Investigating Measurement Invariance Assumption Using Item Parameter Drift Across Grade Levels and ELL Groups

Luke Stanke, Jose Palma, Okan Bulut, & Michael C. Rodriguez
University of Minnesota

Minnesota Youth Development Research Group

April 2013

Paper presented at the annual meeting of the
National Council on Measurement in Education
San Francisco, CA

Citation:

Stanke, L., Palma, J., Bulut, O., & Rodriguez, M.C. (2013, April). *Investigating measurement invariance assumptions using item parameter drift across grade levels and ELL groups*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Investigating Measurement Invariance Assumption

Using Item Parameter Drift Across Grade Levels and ELL Groups

Introduction

Questions regarding the interpretation of measures are central to validity, particularly in the context of increasing school diversity and desire to generalize across subgroups, administration occasions, and for some measures, grades. Across multicultural contexts and grade levels, measurement invariance is an important assumption to facilitate a common interpretation framework. Evidence must be gathered to support inferences from measures where score interpretation is important across subgroups, particularly subgroups for which special interventions are designed.

A common method for assessing the functioning of items across important subgroups is differential item functioning (DIF), which addresses the functioning of items for subgroups conditioned on the total trait level. Zumbo (2003) evaluated the question of measurement invariance and asked whether item-level DIF is the relevant characteristic or is scale-level measurement invariance the issue with respect to translating language tests. He found that the relevant question is at the item level, not at the scale score level. Zumbo (2007) offered important reasons for future DIF analyses. First, we need to know not only whether subgroups perform differentially on some items but also why they perform differently. Second, DIF can be used as part of a research agenda to isolate variables that may affect item performance. Third, cognitive researchers are using item response theory (IRT) to increase understanding of the cognition behind a DIF finding.

Item parameter drift (IPD) is a parallel alternative to DIF for detecting measurement variability. When measurement bias is examined by groups categorized by testing occasions or time-related variables, it is referred to as IPD (Goldstein, 1983). In DIF studies, measurement invariance is examined between groups of examinees, grouped on a characteristic unrelated to the construct being measured. In IPD studies, the examinees are studies by time variables. For this study, items were delivered to students in a single administration, but the students varied by grade-level, therefore grade-level will act as the relevant time variable. Items parameters may drift for several reasons, these include, but not limited to: increased coverage in the topic by the mass media (Donoghue & Isham, 1998), changes to a curriculum (Goldstein, 1983; Bock, Muraki, & Pfeiffenberger, 1988), teaching-to-test preparation (Wu, Li, Ng, & Zumbo, 2012), or total number of item exposures (Smith, 2004), essentially a context condition significant enough to change the construct or some aspect of the construct.

In this study, IPD is being analyzed in the context of explanatory item response modeling (EIRM). EIRM has emerged as the special case of Item Response Theory (IRT) in the context of generalized linear mixed models (GLMMs), where it is possible to specify between-group differences in the latent constructs being measured (De Boeck & Wilson, 2004). Here, parameter variability was evaluated on youth development measures in a large Latino student sample, given their English language learner (ELL) status; current-ELL, exited ELL services, or non-ELL; and grade level. The ELL characteristics are described more completely below.

Youth Development

In their comprehensive review of the theory and research on positive youth development, Benson, Scales, Hamilton, and Sesma (2006) identified six essential principles about which there is broad consensus, including (a) youth have the inherent capacity for positive development; (b) positive development is enabled through relationships, contexts, and environments that nurture development; (c) positive development is enhanced when youth participate in multiple meaningful relationships, contexts, and environments; (d) all youth benefit from these opportunities, the benefits of which generalize across gender, race, ethnicity, and family income; (e) community is a critical delivery system for positive youth development; and (f) youth themselves are major actors in their own development, serving as a central resource for creating the kinds of relationships, contexts, environments (ecologies), and communities that facilitate optimal development.

A positive vision of youth potential has implications for research, education, and social policy (Damon, 2004). "Changes across the life span are seen as propelled by the dynamic relations between the individual and the multiple levels of the ecology of human development (family, peer group, school, community, culture), all changing interdependently across time" (Lerner, 2002, as cited by Benson et al., p. 904). Others have investigated important cultural contexts relevant to the development of ethnic minority youth (McLoyd, 1998; Rodriguez & Morrobel, 2004; Sesma & Roehlkepartain, 2003; Spencer, 1995).

In evaluating the work in this area, it is important to assess measurement invariance and consistency of the inferences regarding the role of assets for minority youth, associated score differences among subgroups, and the magnitudes of correlations with important developmental educational outcomes. The researchers engaging in this work have been examining large scale databases in the ecological context of positive youth development (Albano & Rodriguez, 2012; Cabrera, & Rodriguez, 2011; Cabrera, & Rodriguez, 2010; Palma, Rodriguez, Cabrera, Albano, Vue, Warshawsky, 2012; Warshawsky, Rodriguez, Cabrera, Palma, Albano, & Vue, 2012). In a recent study, Albano and Rodriguez (2012) demonstrated the use of a hierarchical linear model to evaluate the parameter drift over time of a measure of school climate. They found that some items became easier to endorse over time, conditioned on overall school climate, suggesting a model where students may becoming desensitized to issues related to school climate, resulting in shifts in score interpretation.

Latino-English Language Learners

In the context of score and item functioning on language background, DIF analyses are typically used to investigate measurement invariance across groups on measures of academic performance and not as common in other types of measures. In this study, DIF is investigated across three groups of English Language Learners (ELL) on a measure of developmental assets (Positive Identity and Support). All ELL students, in this study identified themselves as Latino and were classified as current-ELL, exited- ELL, or non- ELL. Latino students currently receiving ELL services are those for whom English is not their primary language and scored below 4 on the Woodcock Muñoz assessment and/or below benchmarks on the English OAKS (Oregon assessment of Knowledge Skills). Latino students who have exited ELL services have satisfied their ELL learning goals and no longer quality for ELL services. Latino students who are not ELL students are those for whom English is their primary language.

The Latino population is the largest ethnic minority group in the United States, 15.4% of the total U.S. population (U.S. Census Bureau, 2010). Studies frequently do not breakdown Latino populations into smaller subgroups; however it is important to explore the with-in group variations (Fuligni & Perreira, 2009). This is particularly relevant regarding language proficiency and assessments that are English based for subpopulations with different levels of English proficiency.

