Published in final edited form as:
J Sch Psychol. 2010 April ; 48(2): 163–186. doi:10.1016/j.jsp.2009.12.002.

School profiles of at-risk student concentration: Differential growth in oral reading fluency

Jessica A.R. Logan[a,*] and **Yaacov Petscher**[b,c]

[a]The Ohio State University, United States

[b]The Florida State University, United States

[c]Florida Center for Reading Research, United States

Abstract

The present study provides a data-driven approach to identifying groups of schools based on the concentration of at-risk students the school serves. The percentage of English language learners, minority students, and students eligible for free or reduced priced lunch were used as indicators in a latent profile analysis of 569 schools. The goal of the present study was to determine whether school-level average student reading performance varied as a function of the groups identified in the latent profile analysis. To do so, groups extracted by the latent profile analysis were used as school-level predictors of growth in oral reading fluency, which was modeled at the within-student level of a three-level hierarchical growth curve model. Oral reading fluency was measured at four points during the year in a large cross-sectional sample of first-, second-, and third-grade students. Results indicated that schools were able to be classified into four distinct groups based on their concentrations and types of at-risk students. Further, in all three grades, there were significant differences between the four identified groups observed in average reading fluency scores at the beginning of the year, the end of the year, and growth during the year indicating that groups based on school-concentration of at-risk students were significantly related to average student achievement in reading ability.

Keywords

Education; Hierarchical growth model; Latent class analysis; Literacy; Oral reading fluency

---

The goal of the No Child Left Behind (NCLB) act of 2001 was for all schools to meet an absolute level of performance in reading and math; a goal uniformly applied to all subgroups of students within a school (U.S. Congress, 2001 © 1111 [b][2][G]). The purpose of this goal was to close several achievement gaps, the three largest being that between minority and non-minority students, students living below and above the poverty line, and students who are English Language Learners (ELL) and those who are native speakers of English. Although NCLB has been in the implementation stages for a few years, it appears that these achievement gaps have yet to substantially diminish and that student risk factors may continue to impact student achievement. Schools with high concentrations of minority, poverty, and ELL students (i.e. at-risk students) typically show poorer academic outcomes than schools with smaller percentages of these students (U.S. Department of Education, 2006). However, researchers disagree on what percentage is large enough to label a school as having a high concentration of at-risk students. The purpose of the present study, therefore, is to examine whether distinct

*Corresponding author. PAES Building, office A328. 305 W. 17th Ave, Columbus, OH 43210-1224, United States. Logan.251@osu.edu .

groups of schools exist based on their percentages of at-risk students, and how these student-based risk factors impact average school-level reading performance.

Student-level risk

The continued existence of a discrepancy in achievement between minority and non-minority students is well demonstrated by The National Center for Educational Statistics (NCES). The NCES reported that in 2005, 4th-grade African American students scored 29 points lower on average than Caucasian students in reading achievement based on a 500 point scale and Hispanic students scored 26 points lower than Caucasian students (NCES, 2007). Similarly, California reported that only 32% of Hispanic students reached proficiency in 4th-grade reading on the California Achievement Test, a figure similar to that of African American students (35%). Caucasian students, however, far outperformed these groups, with 68% of their students reading at or above proficiency (California Department of Education [CDOE], 2005). New York's state test showed a similar trend, although they collapsed their results across grades 3–8. On their state test, only 45% of African American students and 46% of Hispanic students scored at or above proficiency, compared to 75% of Caucasian students (New York State Education Department [NYSED], 2007).

Historically, one of the most highly used indicators of student risk has been socio-economic status, with a consistently large gap observed between the scores of high and low poverty students (Murnane, Willett, Bub, & McCartney, 2006; Reardon & Robinson, 2007). Results from Tennessee's 2005 state test of 4th graders, for example, showed that 20% of those students who were identified as economically disadvantaged scored below proficient levels on reading (TDOE, 2005). By contrast, only 6% of students who were not economically disadvantaged scored at this low of a level. In California, 4th-grade language arts test results were very similar to those of Tennessee. Specifically, 32% of economically disadvantaged students scored at or above proficiency, compared to 70% of non-disadvantaged students. These results are also similar to those from the National Assessment of Educational Progress (NAEP, 2007): of those students who were classified as eligible for free or reduced priced lunch, 43% scored at or above a basic level on 4th-grade reading. However, 73% of students who were classified as not-eligible scored above basic level on the same measure (California Department of Education, 2005).

The final achievement gap targeted by NCLB is that between those students for whom English is a second language, and those who speak it as a first language. Only 2% of students in Tennessee are English Language Learners (ELL); of this population, 44% of 4th graders scored at or above the proficiency benchmark in reading according to their 2007 results. By contrast, 90% of students who were non-ELL scored at or above this same level (Tennessee Department of Education [TDOE], 2005). New York State also has a small (i.e., 6%), albeit growing, ELL population. Of those students identified as ELL, only 6% were proficient in reading, compared to 71% of native speakers (NYSED, 2007). California, by contrast, has a higher percentage of ELL students (i.e., 31%); however, the results are quite similar to Tennessee and New York. Only 22% of fourth graders identified as ELL tested as reading proficiently on California's state level test (CDOE, 2005). Of English speaking fourth grade students, 60% were proficient, which included those students who were originally English learners, but who were then classified as English proficient.

School risk concentration

As discussed in the previous section, students who have been identified as minority, ELL, or are receiving FRL are typically considered to be at-risk for reading difficulty, and score consistently lower than their more advantaged peers. As a function of this phenomenon, schools with high concentrations of at-risk students typically have mean student performance that is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

substantially lower than schools with fewer at-risk students (U.S. Department of Education, 2006). This school-level achievement gap has been well documented in the education research literature, with a large body of evidence showing an association between high percentages of minority, poverty, or ELL students and poorer academic outcomes (Bryk & Raudenbush, 1988; Connor, Son, Hindman, & Morrison, 2005; Denton, Foorman, & Mathes, 2003; Goldhaber & Brewer, 1999; Kim & Sunderman, 2005; Mosteller & Moynihan, 1972).

However, researchers use different criteria to classify a school as having "high" percentages of students in these risk categories. Kim and Sunderman (2005), for example, classified the schools in their Virginia sample into three risk categories based upon the percentage of students identified as eligible for free or reduced lunch (low poverty = 0 to 23%; medium poverty = 24% to 45%; high poverty = greater than 45%). The NCES also occasionally groups schools by percentage of students receiving free or reduced price lunch. In the 2005 write up of 4th-grade mathematics results, NCES specifically divided schools with 75% or more of their students eligible for free or reduced lunch as being the highest poverty group, and those with 10% or fewer being the lowest poverty group (NCES, 2005). Raudenbush (2004) classified a school as "high poverty" when more than 50% of their students were eligible for free or reduced price lunch. Kannapel and Clements (2005) used the same criteria as Raudenbush (2004), but the authors termed these schools "high risk" rather than "high poverty." Following the example set by Kannapel and Clements (2005), in the present study we use the term "risk" or "school risk" to refer to the percentages of minority, poverty, and ELL students within a school.

It is interesting to note that most studies have focused exclusively on either SES or race alone when classifying schools into groups, with very few including ELLs alone or in conjunction with other factors. Although the percentages of students meeting the three discussed risk factors tend to correlate highly (Klingele & Warrick, 1990; McDermott, 1995; Sirin, 2005), the use of multiple correlated predictors typically increases the ability to predict outcomes as each variable will account for some unique variance. A primary goal of the present study, therefore, was to use a data-driven approach to classify schools into groups, simultaneously accounting for all three risk factors to ascertain which percentages of student risk should be used in the present sample to classify schools into risk groups.

One method by which all three risk factors can be included within the analyses is Latent Class Analysis (LCA). LCA, discussed later in greater detail, is similar in purpose to cluster analysis in that it uses observed indicators to create groups. In the present study, a type of LCA, Latent Profile Analysis, was used to identify different groups of schools based on their percentages of students who meet each of the aforementioned risk criteria. The first goal was to determine how many different clusters or groups of schools existed, and what factors distinguished the clusters from each other. A second goal of this study was to determine if average student reading performance varied as a function of school classification. Specifically, whether between-group differences existed in average student performance at the beginning of the school year, the end of the school year, and how group membership impacted change during the year.

Latent class analysis

LCA is a technique relatively new to the psychology and education fields. It is typically used to classify individuals into groups based on their individual responses on a single test, or their scores on multiple tests. Thus, most of the work with the LCA technique has been applied to the classification of substance abuse (i.e., Agrawal, Lynskey, Madden, Bucholz, & Heath. 2007; Muthén, 2006) and mental or psychological disorders, such as Depression or Attention Deficit Disorder (i.e., VanLang, Ferdinand, Ormel, & Verhulst, 2006). When indicators of the latent class variable are dichotomous, such as the correct/incorrect responses to test items or the present/absent responses to symptom count variables as were used in the given examples, the analysis is always referred to as *latent class analysis.* When the indicators are continuous, such as standardized test scores, the LCA technique may also be referred to as Latent Profile Analysis (LPA). The current study employed LPA but, as LCA is the more overarching term, it will be used in the present manuscript when discussing generalities about the technique (Muthén & Muthén, 1998).

