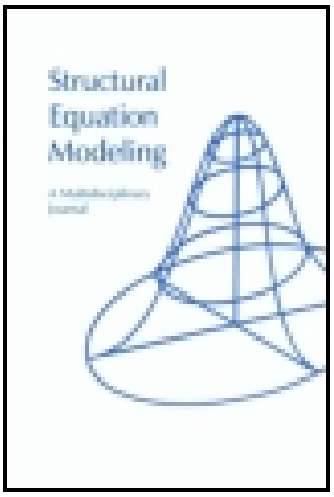

Structural Equation Modeling: A Multidisciplinary Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

The Partial Credit Model and Generalized Partial Credit Model as Constrained Nominal Response Models, With Applications in Mplus

Anne Corinne Huggins-Manley[a] & James Algina[a]

[a] University of Florida

Published online: 06 Jan 2015.

To cite this article: Anne Corinne Huggins-Manley & James Algina (2015): The Partial Credit Model and Generalized Partial Credit Model as Constrained Nominal Response Models, With Applications in Mplus, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2014.937374

To link to this article: http://dx.doi.org/10.1080/10705511.2014.937374

Structural Equation Modeling: A Multidisciplinary Journal, 00: 1–11, 2015

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2014.937374

The Partial Credit Model and Generalized Partial Credit Model as Constrained Nominal Response Models, With Applications in M*plus*

Anne Corinne Huggins-Manley and James Algina
University of Florida

The purpose of this article is to demonstrate constraining the nominal response model in M*plus* software to calibrate data under the partial credit model (PCM) and generalized partial credit model (GPCM). Currently, many researchers are uncertain if the PCM and GPCM can be estimated within M*plus*. Through model constraint commands in M*plus,* we demonstrate that both models can be estimated in recent versions of this software. We present an example of this approach with data from 522 respondents on a subset of items from the Math Self-Efficacy Scale (Betz & Hackett, 1983). It is demonstrated that the presented model code is a viable way of estimating the models in M*plus.*

Keywords: generalized partial credit model, M*plus*, nominal response model, partial credit model

In the mid-1900's, Rasch introduced a latent trait model for fitting data from test items that are scored dichotomously (i.e., 0,1; Rasch, 1960). Lord introduced the foundations of alternative latent trait models for this same type of binary data (Lord, 1953), which were later solidified in Lord and Novick's (1968) seminal book. A host of models have been introduced since this time, several of which were built to analyze data from polytomously scored items (e.g., 1,2,3,4,5). Bock (1972) introduced a model for polytomously scored items on a nominal scale, called the nominal response model (NRM). Others introduced models built for polytomously scored items on an ordered scale. These include, but are not limited to, the partial credit model (PCM; Masters, 1982) and the generalized partial credit model (GPCM; Muraki, 1992). Although many of these models were introduced by different persons, in different decades, from different countries, and within different theoretical frameworks, there are more similarities than differences across the models (Thissen & Steinberg, 1986). This article provides a demonstration of the similarities among the NRM, PCM, and GPCM and demonstrates how those similarities allow practitioners and researchers to estimate the PCM and GPCM using M*plus* (Muthén & Muthén, 2012), a popular statistical package that directly estimates the NRM.

More specifically, this article shows that the PCM and GPCM are nested in the NRM. This indicates that both models are simply constrained versions of the NRM. The constraints account for the ordered nature of items for which the PCM and GPCM are built. Several researchers have demonstrated this nested relationship in previous work (e.g., Mellenbergh, 1995; Ostini & Nering, 2009; Thissen & Steinberg, 1986), but most of the demonstrations are technical and none were connected to a statistical package or coding procedure. This article aims to introduce the models and the necessary constraints in a less technical manner. Additionally, the constraints are demonstrated in a statistical package to further the understanding of the model relationships and provide instructions for applying the constraints.

M*plus* (Muthén & Muthén, 2012) is a popular statistical package. It is widely used both because it is flexible (e.g., it can estimate structural equation models, multilevel models, confirmatory factor analysis models, and item response theory/Rasch models) and because it is known to produce accurate results. Many U.S. universities have M*plus* in some computer labs, and some universities (e.g., the University of Florida) have begun to purchase licenses that allow for free

Correspondence should be addressed to Anne Corinne Huggins-Manley, University of Florida, P.O. Box 117049, Gainesville, FL 32611. E-mail: ahuggins@coe.ufl.edu

student access to M*plus* through apps. Due to these benefits of M*plus,* it is important that users are aware of how they can run popular statistical models such as the PCM and GPCM within the package. The M*plus* user's guide does not discuss the models (Muthén & Muthén, 2012), the online M*plus* discussion posts have several comments indicating that it is unclear if and how the models can be estimated in the package (B. O. Muthén, 2008; L. K. Muthén, 2004, 2007), and several graduate course lecture notes located online indicate uncertainty that M*plus* can estimate these models (e.g., Hoffman, 2010; Templin, 2013). The application presented in this article shows that, indeed, M*plus* can estimate the PCM and GPCM through a constrained NRM. Implementing these models in this package has multiple benefits for practitioners and researchers, including (a) the ability to conduct model comparisons across multiple latent trait models for polytomous data without having to leave the M*plus* package, avoiding possible confounds of model differences and statistical package differences; (b) allowing students and faculty to run the models in this program that is often available on campus and widely known for accuracy in its results; (c) the ability to obtain log-likelihood, Akaike's information criterion (AIC), Bayesian information criterion (BIC), and adjusted BIC of the models; and (d) availability of the wide selection of estimation options provided by the M*plus* package (Muthén & Muthén, 2012).

THE NESTING OF THE PCM AND GPCM IN THE NRM

The NRM (Bock, 1972) is a latent trait model for test item responses that are categorized nominally. For example, a test that measures parenting styles has a series of situational items for which a counselor rates parents as *permissive, authoritative,* or *authoritarian.* There is no inherent ordering to these categories, and the NRM is formulated to handle these nominal item data. The NRM is often used for standard educational multiple-choice items as well (Embretson & Reise, 2000; Ostini & Nering, 2006). An examinee chooses among, for example, five response options (four of which are incorrect), and it might be of interest to estimate the relationship between the examinee's ability level and the probability of selecting each of the individual response options. The NRM allows for such estimation.