The Latino population is not only disproportionally young (38% under the age of 20) but it is also disproportionally poor (28.6% of those under the age of 18 live below the national poverty level) according to the U.S. Census Bureau (2010). Eamon and Mulder (2005) found that Latino youth face many developmental risks at the family, community, and school level that increase the likelihood of involvement in risky behaviors and prevent them from attaining higher levels of academic success. Similarly, researchers have shown that first-generation Latino immigrants tend to have higher levels of psychological, behavioral, and educational adjustment than U.S.-born Latinos in the presence of long-lasting social and economic challenges (Fuligni & Perreira, 2009). This phenomenon is commonly known as the immigrant paradox. Some of the key factors that affect the immigrant paradox are the retention of cultural values (i.e., traditions and native language), ethnic identification, and age of immigration (Fuligni & Perreira, 2009). Similarly immigrants entering the United States at later ages often tend to show higher levels of adjustment; however the reasons for this are not clear (Vega, Sribney, Aguilar-Gaxiola, & Kolody, 2004, as cited by Fuligni & Perreira, 2009). In addition, Latino youth tend to segregate in terms of birthplace and age of immigration. Latinos who immigrated at later ages tend not to join peer groups of Latinos who immigrated at earlier ages because of language barriers and stereotypes (Matute-Bianchi, 1991, as cited by Fuligni & Perreira, 2009).

Finally we recognize the ongoing need to support positive youth development among the fastest growing segment of the US population – Latino youth (see Contreras, Flores-Ragade, Lee, & McGuire, 2011, for a review of research relevant to the growing population of Latino youth in educational contexts). For these reasons and as part of a larger body of research, investigating developmental assets in terms of ELL status among Latino youth across grades is a timely endeavor.

Detecting Parameter Drift

The EIRM framework extends the Rasch (1960) model, which describes the probability of endorsing an item as a function of the difference between person *j*'s ability and item *i*'s difficulty, and can be written as a GLMM. The notation for this is a modified version of the hierarchical generalized linear modeling framework presented by Kamata (2001). The level-1 portion of the Rasch model, the response level can be written as a logistic regression model in terms of the log-odds of correct response:

$$\begin{aligned}\log\left(\frac{P_{ij}}{1-P_{ij}}\right) &= \eta_{ij} = \beta_{0j} + \beta_{1i}X_{1ij} + \cdots + \beta_{kj}X_{kij} \\ &= \beta_{0j} + \sum_{q=1}^{k}\beta_{qj}X_{qij}\end{aligned} \tag{1}$$

where η_{ij} is the log-odds of person *j* endorsing the dichotomous item *i*, β_{0j} is an intercept term and β_{1j} through β_{qj} are coefficients associated with an indicators X_{1ij} through X_{kij}. For these indicators, X_{qij} represents the *q*th dummy variable for person *j*, and is coded as negative one when $q=i$, and zero when $q\neq i$. Because dummy coding is negative one and zero the item parameters are interpreted as item difficulty. Level-2 of the model, the person-level, is then described as

$$\begin{cases}\beta_{0j} = u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \vdots \\ \beta_{kj} = \gamma_{k0}\end{cases} \tag{2}$$

where the intercept is equal to a random effect for persons, u_{0j}, that is $N(0, \tau_j)$. Coefficients β_{1j} through β_{kj} are then equal to the item difficulties γ_{10} through γ_{k0}. When level-1 and level-2 are combined, where the log-odds model is then

$$\eta_{ij} = u_{0j} + \sum_{q=1}^{k} \gamma_{q0} X_{qij} \tag{3}$$

Because a response to a particular item will produce an indicator equal to negative one for a single item, and zeros for all other items, the linear portion of the model can be simplified to

$$\eta_{ij} = u_{0j} - \gamma_{q0} \tag{4}$$

The Rasch model assumes that items show local independence, meaning that after taking into account the parameters of the items and the persons, the responses to items are independent of one another. Parameter invariance is critical for assessing generalizability across populations and test occasions (Rupp & Zumbo, 2006). Occasionally, parameter invariance does not hold, which in the context of this study could occur for developmental reasons and/or language proficiency reasons.

Regardless of cause, IPD threatens measurement applications requiring a stable scale (Wells, Subkovak, & Serlin, 2002). IPD models include a time-related parameter; in this study, the time-related variable is grade level. If drift is present for item parameters, then the items cannot accurately model response probabilities, and person ability estimates will be misestimated for students in different grade levels. Additionally, comparing latent scores of persons across groups would be inappropriate due to the misestimated latent traits.

Goldstein (1983) introduced IPD as a way to measure changes over time in academic achievement exams. Since then, IPD has been examined in a number of contexts. Bock, Muraki, and Pfeiffenberger (1988) investigated linear drift of item location parameters using a time-dependent item response model, and concluded that drift can occur if curricular emphasis changes, that drift can be steady for large populations and can be described as a function of time, and that time-dependent item response models can describe data to maintain scales over an extended period of time. Chan, Drasgow, and Sawin (1999) examined the effect of IPD on test characteristics and found little effect of IPD on performance of the Armed Services Vocational Aptitude Battery. Wells, Subkoviak, and Serlin (2002) examined the effect of item parameter drift on latent trait values using simulated data by varying sample size, test length, type of drift, and percent of drifting items, and found that IPD had a small effect on latent trait estimates for the conditions used in the simulation. DeMars (2004) described IPD in information literacy and global issues, where items on the information literacy scale showed a greater magnitude of IPD, which may have been caused by rapid changes in the field. Wu, Li, Ng, & Zumbo (2012) found that uniform and non-uniform IPD did not exist for trend items on the Trends in International Mathematics and Science Study for examinees from the United States and Singapore.

For this study, item parameters drift occurs not over time but over grade levels. The level-2 of the Rasch model is extended to include grade:

$$\begin{cases} \beta_{0j} = u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Grade}) \\ \vdots \\ \beta_{kj} = \gamma_{k0} + \gamma_{k1}(\text{Grade}) \end{cases} \tag{5}$$

and the full model is equal to

$$\begin{aligned} \eta_{ij} &= u_{0j} - \sum_{q=1}^{k} (\gamma_{q0} X_{qij} + \gamma_{q1} X_{qij}(\text{Grade})) \\ &= u_{0j} - \gamma_{q0} - \gamma_{q1}(\text{Grade}) \end{aligned} \tag{6}$$

where γ_{q1} is a linear deviation from the item parameter γ_{q0} over grade.