LCA draws many parallels to item response theory and confirmatory factor analysis. Across these techniques, several observed variables load onto a single latent variable. The latent variable is continuous in both factor analysis, where it represents a factor score, and item response theory, where it represents each person's true ability. In LCA the latent variable is categorical. This categorical latent variable is constrained to find a given number of groups or clusters based on scores or responses on the observed variables. All paths and parameters are estimated separately for each group. In addition, each person's probability of belonging to each group is reported, along with a categorical representation of their most likely group membership.

To consider an example, LCA has been used to divide children into those suffering from three types of attention deficit disorder (attention deficit only, hyperactivity deficit only, and combined) and those who do not have the problem (e.g. deNijs, Ferdinand, & Verhulst, 2007). The analysis is represented graphically in Fig. 1. The observed variables in the LCA represent items from a single test, representing several of the indicators of attention problems included in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994). The first two boxes on the left side of the figure represent items on the test that measure inattention, and the second two represent those items that measure hyperactivity. In actuality, several items would be used, but for illustrative purposes only four items were used in this example. The observed variables all load onto a latent variable representing class. In this example, there should theoretically be four groups of responders (ADHD — inattentive type only, ADHD — hyperactivity type only, ADHD — combined type, and those students who do not have ADHD). Thus, in the example, the latent class variable was constrained to search for four classes, and found one group with a high probability of endorsing the hyperactivity items, one with a high probability of endorsing the inattention items, one with a high probability on all items, and one with a low probability on all items. The right side of Fig. 1 is a graphical representation of the probabilities of endorsing each of the four items. As the first two items measure inattention, responders with the inattentive subtype and the combined subtype are highly likely to endorse those items. Those with the hyperactivity subtype and those who do not have ADHD are unlikely to endorse these items.

As was discussed previously, most applications of LCA techniques have primarily been performed for diagnostic purposes, as was illustrated in the ADHD example. Also, such applications are typically conducted such that indicators of latent classes consist of items on a single test. The present study differs from typical applications of this technique in three ways. First, the outcomes examined in the present study are academic, not diagnostic, in nature. Second, rather than use LCA for individual test items, the present study used LPA with three continuous variables (percentages of students) as indicators of latent clusters or groups of schools. Third, the final goal of this study was not the identification of groups, but was to identify how school group membership related to average student performance. Thus to address this question a second step was conducted that used the extracted school group membership as variables predicting average student growth.

Some researchers have recently examined latent classes and growth models for academic outcomes, such as reading and math ability, as part of larger growth mixture models (Booth, 2003; McCoach, O'Connell, Reis, & Levitt, 2006; Muthen, Khoo, Francis, & Boscardin, 2003). Growth Mixture Modeling combines growth modeling and LCA to simultaneously estimate intercept and growth parameters for several classes of responders. This at first sounds similar to the technique proposed in the present study; however, a growth mixture model uses both intercepts and growth trajectories when determining group membership. Alternatively, this study employed a two-step analysis, by first classifying schools based on demographic variables, and second, by examining whether school-level initial status, end-of-year status, and growth rate were differentiated by group membership. In this way, the employed methodology allows for a test of the importance or meaningfulness of the groups extracted by the LCA not afforded in the mixture models, and adds to the extant literature on the application of LCA techniques to academic outcomes. In the current study, the academic outcome examined was reading performance, as measured by oral passage reading fluency.

Oral reading fluency

Oral reading fluency, typically measured as the number of words read correctly in 1 min, is an important indicator of reading proficiency (Good & Kaminski, 2002). A standardized set of procedures referred to as Curriculum-Based Measurement (CBM) has been used for years by teachers to assess student progress in oral reading fluency (Fuchs & Deno, 1991). CBM of oral reading fluency is extremely useful, as assessment materials can be administered and scored by teachers to provide quick, reliable, and valid estimates of global reading achievement, especially in the early years of elementary school (Hasbrouck & Tindal, 2006; Fuchs, Fuchs, & Maxwell, 1988).

Fuchs et al. (1988) suggested that any grade-level appropriate passage can provide a valid measure of reading fluency, hence the term “curriculum-based.” One set of commercially available CBM materials — the Dynamic Indicators of Basic Early Literacy (DIBELS) — also provides both criterion (suggestions of low, moderate, and high risk) and norm-referenced (indicating average performance) benchmarks for children’s performance. As a result, DIBELS materials and benchmarks have become widely adopted by school systems and researchers, and were almost universally used as a progress-monitoring measure by states in accordance with *Reading First* requirements. (DIBELS was used by 40 of the 50 states.)

In the present study, students in first, second, and third grades were assessed with the DIBELS oral reading fluency measure four times during the year. This allowed for growth models to be fit for each grade. Because the goal of the present study was to examine the relations between-school risk profiles and reading outcomes, the school clusters extracted from the LPA were added to the growth model as predictors. This was done in such a way that separate estimates were obtained for each identified cluster. For example, if two clusters of schools had been identified by the LPA, then two separate sets of mean growth rates, beginning-of-year intercepts, and end-of-year intercepts would be obtained, one for each cluster.

The hypotheses for the growth models involved the differences of growth and intercept parameters between the groups. There were five likely outcomes for these relations, each of which would have different implications for the relation of school risk to growth. First, the different groups of schools could show no differentiation in either average initial status or growth, which would suggest that school risk was unimportant in student outcomes. Second, a risk group could impact the slope only; groups would be unrelated to beginning-of-year performance, but different slopes could result in differentiation by end-of-year measurement. Third, group membership may impact intercepts only and not growth; school groups would end the year as differentiated as they began. Fourth, group membership may impact both slopes and intercepts, likely resulting in fan-spread growth consistent with a Matthew effect (Stanovich, 1986), such that the schools at high risk would show slower growth than those at lower risk. Finally, and most optimistically, it is also possible that groups would be differentiated at the beginning of the year, but differences in slope would benefit high-risk schools, and as such no differentiation would be observed at the end of the year. Any of these outcomes is possible, and as none has been explicitly explored before, each is important to consider.

Method

Participants

The sample was drawn from the Progress Monitoring and Reporting Network (PMRN), an archival data set containing data on students in every *Reading First* school in the state of Florida. Although this was an archival dataset, approval was granted by the Institutional Review Board to access the data. At the time of extraction, the included 175,857 first through third-grade students were enrolled in 586 *Reading First* schools across the state. Schools with fewer than 100 students representing them in the PMRN (n = 17) were excluded from the analyses, as these schools would not be representative of typical schools in Florida. After applying the exclusionary rule, a total of 569 schools remained in the sample, including 173,485 students in three grades first (n = 58,844), second (n = 56,768), and third (n = 57,873). Within 1st through 3rd grades, 2.8%, 2.9%, and 2.5% of children, respectively, were missing at least one of the four assessment points. Though individual growth curves were estimated for these children, when estimating at the school level, HLM 6.0 deletes cases such that a student who has incomplete data will not be used to estimate the mean effect of their school. A full description of the handling of missing data in a three-level model is included in Raudenbush and Bryk (2002) on pp. 341–343.

On average, the population of Florida *Reading First* schools had 71% minority students, 74% of students receiving free or reduced price lunch, and 16% of students who were ELL. Within each grade, four data points were included. Thus, the data were cross-sectional and longitudinal; three separate grades were measured four times within 1 school year.

Measures

Risk factors—Demographic information for individual students was recorded as part of the PMRN, and typically included gender, ethnicity/minority (hereafter referred to simply as minority), English proficiency, lunch status, migrant status, and exceptionality status. In the current study, minority status, English proficiency, and lunch status were used in the description of school risk, as these factors comprise common indicators of risk in research (Bryk & Raudenbush, 1988; Linn, 2003; Raudenbush, 2004). Minority status was determined by the parent/guardian who designated the student’s status when enrolling each year.

The state of Florida defined English Language Learners (ELL) as students who met one of the following criteria: (a) a student who was not born in the United States and whose native language is other than English; (b) a student who was born in the United States but who comes from a home in which a language other than English is most relied upon for communication, or (c) a student of American Indian or Alaskan Native descent who comes from a home in which a language other than English has had a significant impact on his/her level of English language proficiency *and* who, as a result, has sufficient difficulty speaking, reading, writing, or understanding the English language to deny him/her the opportunity to learn successfully in classrooms in which the language of instruction is English. No distinction was made within the current study as to which criteria the student met to be considered ELL; rather, students were either listed as being an ELL or an English speaker.

