15

Modeling Change with Latent Variables Measured by Ordinal Indicators

In this chapter we combine concepts from the previous two chapters and discuss *item response models*, sometimes referred to as *nonlinear* or *item factor models*, and how they can be combined with longitudinal data analysis and growth models (Curran et al., 2008; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; Ram et al., 2005). As with the incorporation of the common factor model in longitudinal data analysis, we discuss how measurement invariance is tested with longitudinal item-level models. Testing measurement invariance with item response models has a separate history from testing measurement invariance with the common factor model (see Reise, Widaman, & Pugh, 1993). In item response modeling, the term *differential item functioning* (DIF; Thissen, Steinberg, & Gerrard, 1986) is used to refer to a lack of measurement equivalence at the item level (similarly, *differential test functioning* is used to refer to a lack of measurement equivalence at the test level) and is often described as testing for *item* (or *test*) *bias*. Regardless of the terms used, the idea of measurement equivalence or invariance is the same. That is, when measurement equivalence holds, differences (cross-sectionally) or changes (longitudinally) can be examined at the latent (factor or ability) level. We proceed with an examination of measurement invariance in much the same way as discussed in the prior chapter; however, we note that there are several approaches to test for measurement invariance in item response models and that testing for measurement invariance is often done on an item-by-item basis as opposed to studying whether a collection of items, as a whole, demonstrate measurement invariance (see Meredith, 1993; Meredith & Horn, 2001; Reise et al., 1993).

As discussed previously, modeling changes in common factors is a logical extension in the structural equation modeling framework, but not in the multilevel modeling framework and note that the multilevel modeling programs discussed here have difficulty fitting such models. We describe `Mplus` and `OpenMx` code for fitting these models to our illustrative data. As with longitudinal common factor models, `WinBUGS`, a freely available program that utilizes Bayesian estimation, is able to fit growth models that are combined with item response measurement models (see Grimm, Kuhl, & Zhang, 2013; McArdle et al., 2009).

ITEM RESPONSE MODELING

Item response modeling is concerned with understanding and modeling the item response process—that is, modeling the probability of a given response. Item response models, as they were originally developed, can be considered a restricted form of the common factor model for dichotomously and/or polytomously scored outcomes. In the history of common factor models, these same models were described as nonlinear factor models (McDonald, 1967) because the link between the common factor and the dichotomous/polytomous items was a nonlinear function (often the normal ogive or probit link). Item response models were originally developed with several assumptions. First, models assumed *unidimensionality*, such that there was a single latent trait (common factor) underlying all item responses; however, multidimensional models (see Wirth & Edwards, 2007) are now common. Second, item responses were assumed to be *conditionally independent*, such that item responses were uncorrelated after accounting for the latent trait. Third, the item response function was assumed to follow the *normal ogive* or *logistic* functions, which, as seen in Chapter 13, are elongated *s*-shaped curves. Currently, logistic curves are the most commonly used item response function because of estimation complexity with the normal ogive function. Commonly used item response models include the two-parameter logistic model (2PL; Birnbaum, 1968) for dichotomously scored responses and the graded response model (Samejima, 1969) for polytomously scored responses. These models, along with their normal ogive counterparts are subsequently described. Comprehensive overviews of item response models, theory, and approaches are found in Embretson and Reise (2000) and de Ayala (2009).

Dichotomous Response Models

The two-parameter logistic model can be written as

$$\Pr(y_{pi} = 1 | \theta_i, \alpha_p, \omega_p) = \frac{\exp(\alpha_p(\theta_i - \omega_p))}{1 + \exp(\alpha_p(\theta_i - \omega_p))} \tag{15.1}$$

where $\Pr(y_{pi} = 1|\theta_i, \alpha_p, \omega_p)$ is the probability of a correct response (i.e., scoring in the higher of two categories) to item p by individual i conditional upon person and item parameters, α_p is the discrimination parameter for item p, ω_p is the location (difficulty) parameter for item p, and θ_i is the latent trait (ability/aptitude/common factor score) for individual i. Figure 15.1 is a plot of three example item characteristic curves with different discrimination and location parameters. The discrimination parameters were set to 1.03, 0.80, and 2.20 (from the leftmost to the rightmost curve), and the location

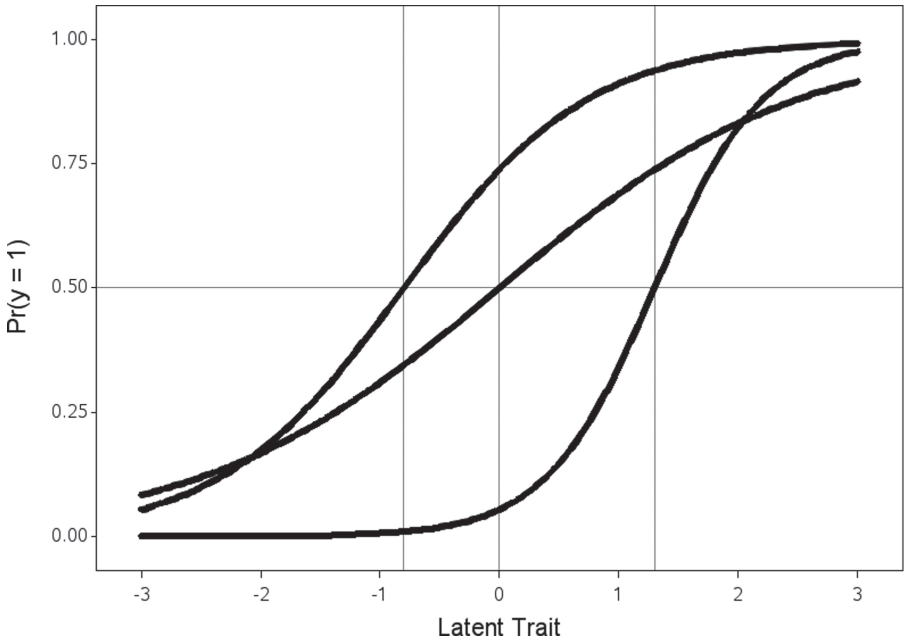

FIGURE 15.1. Item characteristic curves based on the two-parameter logistic model.

parameters were set to –0.80, 0.00, and 1.30, respectively. As seen in these plots, the location parameter refers to the point on the latent trait where an individual is predicted to have a 50% chance of correctly responding to the item. The discrimination parameter references the slope (divided by 4) of the tangent line when the latent trait and the location parameters are equal (i.e., $\theta_i = \omega_p$). Thus, the rate of change in the probability of a correct response when $\theta_i = \omega_p$ is $\alpha_p/4$. Items with greater discrimination parameters have greater changes in the probability for differences in the latent trait when θ_i is near ω_p, indicating that the item can discriminate between individuals with slightly different levels of the underlying ability.

The two-parameter normal ogive model (Lord, 1952) can be written as

$$\Pr(y_{pi} = 1|\eta_i, \lambda_p, \tau_p) = \mathbf{\Phi}(\lambda_p \cdot \eta_i - \tau_p) \tag{15.2}$$

where $\mathbf{\Phi}$ is the standard normal cumulative distribution function, λ_p is the slope parameter for item p, τ_p is the threshold parameter for item p, and η_i is the underlying latent trait for individual i. Alternatively, the two-parameter normal ogive model can be described using factor analytic terminology (Forero & Maydeu-Olivares, 2009), where the standard normal distribution is first divided by a single threshold, such that

$$y_{pi} = \begin{cases} 0 & if \;\; y_{pi}^{*} \leq \tau_p \\ 1 & if \;\; y_{pi}^{*} > \tau_p \end{cases} \tag{15.3}$$

where y_{pi} is the observed dichotomous response for individual i on item p, y^*_{pi} is the underlying response propensity for individual i for item p, and τ_p is the threshold for item p. The p underlying response propensities are then linearly related to the common factor, such that

$$y^*_{pi} = \lambda_p \cdot \eta_i + \epsilon_{pi} \tag{15.4}$$

where λ_p is the factor loading for item p, η_i is the common factor score for individual i, and ϵ_{pi} is the unique factor score of individual i for item p. Typically, the variances of the unique factors are fixed at 1.

Figure 15.2 is a plot of three example item characteristic curves with different slope (factor loading) and threshold parameters based on the two-parameter normal ogive model. The slope (factor loading) parameters were set to 0.76, 0.47, and 1.29 (discrimination parameters from the two-parameter logistic model divided by 1.702) and the threshold parameters were set to –0.81, 0.00, and 1.68 (location parameters from two-parameter logistic model multiplied by λ_p; see Kamata & Bauer, 2008), respectively. The item characteristic curves presented in Figure 15.2 are highly similar to the item characteristic curves from the two-parameter logistic model in Figure 15.1 showing the similarity between the models. Multiplying the slope parameters from the normal ogive model by 1.702 yields the discrimination parameters from the logistic model that are most similar. Based on this specification of the normal ogive model, the threshold parameter refers to the point on the

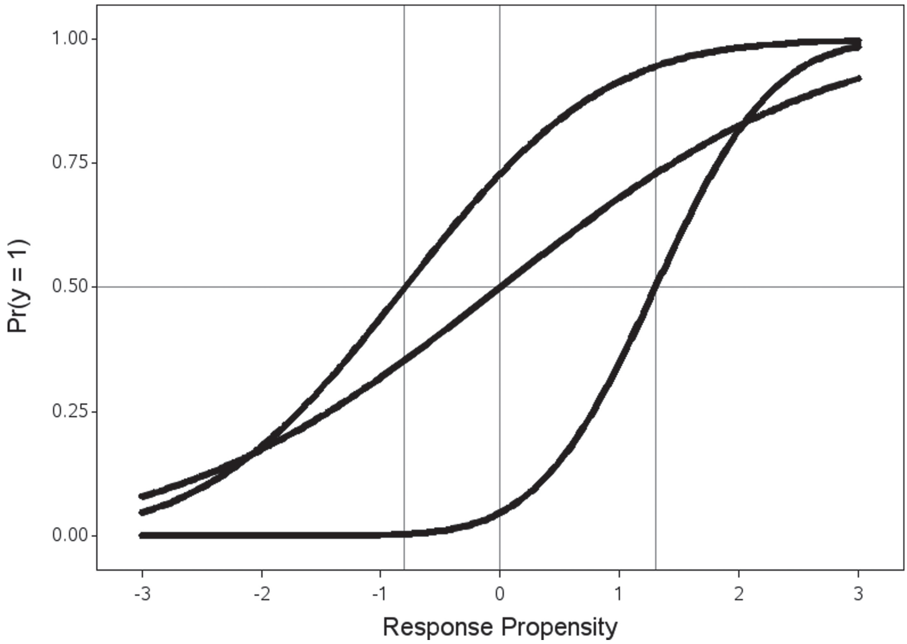

FIGURE 15.2. Item characteristic curves based on the two-parameter normal ogive model.

latent response propensity where an individual has a 50% chance of correctly responding to the item. The slope parameter references the rate of change when the latent response propensity equals the threshold. As before, items with greater slope parameters show greater changes in the probability for differences in the latent response propensity when η_i is near τ_p, indicating that the item can more easily discriminate between individuals with slightly different levels of the underlying ability.

