Journal of Modern Applied Statistical Methods

Volume 6 | Issue 1 Article 28

5-1-2007

Modeling Longitudinal Ordinal Response Variables for Educational Data

Ann A. O'Connell
The Ohio State University, aoconnell@ehe.osu.edu

Heather Levitt Doucette
University of Connecticut

Journal of Modern Applied Statistical Methods
May, 2007, Vol. 6, No. 1, 304-319

1538 – 9472/07/$95.00

Modeling Longitudinal Ordinal Response Variables for Educational Data

Ann A. O'Connell
The Ohio State University

Heather Levitt Doucette
University of Connecticut

This article presents applications for the analysis of multilevel ordinal response data through the proportional odds model. Data are drawn from the public-use *Early Childhood Longitudinal Study*. Results showed that gender, number of family risk characteristics, and age at kindergarten entry were associated with initial reading proficiency (0 to 5 scale). The number of family risks and age were associated with time-slopes. Three issues are highlighted: building multilevel ordinal models, interpretation of multilevel effects; and determination of predicted probabilities based on results of the multilevel proportional odds models.

Key words: Proportional odds models, multilevel models, ordinal data.

Introduction

Prior to the fitting of statistical models to investigate relational characteristics of data, researchers must first consider the much more fundamental process of measurement. Stevens (1946) referred to the measurement process as the development of a model that "represent[s] aspects of the empirical world" (p. 677) that are consistent with the nature of the objects under study. In education as well as the social and behavioral sciences, many outcomes are measured on an ordinal rather than an interval or ratio scale, reflecting of course the underlying nature of the phenomenon under study. As an example of an ordinal scale, consider the Concerns-Based Adoption Model (CBAM), developed to characterize the progression of teacher and administrator concerns regarding implementation of innovations within their classrooms or schools (Hall, George & Rutherford, 1986; van den Berg, Sleegers & Pelkmans, 2002, etc.).

Responses on the CBAM correspond to eight ordinal categories, representing progressive stages ranging from self-concern, task-concern to other-concern. This stage-based model is currently being adapted to characterize agency capacity for implementation of evidence-based HIV prevention interventions (O'Connell, Cornman & Heybruck, 2003). Examples of ordinal scales can be found in many different contexts. Proficiency on statewide educational assessments has been characterized as ordinal, with students identified as below basic, basic, proficient, goal, and advanced in mathematics and reading (Beaudin, 2003). The goals set by No Child Left Behind (http://www.nclb.org/)

Ann A.O'Connell, Ed.D., is an Associate Professor in the School of Educational Policy and Leadership at The Ohio State University. Her research interests include data analysis for ordinal-level response variables and multilevel modeling of health and educational outcomes. Heather Levitt Doucette is at the Connecticut State Department of Education and is a graduate student in the Department of Educational Psychology at the University of Connecticut. Her research interests are in assessment and early childhood education. This research was supported by a grant to the first author from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the US Department of Education's National Center for Education Statistics and the Office of Educational Research and Improvement (now the Institute for Education Sciences), under NSF Grant #REC-9980573. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies.

require 100% of students within schools to attain proficiency in order to demonstrate effectiveness, making an understanding of ordinal measures and their statistical treatment important for schools, teachers, administrators, districts and state personnel.

In fact, most variables that are used to detect educational or behavior change are ordinal in nature. For example, change in proficiency during the kindergarten year in early reading or mathematics can be characterized as ordinal (i.e., achieved or did not achieve a particular level within a hierarchy of proficiency goals, pre- and post-school year); so can frequency of condom use before and after an intervention (never, sometimes, almost always, always). Many health intervention studies have relied on the transtheoretical model to characterize individual change before and after participation (Bowen & Trotter, 1995; Hedeker & Mermelstein, 1998; Lauby et al., 1998; Prochaska & DiClemente, 1983, 1986; Prochaska, DiClemente, & Norcross, 1992; Prochaska, Redding, Harlow, Rossi, & Velicer, 1994; Stark et. al, 1996). Other examples include change in severity of illness or physical condition with scale categories such as mild, moderate, and severe (Knapp, 1999), and the common approach of using endorsement of responses to a particular statement (strongly disagree, disagree, neutral, agree, strongly agree) to assess attitudes before and after an event or period of time.

As these examples suggest, the use of ordinal-level variables in education and the social sciences are abundant. This should not be surprising, as Cliff (2003, 1996, and 1993) has consistently pointed out in much of his work on ordinal measurement that the questions we ask of our data are primarily ordinal in nature as well (Did students perform better after a school-based intervention?). However, there is inconsistency in the fidelity between ordinal measurement of a behavioral or cognitive outcome and how these quantities are analyzed in statistical models (Cliff, 2003, 1996, 1993; O'Connell, 2000; Clogg & Shihadeh, 1994; Long, 1997; Agresti, 1996). The accurate interpretation of relationships among variables is dependent on the application of appropriate statistical techniques, yet the treatment of ordinal responses present challenges for many applied researchers in the educational and behavioral sciences. Similar to the field of biomedical and epidemiological research, the underutilization of ordinal regression models in the educational and behavioral sciences may be partially explained by researcher unfamiliarity with software programs capable of fitting these models, confusion about model assumptions and how to investigate these assumptions, and problems in interpretation of model results (Bender & Benner, 2000). These challenges are multiplied when the study purports to consider change in an ordinal outcome over time. In this paper, the hierarchical generalized linear model (HGLM; Goldstein, 2003; McCullagh & Nelder, 1989; Raudenbush & Bryk, 2002) for ordinal responses is demonstrated and explained, using a small number of potential explanatory variables for illustration purposes.