The model is a multivariate generalization of the logistic regression model (Bock, 1997; Hosmer & Lemeshow, 2000; Ostini & Nering, 2006), and in the social sciences it is often expressed as the probability of an examinee selecting response option x on item i, such that

$$P_{i(x)} = \frac{\exp\left(\zeta_{i(x)} + \lambda_{i(x)}\theta\right)}{\sum_{k=1}^{K} \exp\left(\zeta_{i(k)} + \lambda_{i(k)}\theta\right)}, \tag{1}$$

where $P_{i(x)}$ denotes the probability of selecting response option x on item i, k is an index for the elements of a vector of K possible response options (i.e., $k = 1, \cdots, K$), θ is the unidimensional latent trait measured by the set of items, $\zeta_{i(k)}$ is an intercept for category k, and $\lambda_{i(k)}$ is the slope (i.e., discrimination) of the relationship between response option k and θ.

For instructional purposes, we rewrite Equation 1 for the probability of selecting response option 3 on an item with four response options. In long notation, the probability is expressed as

$$P_{i(3)} = \frac{\exp\left(\zeta_{i(3)} + \lambda_{i(3)}\theta\right)}{\exp\left(\zeta_{i(1)} + \lambda_{i(1)}\theta\right) + \exp\left(\zeta_{i(2)} + \lambda_{i(2)}\theta\right) + \exp\left(\zeta_{i(3)} + \lambda_{i(3)}\theta\right) + \exp\left(\zeta_{i(4)} + \lambda_{i(4)}\theta\right)}. \tag{2}$$

Equations 1 and 2 demonstrate the common designation of the NRM as a divide-by-total model (Thissen & Steinberg, 1986). The exponential term for the category of interest is divided by the sum of all possible exponential terms for item i.

It can be shown that if M is added to each logit—that is $\zeta_{i(x)} + \lambda_{i(x)}\theta$ becomes $\zeta_{i(x)} + \lambda_{i(x)}\theta + M$—the probability of selecting response x is unchanged. Because there are an infinite number of solutions for the logits, an arbitrary constraint must be set. Bock (1972) introduced a constraint such that the logits within each item sum to 0, stated as

$$\sum_{k-1}^{K} \left(\zeta_{i(k)} + \lambda_{i(k)}\theta\right) = 0. \tag{3}$$

This constraint implies that the sum of the intercepts is equal to zero and the sum of the slopes is also equal to zero. Thissen (1991) used an alternative constraint within the MULTILOG package such that λ and ζ parameters associated with the first category in the item (which by nature is an arbitrary choice of category due to the nominal scale of the categories) are fixed to zero. We adopt this approach in the subsequent presentation. For readers less familiar with latent trait models, the necessity of introducing one of these arbitrary constraints is similar to choosing a rotated solution in factor analysis; all solutions are mathematically appropriate and equivalent with respect to fitting the model to the data, but to obtain a single solution one must choose one of the rotated solutions, each of which is associated with constraints on the factor loadings.

There are several other divide-by-total models that are special cases of the NRM, including the PCM (Masters, 1982). The PCM introduces constraints to the NRM to handle response options that are ordered as opposed to nominal. Building on the earlier example of a counselor rating clients in terms of their parenting styles, let's say that the counselor is less interested in the type of parenting but rather is

trying to measure the degree to which a parent uses authoritative parenting techniques with her or his child. For each situational item, the counselor is asked to rate the parent as *never, sometimes, often,* or *always.* These types of item response categories have a clear ordering and the PCM is built to account for this ordered scaling property of item responses.

The PCM model for the category probabilities is

$$P_{i(1)} = \frac{1}{1 + \sum_{c=2}^{K} \exp \sum_{k=2}^{c} \left(\theta - b_{i(k)}\right)} \tag{4a}$$

for response option $x = 1$ and

$$P_{i(x)} = \frac{\exp \sum_{k=2}^{x} \left(\theta - b_{i(k)}\right)}{1 + \sum_{c=2}^{K} \exp \sum_{k=2}^{c} \left(\theta - b_{i(k)}\right)} \tag{4b}$$

for response options $x = 2$ through K, where c represents a vector of possible k responses. The probability of a response of k is equal to the probability of a response of $k - 1$ when $\theta = b_{i(k)}$ and larger than the probability of a response of $k - 1$ when $\theta > b_{i(k)}$. Therefore, the parameter $b_{i(k)}$ is called a step parameter and is interpreted as the value on the θ scale at which a respondent steps from response $k - 1$ to k.

As we did with the NRM, we rewrite Equation 4b with respect to the probability of selecting response option 3 on an item with four response options as

$$P_{i(3)} = \frac{\exp\left(\theta - b_{i(2)} + \theta - b_{i(3)}\right)}{\begin{array}{c}1 + \exp\left(\theta - b_{i(2)}\right) + \exp\left(\theta - b_{i(2)} + \theta - b_{i(3)}\right), \\ +\exp\left(\theta - b_{i(2)} + \theta - b_{i(3)} + \theta - b_{i(4)}\right)\end{array}} \tag{5a}$$

or

$$P_{i(3)} = \frac{\exp\left(2\theta - b_{i(2)} - b_{i(3)}\right)}{\begin{array}{c}1 + \exp\left(\theta - b_{i(2)}\right) + \exp\left(2\theta - b_{i(2)} - b_{i(3)}\right). \\ + \exp\left(3\theta - b_{i(2)} - b_{i(3)} - b_{i(4)}\right)\end{array}} \tag{5b}$$