Lunch status was defined as the student’s eligibility for participation in the free, reduced, or full price lunch program of the district, as evidenced by the submission of an application by the student’s parent/guardian. Applications may be approved for either free or reduced price lunch, or applications may be denied. Similar to the ELL designation, students were either coded as being approved or denied free or reduced price lunch eligibility. As each of these risk factors was recorded at the student level, the school-level percentages of the three risk conditions were calculated as a percentage of students meeting each criterion within the school.

Dynamic indicators of basic early literacy (DIBELS) oral reading fluency (ORF)—*DIBELS ORF* (5th edition; Good, Kaminski, Smith, Laimon, & Dill, 2001) is a set of measurement passages that can be used to assess oral reading fluency in grade-level connected text. *DIBELS ORF* is a standardized, individually administered test of accuracy and reading rate, and was designed to identify children who may need additional instructional support, and to monitor progress toward instructional goals (Good & Kaminski, 2002). ORF scores used as part of this study were derived from passages administered to students during the months of September, December, February and April in the 2005–2006 academic school year. Performance was measured by having students read three separate passages aloud for 1 min, with the cue to, “be sure to do your best reading” given by the individual assessing the student (Good et al., 2001 p. 30). Words omitted, substituted, and hesitations of more than 3 s were scored as errors, but words self-corrected within 3 s were scored as accurate. The administrator noted the number of errors and reported the number of words correct per minute (WCPM), which is the primary outcome measure of ORF. The median score from the three passages was used as the unit of analysis. Median alternate-form reliability for passages has been shown to be stable at .94 (Good et al., 2001). Reporting of ORF scores into the PMRN is conducted at the school level. That is, once a student is administered the three passages, only the median score is reported into the system. The specific probes used for ORF testing across assessments and grades are presented in Table 1.

Analytic techniques

Latent profile analysis—As was previously mentioned, LPA is a type of LCA that is identical in conceptualization to LCA, except that while in LCA the indicators of the latent class are categorical, in LPA those indicators are continuous. The outcome of both types of analyses culminates in group membership suggestions (which are referred to as *group*s, *clusters*, *classes* or in LPA *profiles*). In this study, the LPA technique was used to cluster schools into groups based on their percentages of students meeting three risk categories. If the LPA was asked to find two profiles (or groups) of schools, there would most likely be one group of schools with high percentages of at-risk students and one group with low percentages. The schools with high percentages would be considered high-risk, while those with low percentages would be considered to be at lower risk.

In typical applications of LCA, such as the ADHD example, a given number of classes are typically tested using theory and prior research to guide the specification of classes. Unlike the ADHD example, which has specific diagnostic criteria, there is not a widely adopted theory suggesting the number of different groups of school risk that should exist. When this is the case, a more exploratory analysis can be conducted (Muthén, 2006).

For the present study, six models were compared, each testing for a different number of clusters. Multiple indices were utilized to determine which number of clusters (groups) was the most appropriate for the data, including the Bayesian Information Criterion (BIC; Kaplan, 2000), the Akaike Information Criteria (AIC; Kaplan, 2000), entropy (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993), and two tests reported in the MPlus program (Muthén & Muthén, 1998), the Lo-Mendell-Rubin Likelihood Ratio Test (TECH11; Lo, Mendell, & Rubin, 2001), and a parametric bootstrapped likelihood ratio test (TECH14; McLachlan & Peel, 2000).

The AIC and BIC are popular measures that combine fit and complexity to compare model parsimony, with lower values indicating more parsimonious model fit. When plotted, these values can be interpreted similarly to a scree plot of eigenvalues in exploratory factor analyses. When the slope of the plotted AIC or BIC curve begins to flatten, it is an indication that there is very little information gained relative to the number of degrees of freedom sacrificed for the model to identify additional clusters (groups). The TECH11 and TECH14 tests were used to determine whether the model being tested fit significantly better than a model with one less group (for example, comparing a three-group model to a two-group model). In both cases, a significant *p*-value (p<. 05) indicates that the model tested fits better than the model with fewer clusters. Finally, the entropy statistic was used to determine how separated the identified clusters were from one another; how much differentiation existed in group membership classification. Entropy values greater than .80 indicate a good separation of the identified groups (Ramaswamy et al., 1993). Each model was evaluated using these statistics, and once the best fit was determined, each school's predicted group membership was extracted, and was used as a fixed effect predictor in models of student-level growth in ORF.

Growth curve modeling—In each grade, three-level growth curve models were fit to the data using HLM6.0 (Raudenbush & Bryk, 2002). The full equation for the final model is available in Appendix A. Growth in ORF was modeled within students (level 1 model, Appendix A), between-student, within-school effects were modeled at level 2, and between-school effects were modeled at level 3. Before the effect of the LPA extracted cluster membership could be estimated, the best unconditional fit to the data was established. This was conducted by fitting a series of models varying whether each parameter was fixed (set to the grand mean for all schools) or random (allowed to vary between schools). All models were fit using Maximum Likelihood estimation. Model fit was statistically evaluated with a chi-square deviance test (a −2 log likelihood test). Results indicated that a model with a random intercept, random linear slope, and random quadratic slope was the most appropriate for all grades[1].

As was previously indicated students were assessed with ORF four times during the year, but the intervals of testing were unequal. When this occurs, Raudenbush and Bryk (2002, p 171) recommend that time be recoded to show growth by month using the formula:

$$\text{Growth rate at time } t = \pi_{1i} + 2\pi_{2i}\,(a_{ti} - L) \tag{1}$$

where π_{1i} is the linear slope for person i at time t, π_{2i} is the change in slope for person i at time t, and a_{ij} is the number of time units for person i at time t, which is subtracted from the centered time variable (L). For example, in this study change parameters were estimated based on time centered at the beginning of the school year (i.e. the month of September was coded as 0). To estimate growth during the month of December, which is 3 months from the centering point, the calculation would be *linear* +2*(*quadratic*)*(3–0). In addition to centering at the beginning of the year, it was also of interest to examine performance at the end of the school year. To do so, an additional model was run for each grade, with time recentered at the end of the year. These models provided a second set of intercepts for each grade, representing end-of-year performance.

[1]The five models that were examined included: a) linear growth model with a random intercept and a fixed slope; b) linear growth model with a random intercept and a random slope; c) nonlinear growth model with random intercept, fixed linear slope, fixed quadratic slope; d) nonlinear growth model with random intercept, random linear slope, fixed quadratic slope; and e) nonlinear growth model with random intercept, random linear slope, random quadratic slope.

To address the main study question, whether mean beginning- and end-of-year ORF scores and growth rates would be differentiated by the clusters (groups) extracted from the LPA, a dummy-coded variable (0 or 1) was created to indicate whether a school belonged in each group. Subsequently, these variables were included in a model that estimated a separate intercept, slope, and quadratic parameter for each group (sometimes referred to as a “no-intercept” model, because no grand-mean intercept is calculated; Raudenbush & Bryk, 2002, p 172). In other words, rather than being reported as a deviation from a reference group, a fitted mean value was estimated for each group. Thus, one uncentered parameter was added to the model for each of the clusters extracted from the LPA (See level 3 equation in Appendix A), and the resulting estimated intercepts and slopes represented the predicted mean values for each group.

For each parameter, pair-wise comparisons were statistically compared between the groups of schools identified in the LPA. For example if three groups were found, then three contrasts would need to be made, comparing the mean of each group to the mean of each other group. To control for the inflation of Type 1 error resulting from the multiple statistical tests being conducted, Benjamini and Hochberg’s (1995) Linear Step Up procedure was used. This procedure differs slightly from other Type 1 error control procedures in that it attempts to keep the ratio of false rejections to total rejections at 5%. The benefit to this approach is that it appears to be more powerful than traditional approaches such as the Bonferroni correction (Maxwell & Delaney, 2004).

Results

Latent profile analysis

The indicators of school risk (percentage of minority, ELL, and FRL students within each school) were used as indicators in an LPA of 569 schools. The procedures used to determine the number of clusters that exist in the data are similar to exploratory factor analyses (e.g., principal components, principal axis factoring) in that different numbers of groups are tested, with subsequent model comparisons made to determine which was the best fit for the data. However, unlike methods such as principal components analysis, more empirical statistical evidence than just a scree plot is available to judge the merit of extracted clusters. A total of six cluster models were tested, using the efficiency indices to evaluate the fit of the model to the data.

The results of the indices used to evaluate model fit are presented in Table 2. Entropy was at an acceptable level (i.e., at least .80) for all six models. Additionally, as the number of clusters increased, both the AIC and BIC indices decreased. The BIC was shown to outperform the AIC on efficiency and consistency in a large simulation study (Yu, 2002), thus the BIC results were plotted, and are presented in Fig. 2. Fig. 2 illustrates that the slope of the curve decreases substantially after the 4-cluster model, suggesting initially that a four-group solution be retained. While the TECH11 and TECH14 likelihood statistics were statistically significant for models including two, three, and four groups, both the TECH11 and TECH14 likelihood statistics were non-significant (respectively, 0.121 and 0.129) when testing the five group model. These results indicated that the four different profiles or groups of schools existed in the data.

