as a single dimension. A dependency among two binary attributes may arise for one of two reasons: (1) neither attribute is ever present without the other, or (2) one attribute may be present without the other, but the other is never present without the former. As regards the first reason, the attributes may be the same trait, or they are different traits but the test never presents a situation where one is present without the other. Either way, they are not separable as attributes in a linear attribute hierarchy for the testing occasion. The second reason is the case that applies to attributes in a linear hierarchy, and these attributes are in fact separable traits. For example consider the attribute hierarchy yielding patterns defined by set J_{123}. Suppose Attribute 1 is the ability to add, Attribute 2 is the ability to multiply, and Attribute 3 is the ability to solve expressions involving exponents. It is reasonable to expect such a hierarchy where one will have mastered addition before mastering exponents, which would mean that a student who has mastered exponents and not addition likely will not exist. Thus, in the HDCM, the profiles [001] and [011] will not be estimated. However, this does not mean that adding and solving exponents are the same trait. This example may illuminate why we refute this claim by von Davier and Haberman ([2014](#), see Within Attribute Hierarchies, Conjunctive Attributes Do Not Exist section), which relies on a unidimensional framework of thinking:

> *In addition, under the assumption of attribute hierarchies, items are not allowed or at least are implausible if they require only the higher-order attribute but not include the lower-order attribute, for—by definition of the hierarchy—the higher-order attribute cannot be present without the lower-order attribute.*

For example, the item requiring students to solve for x in the equation $4^3 = x$ measures the exponent attribute without measuring the addition attribute. However, if Attribute 1 was simply a lower stage of mastery of Attribute 3, an item could not measure Attribute 3 without Attribute 1. Nonetheless, the parameterization of the HDCM accounts for the nesting of the mastery of Attribute 1 within 3 in the structural model, even if the item only measures Attribute 3 as specified by the measurement components of the model.

Although we spent most of this section discussing the merits of linear hierarchies, equally, if not more, important and realistic cases would be non-linear attribute hierarchies which have complex nesting structures of preceding and following attributes. Such hierarchies are depicted in Chapter 4 of Rupp, Templin, and Henson ([2010](#)) and have been anticipated in many other context, both in diagnostic psychometric models (see Leighton, Gierl, & Hunka, [2004](#); Tatsuoka, [2002](#), [1983](#)) and in machine learning contexts (see Mislevy, Almond, & Yan, [1999](#); Pardos, Heffernan,

Anderson, Heffernan, & Schools, 2010). The structure of the HDCM allows for its use beyond linear hierarchies, by which we believe the value of the HDCM is not only self-evident but is actually consistent with the values of parsimony espoused by von Davier and Haberman (2014) in their commentary.

3. On the Conjunctive Nature of Attributes in a Hierarchy

Finally, we feel the section of the commentary rhetorically asking if conjunctive attributes are uniquely defined is tangential to the HDCM as discussed in our paper. The research cited and discussed focuses on the DINA model, a model that has long been understood to have identification problems that are not present in more general diagnostic models (see Chiu, Douglas, & Li, 2009 and subsequent forthcoming research). In our paper we showed how if attributes with compensatory behavior followed a linear hierarchy, a DINA model version of the HDCM with strict assumptions of conjunctive attribute behavior could not uncover the attribute structure. The contended situation appears to be when attributes with a strictly conjunctive structure follow a linear hierarchy. The HDCM provides a parameterization that allows for—though it does not strictly assume—two (or more) attributes in a linear hierarchy to have a conjunctive structure for a given item, by the main effect terms for nested attribute(s) being estimated at zero. Although imposing this parameterization on every item would reduce to equivalence with the DINA model in the measurement portion of the model, the structural model would still reflect the linear attribute hierarchy structure, theoretically rendering conjunctive item-level behavior of the trait distinct from hierarchical relationships among the traits. This theory breaks down only for test designs already known to be problematic for the typical DINA model (DeCarlo, 2011), which is all together a separate, though practically important, issue.

4. On Evidence of Multidimensional Data

The preceding sections have demonstrated that it is philosophically possible for multidimensional methods to exist and to be different from what is implied by the methods described by von Davier and Haberman (2014). The next question is whether or not such data may in fact exist. In our experience, we have found that it is possible to develop multidimensional constructs that are measured well by models with ordered categorical latent traits. As an example, we participated in a project to develop a multidimensional assessment of middle grades teachers' knowledge of rational numbers (Bradshaw, Izsàk, Templin, & Jacobson, 2014). The constructs measured by the test developed by this project took a number of years to fully be developed with a great deal of work by scholars in the field of mathematics education. Beyond the conceptual development, we needed a set of methodological tools that would allow us to measure multidimensional latent variables using complex item types with limited information—a set up that called for the use of diagnostic models such as the LCDM and HDCM. The results of the study suggested multidimensional measurement is plausible, but that the analytic tools used to assess the data must be sensitive to nuances in multidimensionality, especially in data with limited information.