Comparing Equation 5b to Equation 2, we can see that for the third response option $\lambda_{i(3)} = 2$. More generally, the PCM is a special case of the NRM with

$$\lambda_{i(k)} = k - 1. \tag{6}$$

That is, the slope for response category k is equal to $k - 1$. In addition, we can see that $\zeta_{i(3)} = -\left(b_{i(2)} + b_{i(3)}\right)$ and generally $\zeta_{i(k)} = -\left(b_{i(2)} + \cdots + b_{i(k)}\right)$. The step parameter for the third response category is $b_{i(3)} = -\left(\zeta_{i(3)} - \zeta_{i(2)}\right)$. For response categories 3 and above

$$b_{i(k)} = -\left(\zeta_{i(k)} - \zeta_{i(k-1)}\right). \tag{7}$$

Because the intercept for the first response option is set equal to zero in the NRM, $\zeta_{i(1)} = 0$ and

$$b_{i(2)} = -\left(\zeta_{i(2)}\right). \tag{8}$$

The constraints on the slopes (Equation 6) and the equations relating the step parameters of the PCM to the intercepts of the NRM (Equations 7 and 8) must be programmed in M*plus* for M*plus* to estimate the PCM as a special case of the NRM.

The GPCM is another model built for polytomous, ordered items (Muraki, 1992) such as the items from our counselor example in which parents are rated as *never, sometimes, often,* or *always* with respect to their authoritative parenting techniques. The category probabilities of the GPCM are defined as

$$P_{i(1)} = \frac{1}{1 + \sum_{c=2}^{K} \exp \sum_{k=2}^{c} a_i \left(\theta - b_{i(k)}\right)} \tag{9a}$$

for response option $x = 1$ and

$$P_{i(x)} = \frac{\exp \sum_{k=2}^{x} a_i \left(\theta - b_{i(k)}\right)}{1 + \sum_{c=2}^{K} \exp \sum_{k=2}^{c} a_i \left(\theta - b_{i(k)}\right)} \tag{9b}$$

for response options $x = 2$ through K. The new parameter a_i is the item discrimination parameter for the ith item. If the discrimination parameters are fixed to 1, Equations 9a and 9b become equivalent to the PCM model shown in Equations 4a and 4b. The parameter a_i varies across items, but is a constant across response options within an item. As we did with the NRM and PCM, we rewrite Equation 9b with respect to the probability of selecting response option 3 on an item with four response options as

$$P_{i(3)} = \frac{\exp\left[a_i\left(\theta - b_{i(2)}\right) + a_i\left(\theta - b_{i(3)}\right)\right]}{\begin{array}{c}1 + \exp\left[a_i\left(\theta - b_{i(2)}\right)\right] \\ \exp\left[a_i\left(\theta - b_{i(2)}\right) + a_i\left(\theta - b_{i(3)}\right)\right] + \\ \exp\left[a_i\left(\theta - b_{i(2)}\right) + a_i\left(\theta - b_{i(3)}\right) + a_i\left(\theta - b_{i(4)}\right)\right]\end{array}}, \tag{10a}$$

which can be rearranged as

$$P_{i(3)} = \frac{\exp\left[\left(2a_i\theta - a_i\left(b_{i(2)} + b_{i(3)}\right)\right)\right]}{\begin{array}{c}1 + \exp\left[\left(a_i\theta - a_i b_{i(2)}\right)\right] + \\ \exp\left[\left(2a_i\theta - a_i\left(b_{i(2)} + b_{i(3)}\right)\right)\right] + \\ \exp\left[\left(3a_i\theta - a_i\left(b_{i(2)} + b_{i(3)} + b_{i(4)}\right)\right)\right]\end{array}}. \tag{10b}$$

Comparing Equations 2 and 10b shows that $\lambda_{i(3)} = 2a_i$ and generally that

$$\lambda_{i(k)} = a_i\,(k-1)\,. \tag{11}$$

Comparing Equations 2 and 10b also shows that, for example, $\zeta_{i(3)} = -a_i\left(b_{i(2)} + b_{i(3)}\right)$ and generally that

$$\zeta_{i(k)} = -a_i\left(b_{i(2)} + b_{i(3)} + \cdots + b_{i(k)}\right) \tag{12}$$

Also, for example, $b_{i(4)} = -\left(\zeta_{i(4)} - \zeta_{i(3)}\right)/a_i$ and generally for $k > 2$

$$b_{i(k)} = \frac{-\left(\zeta_{i(k)} - \zeta_{i(k-1)}\right)}{a_i}. \tag{13}$$

Because $\zeta_{i(1)} = 0$,

$$b_{i(2)} = -\zeta_{i(2)}/a_i. \tag{14}$$

The constraints on the slopes (Equation 11) and the equations relating the step parameters of the GPCM to the intercepts of the NRM (Equations 13 and 14) must be programmed in M*plus* for M*plus* to estimate the GPCM as a special case of the NRM.

ESTIMATING THE PCM IN M*PLUS*

In this section, we demonstrate the aforementioned PCM constraints within the M*plus* package version 7 (Muthén & Muthén, 2012). For simplicity, we only estimate item parameters for four items from a modified version of the Math Tasks subscale of the Math Self-Efficacy Scale (Betz & Hackett, 1983) used by Langenfeld and Pajares (1993). Responses were made on a 5-point scale instead of a 10-point scale and one item ("Work a slide rule") was replaced by "Use a scientific calculator." However, the code presented here can be extended to assessments with a larger number of items. The data in this example come from $N = 522$ respondents. Each item was scored on a Likert scale with five possible response options. The response options are ordered, and hence a PCM is more appropriate than an NRM.

Recall that in the presentation of the relationship between the parameters of the PCM and NRM, we specified that the λ and ζ parameters in the NRM were set equal to zero for the first response option. M*plus* has a default that fixes the λ and ζ parameters for the last response option to 0 (Muthén & Muthén, 2012). To ensure that M*plus* estimates the NRM parameters as we defined them, we reverse coded our data such that a response in Category 5 was awarded a score of 1, a response in Category 4 was awarded a score of 2, and so on. In the remainder of this section we use the term *data score* to refer to how the category is coded in the data and the term *category* to refer to the response category a participant chose when responding to an assessment item. For example, if a person chose Category 1 on Item 4, the data score associated with that person and that item is 5.