Polytomous Item Response Models

There are several item response models appropriate for polytomously scored items (e.g., partial credit model [Masters, 1982]; generalized partial credit model [Muraki, 1992]; rating scale model; graded response model [Samejima, 1969]; modified graded response model). Here, we describe extensions of the two-parameter logistic and normal ogive models because these models can be fit with structural equation modeling programs (i.e., `Mplus` and `OpenMx`).

The graded response model (GRM; Samejima, 1969) is a straightforward extension of the two-parameter logistic model. The GRM can be written as

$$\Pr(y_{pi} \geq c | \theta_i, \alpha_p, \omega_{cp}) = \frac{\exp(\alpha_p(\theta_i - \omega_{cp}))}{1 + \exp(\alpha_p(\theta_i - \omega_{cp}))} \tag{15.5}$$

where $\Pr(y_{pi} \geq c|\theta_i, \alpha_p, \omega_{cp})$ is the probability of individual i responding in or above category c on item p conditional upon person and item parameters. As with the two-parameter logistic model, item parameters include α_p, a discrimination parameter for item p, and ω_{cp}, a location parameter separating category $c - 1$ from c on item p, and the only person parameter is θ_i, the latent trait for individual i.

Figure 15.3 contains operating characteristic curves (OCCs) for two five-category items with parameters $\alpha_1 = 2.4$, $\omega_{11} = -2.0$, $\omega_{21} = -1.0$, $\omega_{31} = 0.3$, and $\omega_{41} = 2.2$ for the first item and $\alpha_2 = 1.6$, $\omega_{12} = -2.3$, $\omega_{22} = -1.1$, $\omega_{32} = 0.0$, and $\omega_{42} = 1.1$ for the second item. The first curve in each plot represents the probability of responding in the second category or higher, the second curve represents the probability of responding in the third category or higher, and so on. First, we note that for each item the slopes of the curves are the same; however, the slopes of the curves can be different across items (items can be differentially related to the latent trait). Second, the spacing of the OCCs does not have to be the same within or between items, which brings up the idea that the spacing between response categories does not have to be equal. For example, the OCC in Figure 15.3a indicates that it takes a larger difference in the underlying latent trait to go from the fourth to the fifth response categories compared to the difference in the underlying latent trait to go from the second to third or the third to fourth response categories. Thus, the categories for this item are not equally spaced, whereas the categories are approximately equally spaced for the item described by the OCC in Figure 15.3b.

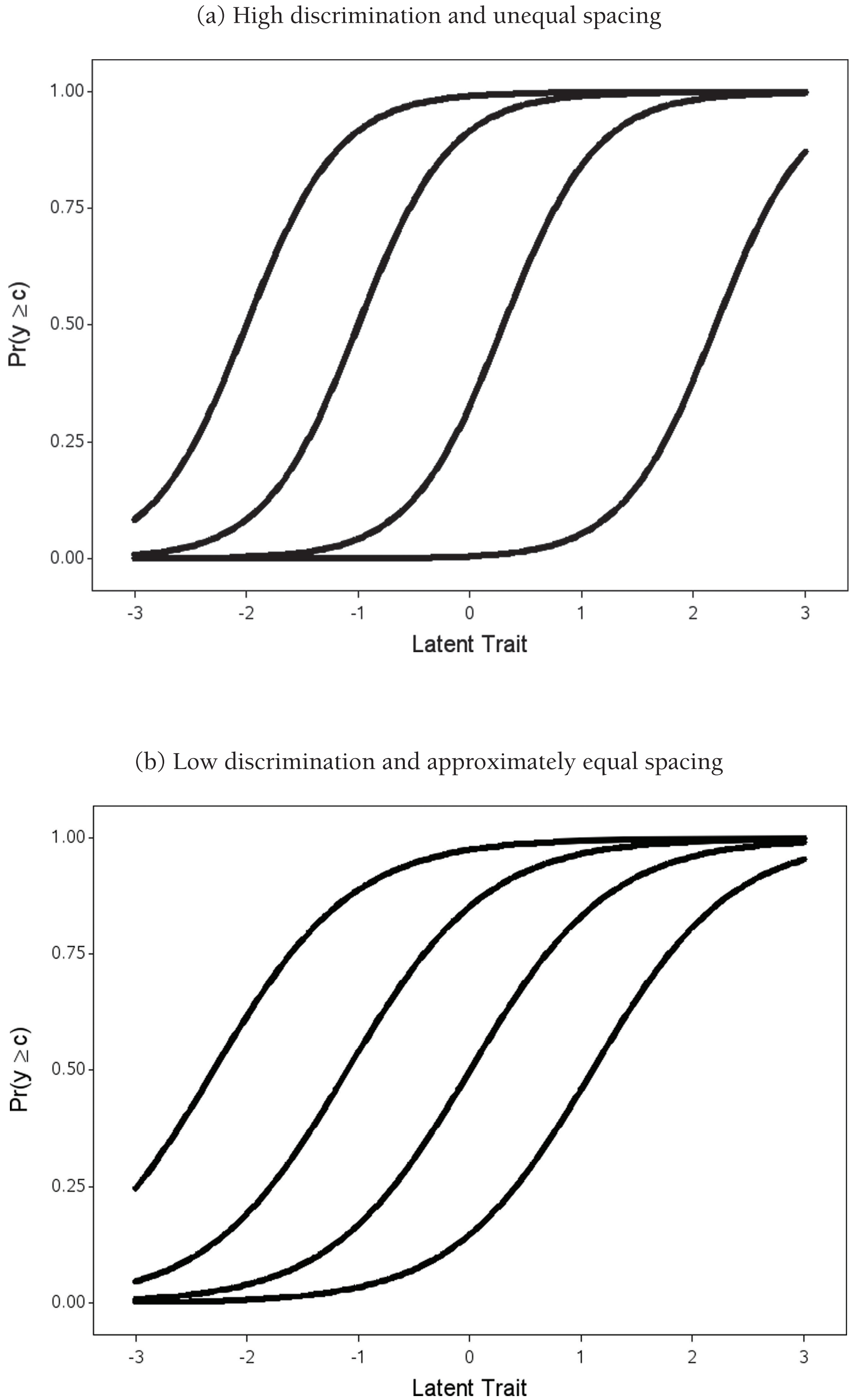

FIGURE 15.3. Example operating characteristic curves for two five-category items.

The probability of responding in each specific category ($\Pr(y_{pi} = c)$) is found through subtraction, such that

$$\Pr(y_{pi} = c|\theta_i, \alpha_p, \omega_{cp}) = \Pr(y_{pi} \geq c|\theta_i, \alpha_p, \omega_{cp}) - \Pr(y_{pi} \geq c + 1|\theta_i, \alpha_p, \omega_{c+1p}) \tag{15.6}$$

where $\Pr(y_{pi} \geq c|\theta_i, \alpha_p, \omega_{cp})$ is the probability of responding in category c or higher and $\Pr(y_{pi} \geq c + 1|\theta_i, \alpha_p, \omega_{c+1p})$ is the probability of responding in category $c + 1$ or higher. Additionally, the probability of responding in the first category (category 0) is $1 - \Pr(y_{pi} \geq 1|\theta_i, \alpha_p, \omega_{1p})$. Plotting the probability of responding in each category yields category response curves (CRCs), and these are plotted for the two example items in Figure 15.4. From these figures it is easy to see which category is the most likely response for each level of the latent trait. Ideally, and as seen in these figures, each category is most likely at some point along the latent trait. We note that a higher discrimination parameter leads to more peaked curves. Thus, the curves in Figure 15.4a are more peaked than the curves in Figure 15.4b.

The GRM can also utilize the probit link following the two-parameter normal ogive model, which is common in the structural equation modeling framework. In much the same way, the graded response normal ogive model is a cumulative probit model tracking the probability of responding in or above specific categories. This model can be written as

$$\Pr(y_{pi} \geq c|\eta_i, \lambda_p, \tau_{cp}) = \mathbf{\Phi}(\lambda_p \cdot \eta_i - \tau_{cp}) \tag{15.7}$$

where $\mathbf{\Phi}$ is the standard normal cumulative distribution function, τ_{cp} is the threshold parameter for category c of item p, λ_p is the slope parameter for item p, and η_i is the underlying latent variable measured for individual i. Equation 15.7 leads to OCCs describing the relation between the response propensity and the probability of responding in or above each category (similar to the GRM with the logit link).