5. On Naming Conventions of Psychometric Models

Although von Davier and Haberman (2014) note that our choice of model names "obscures the fact that the original intent failed to fit a model with multiple attributes" (see Unidimensional Diagnostic Classification Models (DCMs) Are a Misnomer section), we note that our use of such

model names is consistent with not only naming practices in psychometric research, but also with naming conventions used by the authors in the commentary. In particular, the references to von Davier (2011, 2013) noted a recasting of the DINA diagnostic model into a "linear GDM" (see Are Conjunctive Attributes Uniquely Defined section) where GDM refers to General Diagnostic Model (a term used by von Davier, 2005) which could also be described by a general latent class model (e.g. Lazarsfeld and Henry, 1968). Our point in highlighting this is simple: Names of models provide a language-based context that gives the reader an object to remember when continuing within a given paper. Ultimately, model names get in the way and are not as easily extendable as is the general mathematical specification of such models. Our choice of model names was set in the context of our analysis, as were the choices of von Davier (2011, 2013) and of many other authors well beyond this topic. We feel we were clear about the relationships of our models to that of previous models existing in the literature and how they were either more specific or more general and encourage other authors to be as clear in their approach as well.

6. Conclusion

Like von Davier and Haberman (2014), we will conclude with the thoughts of William of Occam: "*Numquam ponenda est pluralitas sine necessitate*," meaning **plurality must never be posited without necessity** (cited in Thorburn, 1918). In the time in which William of Occam was alive, although scholars held a spherical notation of the Earth, the idea that the Earth was flat was a belief of many common people. To this day, the International Flat Earth Society exists, in fact (see http://www.theflatearthsociety.org). That said, the notion of the Earth being flat can be viewed as a simplistic model that is useful for many tasks, from laying concrete for the foundation of houses to calculating a rough measure of distance between two relatively close points. For such tasks, the model of a spherical Earth matters little. Once more complex phenomena are to be studied, such as astronomy or intercontinental travel, however, the model of a flat Earth is no longer sufficient. Herein lies the crux of our argument. We built our models under the necessity of multidimensionality (plurality) because unidimensional models, which may be appropriate in many contexts, are not fully sufficient to measure multifaceted knowledge structures posited by learning theorists. Consider for a moment if the only tools one had were their eyes: One might believe the earth was flat. Similarly, we fear that if the only psychometric tools one has are unidimensional, one might believe cognition, learning, and understanding was unidimensional. Thus, we view plurality as a necessity in the pursuit of objective reality.

Acknowledgements

References

Bradshaw, L., Izsàk, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understanding of rational number: building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practice*.

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis—a general method for cognitive model evaluation and improvement. In *Intelligent tutoring systems* (pp. 164–175). Berlin: Springer.

Chiu, C.-Y., Douglas, J.A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: theory and applications. *Psychometrika*, *74*, 633–665.

Common Core State Standards Writing Team (2012). *Progressions for the common core state standards*. June. http://commoncoretools.files.wordpress.com/2012/06/ccss_progression_g_k6_2012_06_27.pdf. Retrieved 20 September 2013.

DeCarlo, L.T. (2011). On the analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8–26.

Dynamic Learning Maps (2010–2015). *Dynamic learning maps alternate assessment system consortium*. United States Department of Education, IDEA General Supervision Enhancement Grant—Alternate Academic Achievement Standards, Award #H373X100001.

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Leighton, J.P., Gierl, M.J., & Hunka, S.M. (2004). The attribute hierarchy model for cognitive assessment: a variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*, 205–237.

Lindsay, B., Clogg, C.C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96–107.

Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: where the numbers come from. In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 437–446). San Mateo: Morgan Kaufmann.

Pardos, Z.A., Heffernan, N.T., Anderson, B., Heffernan, C.L., & Schools, W.P. (2010). Using fine-grained skill models to fit student performance with Bayesian networks. In *Handbook of educational data mining* (pp. 417–426).

Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: theory, methods, and applications*. New York: Guilford.

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, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *51*, 337–350.

Templin, J., & Bradshaw, L. (2014, in press). Hierarchical diagnostic classification models: a family of models for estimating and testing attribute hierarchies. *Psychometrika*.

Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates in Mplus. *Educational Measurement, Issues and Practice*, *32*(2), 37–50.

Thornburn, W. (1918). The myth of Occam's razor. *Mind*, *27*, 345–353.

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (ETS Research Report RR-05-16).

von Davier, M. (2011). *Equivalency of the DINA model and a constrained general diagnostic model* (Research Report 11-37). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-11-37.pdf.

von Davier, M. (2013). The DINA model as a constrained general diagnostic model—two variants of a model equivalency. *British Journal of Mathematical & Statistical Psychology* doi:10.1111/bmsp.12003.

von Davier, M., & Haberman, S.J. (2014). Hierarchical diagnostic classification models morphing into unidimensional 'diagnostic' classification models—a commentary. *Psychometrika*.