Figure 1 shows the M*plus* input commands. Under the "analysis" command we chose maximum likelihood (ML) as the estimator. Under the "variable" command we named all 18 items in the data set (i.e., names are I1–I18), and then selected only the first four items for analysis (i.e., usevariables are I1–I4). We specified all items to be on a nominal scale (i.e., nominal are I1–I4), although the model constraints we use later in the input result in estimation of PCM parameters.

Under the "model" command, we first name the latent variable "trait" and then use the "by" statement to define this latent trait by 16 indicators (e.g., I1#3@2). For example, a data score of 2 on Item 1 is coded as I1#2. Each category, except the reference category, within an item is an indicator of the trait. Because we have reverse-coded the variables, the first category (data score "5") is the reference category. All indicators used to define the trait can be thought of as dichotomous items loading onto the latent variable, with the category of interest labeled as 1 and data score 5 (the reference category) coded as 0. You will notice that the exclusion of the reference categories as indicators is similar to the exclusion of a dummy variable for the reference category in a regression analysis. As mentioned earlier, the intercepts and slopes for the reference categories are set to 0 in M*plus*; that is, the slope and intercept for a data score of 5 are set equal to zero so no estimation is needed. Due to the reverse coding the slope and intercept are fixed to zero for a response in Category 1, which is consistent with our presentation of the NRM.

The values following the ampersand in the "by" statements are fixed slope values for each data score (see Equation 6 for determining the appropriate fixed slope for any given category). Although the slope of each item is 1 under the PCM, we have to allow each data score to have a slope that provides a weight associated with the within-item category ordering. For example, according to Equation 6 a response in Category 5 on Item 1 should have a slope of 4, a response in Category 4 on Item 1 should have a slope of 3, and so forth. Due to reverse coding, data score 1 is associated with response Category 5, data score 2 is associated with response Category 4, and so on. Therefore data score 1 should have a slope of 4, data score 2 should have a slope of 3, data score 3 should have a slope of 2, and data score 4 should have a slope of 1. As an example, I4#2@3 indicates that we have set the second data score on Item 4 to have a slope of 3.

In the last part of the "model" command, we have assigned a name to each of the NRM intercept (ζ) parameters. For example, the intercept for data score 1 of Item 1 (i.e., I1#1), which is associated with Category 5 of Item 1, is named Int5_I1, indicating it is the NRM intercept for Category 5. Providing names for the parameters is necessary for use of these parameters under the "model constraint" command discussed later. Notice that we named the intercepts with respect to category numbers rather than data

```
analysis:
estimator is ml;
variable:
names are I1 - I18;
usevariables are I1 - I4;
nominal are I1-I4;
model:
trait by I1#1@4 I1#2@3 I1#3@2 I1#4@1 I2#1@4 I2#2@3 I2#3@2 I2#4@1
         I3#1@4 I3#2@3 I3#3@2 I3#4@1 I4#1@4 I4#2@3 I4#3@2 I4#4@1;

[I1#1] (Int5_I1);!label for NRM intercept for response category 5 on Item 1;
[I1#2] (Int4_I1);!label for NRM intercept for response category 4 on Item 1;
[I1#3] (Int3_I1);!label for NRM intercept for response category 3 on Item 1;
[I1#4] (Int2_I1);!label for NRM intercept for response category 2 on Item 1;
[I2#1] (Int5_I2);
[I2#2] (Int4_I2);
[I2#3] (Int3_I2);
[I2#4] (Int2_I2);
[I3#1] (Int5_I3);
[I3#2] (Int4_I3);
[I3#3] (Int3_I3);
[I3#4] (Int2_I3);
[I4#1] (Int5_I4);
[I4#2] (Int4_I4);
[I4#3] (Int3_I4);
[I4#4] (Int2_I4);

model constraint:
new b2_I1 b3_I1 b4_I1 b5_I1 b2_I2 b3_I2 b4_I2 b5_I2
    b2_I3 b3_I3 b4_I3 b5_I3 b2_I4 b3_I4 b4_I4 b5_I4;

b2_I1=-1*Int2_I1;!PCM step parameter for response category 2 on Item 1;
b3_I1=-1*(Int3_I1-Int2_I1);!PCM step parameter for response category 3 on Item 1;

b4_I1=-1*(Int4_I1-Int3_I1);!PCM step parameter for response category 4 on Item 1;

b5_I1=-1*(Int5_I1-Int4_I1);!PCM step parameter for response category 5 on Item 1;


b2_I2=-1*Int2_I2;
b3_I2=-1*(Int3_I2-Int2_I2);
b4_I2=-1*(Int4_I2-Int3_I2);
b5_I2=-1*(Int5_I2-Int4_I2);

b2_I3=-1*Int2_I3;
b3_I3=-1*(Int3_I3-Int2_I3);
b4_I3=-1*(Int4_I3-Int3_I3);
b5_I3=-1*(Int5_I3-Int4_I3);

b2_I4=-1*Int2_I4;
b3_I4=-1*(Int3_I4-Int2_I4);
b4_I4=-1*(Int4_I4-Int3_I4);
b5_I4=-1*(Int5_I4-Int4_I4);
```

FIGURE 1 M*plus* input for partial credit model analysis of 4 items on the Math Self-Efficacy Scale.

scores, which will assist in coding within the model constraint command.

The "model constraint" command allows us to name and create a new set of parameters. These new parameters will be the step parameters of the PCM (i.e., $b_{i(k)}$). With the "new" statement we have given names to 16 new PCM step parameters. For example b2_I1 refers to the step parameter for the second category on the first item. These new PCM parameters are defined at the θ point for which two adjacent categories are associated with an equal probability of selection. For example, b2_I3 is the step parameter for response Category 2 on Item 3. It can be interpreted as the θ value for which a person has an equal chance of obtaining a 1 or a 2 on Item 3.