As above, the graded response normal ogive model can be described using a factor analytic model where the standard normal distribution is first divided by multiple thresholds. For example, the thresholds for a five-category item would be written as

$$y_{pi} = \begin{cases} 0 & \text{if} \quad y^*_{pi} \leq \tau_{1p} \\ 1 & \text{if} \quad \tau_{1p} < y^*_{pi} \leq \tau_{2p} \\ 2 & \text{if} \quad \tau_{2p} < y^*_{pi} \leq \tau_{3p} \\ 3 & \text{if} \quad \tau_{3p} < y^*_{pi} \leq \tau_{4p} \\ 4 & \text{if} \quad y^*_{pi} > \tau_{4p} \end{cases} \tag{15.8}$$

where y_{pi} is the observed dichotomous response for individual i on item p, y^*_{pi} is the underlying response propensity for individual i on item p, and τ_{1p} through τ_{4p} are the

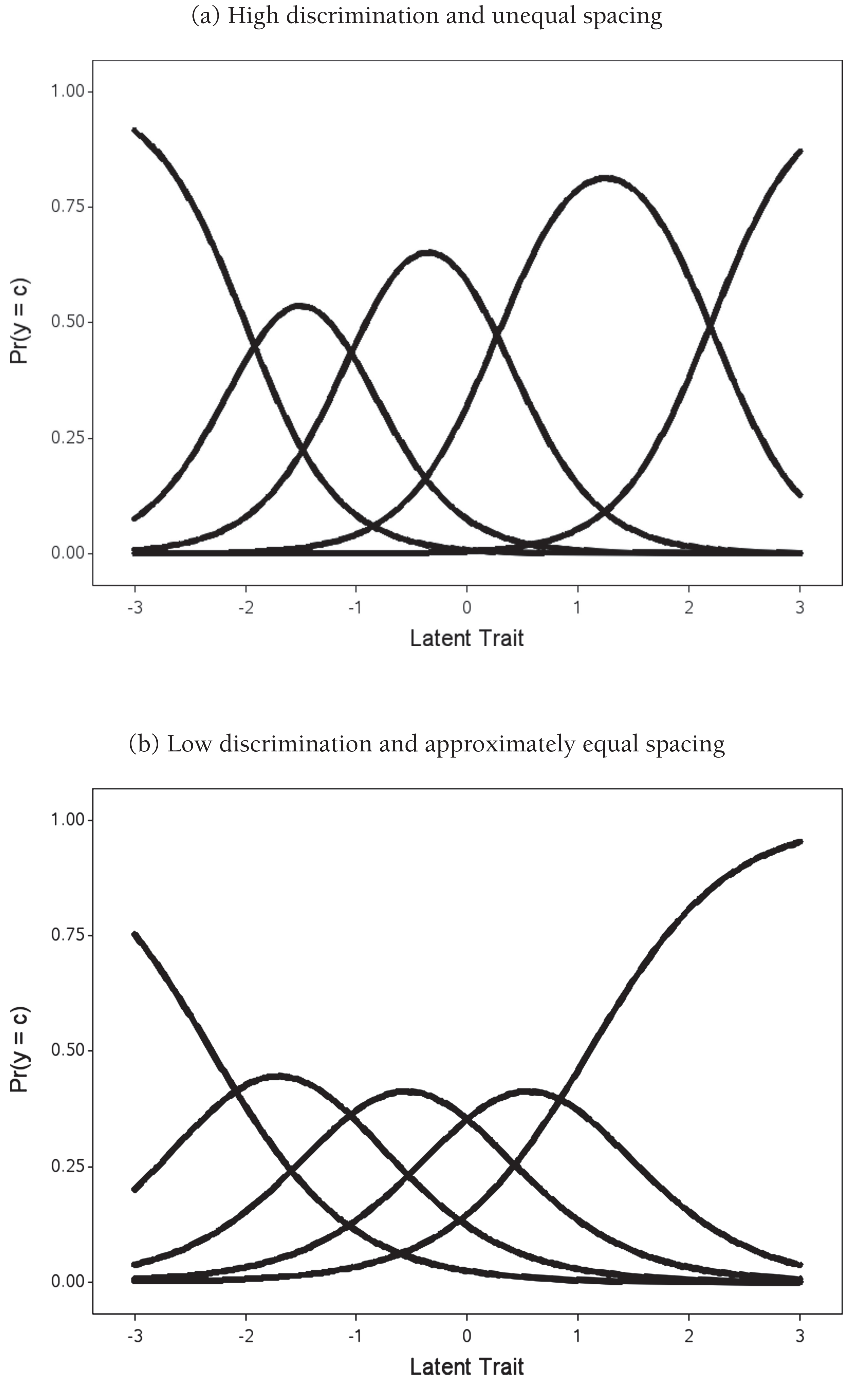

FIGURE 15.4. Example category response curves for two five-category items.

four thresholds for item p, such that $\tau_{1p} < \tau_{2p} < \tau_{3p} < \tau_{4p}$. More generally, this can be specified as

$$y_{pi} = c \quad \text{if} \quad \tau_{cp} < y^*_{pi} < \tau_{c+1p} \tag{15.9}$$

where $\tau_{0p} = -\infty$ and $\tau_{mp} = +\infty$, where m is the number of response categories, which bounds the observed response to the available response categories (Forero & Maydeu-Olivares, 2009). As in the two-parameter normal ogive model, the p underlying response propensities are linearly related to the common factor such that

$$y^*_{pi} = \lambda_p \cdot \eta_i + \epsilon_{pi} \tag{15.10}$$

where λ_p is the factor loading for item p, η_i is the common factor score for individual i, and ϵ_{pi} is the unique factor score of individual i for item p. As with the two-parameter normal ogive model, the variances of the unique factors are typically fixed at 1.

Measurement Invariance in Item Response Models

The item response models described above can and have been extended to longitudinal data (McArdle et al., 2009; Ram et al., 2005). With longitudinal data, the latent trait or underlying response propensity is time dependent (i.e., θ_{ti}, y^*_{pti}, and η_{ti}) as are the item parameters (i.e., α_{pt} and ω_{cpt} from the logistic model and λ_{pt} and τ_{cpt} from the normal ogive model). A first question in longitudinal item response modeling, much like longitudinal factor analysis, is whether the response process is identical at each measurement occasion—specifically, whether or not the item parameters are invariant over measurement occasions. If item parameters are invariant, then changes in the latent trait or common factor can be modeled. If item parameters are not invariant, then the response process has changed and changes in the latent trait or common factor cannot be modeled without some accommodation (e.g., partial measurement invariance).

Differential item functioning, the study of measurement invariance with item response models, is most often conducted with a grouping variable (e.g., gender) to examine whether the groups have different probabilities of responding in each response category (e.g., probability of correctly answering the question), controlling for differences in the underlying latent trait or latent response propensity. The same idea is carried with longitudinal data. That is, controlling for differences in the underlying latent trait or common factor over time, the same probability of each response category should be observed. This is not to say that the probability of a given response must stay the same over time. It is that the probability should remain the same for individuals measured at different measurement occasions who have the same underlying latent trait or common factor score. These ideas relate directly to the invariance of the item characteristic curves over measurement occasion, which maps onto the invariance of the discrimination (factor loading) and location (threshold) parameters over time.

For simplicity, let's consider that the item is from a math test and that the latent trait represents math ability. In order to examine changes in true math ability, the item characteristic curves must be identical across measurement occasions. Thus, the item parameters must be invariant. If item parameters are different at each measurement occasion, then a person (or two different people) with the same math ability (e.g., 0) at the two measurement occasions would have a different probability of correctly answering the question. In this case, the way in which the item is related to the latent trait has changed, and if the properties of the item have changed, it is impossible to determine how the people changed. For example, Figures 15.5a and 15.5b contain item characteristic curves (based on a logistic response function) for the same item at two measurement occasions (e.g., time 1 and time 2). From these figures, it is easy to see that the probability of a correct response differs at the two measurement occasions for individuals who have the same underlying ability. That is, an individual with a low latent trait (< –0.60) has a higher probability of correctly answering the question at the second measurement occasion, whereas a person with a higher latent trait (> 0.60) has a higher probability of correctly answering the question at the first measurement occasion. Thus, the scales of the two latent traits cannot be directly compared because they relate to different response probabilities at each occasion and therefore, the changes in the latent trait cannot be examined.

Testing for measurement invariance in item response models can be carried out in several different ways (see Muthén, 1985; Muthén & Lehman, 1985; Woods, 2011). However, we take an approach that is similar to how measurement invariance was examined with the common factor model in Chapter 14. That is, in a first step, a configural invariance model is fit with no constraints on the item response parameters at each measurement occasion, and the latent traits or factor scores measured at each measurement occasion are allowed to covary. In the second step, the weak invariance model is fit where equality constraints are imposed on the discrimination ($\alpha_{pt} = \alpha_p$) or factor loading ($\lambda_{pt} = \lambda_p$) parameters, and the change in model fit is evaluated. In the third step, the strong invariance model is fit where the location ($\omega_{cpt} = \omega_{cp}$) or threshold ($\tau_{cpt} = \tau_{cp}$) parameters are constrained to be equal and the means of the latent traits or factor scores are estimated at all measurement occasions, except for the first measurement occasion where the mean is fixed, often at 0, for identification.

Item response models are often fit with more indicators than common factor models with continuous indicators simply because there are more items that are indicative of a latent construct than scores from surveys or tests (that are themselves often composed of multiple items). Given the number of items, it is unlikely that full measurement invariance will hold for all items. In these situations, partial measurement invariance, where item parameters are invariant for a collection of items and noninvariant for another collection of items, is a reasonable solution and should be examined. However, when full measurement invariance does not hold, but partial measurement invariance holds, it is important to note that the *changes* in the underlying latent trait are only based on the items for which measurement invariance holds even though items with noninvariant parameters aid the identification of the latent trait or factor score at each measurement occasion. Following this idea, it is also possible to have different (nonoverlapping) items measuring the latent construct at each measurement occasion as long as measurement invariance holds for the