The IPD model assumes that the items are invariant across ELL groups, and that the direction of the drift across grades is the same for all groups. Differences in drift by group can be integrated by extending the IPD model. ELL status was estimated as a group factor, however the model is written to facilitate the interpretation of the results. A multiple group IPD model is written by extending Equation 5, the level-2 for the IPD model, to include group predictors:

$$\begin{cases} \beta_{0j} = u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Exited ELL})_j + \gamma_{12}(\text{ELL})_j \\ \qquad +\gamma_{13}(\text{Grade})_j + \gamma_{14}(\text{Exited ELL})_j(\text{Grade})_j + \gamma_{15}(\text{ELL})_j(\text{Grade})_j \\ \qquad \vdots \\ \beta_{kj} = \gamma_{k0} + \gamma_{k1}(\text{Exited ELL})_j + \gamma_{k2}(\text{ELL})_j \\ \qquad +\gamma_{k3}(\text{Grade})_j + \gamma_{k4}(\text{Exited ELL})_j(\text{Grade})_j + \gamma_{k5}(\text{ELL})_j(\text{Grade})_j \end{cases} \tag{7}$$

In this model, u_{0j} is ability for person j, γ_{q0} is difficulty for item q for someone belonging to the not ELL group, γ_{q1} is the item difficulty in grade 6 for exited ELL students, γ_{q2} is the item difficulty in grade 6 for ELL students, γ_{q3} is the linear deviation in difficulty for non-ELLs on item q across grades, γ_{q4} is the linear deviation in difficulty for students in the exited ELL group on item q across grades, and γ_{q5} the linear deviation in item difficulty for ELL students on item q.

Methods

Data Source

A survey measuring developmental assets was administered to a large urban school district in south-central U.S. This survey employed the Search Institute (2005) Developmental Assets Profile (DAP) that measures students' experiences at the self, family, peer, school and community level. Two scales of interest from this survey include Positive Identity and Support. Positive Identity scale consists of six rating-scale survey items, including issues related to feeling good about one's self and future, dealing with disappointment, and having a sense of purpose. The support scale consists of seven rating-scale survey items, including issues related to having parents that are encouraging and available to talk with, having others in the community that are supportive, and being in a school that is supportive and encouraging. Although the original items were on a four-point scale, they were dichotomized for reducing the complexity of parameter drift analysis. Options “Not at all or rarely” and “Somewhat or sometimes” were coded as zero, and “Very or often” and “Extremely or almost always” were coded as one. The DAP survey was administered to 42,245 middle and high school students from 6th to 12th grade. A subset of 24,322 Latino students was considered for the study. Table 1 describes the sample of Latino students broken down by grade and ELL status. The Latino subset consisted of 35.9% ELL students, 53.0% exited ELL students, and 11.1% non-ELL students.

Table 1

Counts (and Percentages) of Latino Students by Grade Level and ELL Status

	ELL Status			
	ELL	Exited-ELL	Non-ELL	
Grade	*n* (%)	*n* (%)	*n* (%)	Total
6	1,621 (49.7%)	1,089 (33.3%)	554 (17.0%)	3,264
7	2,041 (44.9%)	1,775 (39.0%)	733 (16.1%)	4,549
8	1,595 (37.3%)	2,173 (50.8%)	508 (11.9%)	4,276
9	1,309 (33.4%)	2,279 (58.1%)	337 (8.5%)	3,925
10	961 (27.1%)	2,329 (65.7%)	253 (7.1%)	3,543
11	636 (24.2%)	1,809 (68.8%)	185 (7.0%)	2,630
12	562 (26.3%)	1,436 (67.3%)	137 (6.4%)	2,135
All	8,725 (35.9%)	12,890 (53.0%)	2,707 (11.1%)	24,322

Research Questions

This study addressed the following research questions: (a) Are Positive Identity and Support consistent across grade levels? And (b) do items function similarly across grades and English language status? If the scales are invariant across grade level, then we should find that IPD does not occur. If drift is present, then considerations about parameter variability should not just be considered across grade, but also for the interaction between grade and ELL status. If this interaction exists, then the latent trait estimates for students are incorrect because of both grade and ELL status parameter differences. These questions are evaluated by assessing the item parameter invariance across grades and language groups. Different from previous studies in the field, this study investigates item parameter invariance (i.e., DIF) within the context of item parameter drift (IPD).

Model Fit

Four models were fit to the data for the Positive Identity and Support scales: the Rasch model, which assumes measurement invariance; a model detecting IPD, but treating ELL groups as invariant; a model to detect multiple groups IPD, allowing for separate drift parameters for the three ELL groups; and a model that treats grade as a factor that interacts with ELL status. This results in item parameter estimates for each grade-by-group combination.

Estimation

The models were fit to the Positive Identity and Support scales using the `glmer` function from the package `lme4` (Bates, Maechler, & Bockler, 2012) in `R` (R Core Team, 2012), which uses restricted maximum likelihood.

Analyses

First, both Akaike's information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978) are provided to analyze the relative fit of the three models on the two scales. AIC and BIC were provided for several reasons; first, the analyses completed were generally exploratory. Second, regardless of the criteria of use for both AIC and BIC, readers have a preferred relative fit index. Third, while AIC and BIC answer two different questions, when the criteria agree on the best model, this provides reassurance on the robustness on the model choice (Kuha, 2004). Likelihoods from AIC and BIC can be normalized so that they sum to 1, which allows the fit of the models to be treated probabilistically. These selection weights,

w_i, are useful as the weight of evidence in favor of a model as being the best model in the set, and is written as

$$w_i = \frac{\exp(-\Delta_i / 2)}{\sum_{r=1}^{R} \exp(-\Delta_r / 2)}; \quad (9)$$

where for AIC:

$$\Delta_i = \text{AIC}_i\text{-AIC}_{\min}, \quad (10)$$

and for BIC:

$$\Delta_i = \text{BIC}_i\text{-BIC}_{\min}. \quad (11)$$

The weights are then interpreted as the probability that model *i* is the best model for the data (Burnham & Anderson, 2004).

In addition to comparing the model fit, the parameter estimates will be compared across the models for both scales. These parameters are described to understand the frequency in which multiple-group IPD occurs. Given that Latinos are often treated as a single group during validity studies, we investigate the extent to which the direction of parameter drift varies by ELL status.