To obtain the PCM parameter estimates, we implement Equations 7 and 8 in M*plus*. Namely, for categories 3 and above, we subtract the NRM intercept term of the lower of the two adjacent categories from the NRM intercept term of the higher of the two adjacent categories and multiply the difference by –1 (see Equation 7). For example, the PCM step parameter between Categories 2 and 3 on Item 2 is estimated as b3_I2 = -1*(Int3_I2 -Int2_I2). Consistent with Equation 8, it is not necessary to subtract the intercept for Category 1 from the intercept for Category 2. We only have to multiply intercepts for Category 2 by –1 to obtain the PCM step parameters for Category 2. For example, the PCM step parameter for Category 2 of Item 3 is estimated as b2_I3= -1*Int2_I3.

Figure 2 shows selected output produced from the M*plus* input commands in Figure 1. The log-likelihood provided under the MODEL FIT INFORMATION section can be used for model comparison purposes. Additional fit information (i.e., AIC, BIC, adjusted BIC) can also be used for model comparisons.

The remainder of the output is under the MODEL RESULTS section. Under TRAIT BY, the fixed slope of each item category is shown. For example, data score 1 in Item 1 is associated with a fixed slope of 4. Under "Intercepts" each nonreference category has an estimated NRM intercept parameter. For example, data score 2 of Item 4 (i.e., I4#2) has an NRM intercept of $\hat{\zeta} = 2.68$. Due to reverse coding, this is the estimated NRM intercept parameter associated with Category 4 of Item 4. Because these intercepts are parameterized in reference to Category 1, we can say

```
THE MODEL ESTIMATION TERMINATED NORMALLY


MODEL FIT INFORMATION

Number of Free Parameters                        17

Loglikelihood

          H0 Value                        -2679.563

Information Criteria

          Akaike (AIC)                     5393.125
          Bayesian (BIC)                   5465.506
          Sample-Size Adjusted BIC         5411.544
            (n* = (n + 2) / 24)


MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 TRAIT    BY
    I1#1               4.000      0.000    999.000    999.000
    I1#2               3.000      0.000    999.000    999.000
    I1#3               2.000      0.000    999.000    999.000
    I1#4               1.000      0.000    999.000    999.000
    I2#1               4.000      0.000    999.000    999.000
    I2#2               3.000      0.000    999.000    999.000
    I2#3               2.000      0.000    999.000    999.000
    I2#4               1.000      0.000    999.000    999.000
    I3#1               4.000      0.000    999.000    999.000
    I3#2               3.000      0.000    999.000    999.000
    I3#3               2.000      0.000    999.000    999.000
    I3#4               1.000      0.000    999.000    999.000
    I4#1               4.000      0.000    999.000    999.000
    I4#2               3.000      0.000    999.000    999.000
    I4#3               2.000      0.000    999.000    999.000
    I4#4               1.000      0.000    999.000    999.000

 Intercepts
    I1#1               4.836      0.425     11.378      0.000
    I1#2               5.113      0.413     12.384      0.000
    I1#3               4.484      0.390     11.501      0.000
    I1#4               3.069      0.366      8.389      0.000
    I2#1               6.954      0.564     12.330      0.000
    I2#2               6.336      0.554     11.444      0.000
    I2#3               5.404      0.528     10.237      0.000
    I2#4               3.144      0.517      6.084      0.000
    I3#1               3.774      0.362     10.436      0.000
    I3#2               4.174      0.348     11.993      0.000
    I3#3               4.050      0.320     12.643      0.000
    I3#4               2.465      0.300      8.223      0.000
    I4#1               2.374      0.295      8.037      0.000
    I4#2               2.677      0.278      9.646      0.000
    I4#3               2.683      0.247     10.851      0.000
    I4#4               1.958      0.215      9.106      0.000

 Variances
    TRAIT              1.020      0.128      7.976      0.000

 New/Additional Parameters
    B2_I1             -3.069      0.366     -8.389      0.000
    B3_I1             -1.415      0.182     -7.773      0.000
    B4_I1             -0.629      0.138     -4.550      0.000
    B5_I1              0.277      0.127      2.187      0.029
    B2_I2             -3.144      0.517     -6.084      0.000
    B3_I2             -2.259      0.250     -9.039      0.000
    B4_I2             -0.932      0.157     -5.935      0.000
    B5_I2             -0.618      0.122     -5.074      0.000
    B2_I3             -2.465      0.300     -8.223      0.000
    B3_I3             -1.585      0.173     -9.186      0.000
    B4_I3             -0.124      0.132     -0.938      0.348
    B5_I3              0.399      0.135      2.952      0.003
    B2_I4             -1.958      0.215     -9.106      0.000
    B3_I4             -0.726      0.153     -4.735      0.000
    B4_I4              0.006      0.143      0.041      0.967
    B5_I4              0.303      0.143      2.120      0.034
```