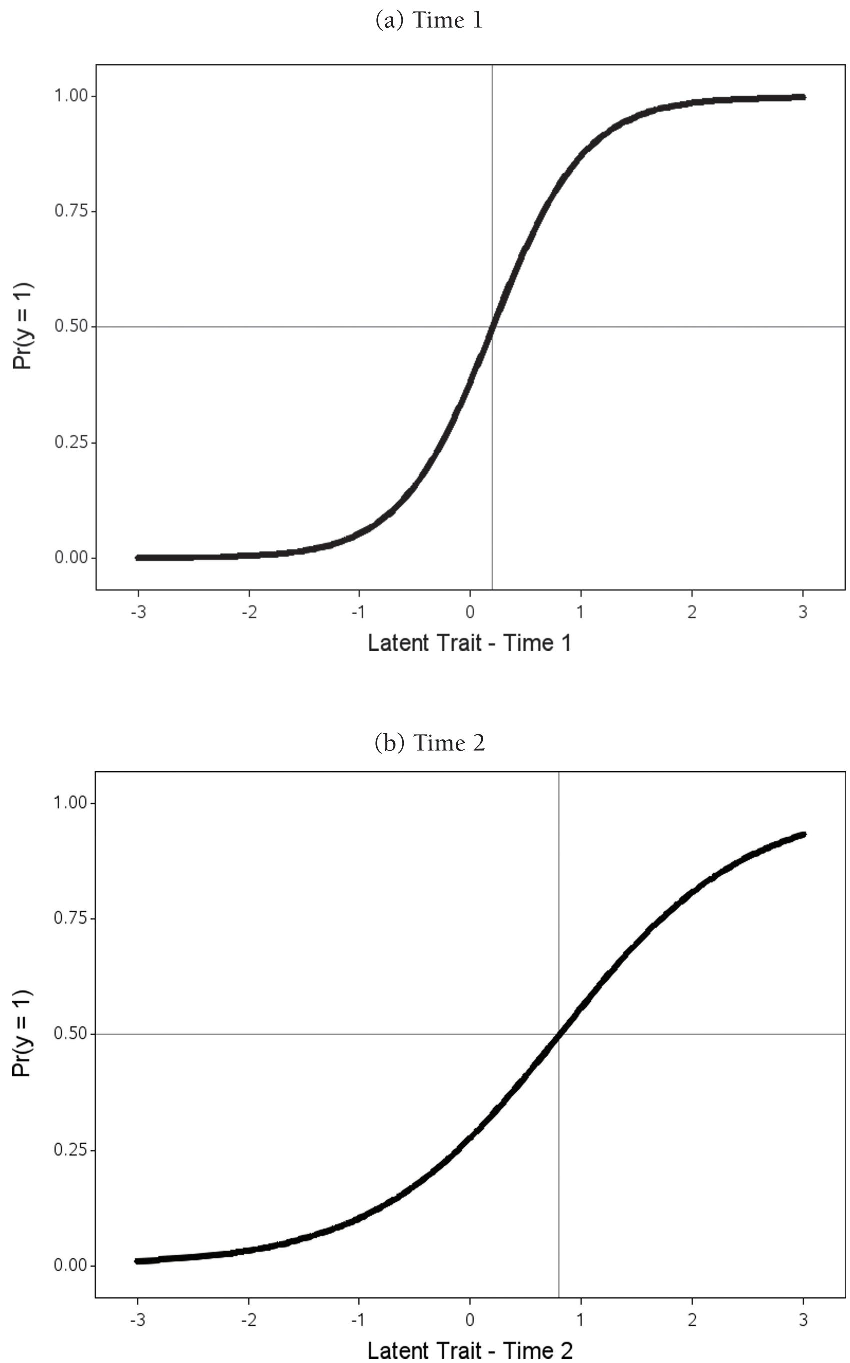

FIGURE 15.5. Example item characteristic curves for a single item at two occasions with different parameters.

overlapping items, which can also vary over measurement occasions (Edwards & Wirth, 2009; McArdle et al., 2009).

SECOND-ORDER GROWTH MODEL

Growth models can be built on item response models (when measurement invariance holds) by replacing the usual observed outcome with the latent trait or factor score measured at each occasion. For example and building on the logistic item response models, a linear growth model for the latent trait can be written as

$$\theta_{ti} = b_{1i} + b_{2i} \cdot t + d_{ti} \tag{15.11}$$

where θ_{ti} is the latent trait for individual i measured at time t, b_{1i} is the intercept or predicted value of θ_{ti} at $t = 0$ for individual i, b_{2i} is the linear slope for individual i relating individual changes in θ_{ti} with changes in t, the chosen time metric, and d_{ti} is the time-dependent disturbance term for individual i at time t. The intercept and linear slope are assumed to be normally distributed with means, variances, and a covariance. The mean of the intercept is fixed at 0 for identification (as in the second-order growth model based on the common factor model and growth models with ordinal indicators), but the mean of the linear slope is freely estimated as well as the intercept and slope variances and their covariance.

Working from the structural equation modeling framework and the probit model, the second-order growth model can be specified as

$$\eta_i = \mathbf{\Gamma}\xi_i + \upsilon_i \tag{15.12}$$

where η_i is a $T \times 1$ vector of common factors for individual i, $\mathbf{\Gamma}$ is a $T \times R$ factor loading matrix defining the latent growth factors (e.g., intercept and linear slope in a linear growth model), ξ_i is an $R \times 1$ vector of second-order latent growth factors, and υ_i is a $T \times 1$ vector of latent variable disturbance terms. The second-order factors are assumed to follow a multivariate normal distribution, such that $\xi_i \sim MVN(\kappa, \mathbf{\Phi})$, where κ is an $R \times 1$ vector of latent variable means and $\mathbf{\Phi}$ is an $R \times R$ covariance matrix. The first-order factor disturbances (υ_i) are assumed to follow a normal distribution, such that $\upsilon_i \sim N(\mathbf{0}, \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a $T \times T$ diagonal disturbance covariance matrix.

ILLUSTRATION

To illustrate the use of the longitudinal item response model and second-order growth model, we use a random sample of observations from the ECLS-K. In the fall and spring of kindergarten as well as the spring of first grade, teachers completed surveys regarding the children's behavior. Specifically, teachers were asked about the children's level of *interpersonal skills*, *self-control*, and *externalizing* behaviors using a four-point (1–4) scale where higher values indicated more of the behavior. The ECLS-K reports average ratings for each behavior, which we have rounded to the nearest whole number to show how item response models can be incorporated into the study of change. The variables include `t1_sc`, `t1_intp`, and `t1_ext` for the self-control, interpersonal skills, and externalizing behaviors, respectively, measured in the fall of kindergarten. Variables collected in the spring of kindergarten and first-grade begin with `t2` and `t4`, respectively.

STRUCTURAL EQUATION MODELING IMPLEMENTATION

Longitudinal Item Factor Model

M*plus*

The M*plus* script for a *longitudinal item factor* model with strong factorial invariance fit to the behavior data is contained in Script 15.1. The script begins with the typical components of a title, datafile, and variable names. In the `VARIABLE` command, the `CATEGORICAL` option is included, and all variables in the analysis are listed here to treat the variables as being ordinal. The `ANALYSIS` command lists `TYPE=MEANSTRUCTURE` with `ESTIMATOR=ML` and `LINK=PROBIT` to use maximum likelihood estimation and the probit link.

The `MODEL` statement begins by defining three latent variables, `eta_1`, `eta_2`, and `eta_3`, which are the latent traits or factors (η_{ti}) measured in the fall of kindergarten, spring of kindergarten, and spring of first grade, respectively. The factor loadings to the self-control items (`t1_sc`, `t2_sc`, and `t4_sc`) are fixed at 1 for identification, and the remaining factor loadings are estimated but constrained to be equal over time by using common labels. For example, the three externalizing items (`t1_ext`, `t2_ext`, and `t3_ext`) have factor loadings that are labeled `lambda_e`. Next, variances and covariances are specified for the latent variables (`eta_1 eta_2 eta_3;` and `eta_1 WITH eta_2 eta_3;` and `eta_2 WITH eta_3;`), which are all freely estimated. The means of the latent variables are then specified. The mean of `eta_1`, the latent variable at the first occasion, is fixed at 0 for identification, but the means of `eta_2` and `eta_3` are estimated. Thresholds are then specified for all items and common labels are supplied for the same threshold for the same item across time. Thus, the first threshold for the self-control items (`t1_sc$1`, `t2_sc$1`, and `t4_sc$1`) are labeled `tau_s1`.

The configural and the weak invariance models can be specified by modifying this script. The weak invariance model is fit by removing the labels on the thresholds (`tau_s1` through `tau_e3`) to allow them to be freely estimated and fixing the means of the latent variables to be 0 (e.g., `[eta_1@0 eta_2@0 eta_3@0];`). The configural invariance model is fit by additionally removing the labels associated with the factor loadings (`lambda_i` and `lambda_e`).

OpenMx

The `OpenMx` script for the longitudinal item factor model with strong factorial invariance is contained in Script 15.2. The script begins by defining the observed variables as ordinal using `mxFactor` and stating the response categories for each item. The model

Script 15.1. *Mplus* Script for the Longitudinal Item Factor Model with Strong Factorial Invariance

```
TITLE: Longitudinal Categorical Factor Model - Strong Invariance;
DATA: FILE= ECLS_Behavior.dat;
VARIABLE:
  NAMES = id t1_sc t1_intp t1_ext t2_sc t2_intp t2_ext
             t4_sc t4_intp t4_ext;
  MISSING = .;
  USEVAR = t1_sc t1_intp t1_ext t2_sc t2_intp t2_ext
           t4_sc t4_intp t4_ext;
  CATEGORICAL = t1_sc t1_intp t1_ext t2_sc t2_intp t2_ext
                t4_sc t4_intp t4_ext;

ANALYSIS: TYPE=MEANSTRUCTURE; ESTIMATOR=ML; LINK=PROBIT;

MODEL:
! Defining Latent Variables
  eta_1 BY t1_sc@1
           t1_intp (lambda_i)
           t1_ext  (lambda_e);
  eta_2 BY t2_sc@1
           t2_intp (lambda_i)
           t2_ext  (lambda_e);
  eta_3 BY t4_sc@1
           t4_intp (lambda_i)
           t4_ext  (lambda_e);

! Latent Variable Variances & Covariances
  eta_1 eta_2 eta_3;
  eta_1 WITH eta_2 eta_3;
  eta_2 WITH eta_3;

! Latent Variable Means
  [eta_1@0 eta_2*0 eta_3*0];

! Observed Variable Thresholds
  [t1_sc$1 t2_sc$1 t4_sc$1] (tau_s1);
  [t1_sc$2 t2_sc$2 t4_sc$2] (tau_s2);
  [t1_sc$3 t2_sc$3 t4_sc$3] (tau_s3);

  [t1_intp$1 t2_intp$1 t4_intp$1] (tau_i1);
  [t1_intp$2 t2_intp$2 t4_intp$2] (tau_i2);
  [t1_intp$3 t2_intp$3 t4_intp$3] (tau_i3);

  [t1_ext$1 t2_ext$1 t4_ext$1] (tau_e1);
  [t1_ext$2 t2_ext$2 t4_ext$2] (tau_e2);
  [t1_ext$3 t2_ext$3 t4_ext$3] (tau_e3);

OUTPUT: RESIDUAL STANDARDIZED;
```