Data applications that characterize an approach to analyzing change over time in ordinal response variables are presented. The data used is drawn from the Early Childhood Longitudinal Study (ECLS), a national database developed and managed through the National Center for Education Statistics (NCES). The ECLS-K (Kindergarten cohort) follows nearly 20,000 students from kindergarten through the first grade, with additional follow-ups in 3rd and 5th grade. The outcome of interest in the models constructed is student proficiency for early reading and literacy assessed across kindergarten and 1st grade, which was measured using six ordinal categories (Table 1). Particular attention is paid to interpretation of the model estimates and assumptions, and the effects of independent variables on proficiency over time. HLM version 6.03 is used for these analyses (Raudenbush, Bryk, Cheong & Congdon, 2004). The goal is to make a contribution to the applied literature on use and interpretation of hierarchical ordinal models, as well as to highlight the methodological challenges of modeling longitudinal ordinal outcomes.

Table 1. Percent of Sample Reaching Reading Proficiency Levels Across Four Waves of ECLS-K.

Proficiency Level	Baseline 0 months n = 3242	8 months n =3346	12 months n =3380	20 months n =3425
0. Did not pass level 1	28.0	4.5	2.0	0.2
1. Identifying upper/lower case letters	34.6	14.8	8.3	1.1
2. Associating letters with sounds at the beginning of words	17.2	23.3	17.6	3.0
3. Associating letters with sounds at the end of words	17.0	40.9	44.0	11.8
4. Recognizing words by sight	2.1	11.3	17.5	37.9
5. Recognizing words in context	1.2	5.2	10.6	46.0

Methodology

Context: Proficiency in Early Literacy

In the ECLS-K, proficiency in early literacy is represented as a series of stepping-stones, which reflect the skills that form the foundation for further learning in reading (West, Denton, & Germino-Hausken, 2000). The categorization of early literacy proficiencies represented in the ECLS-K assessment instrument is consistent with the skills that have been identified as the building blocks of reading mastery: phonemic awareness (the understanding that letters represent spoken sounds), phonics (understanding the sounds of letters in combination), fluency, vocabulary and text-comprehension (CIERA, 2001). Six categories of hierarchical skill levels are used to establish the proficiency scale (Table 1). Mastery is defined as passing 3 out of 4 items in a cluster representing each successive proficiency level.

Research has indicated that children who experience difficulty learning to read in the early primary grades tend to begin school with limited proficiency for early-literacy skills (Burns, Snow & Griffen, 1998). These early skills in reading carry-over to performance at later grades in reading as well as in other subjects, and children who experience difficulties early in school tend to experience continuation of these difficulties as they progress through school (Bayder, Brooks-Gunn, & Furstenberg, 1993; Butler, Marsh, Sheppard & Sheppard, 1985; Juel, 1988; McCoach, O'Connell, Reis, & Levitt, 2006). Even prior to formal schooling, much is happening in the way of literacy skill development via the interaction between life experience and language development. The notion of emergent literacy suggests that children do indeed enter kindergarten with diverse literacy skills that may have an important predictive relationship with later reading abilities (Lonigan, Burgess, & Anthony, 2000).

Initial data summaries from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) cohort indicate that some children do enter kindergarten with greater preparedness and readiness to learn relative to other children, perhaps putting them a step ahead of their peers for the important early grades at school (West, Denton, Germino-Hausken, 2000). ECLS-K studies have shown that children entering kindergarten from families with particular characteristics (living in a single parent household, living in a family that receives welfare payments or food stamps, having a mother with less than a high school education, or having parents whose primary language is not English) tended to be at risk for low reading skills (Zill & West, 2001). Pre-kindergarten experiences related to family life, pre-school or daycare and personal characteristics (e.g., gender, persistence) may relate to children's initial proficiency in reading as well as their potential growth in skills and abilities across the kindergarten year and beyond. For example, girls typically enter kindergarten with slightly greater early literacy ability than boys. Child-focused predictors of success and failure in early reading are helpful for understanding how individual children may be at risk for reading difficulties. From a policy and practice perspective it is clearly desirable that teachers, school administrators, parents, and other stakeholders be aware of these individual factors related to entry-level proficiency as well as to growth in proficiency in order to develop curriculum and instructional practices that can promote achievement for all students relative to their kindergarten entry skills.

School and instructional characteristics have also been shown to be associated with student ability in early literacy, but it is not entirely clear how the differing educational experiences of children across schools (teacher and school effects) might affect growth in proficiency. The National Research Council (1998) reviewed predictors of success and failure in early reading at the neighborhood, school, and community level. In the continuing work using the ECLS-K, the effects of specific school-level variables on proficiency have been modeled separately across the four years of available data. These models included frequency of use of ability-grouping in kindergarten, principals' ratings on the success of various teacher instructional practice, attendance at public versus private schools, school socioeconomic status, and neighborhood climate including the presence of racial tensions, litter, drug/alcohol use in the neighborhood, and extent of crime (Levitt & O'Connell, 2002; McCoach, O'Connell, Levitt & Reis, 2006; O'Connell & Levitt, 2002).

Although instructional, organizational and neighborhood effects on children's entry-level reading ability and growth in reading are critical to understanding how to create and implement effective school-supported teaching strategies, these effects have not been modeled here. Instead, as the purpose of this article is on the methodology for developing and interpreting multilevel models for ordinal responses, the focus herein is on the development and interpretation of two-level models investigating the effect of child-level characteristics on reading growth across four time points (fall and spring of kindergarten, and fall and spring of first grade); extensions to the three-level case are relatively straightforward.

Hierarchical Ordinal Regression Models

Explanatory models for ordinal outcome data collected during a single time frame have been previously reviewed by O'Connell (2000; 2006) and others (e.g., Agresti, 1989, 1990, 1996; Bender & Benner, 2000; Clogg & Shihadeh, 1994; Long, 1997; McCullagh, 1980). This work can be adapted to fit the needs of a hierarchical context. Wong and Mason (1985) and Hedeker and Mermelstein (1998) provided examples of extensions of models for dichotomous and ordinal outcomes for hierarchical data. In addition, the latest version of the HLM program (HLMv6.03; Raudenbush, Bryk, Cheong, and Congdon, 2004) includes options for modeling the cumulative odds for ordinal hierarchical data. An article by Plewis (2002) in the *Multilevel Modeling Newsletter* describes the fitting of multilevel ordinal data using MLwiN.