FIGURE 2 M*plus* output for partial credit model analysis of 4 items on the Math Self-Efficacy Scale.

```
  model:
  trait by I1#1*  I1#2 I1#3 I1#4 (L15 L14 L13 L12);
  trait by I2#1   I2#2 I2#3 I2#4 (L25 L24 L23 L22);
  trait by I3#1   I3#2 I3#3 I3#4 (L35 L34 L33 L32);
  trait by I4#1   I4#2 I4#3 I4#4 (L45 L44 L43 L42);
  trait@1;

  [I1#1] (Int5_I1);!label for NRM intercept for response category 5 on Item 1;
  [I1#2] (Int4_I1);!label for NRM intercept for response category 4 on Item 1;
  [I1#3] (Int3_I1);!label for NRM intercept for response category 3 on Item 1;
  [I1#4] (Int2_I1);!label for NRM intercept for response category 2 on Item 1;
  [I2#1] (Int5_I2);
  [I2#2] (Int4_I2);
  [I2#3] (Int3_I2);
  [I2#4] (Int2_I2);
  [I3#1] (Int5_I3);
  [I3#2] (Int4_I3);
  [I3#3] (Int3_I3);
  [I3#4] (Int2_I3);
  [I4#1] (Int5_I4);
  [I4#2] (Int4_I4);
  [I4#3] (Int3_I4);
  [I4#4] (Int2_I4);

  model constraint:
  new b2_I1 b2_I2 b2_I3 b2_I4 b3_I1 b3_I2 b3_I3 b3_I4
      b4_I1 b4_I2 b4_I3 b4_I4 b5_I1 b5_I2 b5_I3 b5_I4;

  L15=4*L12;
  L14=3*L12;
  L13=2*L12;
  L25=4*L22;
  L24=3*L22;
  L23=2*L22;
  L35=4*L32;
  L34=3*L32;
  L33=2*L32;
  L45=4*L42;
  L44=3*L42;
  L43=2*L42;

  b2_I1=-1*Int2_I1/L12;!GPCM step parameter for response category 2 on Item 1;
  b3_I1=-1*(Int3_I1-Int2_I1)/L12;!GPCM step parameter for response category 3 on
Item 1;
  b4_I1=-1*(Int4_I1-Int3_I1)/L12;!GPCM step parameter for response category 4 on
Item 1;
  b5_I1=-1*(Int5_I1-Int4_I1)/L12;!GPCM step parameter for response category 5 on
Item 1;

  b2_I2=-1*Int2_I2/L22;
  b3_I2=-1*(Int3_I2-Int2_I2)/L22;
  b4_I2=-1*(Int4_I2-Int3_I2)/L22;
  b5_I2=-1*(Int5_I2-Int4_I2)/L22;

  b2_I3=-1*Int2_I3/L32;
  b3_I3=-1*(Int3_I3-Int2_I3)/L32;
  b4_I3=-1*(Int4_I3-Int3_I3)/L32;
  b5_I3=-1*(Int5_I3-Int4_I3)/L32;

  b2_I4=-1*Int2_I4/L42;
  b3_I4=-1*(Int3_I4-Int2_I4)/L42;
  b4_I4=-1*(Int4_I4-Int3_I4)/L42;
  b5_I4=-1*(Int5_I4-Int4_I4)/L42;
```

FIGURE 3 M*plus* input for generalized partial credit model analysis of 4 items on the Math Self-Efficacy Scale.

that a person with $\theta = 2.68$ has an equal chance of selecting nominal Category 4 or nominal Category 1 on Item 4. Under "Variances," we can see that the estimated variance of the latent trait is $\hat{\sigma}^2 = 1.02$.

The final part of the selected output is under the "New/Additional Parameters" section. These are the PCM step parameters, and recall that they were named so that the subscript of b refers to the response category rather than the data score. For example, the first step parameter estimate of Item 2 has a label of b2_I2 and is $\hat{b}_{2(2)} = -3.14$. Therefore, a person with $\theta = -3.14$ has a 50% chance of transitioning from a response of 1 to a response of 2 on Item 2. As another example, the third step parameter estimate of Item 4 is $\hat{b}_{4(4)} = 0.01$. Therefore, a person with $\theta = 0.01$ has an equal chance of selecting either a response of 3 or 4 on Item 4.

ESTIMATING THE GPCM IN M*PLUS*

Figure 3 shows the M*plus* input commands for the GPCM. Recall that the slope for the k*th* response option in the NRM is $\lambda_{i(k)}$ and that in the GPCM $\lambda_{i(k)} = a_i\,(k-1)$. Therefore we need to construct the M*plus* code to (a) estimate a slope for response options 2 to K, (b) identify the slope for the second response option for an item as the a parameter for the item, and (c) constrain the slope for response options $k = 3, \ldots, K$ to be $k - 1$ times the slope for the second response option. These aims are accomplished by providing labels for the slopes in the "by" statements and by constraining the slopes in the "model constraint" section. Before turning to how the code accomplishes this aim, note that the first code element following the first "by" statement is I1#1*. If the asterisk were not included, the slope for the first data score

```
MODEL FIT INFORMATION

Number of Free Parameters                        20

Loglikelihood

          H0 Value                        -2662.346

Information Criteria

          Akaike (AIC)                     5364.693
          Bayesian (BIC)                   5449.846
          Sample-Size Adjusted BIC         5386.362
            (n* = (n + 2) / 24)

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 TRAIT    BY
    I1#1               2.545      0.333      7.645      0.000
    I1#2               1.909      0.250      7.645      0.000
    I1#3               1.273      0.166      7.645      0.000
    I1#4               0.636      0.083      7.645      0.000
    I2#1               8.063      1.397      5.772      0.000
    I2#2               6.047      1.048      5.772      0.000
    I2#3               4.032      0.698      5.772      0.000
    I2#4               2.016      0.349      5.772      0.000
    I3#1               3.731      0.473      7.889      0.000
    I3#2               2.798      0.355      7.889      0.000
    I3#3               1.866      0.236      7.889      0.000
    I3#4               0.933      0.118      7.889      0.000
    I4#1               4.500      0.611      7.361      0.000
    I4#2               3.375      0.459      7.361      0.000
    I4#3               2.250      0.306      7.361      0.000
    I4#4               1.125      0.153      7.361      0.000

 Intercepts
    I1#1               3.865      0.408      9.470      0.000
    I1#2               3.962      0.418      9.482      0.000
    I1#3               3.463      0.400      8.661      0.000
    I1#4               2.436      0.369      6.598      0.000
    I2#1              11.241      1.644      6.837      0.000
    I2#2              10.715      1.640      6.533      0.000
    I2#3               9.002      1.380      6.524      0.000
    I2#4               5.307      0.938      5.655      0.000
    I3#1               3.601      0.411      8.756      0.000
    I3#2               3.952      0.439      9.005      0.000
    I3#3               3.846      0.405      9.496      0.000
    I3#4               2.337      0.338      6.909      0.000
    I4#1               2.525      0.366      6.902      0.000
    I4#2               2.909      0.407      7.144      0.000
    I4#3               2.908      0.377      7.722      0.000
    I4#4               2.106      0.285      7.383      0.000

 Variances
    TRAIT              1.000      0.000    999.000    999.000

 New/Additional Parameters
    B2_I1             -3.827      0.626     -6.115      0.000
    B3_I1             -1.614      0.278     -5.802      0.000
    B4_I1             -0.785      0.212     -3.695      0.000
    B5_I1              0.152      0.185      0.823      0.411
    B2_I2             -2.633      0.303     -8.700      0.000
    B3_I2             -1.833      0.161    -11.413      0.000
    B4_I2             -0.850      0.094     -8.997      0.000
    B5_I2             -0.261      0.093     -2.798      0.005
    B2_I3             -2.506      0.336     -7.448      0.000
    B3_I3             -1.618      0.202     -8.024      0.000
    B4_I3             -0.114      0.139     -0.815      0.415
    B5_I3              0.377      0.142      2.647      0.008
    B2_I4             -1.872      0.206     -9.071      0.000
    B3_I4             -0.713      0.136     -5.231      0.000
    B4_I4             -0.001      0.130     -0.009      0.992
    B5_I4              0.341      0.135      2.538      0.011
```