Script 15.2. `OpenMx` **Script for the Longitudinal Item Factor Model with Strong Factorial Invariance**

```
(Converting Items of Interest to Categorical Variables
ecls$t1_sc   <- mxFactor(ecls$t1_sc,   levels=c(1,2,3,4))
ecls$t1_intp <- mxFactor(ecls$t1_intp, levels=c(1,2,3,4))
ecls$t1_ext  <- mxFactor(ecls$t1_ext,  levels=c(1,2,3,4))
ecls$t2_sc   <- mxFactor(ecls$t2_sc,   levels=c(1,2,3,4))
ecls$t2_intp <- mxFactor(ecls$t2_intp, levels=c(1,2,3,4))
ecls$t2_ext  <- mxFactor(ecls$t2_ext,  levels=c(1,2,3,4))
ecls$t4_sc   <- mxFactor(ecls$t4_sc,   levels=c(1,2,3,4))
ecls$t4_intp <- mxFactor(ecls$t4_intp, levels=c(1,2,3,4))
ecls$t4_ext  <- mxFactor(ecls$t4_ext,  levels=c(1,2,3,4))

# Strong Factorial Invariance Model
lirt.strong.omx <- mxModel('Strong Invariance Model, Path Specification',
 type='RAM', mxData(observed=ecls, type='raw'),
 manifestVars=c('t1_sc','t1_intp','t1_ext','t2_sc','t2_intp','t2_ext','t4_sc','t4_intp','t4_ext'),
 latentVars=c('eta_1','eta_2','eta_3'),

# Residual Variances
mxPath(from=c('t1_sc','t1_intp','t1_ext','t2_sc','t2_intp','t2_ext','t4_sc','t4_intp','t4_ext'),
 arrows=2, free=FALSE, values=1),

# Latent Variable Covariances
mxPath(from=c('eta_1','eta_2','eta_3'),
 connect='unique.pairs', arrows=2,
 free=TRUE, values=c(5,3,3,5,3,5),
 labels=c('psi_11','psi_21','psi_31','psi_22','psi_32','psi_33')),
```

(continued)

Script 15.2. (*Continued*)

```
# Factor Loadings
mxPath(from='eta_1', to=c('t1_sc','t1_intp','t1_ext'),
 arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),

mxPath(from='eta_2', to=c('t2_sc','t2_intp','t2_ext'),
 arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),

mxPath(from='eta_3', to=c('t4_sc','t4_intp','t4_ext'),
 arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),

# Latent Variable Means
mxPath(from='one', to=c('eta_2','eta_3'),
 arrows=1, free=TRUE,
 values=.5, labels=c('alpha_2','alpha_3')),

# Thresholds
mxThreshold(vars=c('t1_sc','t2_sc','t4_sc'), nThresh=3, free=TRUE, values=c(-7,-3,1),
 labels=c('tau_s1','tau_s2','tau_s3')),

mxThreshold(vars=c('t1_intp','t2_intp','t4_intp'), nThresh=3, free=TRUE, values=c(-5,-1,1),
 labels=c('tau_i1','tau_i2','tau_i3')),

mxThreshold(vars=c('t1_ext','t2_ext','t4_ext'), nThresh=3, free=TRUE, values=c(0,2,4),
 labels=c('tau_e1','tau_e2','tau_e3'))

) # Close Model
```

is then specified using mxModel. The mxModel command begins by stating the datafile along with the manifest and latent variables included in the model. A series of mxPath statements follow to specify the model components. The first mxPath statement is for the residual variances of the ordinal items. These two-headed arrows (arrows=2) begin from each observed variable and are fixed (free=FALSE) at 1 (values=1) following the specification of the normal ogive model. The next mxPath statement is for the latent variable variances and covariances. In the strong factorial invariance model, the latent variables (eta_1, eta_2, and eta_3) have estimated variances and estimated covariances. To do this, we use connect='unique.pairs' to specify all variances and covariances of the latent variables in a single mxPath statement.

The factor loadings for the three latent variables are then specified in the next three mxPath statements. For each factor, the factor loading to the self-control item (t1_sc, t2_sc, and t4_sc) is fixed at 1 for identification. The remaining factor loadings are freely estimated but constrained to be equal across time. For example, factor loadings for t1_intp, t2_intp, and t4_intp are constrained to be equal to one another, as are the factor loadings for t1_ext, t2_ext, and t4_ext. As with prior specifications, a common label is used for equality constraints (labels=c(NA, 'lambda_i', 'lambda_e')). Next, the latent variable means are specified. The means of the eta_2 and eta_3 factors are estimated because this is the strong invariance model and the thresholds (specified next) are constrained to be equal. The means are given starting values of 0 and labeled alpha_2 and alpha_3. The thresholds are specified next using three mxThreshold statements. To simplify the mxThreshold statements, we have grouped them according to their respective variable. The first mxThreshold statement is for the three thresholds for the self-control items, which are freely estimated, given starting values, and labeled tau_s1, tau_s2, and tau_s3. Thus, the three thresholds for this item are constrained to be equal following the strong invariance model. Similar mxThreshold statements are then specified for the interpersonal skills and externalizing items. The model is then closed.

The above script can be modified to specify the configural and weak invariance models. For the weak invariance model, the means of the latent variables eta_1, eta_2, and eta_3 are fixed to 0, and the equality constraints imposed on the threshold parameters are removed. Thus, in the mxPath statement for the latent variable means, free is set equal to FALSE and the labels are removed. In the mxThreshold statements, the labels are changed to be specific to the measurement occasion (e.g., tau_s11 through tau_s13 for the three thresholds for the self-control item measured at the first occasion). For the configural invariance model, the equality constraints on the factor loadings are also removed. Thus, in the mxPath statements for the factor loadings, the labels are made to be occasion-specific (e.g., lambda_i1 and lambda_e1 for the t1_intp and t1_ext items at the first measurement occasion).

Output

Output for the longitudinal item factor model with strong factorial invariance from M*plus* and OpenMx is contained in Output 15.1 and 15.2, respectively. As a first point of

Output 15.1. *Mplus* Output for the Longitudinal Item Factor Model with Strong Factorial Invariance

```
MODEL RESULTS
                                                         Two-Tailed
             Estimate       S.E.    Est./S.E.             P-Value
ETA_1     BY
   T1_SC        1.000      0.000      999.000             999.000
   T1_INTP      0.614      0.029       20.969               0.000
   T1_EXT      -0.462      0.022      -20.980               0.000
ETA_2     BY
   T2_SC        1.000      0.000      999.000             999.000
   T2_INTP      0.614      0.029       20.969               0.000
   T2_EXT      -0.462      0.022      -20.980               0.000
ETA_3     BY
   T4_SC        1.000      0.000      999.000             999.000
   T4_INTP      0.614      0.029       20.969               0.000
   T4_EXT      -0.462      0.022      -20.980               0.000
ETA_1     WITH
   ETA_2        5.727      0.465       12.328               0.000
   ETA_3        3.827      0.326       11.727               0.000
ETA_2     WITH
   ETA_3        4.567      0.381       11.974               0.000
Means
   ETA_1        0.000      0.000      999.000             999.000
   ETA_2        0.423      0.049        8.634               0.000
   ETA_3        0.306      0.058        5.299               0.000
Thresholds
   T1_SC$1     -7.813      0.308      -25.400               0.000
   T1_SC$2     -2.967      0.116      -25.675               0.000
   T1_SC$3      0.926      0.063       14.738               0.000
   T1_INTP$1   -4.871      0.112      -43.372               0.000
   T1_INTP$2   -1.514      0.042      -36.411               0.000
   T1_INTP$3    1.090      0.039       27.749               0.000
   T1_EIT$1    -0.167      0.027       -6.169               0.000
   T1_EIT$2     2.011      0.040       50.837               0.000
   T1_EIT$3     3.374      0.062       54.551               0.000
   T2_SC$1     -7.813      0.308      -25.400               0.000
   T2_SC$2     -2.967      0.116      -25.675               0.000
   T2_SC$3      0.926      0.063       14.738               0.000
   T2_INTP$1   -4.871      0.112      -43.372               0.000
   T2_INTP$2   -1.514      0.042      -36.411               0.000
   T2_INTP$3    1.090      0.039       27.749               0.000
   T2_EXT$1    -0.167      0.027       -6.169               0.000
   T2_EXT$2     2.011      0.040       50.837               0.000
   T2_EXT$3     3.374      0.062       54.551               0.000
   T4_SC$1     -7.813      0.308      -25.400               0.000
   T4_SC$2     -2.967      0.116      -25.675               0.000
   T4_SC$3      0.926      0.063       14.738               0.000
   T4_INTP$1   -4.871      0.112      -43.372               0.000
   T4_INTP$2   -1.514      0.042      -36.411               0.000
```

```
   T4_INTP$3     1.090     0.039        27.749            0.000
   T4_EXT$1     -0.167     0.027        -6.169            0.000
   T4_EXT$2      2.011     0.040        50.837            0.000
   T4_EXT$3      3.374     0.062        54.551            0.000
Variances
   ETA_1         6.890     0.584        11.794            0.000
   ETA_2         7.928     0.669        11.850            0.000
   ETA_3         8.001     0.666        12.016            0.000
```

Output 15.2. `OpenMx` Output for the Longitudinal Item Factor Model with Strong Factorial Invariance

```
free parameters:
   name         matrix        row        col      Estimate      Std.Error
1  lambda_i          A    t1_intp      eta_1     0.6122294    0.010915174
2  lambda_e          A     t1_ext      eta_1    -0.4600890    0.005170421
3  psi_11            S      eta_1      eta_1     6.9290607    0.015976345
4  psi_21            S      eta_1      eta_2     5.7565299    0.009902889
5  psi_22            S      eta_2      eta_2     7.9742622    0.030190101
6  psi_31            S      eta_1      eta_3     3.8449325    0.005789941
7  psi_32            S      eta_2      eta_3     4.5955841    0.010414252
8  psi_33            S      eta_3      eta_3     8.0496592    0.016391002
9  alpha_2           M          1      eta_2     0.4238295    0.002193305
10 alpha_3           M          1      eta_3     0.3072734    0.002269799
11 tau_s1  Thresholds           1      t1_sc    -7.8299435    0.151922879
12 tau_s2  Thresholds           2      t1_sc    -2.9724856    0.007081997
13 tau_s3  Thresholds           3      t1_sc     0.9292774    0.010708992
14 tau_i1  Thresholds           1    t1_intp    -4.8688107    0.094284898
15 tau_i2  Thresholds           2    t1_intp    -1.5128191    0.003817751
16 tau_i3  Thresholds           3    t1_intp     1.0899793    0.004588982
17 tau_e1  Thresholds           1     t1_ext    -0.1674869    0.006046656
18 tau_e2  Thresholds           2     t1_ext     2.0100857    0.003166900
19 tau_e3  Thresholds           3     t1_ext     3.3727726    0.044078803
```

discussion, we focus on the fit of this model along with the fits of the configural and weak invariance models. The −2 log likelihood (−2*LL*) for the strong invariance model was 49,114 (reported in `OpenMx` or 49,113 [−2 · −24556.340] reported in `Mplus`). The −2*LL*s for the configural and weak invariance models were 48,921 and 48,926, respectively. Thus, the move from configural to weak invariance did not lead to a significant increase in model misfit ($\Delta -2LL = 5$ and $\Delta parameters = 4$; $\chi^2(4) = 5$, $p = .29$). However, the move from weak to strong invariance resulted in a significant degradation of model fit ($\Delta -2LL = 188$ and $\Delta parameters = 16$; χ^2 (16) = 188, $p < .01$). The significant decrease in model fit was partly due to the large sample size and partly due to the degree of noninvariance. With maximum likelihood estimation, the global and absolute fit indices are unavailable, which makes it difficult to determine how well or poorly the strong invariance model fit the data (or for that matter, how any of these models fit). As an alternative, we estimated the strong invariance model with weighted least squares in `Mplus`, which allows for the estimation of global fit indices. When fitting this model, the TLI equaled 0.977, the CFI was 0.971, and the RMSEA was 0.091, which supports that notion that the strong invariance model fit the data adequately. Therefore, we move onto parameter interpretation.