Posterior probabilities were also examined to determine whether the model fit well to the data. More specifically, the output of the LPA included each school’s probability of being in each group, as well as each school’s most likely group membership. It is possible that the most likely cluster for a school could be one which still has a relatively low probability; for example, a school in a model with four identified groups could have a 25% chance of being in each group. This would indicate poor school-level fit. Although there are no direct guidelines indicating what level of posterior probability fit is acceptable, for our data, the mean probability for a school that was placed in a given group belonging to that group ranged from 90% to 94%.

To determine the nature of the four groups found by the LCA, the average school-level percentages of Minority, FRL, and ELL in each group were compared to the average percentages in these risk categories for the total population of Florida's *Reading First* schools. Specific percentages of students meeting the risk criteria were reported for each extracted group as well as the total population in Table 3. Statewide *Reading First* base rates were 71.3% Minority, 74.5% FRL, and 16.5% ELL. The first extracted group showed a risk level well below the average *Reading First* school enrollment for students in each of the three risk categories, and will hereafter be referred to as the “Low-Risk” group. The second extracted group was generally slightly below the average risk for the population of schools, and so will be referred to as the “Average-Risk” group. The third extracted group showed higher than average enrollment in FRL and minority but average ELL enrollment, and so will be referred to as the “Poverty-Risk” group. The fourth and final group was just as high as the Poverty-Risk group in the minority and FRL percentages, but also showed high percentages of ELL students. Therefore, this group will be referred to as the “Language-Risk” group in order to distinguish it from the Poverty-Risk group. The four clusters were represented by four dummy-coded variables, and were subsequently entered as the school-level predictors of student-level growth in ORF in first, second, and third grade using a three-level hierarchical growth curve model.

Growth curve modeling

First grade—Table 4 shows the estimates of variance for each parameter in both the unconditional model (where no predictors were included), as well as the conditional model (where the school-level cluster predictors were modeled). As an example, the variance estimate for the school-level intercept in the unconditional model (i.e. u_{00j}) was 72.84. The corresponding value in the conditional model was smaller, (u_{00j} = 14.71), indicating that the inclusion of the four predictor variables from the LPA explained 79.8% of the variance in the school-level intercept. The proportion reduction in variance is reported in Table 4 for all random variables. It should be noted that for parameters where the variances were fixed (e.g. student-level growth), it was not possible to calculate a proportion of reduction in variance.

Using methods described by Raudenbush and Bryk (2002; p. 230), the variance components estimates from the first-grade unconditional model (Column 2 of Table 4) were used to calculate intraclass correlations. These figures indicated how much of the variance in beginning-of-year ORF scores was due to differences between students (76%; 589.55/[589.55 + 72.84 + 115.53]) compared to schools (9%; 72.84/[589.55 + 72.84 + 115.53]). The variability in both the linear growth and quadratic parameters was largely uniform across schools, with only 1.3% of the variance in linear growth due to between-student differences (1.55/[1.55 + 1.46 + 115.53), 1.2% due to between-school effects (1.46/[1.55 + 1.46 + 115.53), and less than 1% of the variance in the quadratic parameter could be attributed to between-student and between-school differences.

Conditional modeling of the latent cluster variables indicated that the differences in school mean initial status and growth were significantly related to the school's group membership. The results of the conditional models, including school-level means, are presented in Table 5. At the beginning of the year, students who were enrolled in schools in the Low-Risk group correctly read, on average, 18 WCPM, while students in schools in the Average-Risk group correctly read nearly 17 WCPM. Conversely, the two groups of schools who were high on risk factors (the Poverty-Risk and Language-Risk groups) demonstrated initial status scores of 15 and 14 WCPM, respectively (a graphical representation of the first-grade conditional model is presented in Fig. 3). The results of the pairwise contrasts conducted to determine whether these

NIH-PA Author Manuscript

NIH-PA Author Manuscript

values were significantly different from one another are presented in Table 6. For example, in first grade, the difference between the Average-Risk and Low-Risk groups on BOY intercept was a statistically significant 1.64 WCPM/month. The largest observed difference in first grade was that between the Language-Risk and Low-Risk group EOY scores, with the Low-Risk group reading 11.18 WCPM/month more than the Language-Risk group.

Linear growth rates were also reliably differentiated based on the school's cluster membership for three of the four clusters (the Poverty-Risk and Language-Risk groups were statistically equivalent, Table 6). The magnitude of the linear slopes followed the same pattern as the intercepts; the Low-Risk group had the strongest linear slope (4.01 WCPM/month) and the Language-Risk group had the weakest slope (3.08 WCPM/month). This resulted in significant differentiation between the four groups at end of the school year (see Table 6). The fitted mean acceleration rates for all four groups were nearly identical at .20 (differences between groups were either 0.01 or 0.0; Table 6).

Second grade—ICCs based on the variance components from the unconditional model (Table 4) indicated that most of the variability in ORF scores at the beginning of the year was due to between-student differences (87.3%), as opposed to between-school differences (5.5%). Approximately 2% of the variance in linear slopes was due to both between-student and school differences; and less than 1% of the variance in the quadratic growth rate was due to student or school differences.

Also reported in Table 4, the inclusion of the latent profile variables as predictors of schools means had little impact on the student-level effects; little reduction was observed for the between-student intercept (0%), linear slope (2%), or quadratic slope (0%). However, the addition of the predictors explained 24% of the variance in school-level intercept and 7% of the variance in the school-level linear trend. The conditional model is represented graphically in Fig. 4 and demonstrates the differential performances at the beginning of the year, as well as the growth patterns across the four clusters of schools.

From Table 5, schools that were identified as Low-Risk began the year with an average student score of 58 WCPM, which was significantly better than the Average-Risk group (53 WCPM; see Table 6 for significance tests). Both the Low-Risk and Average-Risk groups performed significantly better than the Poverty-Risk and Language-Risk groups, correctly reading 46 and 48 WCPM, respectively (Table 6). Regarding linear slope differences, the Low-Risk group demonstrated the largest linear trend at 4.4 WCPM in growth, followed by the Average-Risk group at 3.9 WCPM, and the two groups at highest risk displayed slightly slower growth, at 3.6 (Language-Risk) and 3.7 (Poverty-Risk) WCPM.

Third grade—ICCs from the third grade unconditional model indicated that 87% of the variability in third grade initial status was due to differences between students and 6% was due to differences between schools. The linear slopes were not as differentiated, with only 1.0% and 1.9% due to between-student and between-school differences, respectively. Less than 1% of the variance in the quadratic slope could be attributed to differences between schools or students. Neither the linear or quadratic variance estimates significantly varied between students (Table 4), indicating that student growth was fairly homogenous in the sample.

From Table 4, the inclusion of the school-level predictors explained 29% of the variability in school-level mean initial status, 5% of the linear trend, and 7% of the quadratic growth variances. As previously mentioned, no variability was observed in student slopes; thus, only intercept variability could be explained, with an estimated 1% reduction.

The parameter estimates of the conditional models for Third grade, including school-level mean ORF intercepts and growth rates, are presented in Table 5, with significance tests of observed differences reported in Table 6. Significant differences were observed between the clusters of schools on their beginning-of-year scores (Table 6), with Low-Risk schools outperforming the others (78 WCPM), followed by the Average-Risk (73 WCPM), Poverty-Risk (67 WCPM) and Language-Risk (66 WCPM) schools. Differences in linear slopes were also observed: Language- and Poverty-Risk groups demonstrated stronger linear growth (5.50 and 5.35 WCPM/month) than either the Low-Risk or Average-Risk groups (5.03 and 5.09 WCPM/month). No other significant differences were observed for linear or quadratic slopes (Table 6).

Though no significant between-group differences were observed, it is interesting to note that a negative quadratic change, or deceleration effect was observed (approximately −0.16, Table 5).This indicates that while school-level average gains in ORF increased over the course of the school year, the amount of gain made decreased over time. This phenomenon may be observed in the graphical depiction of the third-grade conditional model in Fig. 5. The plots demonstrated that schools grew over time, yet the rate of change, evidenced by the slope of the line, levels off towards the end of the year.

Discussion

In the present study, the hypothesis was tested that school-level percentages of students eligible for free or reduced priced lunch, minority students, and students learning English could be used to cluster together schools into clearly defined risk groups using LPA. The results suggested that in this selected sample of primarily high-risk *Reading First* schools, four distinct groups of schools could be identified. The Low-Risk group of schools had below-average enrollment of students in the three risk categories, schools in the Average-Risk group contained just below the average percentages of at-risk students, the Poverty-Risk-schools had higher than average enrollment of FRL and minority students but average ELL enrollment, and the Language-Risk group showed high enrollment percentages on all three risk factors.