FIGURE 4 M*plus* output for generalized partial credit model analysis of 4 items on the Math Self-Efficacy Scale.

for Item 1 would be set equal to 1; that is, it would not be estimated. Including the asterisk instructs M*plus* to estimate the slope, and the latent scale is then defined by the command trait@1, which sets the variance of trait equal to one.

There are four "by" statements, one for each item. Using multiple "by" statements rather than just one was done to organize the program. Following the list of indicators in each "by" statement are labels in parentheses. These serve as labels for the slopes (with the L representing lambda). For example L15, L14, L13, and L12 are the labels for the slopes for data scores 1 to 4 on Item 1. For example, L12 is the label for the slope of data score 4 on Item 1. Due to the reverse coding described earlier, L12 is the label for the slope of response Category 2. L12 is the label for the a parameter for Item 1 because according to Equation 11, slopes for response Category 2 are equal to $\lambda_{i(2)} = a_i\,(2-1) = a_i$. Similarly, L22, L32, and L42 are the labels for the a parameters of response Catgory 2 for Items 2, 3, and 4, respectively. For the remainder of the k categories, we need to multiply the slope of response category 2 by $k-1$, which is achieved under the model constraint section. The code that begins with L15=4*L12 and ends with L43=2*L42 constrains the slopes for the indicators such that $\lambda_{i(k)} = a_i\,(k-1)$. For example, L35=4*L32 constrains the slope for the fifth response option on Item 3 to be four times the slope for the second response option on Item 3.

All that remains in the input is to provide code that calculates the b parameters for the GPCM. This code is modified from the code used in the PCM. As an example, the command b3_I2= -1*(Int3_I2-Int2_I2)/L22 computes the b parameter for response option 3 on Item 2. As required by Equation 13, the difference between the intercepts for response options 3 and 2 on Item 2 is multiplied by –1 and divided by the discrimination parameter for Item 2. For the PCM, the corresponding expression is b3_I2= -1*(Int3_I2-Int2_I2). This code is consistent with the fact that the PCM is a special case of the GPCM in which the discrimination parameter is fixed to one for all items. As indicated by Equation 14, the first step parameter ($b_{i(2)}$) for each item needs only to be multiplied by –1 and divided by the a parameter of the item. For example, b2_I4= -1*Int2_I4/L42 estimates $b_{4(2)}$.

Figure 4 shows the M*plus* output for the GPCM code. The a_i parameters are associated with response Category 2, which is associated with data score 4. Hence the estimates for I1#4, I2#4, I3#4, and I4#4 are the a_i parameter estimates for Items 1, 2, 3, and 4, respectively. For example, the estimated a parameter for item 2 is $\hat{a}_2 = 2.02$. The GPCM step parameters are labeled in the same manner as the PCM step parameters in Figure 2. For example, the estimate of the first step parameter (associated with response Category 2) of Item 1 is $\hat{b}_{1(2)} = -3.83$. This indicates that an individual with $\theta = -3.83$ has an equal chance of responding in Category 1 or Category 2 of Item 1.

DEMONSTRATING THE APPROPRIATENESS OF THE M*PLUS* PCM CODE

A statistical package that directly estimates the PCM and GPCM and provides log-likelihood information is IRTPRO (Scientific Software International, 2011). In addition, SAS software (SAS Institute, 2006–2010) can directly estimate the PCM by programming Equations 4a and 4b and the log-likelihood in PROC NLMIXED. See, for example, Sheu, Chen, Su, and Wang (2005) or Hoffman (2010). The GPCM can similarly be estimated in SAS software. For comparison of results across programs, we used SAS 9.3 and IRTPRO version 2.1.21111.16001. We present comparative results of log-likelihoods, ability (θ) estimates, and step parameter estimates.

Table 1 shows the –2 log-likelihood values of the PCM and GPCM estimations from each of the three packages. M*plus* and SAS software have identical –2 log-likelihood values for the PCM, and IRTPRO's PCM value differs from M*plus* by only .03 units. For the GPCM, both SAS and IRTPRO differ from the M*plus* –2 log-likelihood value by –.01 units. The bivariate Pearson correlations between θ estimates from the two packages and M*plus* are shown in Table 2. When rounded to two decimal places, all correlations are statistically significant and observed as $r = 1.00$.

IRTPRO reports average item difficulty and thresholds for both the PCM and GPCM, so the step parameters were obtained by subtracting the latter from the former. Table 3 shows the PCM results in which all step parameter estimates from SAS software and IRTPRO are within |.02| units of the M*plus* step parameter estimates. Table 4 shows the

TABLE 1
–2 Log-Likelihood of Partial Credit Model (PCM) and Generalized Partial Credit Model (GPCM) With 4 Items on the Math Self-Efficacy Scale

Statistical Package	*–2 Log-Likelihood of PCM*	*–2 Log-Likelihood of GPCM*
M*plus*	5359.12	5324.70
SAS	5359.12	5324.69
IRTPRO	5359.15	5324.69

TABLE 2
Bivariate Pearson's Correlations Between θ Estimates From 4 Items on the Math Self-Efficacy Scale

Statistical Package	*Correlation With Mplus Partial Credit Model θ Estimates*	*Correlation With Mplus Generalized Partial Credit Model θ Estimates*
SAS	1.00*	1.00*
IRTPRO	1.00*	1.00*

* $p < .001$.