The most common ordinal outcome model is the regression-type proportional or cumulative odds (PO) model (Agresti, 1996; Armstrong & Sloan, 1989; Long, 1997; McCullagh, 1980). In this approach, the (log of the) odds of a response at or below each of the ordinal categories form the quantities of interest. For example, with a six-category ordinal outcome (K=6), the K-1 formulas shown in Table 2 would be used to compute the cumulative probabilities and consequently the cumulative odds (note: consistent with the ECLS-K categories, the possible outcomes are 0 through 5). The cumulative probabilities are the probabilities that the response for the i^{th} student nested within the j^{th} school (or, for longitudinal data, the i^{th} student at the t^{th} time point) is at or

Table 2. Cumulative Odds Model for K=6 (K=0, 1, …5), Where R_{ti} Represents the Proficiency Outcome (Response) for the Ith Student at the Tth Wave.

Category	Cumulative Probability	Cumulative Odds [Y'_{kti}]	Probability Comparison
k=0 (Proficiency 0)	$P(R_{ti} \le 0)$	$\frac{P(R_{ti} = 0)}{P(R_{ti} > 0)}$	Proficiency 0 versus all levels above
k=1 (Proficiency 1)	$P(R_{ti} \le 1)$	$\frac{P(R_{ti} \le 1)}{P(R_{ti} > 1)}$	Proficiency 0 and 1 combined versus all levels above
k=2 (Proficiency 2)	$P(R_{ti} \le 2)$	$\frac{P(R_{ti} \le 2)}{P(R_{ti} > 2)}$	Proficiency 0,1,2 combined versus 3, 4, 5 combined
k=3 (Proficiency 3)	$P(R_{ti} \le 3)$	$\frac{P(R_{ti} \le 3)}{P(R_{ti} > 3)}$	Proficiency 0,1,2,3 combined versus 4,5 combined
k=4 (Proficiency 4)	$P(R_{ti} \le 4)$	$\frac{P(R_{ti} \le 4)}{P(R_{ti} > 4)}$	Proficiency 0,1,2,3,4 versus proficiency 5

below a given proficiency level. The odds is a ratio of the probability of an event occurring to the probability of an event not occurring. Accordingly, the cumulative odds $[Y'_{kti}]$ represent the odds that any given response would be in at most category *k* (rather than beyond category *k*), for the i^{th} child at the t^{th} wave of data collection. From Table 2, it may be seen that the cumulative odds, in order, correspond to the probability of being in proficiency level 0 relative to all categories above it; the probability of being in proficiency level 0 or 1 relative to all above it; and so on until arriving at the probability of being in categories 0, 1, …4 relative to being in category 5. The Kth or final cumulative probability would always be 1.0 (being at or below the last possible level), and its probability and associated odds are therefore not included in the table. It is common to refer to the value marking each of these binary comparisons as cutpoints or cumulative splits. For example, the cutpoint for the first comparison is 0 (proficiency level 0 versus above 0); the cutpoint for the second comparison is 1 (proficiency 0 and 1 versus above 1), etc.

To better understand how the PO model works, imagine if the separate comparisons indicated in the last column of Table 2 were investigated using corresponding binary (hierarchical) logistic regressions at each of the associated cumulative splits. The simultaneous fitting of each of these separate K-1 (in this example, K-1=5) logistic models represents the overall PO approach. For this approach to be valid, a critical assumption must be made of the data. This assumption of proportionality states that the effects of the explanatory variables cannot be statistically different across these cutpoint comparisons. This is also called the cumulative odds assumption or the equal slopes assumption and can be restrictive but is the most common choice for ordinal regression models (Hedeker & Gibbons, 2006; O'Connell, 2006). For non-hierarchical data, the assumption of equal slopes can easily be tested within SAS or SPSS, for example. However, in a multi-level context direct tests of this assumption are not currently available. Interaction terms can be used to test for non-proportionality of some or all of the predictors, or an ad hoc approach can be applied that investigates the consistency of slope estimates across the cumulative splits described in Table 2. Space does not allow for a demonstration of this assessment here; interested readers can find further discussion and examples in O'Connell, Goldstein, Rogers & Peng (in press), as well as in Hedeker, et al., 2006).

General Model: Students Nested Within Schools.

A brief description of the ordinal HGLM is presented for analyses focused at one point in time; in the next section it is expanded this to cover repeated ordinal measures. For the i^{th} student in the j^{th} school, the hierarchical proportional odds model is fit according to the following equations (Raudenbush & Bryk, 2002):

Student level:

$$\ln(\mathrm{Y}'_{\mathrm{kij}}) = \ln\left(\frac{\mathrm{P}(\mathrm{R}_{\mathrm{ij}} \le \mathrm{k})}{\mathrm{P}(\mathrm{R}_{\mathrm{ij}} > \mathrm{k})}\right) = \beta_{0\mathrm{j}} + \sum_{\mathrm{q}=1}^{\mathrm{Q}} \beta_{\mathrm{qj}} \mathrm{X}_{\mathrm{qij}} + \sum_{\mathrm{k}=2}^{\mathrm{K}-1} \mathrm{D}_{\mathrm{kij}} \delta_{\mathrm{k}} \tag{1}$$

School or Context level:

$$\beta_{\mathrm{qj}} = \gamma_{\mathrm{q0}} + \sum_{\mathrm{s}=1}^{\mathrm{S_q}} \gamma_{\mathrm{qs}} \mathrm{W}_{\mathrm{sj}} + \mathrm{u}_{\mathrm{qj}} \tag{2}$$

where $[Y'_{kij}]$ represents the cumulative odds for each category k, with k=1…K-1 levels of the ordinal response and q = 1…Q independent variables at the student level. For these models, the term on the left side of equation (1) is the log of the cumulative odds for each category k, and is referred to as the logit for the cumulative distribution. The terms on the right can be interpreted similar to any logistic regression model, with the β_{qj} representing the expected change in the logit for each one unit change in the q^{th} explanatory variable, X_q. Its exponentiation will provide the estimate of the cumulative odds for that variable. However, an important difference between an ordinal model and a binary logistic regression model is that with K-1 ways to characterize the cumulative odds, the slope parameters for each of the independent variables are restricted to be constant across all the separate possible cumulative splits derived according to the second column of Table 2. That is, the model assumes that the effect of any independent variable can be represented by a common cumulative odds ratio, exp(β); this is the assumption of proportional odds. If this assumption does not hold, then the PO model is not a plausible one for the data and less restrictive models should be investigated.