Finally, a visual analysis will be completed comparing the multiple group IPD parameter estimates to the group-by-item factor model estimates. The appropriateness of 39 linear multiple group IPD drift parameters were visually analyzed (3 drift parameters across the 13 items on the 2 scales), parameters were either judged as following linear drift or not following linear drift. If multiple group linear IPD estimates are appropriate, then the grade-by-group factor estimates will deviate across years in a consistent and linear manner. If the grade-by-group factor estimates for an item do not follow a linear patter for a particular item and group, then it is inappropriate to use linear IPD models. This may occur if a parameter is drifting in one direction during the middle school years, but the drift changes directions during high school years. Such a model would suggest that linear IPD is not appropriate.

Results

Table 2 displays the number of parameters estimated, deviance, AIC, and BIC for the three models on the Positive Identity scale. For BIC values, the multiple group IPD model fit the data best, where for the AIC, the grade-by-group factor model fits best. By examining w_{AIC} and w_{BIC}, it is very unlikely that the Rasch model or the single group IPD models would fit this data best. The model fit results for the Support scale, shown in Table 3, show similar results to the model fit for the Positive Identity scale. For BIC, the multiple group IPD model fit the Support scale best, and for AIC the grade-by-group factor model fits best. Based on both AIC and BIC the Rasch model has the worst fit, which suggests that the parameters are not invariant.

Table 2

Model Fit Results for Positive Identity Scale: Deviance, AIC, and BIC

Model	Parameters Estimated	Deviance	AIC	w_{AIC}	BIC	w_{BIC}
Concurrent	7	139,406	139,420	4.29×10^{-99}	139,477	3.49×10^{-19}
IPD	13	139,293	139,319	3.67×10^{-77}	139,424	1.13×10^{-7}
Multiple Group IPD	37	139,018	139,092	7.19×10^{-28}	139,392	.999
Group x Grade Factor	127	138,713	138,967	.999	140,227	4.81×10^{-182}

Table 3

Model Fit Results for Support Scale: Deviance, AIC, and BIC

Model	Parameters Estimated	Deviance	AIC	w_{AIC}	BIC	w_{BIC}
Concurrent	8	168,441	168,457	2.14×10^{-295}	168,522	8.35×10^{-202}
IPD	15	167,585	167,615	1.48×10^{-112}	167,736	3.98×10^{-31}
Multiple Group IPD	43	167,162	167,248	7.28×10^{-33}	167,596	.999
Group x Grade Factor	148	166,804	167,100	.999	168,591	8.68×10^{-217}

Tables 4, 5, and 6 display the estimates, standard errors, and description of the coefficients in the Rasch, IPD, multiple group IPD models for the Positive Identity scale. A Wald test assessed the significance of each parameter being different than zero. Ten of the 18 drift parameters from the multiple groups IPD model, γ_{q3} through γ_{q5}, on the positive identity scale, had *p*-values less than .05; 5 of 6 drift parameters for ELLs, 2 of 6 drift parameters for Exited-ELLs, and 3 of 6 for non-ELLs. Additionally, the IPD drift parameters, γ_{q1}, are equal to the weighted average of the drift parameters of the drift parameters from the multiple groups IPD model.

Table 4

Estimates, Standard Errors, and Descriptions for the Positive Identity Scale Parameters Using the Rasch Model.

Parameter	Estimate	SE	Description
γ_{10}	-2.059*	0.022	Difficulty for item 1 across all groups
γ_{20}	-2.359*	0.023	Difficulty for item 2 across all groups
γ_{30}	-2.447*	0.023	Difficulty for item 3 across all groups
γ_{40}	-0.358*	0.018	Difficulty for item 4 across all groups
γ_{50}	-1.407*	0.020	Difficulty for item 5 across all groups
γ_{60}	-2.538*	0.024	Difficulty for item 6 across all groups

Table 5

Estimates, Standard Errors, and Descriptions for the Positive Identity Scale Parameters Using the IPD Model.

Parameter	Estimate	SE	Description
γ_{10}	-1.982*	0.038	Difficulty for item 1 across all groups; baseline = 6th grade
γ_{11}	-0.029*	0.012	Linear deviation for all groups from the baseline estimate across grade for item 1
γ_{20}	-2.251*	0.040	Difficulty for item 2 across all groups; baseline = 6th grade
γ_{21}	-0.042*	0.013	Linear deviation for all groups from the baseline estimate across grade for item 2
γ_{30}	-2.645*	0.043	Difficulty for item 3 across all groups; baseline = 6th grade
γ_{31}	0.072*	0.013	Linear deviation for all groups from the baseline estimate across grade for item 3
γ_{40}	-0.310*	0.032	Difficulty for item 4 across all groups; baseline = 6th grade
γ_{41}	-0.018	0.010	Linear deviation for all groups from the baseline estimate across grade for item 4
γ_{50}	-1.278*	0.035	Difficulty for item 5 across all groups; baseline = 6th grade
γ_{51}	-0.049	0.011	Linear deviation for all groups from the baseline estimate across grade for item 5
γ_{60}	-2.350*	0.041	Difficulty for item 6 across all groups; baseline = 6th grade
γ_{61}	-0.072*	0.013	Linear deviation for all groups from the baseline estimate across grade for item 6

Table 6

Estimates, Standard Errors, and Descriptions for the Positive Identity Scale Parameters Using the Multiple Groups IPD Model.