TABLE 3
Partial Credit Model Step Parameter Estimates of 4 Items on the Math Self-Efficacy Scale

	Package		
Step Parameter	*Mplus*	*SAS*	*IRTPRO*
Item 1 Step 1	−3.07	−3.07	−3.05
Item 1 Step 2	−1.42	−1.42	−1.40
Item 1 Step 3	−0.63	−0.63	−0.62
Item 1 Step 4	0.28	0.28	0.28
Item 2 Step 1	−3.14	−3.15	−3.13
Item 2 Step 2	−2.26	−2.26	−2.25
Item 2 Step 3	−0.93	−0.93	−0.93
Item 2 Step 4	−0.62	−0.62	−0.62
Item 3 Step 1	−2.47	−2.47	−2.46
Item 3 Step 2	−1.59	−1.58	−1.58
Item 3 Step 3	−0.12	−0.12	−0.12
Item 3 Step 4	0.40	0.40	0.40
Item 4 Step 1	−1.96	−1.96	−1.95
Item 4 Step 2	−0.73	−0.72	−0.72
Item 4 Step 3	0.01	0.01	0.01
Item 4 Step 4	0.30	0.30	0.30

TABLE 4
Generalized Partial Credit Model Step Parameter and Item Discrimination Estimates of 4 Items on the Math Self-Efficacy Scale

	Package		
Step Parameter	*Mplus*	*SAS*	*IRTPRO*
Item 1 Step 1	−3.83	−3.83	−3.83
Item 1 Step 2	−1.61	−1.62	−1.62
Item 1 Step 3	−0.79	−0.79	−0.79
Item 1 Step 4	0.15	0.15	0.15
Item 2 Step 1	−2.63	−2.63	−2.64
Item 2 Step 2	−1.83	−1.83	−1.84
Item 2 Step 3	−0.85	−0.85	−0.86
Item 2 Step 4	−0.26	−0.26	−0.27
Item 3 Step 1	−2.51	−2.51	−2.51
Item 3 Step 2	−1.62	−1.62	−1.62
Item 3 Step 3	−0.11	−0.12	−0.12
Item 3 Step 4	0.38	0.37	0.37
Item 4 Step 1	−1.87	−1.88	−1.87
Item 4 Step 2	−0.71	−0.71	−0.71
Item 4 Step 3	0.00	0.00	0.00
Item 4 Step 4	0.34	0.34	0.34
Item 1 discrimination	0.64	0.64	0.64
Item 2 discrimination	2.02	2.02	2.02
Item 3 discrimination	0.93	0.93	0.93
Item 4 discrimination	1.13	1.12	1.12

GPCM results in which all step parameter estimates from the two packages are within |.01| units of the M*plus* step parameter estimates. The a_i parameters of the GPCM from SAS software and IRTPRO are within |.01| units of the M*plus* a_i parameters.

CONCLUSION

The model presentations, M*plus* codes aligned with model constraints, and statistical package comparisons demonstrate that the M*plus* input commands in Figures 1 and 2 are viable ways to estimate the PCM and the GPCM. This provides M*plus* users with additional psychometric analysis options. Practitioners and researchers can now benefit from the advantages of being able to conduct more psychometric model comparisons within M*plus,* to implement the PCM and GPCM in a package that is typically available for students and known to produce accurate results, to obtain multiple pieces of model fit information of the PCM and GPCM, and to implement a variety of estimation options that are available in M*plus.* In addition, by aligning model constraint commands in M*plus* with the nested nature of the NRM, PCM, and GPCM, practitioners and students can become more familiar with the relationships between the various models used for polytomous item calibration.

REFERENCES

Betz, N., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior*, *23*, 329–345.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51.

Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33–49). New York, NY: Springer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Hoffman, L. (2010). IRT models in SAS NLMIXED [PDF document]. Retrieved from http://psych.unl.edu/psycrs/948_2011/15b_IRT_Models_in_SAS_NLMIXED.pdf

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York, NY: Wiley.

Langenfeld, T. E., & Pajares, F. (1993, April). *The mathematics self-efficacy scale: A validation study*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, *13*, 517–548.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, *19*, 91–100.

Muraki, E. M. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.

Muthén, B. O. (2008, July 4). Partial credit model ["I am not sure if the M*plus* Model Constraint feature could accomplish this, but it would probably be complicated"]. Retrieved from http://www.statmodel.com/discussion/messages/23/3360.html

Muthén, L. K. (2004, January 27). IRT models in M*plus* ["No, we don't do these models as far as I know."]. Retrieved from http://www.statmodel.com/discussion/messages/23/35.html

Muthén, L. K. (2007, October 17). IRT models in M*plus* ["This may be possible under MODEL CONSTRAINT"]. Retrieved from http://www.statmodel.com/discussion/messages/23/35.html

Muthén, L. K., & Muthén, B. O. (2012). *M*plus *user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danmarks Paedogogiske Institut.

SAS Institute. (2006–2010). *SAS enterprise guide version 4.3*. Cary, NC: Author.

Scientific Software International. (2011). *IRTPRO application*. Skokie, IL: Author.

Sheu, C. F., Chen, C. T., Su, Y. H., & Wang, W. C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods*, *37*, 202–218.

Templin, J. (2013). *IRT models for polytomous response data* [PDF document]. Retrieved from http://jtemplin.coe.uga.edu/files/irt/irt11icpsr/irt11icpsr_lecture04.pdf

Thissen, D. (1991). *MULTILOG, 6.0*. Chicago, IL: Scientific Software.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*, 546–577.