The collection of estimates at the far right of equation (1) are referred to as thresholds or delta coefficients, and they operate as deviations from the baseline intercept for each of the K-1 separate binary comparisons beyond the first, with β_{0j} as the baseline intercept (i.e., for the first cumulative comparison). D_{kij} is the indicator variable for each category beyond the first. In other words, each cumulative comparison has its own intercept, while the effects of the explanatory variables are assumed to be constant across each comparison.

Changes Over Time in an Ordinal Response.

When data are gathered over time, methodologies for the treatment of ordinal outcomes need to be combined with methods that address the multilevel nature of longitudinal data. As with other studies of growth, change was modeled in the logit as a linear effect. With only four time points, this approach is reasonable (Murray, 1998). At level one, the repeated measures are modeled over time, and at level two student characteristics are used to look at changes in intercepts or growth trajectories across children. For demonstration purposes, the focus is on the two-level model in this article rather than include a third level for modeling school effects. To investigate child-level variability in baseline (entry) proficiency and in the trajectory of change, we considered the following child-level variables: age at kindergarten entry, gender (boys = 1), attending half-day rather than full-day kindergarten (half-day = 1), previously attending any center-based care (yes = 1), frequency with which parents read books to their child, socio-economic status, count of family risks, and a model-based approach was used to adjust for oversampling of Asian and Pacific Islanders (API) by including API (yes = 1) in all preliminary analyses. The general level one and level two models are provided below.

Time level:

$$ln(Y'_{kti}) = \ln\left(\frac{P(R_{ti} \le k)}{P(R_{ti} > k)}\right)$$
$$= \pi_{0i} + \pi_{1i}T_{ti} + \sum_{k=2}^{K-1} D_{kti}\delta_k \tag{3}$$

Student level:

$$\pi_{qj} = \beta_{q0} + \sum_{s=1}^{S_q} \beta_{qs} X_{si} + u_{qi} \tag{4}$$

The last term in equation 3 is used to estimate the increasing intercepts for each of the underlying cumulative models, and is described in depth in the next section. Not unsurprisingly, these multilevel ordinal models were difficult to converge. Therefore, each independent variable was considered separately, consistent with Raudenbush and Bryk's (2002) suggestions regarding strategies for building complex multilevel models. Based on these preliminary analyses, two were selected that were found to be associated with proficiency in the simpler (univariate) models, and known to be associated with early literacy: gender and family-risk characteristics. The sum of the number of family-risk characteristics was used as a contextual variable in models predicting baseline proficiency (intercept) as well as change in proficiency (slope) over time. Once the contextual model was derived, age at kindergarten entry was included to control for age-effects. These few variables were selected to illustrate how contextual models may be developed and interpreted for longitudinal ordinal outcomes, and below an application of the PO model is presented in the prediction of change in reading proficiency across four years of data using the ECLS-K.

The following section describes the process by which the repeated measures and hierarchical ordinal models were developed.

Procedures

A sample of n=3440 children were selected from the ECLS-K. Since the primary purpose of this presentation is to illustrate the application of a multilevel approach to ordinal data, the sample was limited to children who did not change schools from kindergarten to first-grade, had four waves of data (a 30% subsample of the original data were included in a fall first-grade wave of data collection), were first-time kindergarteners only (no repeaters were included), and had no missing observations on the child-level (level-2) characteristics investigated for this study (gender, family-risk, and age at kindergarten entry). These criteria were applied to minimize complexity of the statistical design regarding number of data points available per child, convergence issues, and concerns regarding the impact of cross-classification of children changing schools during the study period. The resulting data set represents a sample of first-time kindergarteners assessed twice in kindergarten and twice in first grade.

HGLM, the non-linear counterpart to hierarchical linear modeling (HLM), was used to model the ordinal outcomes (Raudenbush and Bryk, 2002). The most general case of an HGLM for ordinal data assumes proportional odds across successive cumulative categories. Proportionality implies that the effect of an independent variable remains constant across the cumulative categories of the outcome variable.

In the PO model, the likelihood (or odds) of an observation falling into category *k* or below is assessed over time. Similar to the familiar logistic regression model, the PO analysis predicts a transformation of the odds, i.e., the logit, which is the log of the odds. A logit of zero corresponds to an odds of 1.0, which implies that there is no difference between the probability of being in a certain category (or below) and being above that category (.5/.5 = 1.0, ln(1.0) = 0). A positive logit implies that the likelihood of being in lower categories is greater (e.g., .7/.3 = 2.33, log(2.33) = .847); and a negative logit implies that the likelihood of being *in higher* categories is greater (e.g., .3/.7 = .429, log(.429) = -.847).

Using the HLM program, the desired data structure is similar to that in other multilevel analyses of longitudinal data. The level-one data file represents the repeated measures outcomes, and contains the proficiency score as an ordinal-level response variable for each child at each of the four time points. With 3440 children, there would be at most 4x3440 or 13,760 observations at level one. Some children were missing proficiency scores at some point during the four waves of data collection; thus there were 13,393 observations overall at level one for the analytic sample. The level-two data contains the child-level characteristics, including gender, the number of family risk characteristics, and age at kindergarten entry. Although three level ordinal models are now available in HLMv6.03 (Raudenbush et al., 2004), the models presented in this article illustrate the assessment of child-level effects (level two) on changes in proficiency over time (level one), and work is continuing on how these models might be extended to incorporate school effects as a third level.