Parameter	Estimate	SE	Description
γ_{10}	-2.330*	0.063	Difficulty for item 1 for non-ELL; baseline = 6^{th} grade
γ_{11}	-2.371*	0.104	Difficulty for item 1 for exited ELL; baseline = 6^{th} grade
γ_{12}	-1.634*	0.054	Difficulty for item 1 for ELL; baseline = 6^{th} grade
γ_{13}	0.017	0.018	Linear deviation for non-ELLs from the baseline estimate across grade for item 1
γ_{14}	0.063	0.038	Linear deviation for exited ELLs from the baseline estimate across grade for item 1
γ_{15}	-0.043*	0.019	Linear deviation for ELLs from the baseline estimate across grade for item 1
γ_{20}	-2.425*	0.065	Difficulty for item 2 for non-ELL; baseline = 6^{th} grade
γ_{21}	-2.458*	0.106	Difficulty for item 2 for exited ELL; baseline = 6^{th} grade
γ_{22}	-2.055*	0.059	Difficulty for item 2 for ELL; baseline = 6^{th} grade
γ_{23}	-0.012	0.018	Linear deviation for non-ELLs from the baseline estimate across grade for item 2
γ_{24}	0.095*	0.038	Linear deviation for exited ELLs from the baseline estimate across grade for item 2
γ_{25}	-0.091*	0.021	Linear deviation for ELLs from the baseline estimate across grade for item 2
γ_{30}	-2.903*	0.070	Difficulty for item 3 for non-ELL; baseline = 6^{th} grade
γ_{31}	-3.129*	0.120	Difficulty for item 3 for exited ELL; baseline = 6^{th} grade
γ_{32}	-2.330*	0.061	Difficulty for item 3 for ELL; baseline = 6^{th} grade
γ_{33}	0.107*	0.019	Linear deviation for non-ELLs from the baseline estimate across grade for item 3
γ_{34}	0.201*	0.041	Linear deviation for exited ELLs from the baseline estimate across grade for item 3
γ_{35}	0.043*	0.021	Linear deviation for ELLs from the baseline estimate across grade for item 3
γ_{40}	-0.306*	0.050	Difficulty for item 4 for non-ELL; baseline = 6^{th} grade
γ_{41}	-0.384*	0.083	Difficulty for item 4 for exited ELL; baseline = 6^{th} grade
γ_{42}	-0.298*	0.048	Difficulty for item 4 for ELL; baseline = 6^{th} grade
γ_{43}	-0.023	0.014	Linear deviation for non-ELLs from the baseline estimate across grade for item 4
γ_{44}	0.047	0.031	Linear deviation for exited ELLs from the baseline estimate across grade for item 4
γ_{45}	-0.025	0.017	Linear deviation for ELLs from the baseline estimate across grade for item 4
γ_{50}	-1.357*	0.055	Difficulty for item 5 for non-ELL; baseline = 6^{th} grade
γ_{51}	-1.567*	0.092	Difficulty for item 5 for exited ELL; baseline = 6^{th} grade
γ_{52}	-1.145*	0.051	Difficulty for item 5 for ELL; baseline = 6^{th} grade
γ_{53}	-0.051*	0.016	Linear deviation for non-ELLs from the baseline estimate across grade for item 5
γ_{54}	0.033	0.034	Linear deviation for exited ELLs from the baseline estimate across grade for item 5
γ_{55}	-0.039*	0.018	Linear deviation for ELLs from the baseline estimate across grade for item 5
γ_{60}	-2.540*	0.068	Difficulty for item 6 for non-ELL; baseline = 6^{th} grade
γ_{61}	-2.514*	0.111	Difficulty for item 6 for exited ELL; baseline = 6^{th} grade
γ_{62}	-2.182*	0.060	Difficulty for item 6 for ELL; baseline = 6^{th} grade
γ_{63}	-0.060*	0.019	Linear deviation for non-ELLs from the baseline estimate across grade for item 6
γ_{64}	-0.058	0.042	Linear deviation for exited ELLs from the baseline estimate across grade for item 6
γ_{65}	-0.042*	0.021	Linear deviation for ELLs from the baseline estimate across grade for item 6

* = p-value < .05

Tables 7, 8, and 9 displays the estimates and standard errors for the coefficients in the Rasch, IPD, and multiple group IPD models for the Support scale. Eighteen of the 21 drift parameters for the multiple groups IPD on the support scale had *p*-values less than .05, suggesting item locations are drifting based on grade level and language status. Additionally, the drift estimates for each of the groups on the Support scale showed an interesting phenomenon, generally, as grade increased, the likelihood of endorsing an item became more difficult, regardless of group status. This phenomenon is displayed in Figure 1. This suggests that if item difficulty parameters are treated as invariant, the latent trait estimates of students in higher grades are likely to be underestimated for the Support scale. This phenomena that occurs across all seven items on the Support scale suggests that when item difficulty parameters are treated as measurement invariant, then 12th grade students Support trait is being underestimated and 6th grade students Support trait are being overestimated.

Table 7

E*stimates, Standard Errors, and Descriptions for the Support Scale Parameters Using the Rasch Model.*

Parameter	Estimate	SE	Description
γ_{10}	-0.672*	0.018	Difficulty for item 1 across all groups
γ_{20}	-3.119*	0.027	Difficulty for item 2 across all groups
γ_{30}	-0.183*	0.017	Difficulty for item 3 across all groups
γ_{40}	-1.252*	0.019	Difficulty for item 4 across all groups
γ_{50}	-1.443*	0.019	Difficulty for item 5 across all groups
γ_{60}	-2.910*	0.026	Difficulty for item 6 across all groups
γ_{70}	-1.779*	0.020	Difficulty for item 7 across all groups

Table 8

Estimates, Standard Errors, and Descriptions for the Support Scale Parameters Using the IPD Model.

Parameter	Estimate	SE	Description
γ_{10}	-0.847*	0.032	Difficulty for item 1 across all groups; baseline = 6th grade
γ_{11}	0.066*	0.010	Linear deviation for all groups from the baseline estimate across grade for item 1
γ_{20}	-3.683*	0.054	Difficulty for item 2 across all groups; baseline = 6th grade
γ_{21}	0.201*	0.015	Linear deviation for all groups from the baseline estimate across grade for item 2
γ_{30}	-0.825*	0.031	Difficulty for item 3 across all groups; baseline = 6th grade
γ_{31}	0.238*	0.010	Linear deviation for all groups from the baseline estimate across grade for item 3
γ_{40}	-1.759*	0.035	Difficulty for item 4 across all groups; baseline = 6th grade
γ_{41}	0.186*	0.010	Linear deviation for all groups from the baseline estimate across grade for item 4
γ_{50}	-1.571*	0.035	Difficulty for item 5 across all groups; baseline = 6th grade
γ_{51}	0.050*	0.011	Linear deviation for all groups from the baseline estimate across grade for item 5
γ_{60}	-3.358*	0.050	Difficulty for item 6 across all groups; baseline = 6th grade
γ_{61}	0.163*	0.014	Linear deviation for all groups from the baseline estimate across grade for item 6
γ_{70}	-2.212	0.038	Difficulty for item 7 across all groups; baseline = 6th grade
γ_{71}	0.159	0.011	Linear deviation for all groups from the baseline estimate across grade for item 7

Table 9

Estimates, Standard Errors, and Descriptions for the Support Scale Parameters Using the Multiple Groups IPD Model.