Similar to other findings in this literature (Klingele & Warrick, 1990; McDermott, 1995; Sirin, 2005), schools with a large percentage of students living in poverty also had high minority enrollment. This may indicate that in this sample the two were redundant in creating qualitatively different groups of school risk. If only two groups had been observed, one with all high indicators and one with all low indicators, it would suggest that all three variables were not necessary in the prediction of school risk; any one variable would have worked the same way. The amount each variable contributed to the classification of schools is a direct function of the percentages of students meeting the three risk criteria in each group. As such, our findings suggest that poverty and minority percentages were redundant with one another, but both were distinct from ELL percentage. While the reviewed studies consider schools to be high poverty or high risk when they contain high percentages of students meeting just one risk criterion (e.g., FRL students, Raudenbush, 2004), the results of the LCA suggest that the use of two separate factors was needed to do so. Thus, even in this selected sample, the classification of schools is more complex than what is currently being used in the literature.

After determining that schools were able to be grouped based on their demographic makeup, we next determined whether these clusters could be used to predict average student outcomes. Growth modeling results indicated that a school's risk group membership, as determined in the LPA, was significantly related to average student ORF scores at the beginning of the school year in all three grades. This finding is consistent with others in the literature; that test scores typically reflect background characteristics present before children enter school (Bryk & Raudenbush, 1988; Linn, 2003; Raudenbush, 2004). As ORF scores were differentiated at the beginning of the year, some argue that teachers should not be held accountable for factors they cannot control, such as economic background, as it would require differential improvement, with average students making a year of growth, but below-average students requiring stronger growth to be considered on grade level (Elmore, Abelmann, & Fuhrman, 1996). However, Elmore et al. argue that controlling for these types of demographics makes acceptable, and even institutionalizes, low expectations for poor, minority, and non-English speaking students.

If only beginning-of-the-year differentiation was observed, the argument could have been made for controlling for demographic information such as percentage of minority or ELL students. However, there were also significant differences between groups observed in linear slope in first and second grades. These differences were consistent with the Matthew effect (Stanovich, 1986), such that schools in the lower risk groups not only had students who began the year higher, but they also showed stronger average growth than higher risk groups. In third grade, however, this trend was not present; no between-group differences were observed in slopes (Table 6).

Interestingly, in both first and second grades there were significant quadratic trends in oral reading scores, such that student growth accelerated during the school year. In third grade, however, students showed the opposite trend, slowing in growth towards the end of the year. Although difficult to determine why such a trend is observed from the present data, it could be due to differential difficulty in the ORF passages across the three grades. Some recent work in oral reading fluency found significant differences between several ORF passages presumed to be equivalent based on readability and difficulty estimates (Ardoin & Christ, 2008). Further, evidence suggests that a lack of equated passages can lead to incongruent ORF scores (Francis, Santi, Barr, Fletcher, Varisco, & Foorman, 2008). To determine whether this is occurring, future studies could consider explicitly measuring the text difficulty of these passages, and include it as a predictor in analyses.

There are some aspects of this study that limit its generalizability. First, although this study included several thousand students all students were attending elementary schools in Florida. Therefore, the findings may not generalize to other geographical areas. Second, as the included schools were all part of *Reading First*, all students were enrolled in relatively high risk, poor performing schools. Thus, the criteria used in this study for a school to be considered high risk in the present study should not be applied to other settings or samples drawn from different populations. Furthermore, the higher than average percentages of at-risk students enrolled in the schools used in this study impact the clusters of schools obtained. It is likely that if additional schools with lower concentrations of at-risk students were included in the sample, different groups of schools and likely larger differentiation between the groups would be observed.

Also, it is important to note that classrooms were not included in this analysis, thus any variance due to teachers is being attributed to students. Therefore, while it may appear that nearly all of the variance (for example, 73% in first grade) is attributable to between-student differences, this statistic also includes any variance that could be due to between-teacher differences. Therefore, the amount of variance within classrooms and between students is still unknown, and would be an interesting direction to take these findings in the future.

Implications

The findings of the present study have strong implications for both policy and applied research. First, the fact that both intercepts and growth rates were reliably differentiated based on school group membership highlights the importance of the school setting in which students receive their instruction. When developing standards or expectations for student end-of-year achievement, policymakers and school personnel often rely on results from national normative samples such as those reported in the DIBELS despite the previously mentioned ease in development of local CBM norms. Similarly, when evaluating rates of growth, schools typically develop goals for individual students using rates of growth reported as standard within the CBM literature as opposed to local standards. This study suggests that this application may not always be appropriate. Alternatively, norms could be locally developed based on the populations from which students are drawn using the techniques outlined in the CBM literature, as suggested by Fuchs et al. (1988).

Finally, growth across the academic year is often assumed to be linear, meaning that students are assumed to grow at a constant rate throughout the year. The significant acceleration and deceleration effects observed in this study, however, suggest that rates of growth may actually change across the year. This finding replicates and extends recent research suggesting that growth across the academic year is not linear (Ardoin & Christ, 2008; Hasbrouck & Tindal, 2006), providing evidence from a large sample and data from across multiple grades to illustrate the nonlinearity of growth across the academic year.

In sum, the present study has described a new technique for identifying distinct groups of school-level concentrations of at-risk students; a technique which could be applied to many other facets of educational practices. The results suggest that, similar to other findings in this literature, school-level demographics are important determinants of student achievement. It was also concluded that different profiles of school risk exist, not simply high and low poverty, thus suggesting that the methods typically used to determine school-level risk may be insufficient to appropriately classify schools. Finally, future studies of at-risk schools should consider using more than one indicator of risk, looking for different types of risk, and most importantly, comparing the demographic information of their sample to that of the population from which it is drawn before determining risk level.

Acknowledgments

One of the authors (J.A.R. Logan) received pre-doctoral training support during the conduct of this research through grant R305B04074 from the Institute of Education Sciences to Florida State University and was supported in part by the National Institute of Child Health and Human Development (NICHD) Grant HD038075 to Ohio State University. We would like to thank Scott Ardoin and several anonymous reviewers for their help preparing the manuscript for publication.

Appendix A: Structural multilevel model

Level 1 Model:

$$\begin{aligned} Y_{ijt} = {} & \pi_{0ij} + \pi_{1ij}(\text{Month})_{ijt} \\ & + \pi_{2ij}(\text{Month})^2_{ijt} + e_{ijt} \end{aligned}$$

Level 2 Model:

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + r_{0ij} \\ \pi_{1ij} &= \beta_{10j} + r_{1ij} \\ \pi_{2ij} &= \beta_{20j} + r_{2ij} \end{aligned}$$

Level 3 Model:

$$\beta_{00j}=\gamma_{001}(\text{Low})_j+\gamma_{002}(\text{Average})_j+\gamma_{003}(\text{Poverty})_j+\gamma_{004}(\text{Language})_j+u_{00j}$$
$$\beta_{10j}=\gamma_{101}(\text{Low})_j+\gamma_{102}(\text{Average})_j+\gamma_{103}(\text{Poverty})_j+\gamma_{104}(\text{Language})_j+u_{10j}$$
$$\beta_{20j}=\gamma_{201}(\text{Low})_j+\gamma_{202}(\text{Average})_j+\gamma_{203}(\text{Poverty})_j+\gamma_{204}(\text{Language})_j+u_{20j}$$

References

Agrawal A, Lynskey M, Madden P, Bucholz KK, Heath AC. A latent class analysis of illicit drug use/dependence: Results from the National Epidemiological Survey on Alcohol and Related Conditions. Addiction 2007;101:94–104. [PubMed: 17207127]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth Edition. American Psychiatric Association; Washington, DC: 1994.

Ardoin SP, Christ TJ. Evaluating curriculum-based measurement slope estimates using data from tri-annual universal screenings. School Psychology Review 2008;37:109–125.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 1995;57:289–300.

Booth KM. Evaluating academic outcomes of head start: An application of general growth mixture modeling. Early Childhood Research Quarterly 2003;18:238–254.

Bryk T, Raudenbush SW. Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. Psychological Bulletin 1988;104:396–404.

Connor CM, Son S, Hindman AH, Morrison FJ. Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. Journal of School Psychology 2005;43:343–375.

California Department of Education. California Standardized Testing and Reporting (STAR). [Retrieved December 4, 2007]. 2005 from http://star.cde.ca.gov/star2005/Viewreport.asp

deNijs PFA, Ferdinand RF, Verhulst FC. No hyperactive–impulsive subtype in teacher-rated attention-deficit/hyperactivity problems. European Child & Adolescent Psychiatry 2007;16:25–32. [PubMed: 16972118]

Denton CA, Foorman BR, Mathes PG. Perspective: Schools that "beat the odds" — Implication for reading instruction. Remedial and Special Education 2003;24:258–272.