Although many different models were investigated, only three are reported here. The final models include a random coefficients model (Table 4), with time in months as the sole predictor of proficiency (more precisely, as the predictor of the logits for the cumulative odds for proficiency). Next, a contextual model was developed using gender and the number of risk factors as the explanatory child-level variables at level 2 (Table 4). This contextual model was designed to illustrate how the effects of gender and the number of family risk factors may moderate the change in cumulative odds over time. These effects were included as predictors of the intercepts or baseline values and as predictors of the slope for time. This model was then adjusted to include age at kindergarten entry (grand mean centered) as a control variable for predicting both the intercept and the slope from level one, as well as deleted non-statistically significant predictors. Results of this final model are provided in Table 5.

The random coefficients analysis looks at the thresholds between (cumulative) adjacent proficiency levels and estimates the odds of a person being in proficiency level k or below over time. If changes in proficiency can be reliably detected over time, the effect of time on the logit should be negative, so that the likelihood of being in higher categories increases over time. With a six-category outcome (k = 0, 1, 2, 3, 4, 5) and time measured in months from baseline (t = 0, 8, 12, 20), five models are fit simultaneously, as shown below.

Level one:

$$\begin{aligned}
\ln(Y'_{0ti}) &= \pi_{0i} + \pi_{1i}(time)_{ti} \\
\ln(Y'_{1ti}) &= \pi_{0i} + \pi_{1i}(time)_{ti} + \delta_2 \\
\ln(Y'_{2ti}) &= \pi_{0i} + \pi_{1i}(time)_{ti} + \delta_3 \\
\ln(Y'_{3ti}) &= \pi_{0i} + \pi_{1i}(time)_{ti} + \delta_4 \\
\ln(Y'_{4ti}) &= \pi_{0i} + \pi_{1i}(time)_{ti} + \delta_5
\end{aligned} \tag{5}$$

Level two:

$$\begin{aligned}
\pi_{0i} &= \beta_{00} + u_{0i} \\
\pi_{1i} &= \beta_{10} + u_{1i}
\end{aligned} \tag{6}$$

In the collection of equations for level one, the terms on the left, $\ln(Y'_{3ti})$ for example, represents the log of the odds for being in category 3 or below (rather than beyond category 3), consistent with the approach described in Table 2.

The critical assumption of proportional odds implies that the effect of time is constant across the cumulative splits identified through the level one model. The level one effects, π_{0i} and π_{1i}, represent, respectively, the baseline estimates (at the first wave of data collection (entry into kindergarten)) for the log of the odds of being in category k or below, and the effect of time (slope) on these logits. These intercepts and slopes are free to vary from person to person. This variability is captured by the level two random effects, u_{0i} and u_{1i}, with variance components, respectively, of τ_{00} and τ_{11} (var(u_{0i}) = τ_{00} and var(u_{1i}) = τ_{11}). The thresholds, δ_2 to δ_5, represent the differences in the logit for each successive cumulative category relative to the first logit; for example, in this sample the estimate at baseline for the log(odds) of being in category 3 or below would be $\beta_{00} + \delta_3$.

The first contextual model analysis considers the effects of gender (1=male) and the number of risk characteristics (0 through 4) on the baseline logits and the slopes. The level one model remains the same as (5), but now the level-two models used to describe the effects of gender and number of family risks on the intercept and slope are:

$$\pi_{0i} = \beta_{00} + \beta_{01}\left(gender\right)_i + \beta_{02}\left(risknum\right)_i + u_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}\left(gender\right)_i + \beta_{12}\left(risknum\right)_i + u_{1i} \tag{7}$$

Finally, in the second contextual model analysis age at kindergarten entry was included (grand mean centered) in the level two models for both the intercepts and the slopes. The gender effect was deleted from the model for time-slopes due to lack of statistically significant results for gender in a preliminary run.

Results

Table 1 contains the proportion of children classified into each literacy proficiency level from kindergarten through first grade. Table 3 shows the proportion of children making specific transitions in literacy proficiency across the four waves. Most children made a positive change across the kindergarten year; most did not change during the summer between kindergarten and first grade, but then children tended to increase again by one or two proficiency levels across the first grade year.

Results of the random coefficients model are provided in Table 4. These results show that overall across children, the expected log odds of being in proficiency level 0 at baseline is negative (β_{00} = -1.73, $p < .01$), which implies that at baseline it is more likely for a child to be at least in level 1 or higher. There is a statistically significant linear trend in the cumulative logits for time (β_{10} = -.41, $p < .01$), indicating that as a child progresses in school, the likelihood of being at or below category 0 decreases (stated differently, the negative slope for time implies that the probability of being beyond category 0 is increasing with time). This is consistent with what we see in Tables 1 and 3. At baseline, children are more likely to be beyond category 0, and this likelihood increases over time. The model estimates are predicted logits. To transform to odds and then to probabilities, odds = exp(β), and probability = odds/(1 + odds) are used. For this example, the odds at baseline of a child being in proficiency level 0 or below is exp(-1.73) = .1773; this corresponds to a probability of .1773/(1+.1773) = .15. For this random coefficients model containing no child-level predictor variables, 15% of children would be predicted to be at or below category 0 at baseline. For the predicted logit of being at or below category 0 at time 2 (8 months), the model estimates the logit as: -1.73 + (-.41)(8) = -5.01. Thus, at the end of kindergarten, the model predicts that the odds of being in category 0 or below is decreased (exp(-5.01) = .0067), and the associated probability of being at or below proficiency category 0 at the end of kindergarten is .007, or .7%.

To examine the model predictions at other splits in the cumulative hierarchy, for example, model predictions for being at or below category three at baseline, δ_4 is used in addition to the baseline intercept and the slope (see equation 5). For this data, the new intercept, or threshold, becomes $\beta_{00} + \delta_4$ = -1.73 + 7.86 = 6.13. Accordingly, the probability of a child being at or below proficiency level 3 at baseline is .998 pr 99.8%. At time 2 (8 months), the predicted logit is $\beta_{00} + \beta_{10}$*(8 months) + δ_4 = -1.73 + (-.41)*(8) + 7.86 = 2.85, where exp(2.85) = 17.29, and the predicted probability of being at or below proficiency category three at the end of kindergarten is .945, or 94.5%. These predictions, based on a model with no explanatory variables, are reasonably consistent with the data in Table 1.