Parameter	Estimate	SE	Description
γ_{10}	-0.693*	0.049	Difficulty for item 1 for non-ELL; baseline = 6^{th} grade
γ_{11}	-0.721*	0.082	Difficulty for item 1 for exited ELL; baseline = 6^{th} grade
γ_{12}	-1.001*	0.049	Difficulty for item 1 for ELL; baseline = 6^{th} grade
γ_{13}	0.060*	0.014	Linear deviation for non-ELLs from the baseline estimate across grade for item 1
γ_{14}	0.163*	0.030	Linear deviation for exited ELLs from the baseline estimate across grade for item 1
γ_{15}	-0.001	0.017	Linear deviation for ELLs from the baseline estimate across grade for item 1
γ_{20}	-3.811*	0.085	Difficulty for item 2 for non-ELL; baseline = 6^{th} grade
γ_{21}	-3.828*	0.148	Difficulty for item 2 for exited ELL; baseline = 6^{th} grade
γ_{22}	-3.544*	0.080	Difficulty for item 2 for ELL; baseline = 6^{th} grade
γ_{23}	0.222*	0.022	Linear deviation for non-ELLs from the baseline estimate across grade for item 2
γ_{24}	0.213*	0.049	Linear deviation for exited ELLs from the baseline estimate across grade for item 2
γ_{25}	0.185*	0.025	Linear deviation for ELLs from the baseline estimate across grade for item 2
γ_{30}	-0.768*	0.049	Difficulty for item 3 for non-ELL; baseline = 6^{th} grade
γ_{31}	-0.693*	0.081	Difficulty for item 3 for exited ELL; baseline = 6^{th} grade
γ_{32}	-0.934*	0.048	Difficulty for item 3 for ELL; baseline = 6^{th} grade
γ_{33}	0.222*	0.014	Linear deviation for non-ELLs from the baseline estimate across grade for item 3
γ_{34}	0.230*	0.030	Linear deviation for exited ELLs from the baseline estimate across grade for item 3
γ_{35}	0.266*	0.016	Linear deviation for ELLs from the baseline estimate across grade for item 3
γ_{40}	-1.692*	0.054	Difficulty for item 4 for non-ELL; baseline = 6^{th} grade
γ_{41}	-1.694*	0.090	Difficulty for item 4 for exited ELL; baseline = 6^{th} grade
γ_{42}	-1.839*	0.054	Difficulty for item 4 for ELL; baseline = 6^{th} grade
γ_{43}	0.180*	0.015	Linear deviation for non-ELLs from the baseline estimate across grade for item 4
γ_{44}	0.197*	0.032	Linear deviation for exited ELLs from the baseline estimate across grade for item 4
γ_{45}	0.174*	0.018	Linear deviation for ELLs from the baseline estimate across grade for item 4
γ_{50}	-1.601*	0.055	Difficulty for item 5 for non-ELL; baseline = 6^{th} grade
γ_{51}	-1.614*	0.091	Difficulty for item 5 for exited ELL; baseline = 6^{th} grade
γ_{52}	-1.574*	0.052	Difficulty for item 5 for ELL; baseline = 6^{th} grade
γ_{53}	0.028	0.015	Linear deviation for non-ELLs from the baseline estimate across grade for item 5
γ_{54}	0.021	0.034	Linear deviation for exited ELLs from the baseline estimate across grade for item 5
γ_{55}	0.118*	0.018	Linear deviation for ELLs from the baseline estimate across grade for item 5
γ_{60}	- 3.515*	0.080	Difficulty for item 6 for non-ELL; baseline = 6^{th} grade
γ_{61}	-3.607*	0.137	Difficulty for item 6 for exited ELL; baseline = 6^{th} grade
γ_{62}	-3.181*	0.073	Difficulty for item 6 for ELL; baseline = 6^{th} grade
γ_{63}	0.184*	0.021	Linear deviation for non-ELLs from the baseline estimate across grade for item 6
γ_{64}	0.245*	0.045	Linear deviation for exited ELLs from the baseline estimate across grade for item 6
γ_{65}	0.141*	0.024	Linear deviation for ELLs from the baseline estimate across grade for item 6
γ_{70}	-2.141*	0.059	Difficulty for item 7 for non-ELL; baseline = 6^{th} grade
γ_{71}	-2.044*	0.096	Difficulty for item 7 for exited ELL; baseline = 6^{th} grade
γ_{72}	-2.347*	0.059	Difficulty for item 7 for ELL; baseline = 6^{th} grade
γ_{73}	0.148*	0.016	Linear deviation for non-ELLs from the baseline estimate across grade for item 7
γ_{74}	0.207*	0.034	Linear deviation for exited ELLs from the baseline estimate across grade for item 7
γ_{75}	0.149*	0.020	Linear deviation for ELLs from the baseline estimate across grade for item 7