Elmore, RF.; Abelmann, C.; Fuhrman, S. The new accountability in state education reform: From process to performance. In: Ladd, Helen F., editor. Holding schools accountable: Performance-based reform in education. Brookings Institution; Washington, D.C: 1996. p. 65-98.

Francis DJ, Santi KL, Barr C, Fletcher JM, Varisco A, Foorman BR. Form effects on the estimation of students' oral reading fluency using DIBELS. Journal of School Psychology 2008;46:315–342. [PubMed: 19083362]

Fuchs LS, Deno SL. Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children 1991;57:488–501.

Fuchs LS, Fuchs D, Maxwell L. The validity of informal reading comprehension measures. Remedial and Special Education 1988;9(2):20–28.

Goldhaber, D.; Brewer, D. Teacher licensing and student achievement. In: Kanstoroom, Marci; Finn, Chester E., Jr., editors. Better teachers, better schools. The Thomas B. Fordham Foundation; Washington, DC: 1999. p. 83-102.

Good, RH., III; Kaminski, RA. DIBELS oral reading fluency passages for first through third grades (Technical Report No. 10). University of Oregon; Eugene: 2002.

Good, RH.; Kaminski, RA.; Smith, S.; Laimon, D.; Dill, S. Dynamic indicators of basic early literacy skills. 5th ed. University of Oregon; Eugene, OR: 2001.

Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. Reading Teacher 2006;59:636–644.

Kannapel, PJ.; Clements, SK. Inside the black box of high-performing high-poverty schools: A report from the Prichard Committee for Academic Excellence. Prichard Committee for Academic Excellence; Lexington, KY: 2005.

Kaplan, D. Structural equation modeling: Foundations and extensions. Sage Publications; Thousand Oaks, CA: 2000.

Kim JS, Sunderman GL. Measuring academic proficiency under the no child left behind act: Implications for educational equity. Educational Researcher 2005;34(8):3–13.

Klingele WE, Warrick BK. Influence of cost and demographic factors on reading achievement. Journal of Educational Research 1990;83:279–282.

Linn RL. Accountability: Responsibility and reasonable expectations. Educational Researcher 2003;32:3–13.

Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika 2001;88:767–778.

Maxwell, SE.; Delaney, HD. Designing experiments and analyzing data: A model comparison perspective. 2nd ed. Erlbaum; Mahwah, NJ: 2004.

McCoach DB, O'Connell AA, Reis SM, Levitt HA. Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. Journal of Educational Psychology 2006;98:14–28.

McDermott PA. Sex, race, class, and other demographics as explanations for children's ability and adjustment: A national appraisal. Journal of School Psychology 1995;33:75–91.

McLachlan, GJ.; Peel, D. Finite mixture models. John Wiley; New York: 2000.

Mosteller, F.; Moynihan, DP. A pathbreaking report. In: Mosteller, F.; Moynihan, DP., editors. On equality of educational opportunity. Random House; New York: 1972. p. 3-66.

Murnane, RJ.; Willett, JB.; Bub, KL.; McCartney, K. Understanding trends in the black–white achievement gaps during the first years of school. Brookings-Wharton Paper on Urban Affairs. 2006.

Muthén B. Should substance use disorders be considered as categorical or dimensional? Addiction 2006;101:6–16.

Muthén, LK.; Muthén, BO. Mplus user's guide. Fourth Edition. Muthén & Muthén; Los Angeles, CA: 1998–2006.

Muthen, B.; Khoo, S.; Francis, DJ.; Boscardin, CK. Analysis of reading skills development from kindergarten through first grade: An application of growth mixture modeling to sequential processes. In: Reise, SP.; Duan, N., editors. Multilevel modeling: Methodological advances, issues, and applications. Lawrence Erlbaum Associates; Mahwah: NJ: 2003.

NAEP Assessment Tool. NAEP questions. [Retrieved September 11, 2009]. Jun 152007 from http://nces.ed.gov/nationsreportcard/itmrls/startsearch.asp

National Center for Educational Statistics. Number of public school districts by 2003 Rural–Urban Continuum (Beale) county code and state: 2003–04. [Retrieved December 4, 2007]. 2005 from http://nces.ed.gov/surveys/ruraled/TablesHTML/1bealedistricts.asp

National Center for Education Statistics. Average reading score for 4th-, 8th-, and 12th-graders, by selected student and school characteristics: 1992 and 2005. [Retrieved December 4, 2007]. 2007 from http://nces.ed.gov/programs/coe/2007/section2/table.asp?tableID=678

New York State Education Department. Percent distribution of public school students by racial/ethnic origin. [Retrieved December 4, 2007]. 2007 from http://www.emsc.nysed.gov/irts/educationstats/edstats-07/table4.pdf

Ramaswamy V, DeSarbo WS, Reibstein DJ, Robinson WT. An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Marketing Science 1993;12(1):103–124.

Raudenbush, SW. Schooling, statistics, and poverty: Can we measure school improvement?. Educational Testing Service, Policy Evaluation and Research Center; 2004. William H. Angoff Memorial Lecture Series

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd edition. Sage; Newbury Park, CA: 2002.

Reardon, SF.; Robinson, J. Patterns and trends in racial/ethnic and socioecomonic academic achievement gaps. In: Ladd, H.; Fiske, E., editors. Handbook of research in education finance and policy. Prentice Hall; New York, NY: 2007.

Sirin SR. Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research 2005;75:417–453.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Stanovich KE. Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly 1986;21:360–407.

Tennessee Department of Education. Tennessee statewide report card 2005. [Retrieved December 4. 2007]. 2005 from http://www.k-12.state.tn.us/rptcrd05/state1.asp

Congress, U. S.. No Child Left Behind Act of 2001. Public Law 107–110. 107th Congress. Washington, DC: Government Printing Office; 2001.

U.S. Department of Education. National Center for Education Statistics. The condition of education 2006, NCES 2006-071. Washington, DC: U.S. Government Printing Office; 2006.

VanLang ND, Ferdinand RF, Ormel J, Verhulst FC. Latent class analysis of anxiety and depressive symptoms of the youth self-report in a general population sample of young adolescents. Behaviour Research and Therapy 2006;44:849–860. [PubMed: 16122697]

Yu, CY. Doctoral dissertation. University of California; Los Angeles: 2002. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.

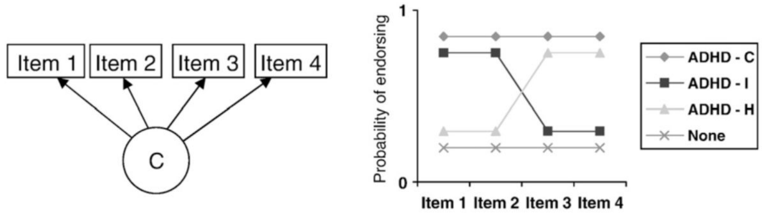

Fig. 1.
Example of a LCA (figures adapted from Muthén, 2006). The left side of the figure shows four observed items loading on a latent class variable. The right side of the figure shows the probability of endorsing each of the four items as a function of four extracted classes.