Parameter estimates include the factor loadings for interpersonal skills (`t1_intp`, `t2_intp`, and `t4_intp`) and externalizing behavior (`t1_ext`, `t2_ext`, and `t4_ext`), category thresholds for each item, latent variable variances and covariances, and latent variable means for `eta_2` and `eta_3`. The factor loadings for interpersonal skills (`ETA_1 BY T1_INTP`; `lambda_i`) were constrained to be equal and estimated to be 0.61. Similarly, the factor loadings for externalizing behavior (`ETA_1 BY T1_EXT`; `lambda_e`) were constrained to be equal and estimated to be –0.46. Given the factor loading pattern, higher factor scores indicate more positive behaviors (self-control and interpersonal skills) and less negative behaviors (externalizing behavior). The standardized factor loadings (reported by M*plus*) were all greater than |0.77|, indicating the items were strongly associated with their respective latent variable. We note that the equality constraint is imposed on the *unstandardized* factor loadings. Thus, the standardized factor loadings do, in fact, vary over time.

Next, we focus on the threshold parameters. As with the factor loadings, threshold parameters were constrained to be equal over time. For example, estimates of the three thresholds for the self-control (`Thresholds` for `T1_SC$1` through `T1_SC$3`; `tau_s1` through `tau_s3`) items were –7.81, –2.97, and 0.93 at each measurement occasion. Similarly, the thresholds for interpersonal skills were –4.87, –1.51, and 1.09, and the thresholds for externalizing behavior were –0.17, 2.01, and 3.37. As discussed above, the threshold parameters divide the underlying response propensity, but in this model the underlying response propensity is not in a standard metric (not standard normal). Thus, interpreting these coefficients in isolation is difficult. We discuss using these parameters in conjunction with the factor loading to derive the predicted proportion of the sample responding in each category after discussing parameter estimates. Since the factor loadings and thresholds were invariant with respect to time, the measurement scale remained constant over time, which allows for changes in response patterns to be purely related to changes in the common factor. The mean of the common factor at the first measurement occasion was fixed at 0 for identification, but the means of the common factor at the subsequent measurement occasions were estimated. The estimated means of `eta_2` and `eta_3` were 0.42 (`Means` of `ETA_2`; `alpha_2`) and 0.31 (`Means` of `ETA_3`; `alpha_3`) and significantly different from 0, indicating that the mean of the latent variable significantly changed from the first to the second and from the first to the third measurement occasion. The variances of the latent variables were freely estimated and allowed to vary over time. The variances of the three latent variables were 6.89 (`Variances` of `ETA_1`; `psi_11`), 7.93 (`Variances` of `ETA_2`; `psi_22`), and 8.00 (`Variances` of `ETA_3`; `psi_33`) for the first, second, and third measurement occasions, respectively. Therefore, it appeared that the amount of between–person differences in behavior increased over time. The final estimates include the latent variable covariances. The covariance between the common factors at the first and second occasions was 5.73 (r = .78; `ETA_1 WITH ETA_2`; `psi_21`); the covariance between the first and third occasions was 3.83 (r = .52; `ETA_1 WITH ETA_3`; `psi_31`); and the covariance between the second and third occasions was 4.57 (r = .57; `ETA_2 WITH ETA_3`; `psi_32`). The covariances were all positive, significantly different from zero, and strong (correlations were all greater than .50) indicating that individual differences in behavior measured by teachers over time were relatively stable.

The parameters of the model can be utilized to highlight the predicted proportion of the sample responding in each category at each measurement occasion. In Chapter 13 we discussed how estimated proportions can be calculated based on the thresholds and the mean and variance of the underlying latent trait or response propensity at each measurement occasion. The same calculations can be made here with a slight variation. Specifically, the factor loading needs to be taken into account when calculating the variance (or standard deviation) and mean of the response propensity. The variance of the response propensity is equal to the variance of the common factor multiplied by the square of the respective factor loading plus 1 (the residual variance of the probit model) and the mean of the response propensity is equal to the mean of the common factor multiplied by the factor loading. For example, the variance of the response propensity for the externalizing behavior item at the first measurement occasion is the variance of `eta_1` (6.89) multiplied by the square of the factor loading (-0.46^2) plus 1, which equals 2.47. The mean of the response propensity for `t1_ext` is the mean of `eta_1` (0) multiplied by the factor loading (–0.46), which equals 0.00. This mean and variance can then be used in conjunction with the thresholds to estimate the expected proportion of the sample responding in each category for each item.

To present this information in a visual format, we embed the distribution of the common factor at each measurement occasion over the CRC for the externalizing behavior item in Figure 15.6. The distribution of the common factors are dashed lines, and the CRC is represented by the solid lines. First, we note that the factor loading for externalizing behavior was negative, which translates into the CRC for the first category appearing on the right-hand side of the figure (higher latent trait corresponds with higher probabilities of lower scores). From this figure, we can determine that the majority of the sample at the first occasion was expected to be rated in the first two categories (dashed distribution on the left) because most of the distribution corresponds with high probabilities for category 1 and 2. This plot also highlights the similarity of distributions for the common factor at the second and third occasions (two dashed distributions to the right), which are nearly coincident.

Second-Order Growth Model

M*plus*

The MODEL statement from the M*plus* script for the second-order linear growth model with a first-order item factor model with strong measurement invariance is contained in Script 15.3. The `MODEL` statement begins with defining the first-order latent variables. The first-order factors (`eta_1`, `eta_2`, and `eta_3`) are indicated by the behavior rating items collected at each measurement occasion. Common labels are applied to the factor loadings to constrain them to be equal across time. Each latent variable is identified by fixing the factor loading to the self-control item at each measurement occasion (`t1_sc`, `t2_sc`, and `t4_sc`) to 1. First-order factor variances and covariances are then specified. The covariances among the first-order factors are fixed at 0 because these associations are

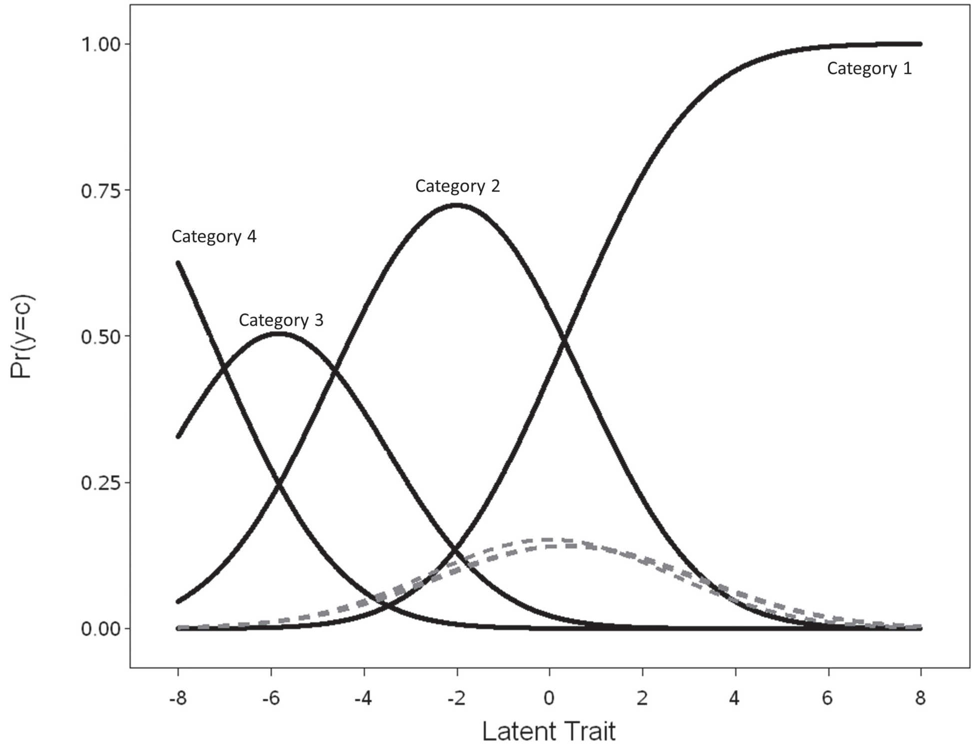

FIGURE 15.6. Distributions of the common factor and response characteristic curves for externalizing behavior.

expected to come from the second-order growth factors. The variances of the first-order factors are constrained to be equal because constraining the residual variances to be equal across time is typical in (first-order) growth models. Next, the first-order factor means and thresholds are specified. The first-order factor means are set equal to 0 because changes in the first-order latent variable means are modeled by the second-order growth factors. The three thresholds per item are specified with equality constraints across time as in the strong invariance model presented above.

In the final part of the script, a linear growth model is specified for the first-order factors. Specifically, the latent variable intercept, `xi_1`, is indicated BY the three first-order factors with factor loadings equal to 1, and the linear slope, `xi_2`, is indicated `BY` the three first-order factors with factor loadings that change linearly with time. The factor loading to `eta_1` is fixed at 0 to center the intercept at the first measurement occasion (fall of kindergarten), the factor loading to `eta_2` is fixed at 1, and the factor loading to `eta_3` is fixed at 3. The linear slope represents the constant rate of change over a 6-month period. Next, the variances of the intercept and slope are specified by listing the variable names, and their covariance is specified using the `WITH` statement. Finally, the means of the intercept and slope are specified. The mean of the latent variable intercept is fixed at 0 for identification, and the mean of the slope is freely estimated.