* = p-value < .05

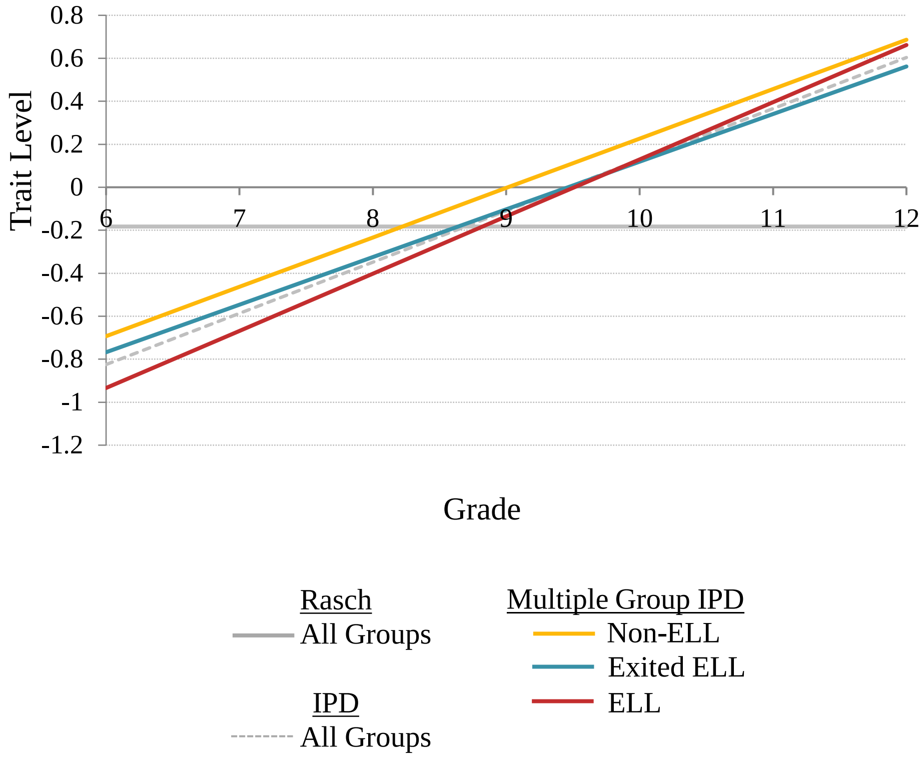

Figure 1. An example of the multiple group drift parameters showing a decrease in item difficulty on the Support scale.

To investigate the extent IPD is linear, a visual analysis of the drift parameters was compared to the grade-by-group factor model (See Appendix A). Based on the review of the graphs, a total of 12 of the 18 drift parameter for the Positive Identity scale using the multiple groups IPD model showed visual evidence for *linear* drift. Visual misfit (nonlinearity) occurred on two of the six items on the Positive Identity scale. The item difficulty locations for one of the two items showing non-linear drift on the Positive Identity scale is shown in Figure 2. Both items were related to dealing with difficult situations. Figure 2 contains the item difficulty estimates for the Rasch model, the IPD model, the multiple groups IPD model and the grade-by-group factor model, which are graphed against grade level. For item in Figure 2 it is clear that the item parameters for grade-by-group factor models for all three groups do not follow the linear drift estimates from the multiple group IPD model. The points in Figure 2 represent the grade-by-group factor model item estimates. Since the Rasch model will have a single parameter estimated for all groups, which is invariant, the line on the graph in Figure two is the same value across all grades. The IPD model produces two parameters for each item, an item difficulty, which is centered at 6th grade, and a linear drift parameter, on the graph in Figure 2, the IPD model is represented by a single line for all groups. Since the multiple group IPD model estimates difficulty and drift parameters for each of the three groups, the model is represented by three lines, one for each group. The grade-by-group factor model has empirical estimates for each ELL group at each grade, these estimates represented by points on the graph.

For the item in Figure 2, the factor model suggests item parameter estimates for all groups item become easier to endorse through the middle school years (grades 6 through 8), and then more difficult to endorse during the high school years (grades 9 through 12). For the item in Figure 2, conditioned on trait level, 6th grade and 12th grade students are more likely to endorse this item than 9th graders. This means if the Rasch parameter estimate for the item in Figure 2 is used to estimate a student's trait level, the trait level for 9th grade students will be underestimated and overestimated for the 6th and 12th grade students.

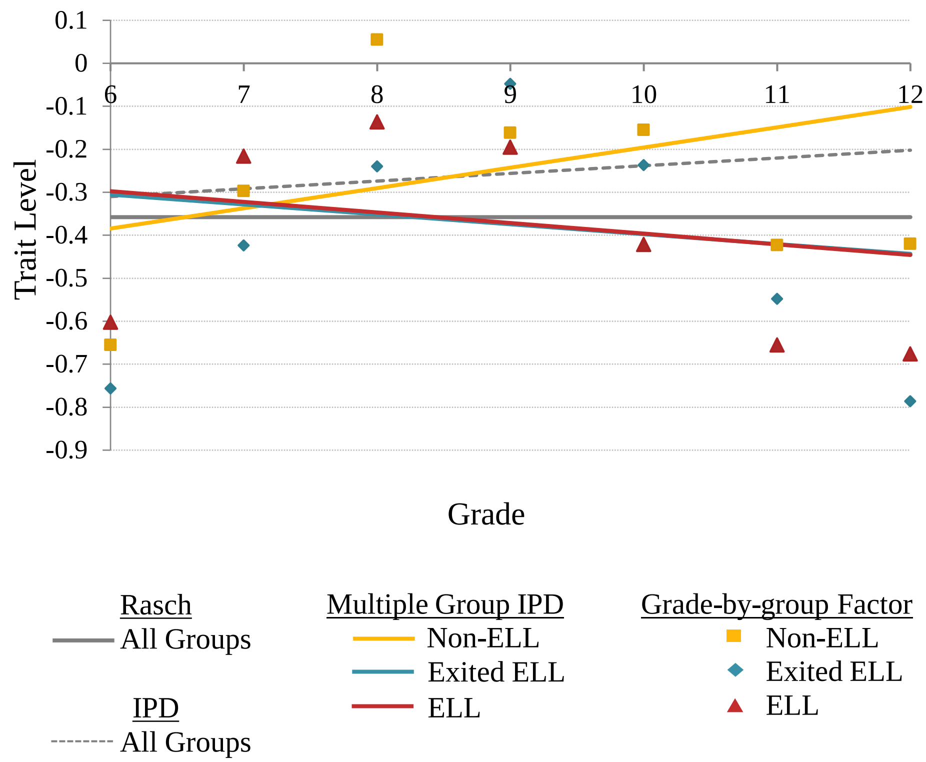

Figure 2. An example of a poor fitting multiple group linear IPD model.

The visual fit for the multiple groups IPD linear model is better for the Support scale, where only three parameters do not follow linear drift, all on a single item. The visual analysis showed that the drift parameters for all three groups either followed linear drift for an item or did not follow linear drift. If the empirical estimates were non-linear, then they were non-linear for all three groups across all items. Despite the non-linear shift, the empirical item difficulties across groups seem to follow a pattern, and because of the role of the item in the construct, this might suggest that the construct is shifting across grades.

Figure 3 shows the item difficulty estimates across year for the Rasch, IPD, multiple group IPD, and grade-by-group factor models from the Positive Identity scale. For the item in Figure 3, the multiple group IPD item difficulty estimates for ELL, exited-ELL, and non-ELL students appear to visually follow the grade-by-group item difficulty estimates for the three groups This is a relatively easy item to endorse, as it requires low trait levels for students to endorse the item. The figure also shows that as the parameter drifts across grade levels; the item difficulty of the non-ELL and exited-ELL groups is converging with the item parameter for ELL students. ELL students require more trait level to endorse this item in the earlier grades than exited-ELL and non-ELL students and by 12th grade, all students require similar trait level to endorse the item. That is, conditioned on overall Positive Identity, students in 12th grade have approximately the same probability of endorsing this item. In fact, many of the items on the Positive Identity scale show converging item parameters.