Finally, reviewing the variance components for the model, it may be seen that considerable variation remains in the intercepts, τ_{00} = 8.35, $p < .01$, as well as in the slopes, τ_{11} = .003, $p < .01$.

The first contextual model (Table 4) describes the effect of gender and the number of family risk factors on the baseline logits and the

Table 3. Change in Proficiency Across the Kindergarten (K1 & K2) and First Grade (FG1 & FG2) Years.

Raw Change in Proficiency	K2-K1	FG1-K2	FG2-FG1
-3	0.0	0.1	0.0
-2	0.3	0.8	0.1
-1	1.7	7.7	0.7
0	21.3	46.5	18.5
1	33.6	32.6	40.9
2	27.5	8.3	29.9
3	9.0	1.0	7.1
4	0.5	0.0	0.8
5	0.01	0.0	0.0

Table 4. Multilevel Ordinal Models for Prediction of Proficiency Using Four Waves Of ECLS-K; Ivs are Gender and Number of Family Risks.

Effect	Coeff. (s.e.)	t (df)	Coeff. (s.e.)	t (df)
Intercept (π_{0i})				
β_{00}	-1.73 (.068)	-25.41 ** (3439)	-2.56 (.097)	-26.48 ** (3437)
β_{01} (gender (M=1))			0.62 (.114)	5.48 ** (3437)
β_{02} (number of risks)			1.07 (.078)	13.75 ** (3437)
Time Slope (π_{1i})				
β_{10}	-.41 (.004)	-98.46 ** (3439)	-.41 (.005)	-77.45 ** (3437)
β_{11} (gender (M=1))			-.001 (.005)	-0.21 (3437)
β_{12} (number of risks)			-.01 (.003)	-2.18 * (3437)
For Thresholds:				
δ_2	2.75 (.053)	51.71 ** (13387)	2.78 (.054)	51.58 ** (13383)
δ_3	4.69 (.060)	77.28 ** (13387)	4.71 (.061)	77.03 ** (13383)
δ_4	7.86 (.077)	101.46 ** (13387)	7.88 (.079)	101.17 ** (13383)
δ_5	10.32 (.091)	112.88 ** (13387)	10.35 (.092)	112.61 ** (13383)

Note: * p < .05; ** p < .01

	Random Coefficients Model			Contextual Model		
Random Effects	*Variance*	*df*	*Chi-square*	*Variance*	*df*	*Chi-square*
Variance in Base- K1 (τ_{oo})	8.346	3391	10350.03 **	7.75	3389	10025.82 **
Variance in Time slope (τ_{11})	.003	3392	3615.27 **	.003	3392	3626.38 **

Note: * p < .05; ** p < .01

slopes for time. Gender has a statistically significant effect on the baseline logits (β_{01} = .62, p < .01). Being a boy tends to increase the logit, making the likelihood of being in higher proficiency categories lower for boys relative to girls. The number of risk factors also has a statistically significant effect on the baseline logit (β_{02} = 1.07, p < .01). Because the logit is positive, it may be seen that as the number of family-risk characteristics increases, the likelihood that a child would be in lower proficiency categories (i.e., at or below any category *k*) increases, relative to a child with fewer risks.

Attention is now turned to interpretation of the effects of gender and the number of risk characteristics on the slope for time. β_{10} = -.41 may be interpreted as the estimated slope for girls with out any family risks. Controlling for the number of risk factors, gender has no effect on the slopes (β_{11} = -.001, p > .05); thus gender does not affect the rate of change in proficiency. The number of risk factors does impact rate of change (β_{12} = -.01, p < .05). On the surface this would suggest that the likelihood is greater that a child with more risk characteristics improves over time even beyond that of a child with fewer risks. However, on closer inspection of the model predictions – particularly in terms of predicted probabilities across the four time points of being at or below any category *k* – it is seen that children with increased family risks tend not to improve as readily over time as their non-risk peers.

This complexity of ordinal model interpretation can be overcome by estimating outcomes for discrete cases of children. For example, substituting into the prediction model, a female child (gender = 0) from a family with 0 risk characteristics would be expected to have a predicted logit for the first cumulative comparison (proficiency level 0 or below) at baseline (time=0) of -2.56, which corresponds to a cumulative odds of exp(-2.56) = .08 and cumulative probability of being at or below proficiency category 0 of .072, or 7.2%. For a girl at baseline from a family with 1 risk characteristic, the predicted logit is -1.49, corresponding to a cumulative odds of .23, and a probability of .187 or 18.7%. This is a large proportion of girls estimated to be at or below proficiency level 0 (rather than beyond category 0), given the addition of just one risk factor. In fact, the odds ratio for the variable number of risks is exp(1.07) = 2.92. The model suggests that, at baseline, the odds of being at or below any category increases by a factor of 2.92 for every one unit increase in a child's number of family risks. Baseline is the simplest case for making predictions; moving to time 2 at 8 months, the model estimates now need to include gender and family risk effects on the effect of time, but the process of estimating outcomes is similar to the process demonstrated above. Based on the parameter estimates from the model, probability predictions for being at or below proficiency category 0 at time 2 (8 months) are .29%, 1.56%, and 13.24% for girls with 0, 1, and 4 family risk factors, respectively.

The variance estimates for this contextual model indicates that variability in the baseline logits and in the time slopes continues to be statistically different from zero, which suggests that additional variables may be useful in understanding proficiency growth (initial status and rate of change). Table 5 provides the model estimates for an adjusted contextual model. In this modified model, age at kindergarten entry (grand-mean centered) is included in the models, and gender is removed from the level 2 models for the slope due to its lack of contribution to that model. The predictions for baseline or initial proficiency remain fairly similar to the contextual model estimates in Table 4. All three predictors contribute to the prediction of the baseline logits, with age at kindergarten entry having a negative effect (β_{03} = -.13, p < .01). This implies that for older children at kindergarten entry, the probability of being in higher categories of proficiency increases. After adjusting for age at kindergarten entry, the number of family risks is still a statistically significant predictor of the trajectory (slope) in the proficiency logits from baseline through the end of first grade (β_{11} = -.01, p < .05), with little change in magnitude from the previous model. In addition, age at kindergarten entry is positively related to the time slopes (β_{12} = .002, p < .01); based on model predictions, older children tend to improve over time more readily than their younger peers.