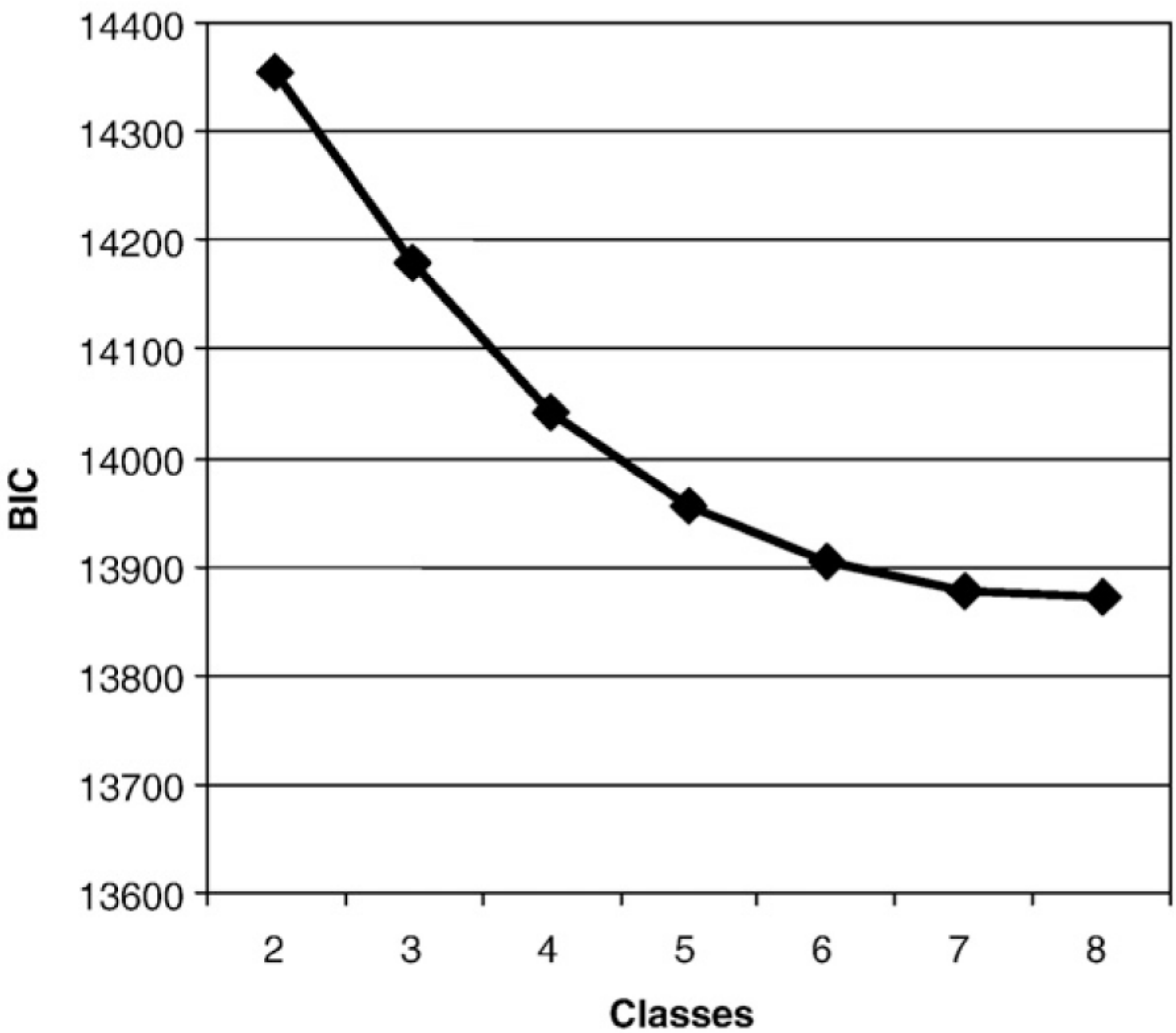

Fig. 2.
The Bayesian Information Criterion of the 7 different LPA models comparing the number of groups that best fits the data.

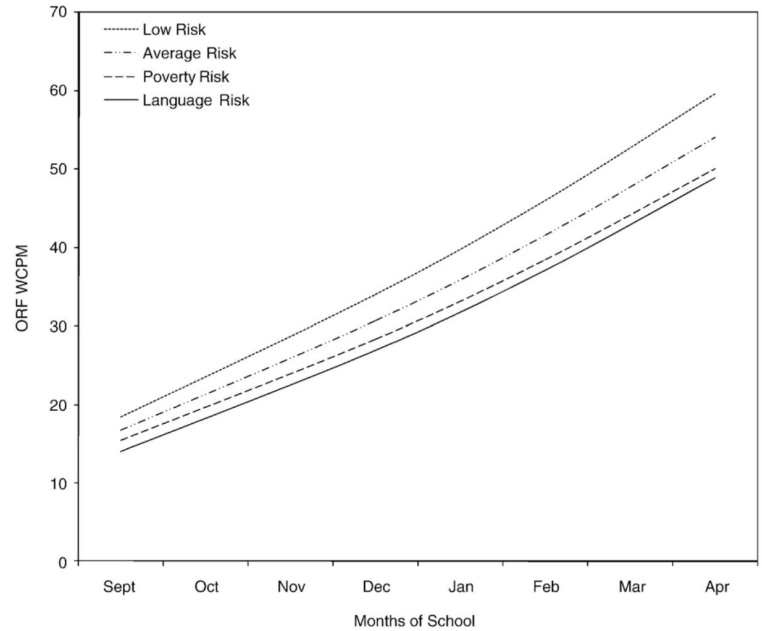

Fig. 3.

Average first-grade ORF performance across the school year as a function of the four identified groups of schools.

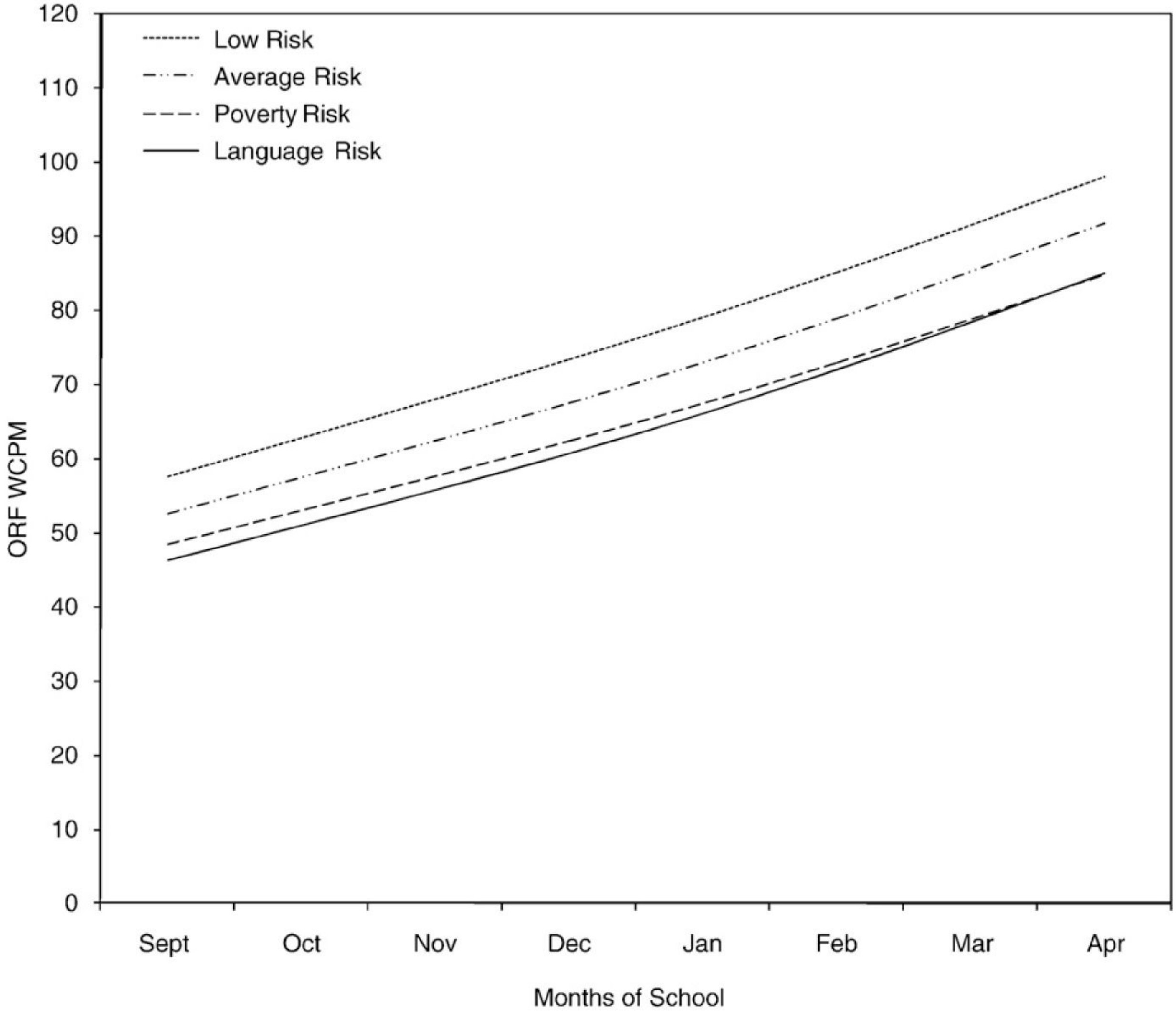

Fig. 4.
Average second-grade ORF performance across the school year as a function of the four identified groups of schools.