Script 15.3. `Mplus` Script for the Second–Order Growth Model

```
MODEL:
! First-Order Latent Factors
    eta_1 BY t1_sc@1
             t1_intp (lambda_i)
             t1_ext  (lambda_e);
    eta_2 BY t2_sc@1
             t2_intp (lambda_i)
             t2_ext  (lambda_e);
    eta_3 BY t4_sc@1
             t4_intp (lambda_i)
             t4_ext  (lambda_e);

! First-Order Latent Variable Variances & Covariances
    eta_1 eta_2 eta_3 (psi);
    eta_1 WITH eta_2@0 eta_3@0;
    eta_2 WITH eta_3@0;

! First-Order Latent Variable Means
    [eta_1@0 eta_2@0 eta_3@0];

! Observed Variable Thresholds
    [t1_sc$1 t2_sc$1 t4_sc$1] (tau_s1);
    [t1_sc$2 t2_sc$2 t4_sc$2] (tau_s2);
    [t1_sc$3 t2_sc$3 t4_sc$3] (tau_s3);

    [t1_intp$1 t2_intp$1 t4_intp$1] (tau_i1);
    [t1_intp$2 t2_intp$2 t4_intp$2] (tau_i2);
    [t1_intp$3 t2_intp$3 t4_intp$3] (tau_i3);

    [t1_ext$1 t2_ext$1 t4_ext$1] (tau_e1);
    [t1_ext$2 t2_ext$2 t4_ext$2] (tau_e2);
    [t1_ext$3 t2_ext$3 t4_ext$3] (tau_e3);

! Linear Growth Model
    xi_1 BY eta_1@1 eta_2@1 eta_3@1;
    xi_2 BY eta_1@0 eta_2@1 eta_3@3;
    xi_1 xi_2;
    xi_1 WITH xi_2;
    [xi_1@0 xi_2];
```

OpenMx

The `OpenMx` script to fit the second-order linear growth model built on the longitudinal item factor model is contained in Script 15.4. The script begins by calling the `ecls` data, listing the manifest variables in `manifestVars` (which are the categorical versions of the variables; see Script 15.2), and the latent variables in `latentVars`. The latent variables include both the first-order factors (`eta_1` through `eta_3`) and the second-order factors (`xi_1` and `xi_2`). The model is then specified with a series of `mxPath` statements along with `mxThreshold` statements for the item thresholds.

Script 15.4. `OpenMx` **Script for the Second-Order Growth Model**

```
lirt.growth.omx <- mxModel('Second Order Growth Model, Path Specification', type='RAM',
 mxData(observed=ecls, type='raw'),
 manifestVars=c('t1_sc','t1_intp','t1_ext','t2_sc','t2_intp','t2_ext','t4_sc','t4_intp','t4_ext'),
 latentVars=c('eta_1','eta_2','eta_3','xi_1','xi_2'),

# Residual Variances
mxPath(from=c('t1_sc','t1_intp','t1_ext','t2_sc','t2_intp','t2_ext','t4_sc','t4_intp','t4_ext'),
 arrows=2, free=FALSE, values=1),

# First-Order Latent Variable Variances
mxPath(from=c('eta_1','eta_2','eta_3'), arrows=2, free=TRUE, values=1, labels='psi'),

# Factor Loadings
mxPath(from='eta_1', to=c('t1_sc','t1_intp','t1_ext'), arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),

mxPath(from='eta_2', to=c('t2_sc','t2_intp','t2_ext'), arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),

mxPath(from='eta_3', to=c('t4_sc','t4_intp','t4_ext'), arrows=1, free=c(FALSE,TRUE,TRUE),
 values=c(1,1,-1), labels=c(NA,'lambda_i','lambda_e')),
```

```
# Thresholds
mxThreshold(vars=c('t1_sc','t2_sc','t4_sc'), nThresh=3, free=TRUE, values=c(-7,-2,0),
 labels=c('tau_s1','tau_s2','tau_s3')),

mxThreshold(vars=c('t1_intp','t2_intp','t4_intp'), nThresh=3, free=TRUE, values=c(-5,-2,0),
 labels=c('tau_i1','tau_i2','tau_i3')),

mxThreshold(vars=c('t1_ext','t2_ext','t4_ext'), nThresh=3, free=TRUE, values=c(0,2,4),
 labels=c('tau_e1','tau_e2','tau_e3')),

# Growth Model Factor Loadings
mxPath(from='xi_1', to=c('eta_1','eta_2','eta_3'), arrows=1, free=FALSE, values=1),

mxPath(from='xi_2', to=c('eta_1','eta_2','eta_3'), arrows=1, free=FALSE, values=c(0,1,3)),

# Latent Variable Covariances
mxPath(from=c('xi_1','xi_2'), connect='unique.pairs', arrows=2, free=TRUE, values=c(1,0,1),
 labels=c('phi_11','phi_21','phi_22')),

# Latent Variable Means
mxPath(from='one', to='xi_2', arrows=1, free=TRUE, values=.5, labels='kappa_2')

) # Close Model
```

The first mxPath statement specifies the residual variances of the polytomous variables measured at each time point. These paths are two-headed arrows and fixed at 1 following the typical specification of the probit model. Next, the latent variable variances for the first-order latent variables are specified (first-order covariances are accounted for by the second-order growth model). In this `mxPath` statement, the first-order latent variables are listed as the originating variables; the paths are two-headed arrows, given starting values of 1, and labeled `psi` to constrain them to be equal over time. The factor loadings are then specified in the next three `mxPath` statements. These one-headed paths originate from each first-order latent variable and go to the items measured at the same occasion. The factor loadings to the self-control items (`t1_sc`, `t2_sc`, and `t4_sc`) are fixed at 1, whereas the factor loadings to the interpersonal skills and externalizing behavior are estimated, but constrained to be equal over time using common labels (`lambda_i` for interpersonal skills and `lambda_e` for externalizing behaviors). The thresholds for the items are then specified with three `mxThreshold` statements. As with the strong invariance script, the `mxThreshold` statements are used to define the thresholds for self-control, interpersonal skills, and externalizing behaviors, and the thresholds are held invariant across time. This completes the specification of the first-order (measurement) model.

The second-order growth model is specified with four `mxPath` statements. First, the factor loadings for the latent variable intercept and slope are specified. The factor loadings for the latent variable intercept orginate at `xi_1`, go to the first-order factors (`eta_1` through `eta_3`), and are fixed at 1. The factor loadings for the linear slope originate at `xi_2`, go to the first-order factors, and are fixed values that change linearly with time beginning with 0 to center the intercept at the first measurement occasion (fall of kindergarten). Next, the paths for the second-order latent variable covariance matrix are specified by listing `xi_1` and `xi_2` as the originating variables and using `connect='unique.pairs'`. These paths are freely estimated and given starting values and labels. The final `mxPath` statement is for the mean of the slope. This one-headed path goes from the constant (`one`) to `xi_2`, is freely estimated, given a starting value of .5, and labeled `kappa_2`. The model is then closed.

Output

M*plus* and `OpenMx` output from fitting the second-order growth model with a first-order item factor model with strong measurement invariance is contained in Output 15.3 and 15.4, respectively. As in Chapter 14, a first question is whether the second-order growth model fit significantly worse than the strong invariance model. The difference in the $-2LL$ was 132 and the difference in the number of estimated parameters was 3, which indicates that the second-order growth model fit significantly worse than the strong invariance model. However, given our sample size, we have the power to detect minor changes in model misfit. Thus, we fit the second-order growth model using `ESTIMATOR=WLSMV` in M*plus* and found that the model adequately accounted for the observed data, with a TLI of 0.974, a CFI of 0.966, and an RMSEA of 0.095. We therefore discuss the model parameters.

Output 15.3. *Mplus* Output for the Second-Order Growth Model

```
                                                 Two-Tailed
             Estimate     S.E.   Est./S.E.          P-Value
ETA_1     BY
   T1_SC        1.000    0.000     999.000          999.000
   T1_INTP      0.600    0.029      20.661            0.000
   T1_EXT      -0.456    0.022     -20.475            0.000
ETA_2     BY
   T2_SC        1.000    0.000     999.000          999.000
   T2_INTP      0.600    0.029      20.661            0.000
   T2_EXT      -0.456    0.022     -20.475            0.000
ETA_3     BY
   T4_SC        1.000    0.000     999.000          999.000
   T4_INTP      0.600    0.029      20.661            0.000
   T4_EXT      -0.456    0.022     -20.475            0.000
XI_1      BY
   ETA_1        1.000    0.000     999.000          999.000
   ETA_2        1.000    0.000     999.000          999.000
   ETA_3        1.000    0.000     999.000          999.000
XI_2      BY
   ETA_1        0.000    0.000     999.000          999.000
   ETA_2        1.000    0.000     999.000          999.000
   ETA_3        3.000    0.000     999.000          999.000
XI_1      WITH
   XI_2        -0.742    0.096      -7.767            0.000
Means
   XI_1         0.000    0.000     999.000          999.000
   XI_2         0.066    0.020       3.359            0.001
Thresholds
   T1_SC$1     -8.094    0.325     -24.925            0.000
   T1_SC$2     -3.163    0.125     -25.297            0.000
   T1_SC$3      0.781    0.061      12.765            0.000
   T1_INTP$1   -4.942    0.113     -43.728            0.000
   T1_INTP$2   -1.599    0.043     -37.598            0.000
   T1_INTP$3    0.986    0.038      26.151            0.000
   T1_EXT$1    -0.095    0.027      -3.501            0.000
   T1_EXT$2     2.087    0.040      51.901            0.000
   T1_EXT$3     3.453    0.063      55.150            0.000
   T2_SC$1     -8.094    0.325     -24.925            0.000
   T2_SC$2     -3.163    0.125     -25.297            0.000
   T2_SC$3      0.781    0.061      12.765            0.000
   T2_INTP$1   -4.942    0.113     -43.728            0.000
   T2_INTP$2   -1.599    0.043     -37.598            0.000
   T2_INTP$3    0.986    0.038      26.151            0.000
   T2_EXT$1    -0.095    0.027      -3.501            0.000
   T2_EXT$2     2.087    0.040      51.901            0.000
   T2_EXT$3     3.453    0.063      55.150            0.000
   T4_SC$1     -8.094    0.325     -24.925            0.000
   T4_SC$2     -3.163    0.125     -25.297            0.000
   T4_SC$3      0.781    0.061      12.765            0.000
   T4_INTP$1   -4.942    0.113     -43.728            0.000
```

(continued)

Output 15.3. (*Continued*)