Table 5. Multilevel Ordinal Model for Prediction of Proficiency Controlling for Kindergarten Entry Age; Ivs are Gender, Number of Family Risks, and Age at Kindergarten Entry.

Contextual Model 2

Effect	Coeff (s.e.)	t(df)	p
Intercept (π_{0i})			
β_{00}	-2.58 (.089)	-28.91 ** (3436)	.000
β_{01} (gender (M=1))	.67 (.089)	7.514 ** (3436)	.000
β_{02} (number of risks)	1.05 (.077)	13.65 ** (3436)	.000
β_{03} (age at K entry)	-.13 (.014)	-9.09 ** (3436)	.000
Time Slope (π_{1i})			
β_{10}	-.41 (.005)	-90.98 ** (3437)	.000
β_{11} (number of risks)	-.01 (.004)	-2.08 * (3437)	.037
β_{12} (age at K entry)	.002 (.001)	3.72 ** (3437)	.000
For Thresholds:			
δ_2	2.76 (.054)	51.54 ** (13382)	.000
δ_3	4.71 (.061)	76.99 ** (13382)	.000
δ_4	7.88 (.078)	101.10 **(13382)	.000
δ_5	10.35 (.092)	112.53 ** (13382)	.000

Note: * p < .05; ** p < .01

Random Components

Random Effects	*Variance*	*df*	*Chi-square (p)*
Variance in Base- K1 (τ_{oo})	7.47	3388	9818.72 (p=.000) **
Variance in Time slope (τ_{11})	.003	3392	3611.19 (p=.005) **

Note: * p < .05; ** p < .01

Despite the addition of entry age to both the intercept and slope models, however, significant variability remains in the initial status and the growth trajectories across children (τ_{00} = 7.47, p < .01; τ_{11} = .003, p < 01).

Table 6 provides predictions based on the random coefficients model and the final contextual model for the probability of a child being at or below proficiency level 3 across all four waves, and contains the actual proportion of children for comparison. Probabilities decline over time, as expected, because it is hoped that children are moving beyond category three by the end of first grade. Among the notable comparisons possible based on this simple table is the predicted probability at the end of first grade for a hypothetical male child of average age with no family risk characteristics (prob = .097) relative to the predicted probability for a male average-age child with four family risk characteristics (prob = .763). Recall that these probabilities are cumulative, and represent the probabilities of being at or below proficiency category 3. These differences are quite large. Further, at the end of first grade, the likelihood that boys do not achieve proficiency in the highest categories in comparison to girls' likelihood is large as well. These predicted probabilities help to make clear the utility of hierarchical ordinal models for understanding effects of child-demographic variables on growth in proficiency for early literacy skills in a way that the basic interpretation of parameter estimates from the models in Tables 5 and 6 cannot easily do.

Table 6. Probability Predictions (at or Below Category 3) for Each Time Point Based on Models in Tables 5 And 6 (Age is Grand Mean Centered).

	K-entry (0 months)	**K-completion (8 months)**	**FG-entry (12 months)**	**FG-completion (20 months)**
At or below Category 3: Actual Data	**.967**	**.835**	**.719**	**.161**
Random Coefficients Model	**.998**	**.945**	**.770**	**.112**
Contextual Model 2 Female Average age Family Risks = 0	**.995**	**.883**	**.594**	**.052**
Female Average age Family Risks = 4	**.999**	**.997**	**.984**	**.622**
Male Average age Family Risks = 0	**.997**	**.936**	**.741**	**.097**
Male Average age Family Risks = 4	**.999**	**.997**	**.992**	**.763**

Conclusion

These examples illustrate the application and interpretation of ordinal regression models to longitudinal data. Given that ordinal responses are best analyzed using ordinal methods, it is important that educational statisticians add these techniques to their toolkit. The ECLS provides a rich data set for investigating many challenging statistical issues. However, some issues need more clarity before these models can be effectively applied.

In this article, the focus has been on the cumulative odds or proportional odds model; however, this assumption may not always hold. Other options are routinely available for researchers dealing with single-level ordinal response data such as the continuation ratio model or non-proportional odds models (Agresti, 1989, 1990; Armstrong & Sloan, 1989; Cox, 1972; Greenland, 1994; Goodman, 1983; McCullagh, 1980; O'Connell, 2000, 2006). In addition, multilevel software programs are somewhat limited in terms of ordinal model methodology, and the default model may often be based on the (untested) assumption of proportional odds. Ultimately, the choice for what approach to take should be guided by theory or an a-priori expectation of which approach would be most appropriate for a given situation (Agresti, 1990; Armstrong & Sloan, 1989). It is hoped that this article has helped to familiarize applied researchers with some of these issues as well as with the interpretation of multilevel ordinal models. Yet, further work is necessary to clarify model fitting for multilevel ordinal data when the assumption of proportional odds is violated, and for when three-level models might offer the best structure for the research data being analyzed.

References

Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin, 105*, 290-301.

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: John Wiley & Sons.

Armstrong, B. G. & Sloan, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology, 129*(1), 191-204.

Beaudin, B. (2003). "NCLB Act of 2001: Connecticut State Department of Education." Invited presentation to the Neag School of Education, University of Connecticut.

Bender, R. & Benner, A. (2000). Calculating ordinal regression models in SAS and S-PLUS. *Biometrical Journal, 42*(6), p. 677-699.

Bayder, N., Brooks-Gunn, J., & Furstenberg, F. F. (1993). Early warning signs of functional illiteracy: Predictors in childhood and adolescence. *Child Development, 64,* 815-829.