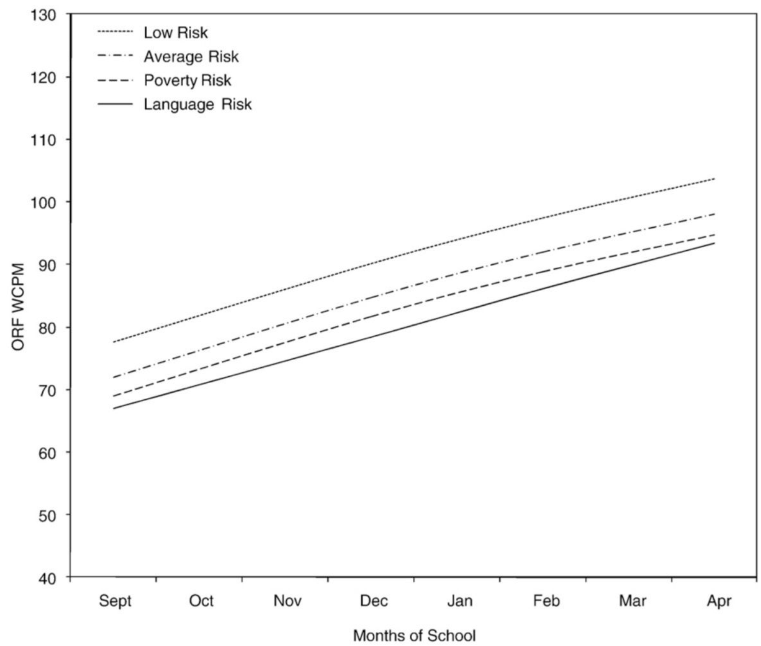

Fig. 5.
Average third-grade ORF performance across the school year as a function of the four identified groups of schools.

Table 1

DIBELS ORF passage name and reported spache difficulty.

		September		December		February		April	
Grade	Passage order	Passage	Spache	Passage	Spache	Passage	Spache	Passage	Spache
1st	1	Party	2.0	Nest	2.0	Aunt Rose	2.3	Spring	2.0
	2	Ice cream	2.2	Camping	2.1	Big Sister	2.1	Sand castle	2.2
	3	Kitty	2.2	Lemonade	2.2	Rocks	2.3	Check-up	2.3
2nd	1	Story	3.0	Job	2.0	Robot	3.0	Rollercoaster	3.0
	2	Babysitter	3.0	Handprints	3.0	Grandpa	3.0	Moving day	3.0
	3	Shuffleboard	3.0	Meals	3.0	Bottle	3.0	Stars	3.0
3 rd	1	Dreams	3.0	Field Trip	3.0	Pots	3.0	Friend	3.0
	2	Clouds	3.0	Whale	3.0	Tracks	3.0	Camping	3.0
	3	Firefighters	2.9	Email	3.0	Parents	2.9	Garden	3.0

Note. Party=The Block Party, Kitty=Our Sick Kitty, Story=Writing my Life Story, Babysitter=I'm a Good Babysitter, Dreams=Dream Catchers, Clouds=Clouds and Weather, Nest=The Robin's Nest, Camping=Camping at Home, Lemonade=My Lemonade Stand, Job=Mom's New Job, Meals=Means on Wheels, Whale=Keiko the Killer Whale, Aunt Rose=Visiting Aunt Rose, Rocks=The Rock Collection, Robot=If I had a Robot, Grandpa=My Grandpa Snores, Bottle=My Drift Bottle, Tracks=Animal Tracks, Spring=Spring is Coming, Rollercoaster=Riding the Rollercoaster, Stars=Stars of the Sea, Camping=Going to Family Camp, Garden=Planting a Garden.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Fit indices from model testing.

Classes	AIC	BIC	TECH11	TECH14	Entropy
2	14,311.2	14,354.6	0.000	0.000	0.865
3	14,119.9	14,180.8	0.000	0.000	0.847
4	13,965.4	14,043.6	0.001	0.000	0.865
5	13,860.2	13,955.9	0.121	0.129	0.888
6	13,791.6	13,904.6	0.128	0.129	0.882
7	13,724.6	13,872.3	0.394	0.401	0.888

AIC=Akaike Information Criteria, BIC=Bayesian Information Criterion.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Average percentages of students meeting risk criteria within schools for the four classes found by the LPA, and for the population of Reading First schools.

	Class 1	Class 2	Class 3	Class 4	All *Reading First* schools
	"Low-Risk"	"Average-Risk"	"Poverty-Risk"	"Language-Risk"	
Minority	27.62	58.16	93.35	89.90	71.3
FRL	57.37	67.51	84.10	83.61	74.5
ELL	3.73	13.20	10.52	45.70	16.5

Note: FRL=Free or Reduced Price Lunch, ELL=English Language Learners.

Table 4

Variances for random effects of the unconditional (no predictors) and conditional models (with predictors), and variance explained by the inclusion of school-level predictors.

Random effect	Unconditional	Conditional	Variance explained
First grade			
Level 1, e_{ijt}	115.53	52.47	
Level 2 intercept, r_{0ij}	589.55	429.10	27.8%
Level 2 linear slope, r_{1ij}	1.55 [a]	–	–
Level 2 quadratic slope, r_{2ij}	0.011 [a]	–	–
Level 3 intercept, u_{00j}	72.84	14.71	79.8%
Level 3 linear slope, u_{10j}	1.46	1.11	24.0%
Level 3 quadratic slope, u_{20j}	0.011	0.010	9%
Second grade			
Level 1, e_{ijt}	73.67	72.27	
Level 2 intercept, r_{0ij}	882.13	882.02	<001%
Level 2 linear slope, r_{1ij}	1.60	1.57	1.9%
Level 2 quadratic slope, r_{2ij}	0.009 [a]	–	–
Level 3 intercept, u_{00j}	55.20	41.71	24.4%
Level 3 linear slope, u_{10j}	1.50	1.40	6.7%
Level 3 quadratic slope, u_{20j}	0.028	0.028	0%
Third grade			
Level 1, e_{ijt}	86.82	79.72	
Level 2 intercept, r_{0ij}	1012.06	1001.55	1.0%
Level 2 linear slope, r_{1ij}	0.94 [a]	–	–
Level 2 quadratic slope, r_{2ij}	0.003 [a]	–	–
Level 3 intercept, u_{00j}	64.19	45.64	28.9%
Level 3 linear slope, u_{10j}	1.69	1.61	4.7%
Level 3 quadratic slope, u_{20j}	0.030	0.028	6.7%

[a] Variance was not statistically significant (as tested by a χ^2 test), thus the effect was fixed in the conditional model. This model was centered at the beginning of the school year.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Parameter estimates for growth curve models (beginning- and end-of-year means, linear and quadratic growth rates).

	Low-Risk		Average-Risk		Poverty-Risk		Language-Risk	
Fixed effect	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**	**Estimate**	**S.E.**
1st grade								
BOY intercept	18.37	0.50	16.73	0.32	15.42	0.32	14.07	0.37
EOY intercept	56.51	0.50	50.84	0.32	46.63	0.32	45.33	0.37
Linear growth	4.01	0.13	3.45	0.09	3.15	0.09	3.08	0.10
Acceleration	0.20	0.02	0.20	0.01	0.19	0.01	0.20	0.01
2nd grade								
BOY intercept	57.51	0.78	52.65	0.57	48.37	0.57	46.31	0.64
EOY intercept	95.70	0.78	88.83	0.57	82.04	0.57	81.02	0.64
Linear growth	4.39	0.14	3.89	0.10	3.7	0.1	3.55	0.15
Acceleration	0.15	0.02	0.18	0.01	0.16	0.01	0.20	0.02
3rd grade								
BOY intercept	78.47	0.78	72.69	0.62	67.42	0.62	65.62	0.69
EOY intercept	107.10	0.78	101.23	0.62	96.21	0.62	95.82	0.69
Linear growth	5.03	0.16	5.09	0.11	5.35	0.11	5.50	0.16
Acceleration	−0.14	0.02	−0.15	0.01	−0.18	0.01	−0.17	0.02

Note: BOY=Beginning-of-year, EOY=End-of-year.

Table 6

Mean between-group differences for intercepts (beginning- and end-of-year) and growth.

	First grade			Second grade			Third grade		
	Average	**Poverty**	**Language**	**Average**	**Poverty**	**Language**	**Average**	**Poverty**	**Language**
Intercept (beginning-of-year)									
Low-risk	1.64*	2.95*	4.3*	4.86*	9.14*	11.2*	5.78*	11.05*	12.85*
Average-risk		1.31*	2.66*		4.28*	6.34*		5.27*	7.07*
Poverty-risk			1.35*			2.06*			1.8*
Intercept (end-of-year)									
Low-risk	5.67*	9.88*	11.18*	6.87*	13.66*	14.68*	5.87*	10.89*	11.28*
Average-risk		4.21*	5.51*		6.79*	7.81*		5.02*	5.41*
Poverty-risk			1.3*			1.02*			0.39
Linear growth									
Low-risk	0.56*	0.86*	0.93*	0.5	0.69*	0.84*	−0.06	−0.32	−0.47
Average-risk		0.3*	0.37*		0.19	0.34		−0.26	−0.41
Poverty-risk			0.07			0.15			−0.15
Quadratic growth									
Low-risk	0	0.01	0	−0.03	−0.01	−0.05	0.01	0.04	0.03
Poverty-risk		0.01	0		0.02	−0.02		0.03	0.02
Average-risk			−0.01			−0.04			−0.01

Note. Column values were subtracted from row values to obtain difference scores.

* $p<.05$.