T4_INTP$2	-1.599	0.043	-37.598	0.000
T4_INTP$3	0.986	0.038	26.151	0.000
T4_EXT$1	-0.095	0.027	-3.501	0.000
T4_EXT$2	2.087	0.040	51.901	0.000
T4_EXT$3	3.453	0.063	55.150	0.000
Variances				
XI_1	6.312	0.537	11.756	0.000
XI_2	0.491	0.054	9.171	0.000
Residual Variances				
ETA_1	1.889	0.196	9.640	0.000
ETA_2	1.889	0.196	9.640	0.000
ETA_3	1.889	0.196	9.640	0.000

Output 15.4. `OpenMx` Output for the Second-Order Growth Model

free parameters:

	name	matrix	row	col	Estimate	Std.Error
1	lambda_i	A	t1_intp	eta_1	0.59883090	0.0002106842
2	lambda_e	A	t1_ext	eta_1	-0.45578587	0.0001918320
3	psi	S	eta_1	eta_1	1.89694401	0.0086517445
4	phi_11	S	xi_1	xi_1	6.32224295	0.0008333518
5	phi_21	S	xi_1	xi_2	-0.74209761	0.0002285194
6	phi_22	S	xi_2	xi_2	0.49229792	0.0001967007
7	kappa_2	M	1	xi_2	0.06652350	0.0001405757
8	tau_s1	Thresholds	1	t1_sc	-8.10276182	0.1473712963
9	tau_s2	Thresholds	2	t1_sc	-3.16654612	0.0367364219
10	tau_s3	Thresholds	3	t1_sc	0.78273372	0.0278449740
11	tau_i1	Thresholds	1	t1_intp	-4.94119753	0.0006513041
12	tau_i2	Thresholds	2	t1_intp	-1.59804869	0.0002106411
13	tau_i3	Thresholds	3	t1_intp	0.98559336	0.0002604351
14	tau_e1	Thresholds	1	t1_ext	-0.09524348	0.0001443622
15	tau_e2	Thresholds	2	t1_ext	2.08714471	0.0002751055
16	tau_e3	Thresholds	3	t1_ext	3.45291148	0.0426773544

We briefly discuss aspects of the measurement model first before discussing model parameters associated with change. Factor loadings, which were constrained to be equal over time, were estimated to be 0.60 for interpersonal skills (`ETA_1 BY T1_INTP`; `lambda_i`) and –0.46 for externalizing behavior (`ETA_1 BY T1_EXT`; `lambda_e`). The magnitude of these estimates was nearly identical to the estimates from the strong invariance model. Similarly, estimates of the threshold parameters, which were also constrained to be equal over time, were close in magnitude to the estimates from the strong invariance model. Their magnitudes highlight how few participants were rated using the low categories for the interpersonal skills and self–control items and in the high categories for the externalizing behavior item.

Turning to the estimates of change in behavior, we find that the mean of the linear slope was 0.07 (`Mean of XI_2`; `kappa_2`), which was positive and significantly different from zero highlighting how children's positive behaviors were expected to grow from the beginning of kindergarten to the spring of first grade, on average. The variance of the latent variable intercept was 6.31 (`Variance` of `XI_1`; `phi_11`) and represents the magnitude of between-child differences in behavior skills in the fall of kindergarten. The variance of the slope was 0.49 and significantly different from zero, indicating that children differed in how their behavior skills changed from kindergarten through first grade. The intercept-slope covariance was –0.74 (`XI_1 WITH XI_2`; `phi_21`), indicating that children who had poorer behavior in the fall of kindergarten tended to show greater increases in their positive behaviors from kindergarten through first grade. Finally, the disturbance variance of the latent factor was 1.89 (`Residual Variance` of `ETA_1`; `psi`) and represents true variability in the first-order factors that was unique from the growth process.

When fitting second-order models, such as these, the estimates of growth can be difficult to interpret, especially because the mean of the intercept was fixed at 0 (for identification, see also Koran & Hancock, 2010). To understand the estimates from such models, we begin by examining the threshold parameters, which provide information regarding the frequency of responses in the fall of kindergarten because the mean of the intercept was fixed at 0. Given the magnitudes of the threshold parameters, it is clear that, in the fall of kindergarten, self-control and interpersonal skills were highly rated, with the majority of responses falling in the highest two categories, and externalizing behaviors had low ratings, with most responses in the first two categories. Based on the growth model, the mean of η_i was predicted to increase slightly (0.07 per 1/2 year) and the standard deviation of η_i was predicted to first decrease slightly (2.51 at the fall of kindergarten to 2.31 at the spring of kindergarten) and then increase slightly (2.51 at the spring of first grade). The magnitude of mean change can be difficult to quantify because we are modeling change in an unobserved entity. One way to contextualize change is by comparing the mean change to the standard deviation of the latent variable intercept. Compared to the standard deviation of the latent variable intercept, the mean change was quite small, and this can be seen in Figure 15.7 where approximate distributions of η_i based on the growth model are overlaid on the thresholds for the interpersonal skills items (as an example). From this figure, it can be seen that the relative frequencies of each response category did not change much from the fall of kindergarten (0 on the *x*-axis) to the spring of first grade (3 on the *x*-axis).

IMPORTANT CONSIDERATIONS

Jointly estimating the second-order growth model with a first-order item factor model presented here is often seen as optimal (Grimm et al., 2013) because the scores on the construct of interest do not have to be estimated. However, there are times when this approach is untenable. That is, the item factor model fit here had relatively few items at each measurement occasion, and there were few measurement occasions. Having few items at each measurement occasion and few measurement occasions makes it possible to estimate such models in the structural equation modeling framework, but adding more

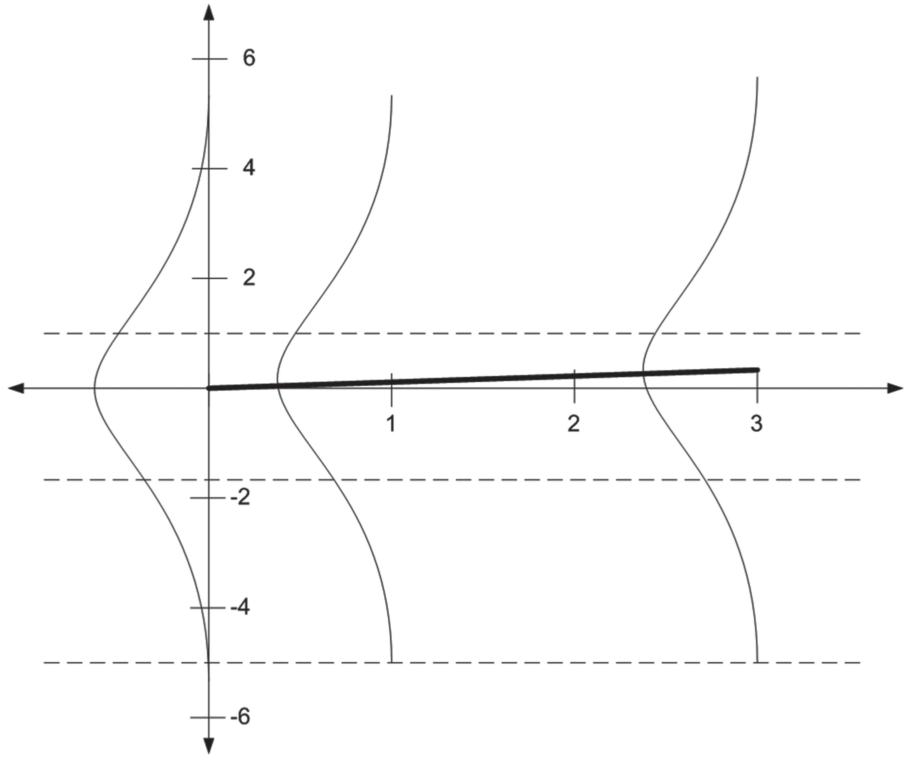

FIGURE 15.7. Approximate distributions of η_i based on the linear growth model plotted against the thresholds for interpersonal skills.

measurement occasions and/or items per measurement occasion will likely lead to nonconvergence. In these situations, a sequential modeling strategy (e.g., Curran et al., 2008; McArdle et al., 2009) where factor (ability) scores are estimated first and then treated as observed data in subsequent analyses is a viable alternative.

There are multiple approaches to using item response models to estimate latent variable scores prior to fitting longitudinal (growth) models. The different approaches arise from the question of how to deal with the clustered nature of the longitudinal item response data. The most common approach is to randomly select a measurement occasion for each individual to create a *calibration* sample. An item response model is fit to the calibration sample, item parameters are saved, and then data from all measurement occasions are scored utilizing the previously saved item parameters (see Curran et al., 2008, and Grimm et al., 2013, for examples). This approach is common because each person appears in the calibration sample once (nesting of individuals over time is avoided), data from all measurement occasions are sampled (if items change over time, item parameters from all items can be estimated), and scoring item data using previously estimated item parameters is common.

A second approach is to fit an item response model to all of the longitudinal item-level data without accounting for the clustered nature of the data and estimate factor scores for all persons at each measurement occasion in this single model run (e.g., McArdle et al., 2009). Technically, this model is misspecified because the data are not independently and identically distributed because the same participants appear in the dataset multiple times. However, this approach is able to generate scores for all participants at all occasions and is a viable approach when data are largely incomplete due to the length of the longitudinal study and/or differences in the measurement instruments used to measure the construct (e.g., McArdle et al., 2009).

A third approach is to fit a multilevel item response model to the longitudinal item-level data and estimate factor scores for all persons at all measurement occasions in a single run. The benefit of this approach is that the model accounts for the clustered nature of the observations. Additionally, aspects of measurement invariance can be studied (comparing the within-cluster factor structure to the between-cluster factor structure). However, this approach is not without limitations. The biggest limitation is that the estimated scores, in addition to being based on an individual's item response pattern (as expected), are partially based on the person's individual distribution of factor scores over time. Lastly, we note that we focused on fitting item factor models to binary and ordinal items, but this general approach to modeling is possible with a combination of continuous, count, ordinal, nominal, and binary items (see Bauer & Hussong, 2009; Muthén, 1984, 2001).

MOVING FORWARD

This chapter brings our discussion of modeling change in latent entities to a close. The next collection of chapters focuses on modeling change with latent change (difference) scores and modeling nonlinear change with a focus on the individual rate of change and how the individual rate of change varies over time within and between individuals. We note that the latent variable models discussed in this section can be combined with the modeling frameworks discussed subsequently (see Ferrer, Balluerka, & Widaman, 2008; Grimm, 2006).