Burns, M. S, Griffen, P., & Snow, C.E.(Eds.) (1998). *Preventing reading difficulties in young children.* Washington, D.C.: National Academy of Sciences - National Research Council.

Butler, S., Marsh, H. W., Sheppard, M. J, & Sheppard, J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology, 77*(3), 349-361.

Center for the Improvement of Early Reading Achievement (CIERA). (2001). *Put reading first: The research building blocks for teaching children to read, kindergarten through grade 3.* Washington, DC: Government Printing Office.

Cliff, N. (1993). What is and isn't measurement. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the social and behavioral sciences: Methodological issues.* Hillsdale, NJ: LEA, p. 59-93.

Cliff, N. (1996). Ordinal methods for behavioral data analysis. Mahwah, NJ: Lawrence Erlbaum Associates.Cliff, N. (1996). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research, 3*(3), 331-350.

Cliff, N. & Keats, J. A. (2003). *Ordinal mesurement in the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Clogg, C. C., Shihadeh, E. S. (1994). *Statistical models for ordinal variables.* Thousand Oaks, CA: Sage.

Cox, D. R. (1972). Regression Models and life tables (with discussion). *Journal of the Royal Statistical Society B, 74*, 187-220.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.

Goodman, L. A. (1983). The analysis of dependence cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics, 39*, 149-160.

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine, 13,* 1665-1677.

Hall, G., George, A. & Rutherford, W. (1986). *Measuring stages of concern about the innovation: A manual for use of the SoC Questionnaire.* Austin, TX: Southwest Educational Development Laboratory.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. New York: John Wiley and Sons.

Hedeker, D. Mermelstein, R. J.(1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research, 33*(4), 427-455.

Huynh, C-L (2002). "Regression models of ordinal response data: Analytical methods and goodness-of-fit tests." Paper presented at the annual meeting of the AERA, New Orleans, LA, April 1-5, 2002.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4) 437-447.

Lauby, J. L., Semaan, S., Cohen, A., Levition, L., Gielen, A., Pulley, L., Walls, C., and O'Campo, P. (1998) Self-efficacy, decisional balance and stages of change for condom use among women at risk for HIV infection. *Health Education Research: Theory and Practice, 13*(3), 343-356.

Levitt, H. & O'Connell, A. A. (2002). "Differential Outcomes on Measures of Reading Ability for Full-Day Versus Half-day Kindergarteners." Paper presented at the 33rd Annual Northeastern Educational Research Association Annual Meeting, October 23-25, Kerhonkson, NY.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Lonigan, C. J. Burgess, S. R., Anthony, J. L.(2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology,* 36(5), 596-613.

McCoach, D. B., O'Connell., A. A., Levitt, H. A., & Reis, S. M. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *Journal of Educational Research, 99*(6), 339-346.

McCoach, D. B., O'Connell., A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first two years of school. *Journal of Educational Psychology, 98*(1), 14-28.

McCullagh, P. (1980). Regression models with ordinal data (with discussion). *Journal of the Royal Statistical Society, B, 42*, 109-142.

McCullagh, P. & Nelder, J. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

National Research Council (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

O'Connell, A. A., (2000). Methods for modeling ordinal outcome variables. *Measurement and Evaluation in Counseling and Development, 33*(3), 170-193.

O'Connell, A.A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks: Sage.

O'Connell, A. A., Cornman, D., & Heybruck, K. (2003). "Measuring CBO capacity for implementation of evidence-based HIV prevention interventions." Paper submitted to APHA.

O'Connell, A. A. & Levitt, H. (2002). "Modeling Ordinal Response Variables for Multilevel Analyses." Paper presented at the 33rd Annual Northeastern Educational Research Association Annual Meeting, October 23-25, Kerhonkson, NY.

O'Connell, A. A., Goldstein, J., Rogers, H. J., & Peng, C. Y. J. (in press). Multilevel logistic models for dichotomous and ordinal data. In A. A. O'Connell and D. B. McCoach (Eds.), *Multilevel modeling of educational data.* Greenwich, CT: Information Age Publishing.

Plewis, I. (2002). *Modeling ordinal data using* MlwiN. Multilevel Modelling Newsletter, 14(1). Available online at http://multilevel.ioe.ac.uk/publref/newsletters.html.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model. *Journal of Consulting and Clinical Psychology, 51*(3), 390-395.

Prochaska, J. O. & DiClemente, C. C. (1986). Towards a comprehensive model of change. In W.R. Miller & N. Heather (Eds.), *Treating addictive behaviors: Processes of change* (pp. 3-27). New York: Plenum Press.

Prochaska, DiClemente, & Norcross (1992). In search of how people change: Applications to addictive behavior. *American Psychologist, 47*(9), 1102-1114.

Prochaska, J. O., Redding, C. A., Harlow, L. L., Rossi, J. S., Velicer, W. F. (1994). The transtheoretical model of change and HIV prevention: A review. *Health Education Quarterly, 21*(4), 471-486.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Stevens, S. S. (1946) On the theory of scales of measurement. *Science, 103*(2684), 677-680.

Stark, M. J., Tesselaar, H. M., O'Connell, A. A., Cohen, A., Person, B., Walls, C. & Galvotti. (1996). Psychosocial factors associated with the stages of change for condom use among women at risk for HIV/STDs: Implications for intervention development. *Journal of Consulting and Clinical Psychology, 66*(6), 967-978.

van den Berg, R., Sleegers, P.& Pelkmans, T. (2002). Research into interventions from a cultural-individual innovation perspective. *Studies in Educational Evaluation, 28*, 147-175.

West, J., Denton, K., Germino-Hausken, E. (2000). *America's kindergartners: findings from the early childhood longitudinal study, kindergarten class of 1998-99: Fall 1998*. Washington, DC: United States Department of Education, National Center for Education Statistics. (NCES 2000-070)

Wong, G. Y., Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association, 80*(391), 513-524.

Zill, N. & West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school.* Washington, DC: United States Department of Education, National Center for Education Statistics. (NCES 2001-035).