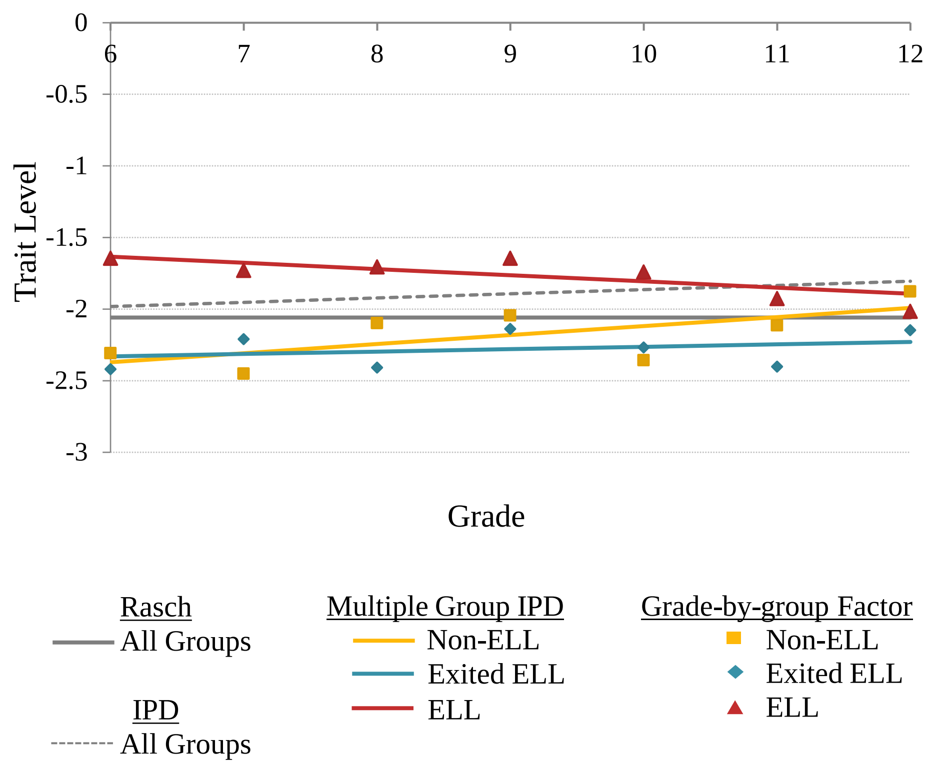

Figure 3. An example of an item showing good linear fit for the multiple group linear IPD model.

Implications

Initial results suggest that items are functionally different across language groups and grade levels, indicating that item functioning not only depends on occasion, but also language status –with implications for the consistent interpretation across grades and language status. For instance, it was found generally that items in the Support scale are more likely to be endorsed by students in later grades than in earlier grades, conditioned on overall trait level, and the drift was similar for all groups (see Figure 1). This suggests that all groups require similar and higher levels of perceived support at later grades to endorse an item in the Support scale compared lower grades. Because of parameter drift, estimates of perceived Support levels of students across grades are biased.

The Positive Identity scale also appears to show parameter drift overall; however, this shift varied across groups and items and for one item the drift was non-linear. The overall drift of this construct was of a convergence nature typically requiring ELL to have more of the positive identity to endorse an item in earlier grades than the other two groups. All groups require similar trait levels to endorse an item for later grades (see Figure 3). This item is about feeling of being in control of one's life and the convergence suggest that groups need different overall levels of Positive Identity in order to feel some control over one's life. A particular item that is about dealing with disappointment showed a non-linear drift. The drift appears to be systematic in a way that it is positive for groups from 6th to 8th grade and then negative from 9th to 12th grades where ELL and exited-ELL students exhibit more dramatic shift. This suggests that something triggers the drift in 9th grade; a potential reason could be that this is the time when students transition from middle to high school toward more independence. It is also important to recognize that the dropout rate is high among Latino youth during the high school years making the sample more selective.

IPD across grade was applied to two asset scales, an internal asset, Positive Identity, and an external asset, Support, it could be applied in a variety of settings where a data was collected at a single time point. Most items on both scales appeared to follow linear drift, but there were some items where drift was not linear. Methodologically, the use of IPD to measure drift across grade levels is a new concept for examining parameter drift when data are collected cross-sectionally, but this concept could be applied to a variety of cross-sectional settings.

One assumption of IPD models in a cross-sectional setting is that the parameters across the continuum, in the case of this study, grade, are invariant in subpopulations. It may be difficult to treat subgroups as sampled from the same population across grades as high dropout rates in Latino youth during high school years will change the population.

There are other limitations with this paper. First, if drift is non-linear than the use of linear drift parameters is inappropriate and will lead to incorrect conclusions about the data. Given that three items on the two scales did not visually follow linear IPD, these items should have been treated differently analytically. Second, it would be very difficult to make conclusions about the direction of drift for any of the subgroups of Latino students and generalize the results to all Latino students for that particular subgroup. There may be certain contextual effects within the school district where the data were collected that may affect the direction of the drift for the given items.

There are several areas recommended for future research. First, studies could be completed to examine the patterns of drift across grade levels for ELL, exited-ELL, and non-ELL students using different ethnic populations and/or subscales of asset profiles. Second, the research could be extended to include predictors that could explain the drift parameters. For instance, for items where drift is converging on the Support scale, that might be explained by the amount of time a the student is interacting with other adults. This type of study could be completed in the EIRM framework. Third, simulation studies should be completed to look at the effect of fitting linear IPD models in non-linear situations, as shown in Figure 3. This simulation research could examine how the bias and fit of these parameters is affected in various scenarios. Finally, we hope to extend this model to the polytomous case, taking advantage of the full information in the rating-scale responses.

This study serves as a unique example of examining parameter invariance across time and groups. Given the heterogeneity of the Latino population and the diverse non-cognitive assessments available for this study, the multiple groups IPD model was a unique opportunity to explore parameter variability. IPD models are flexible, powerful, and easy to interpret. As this study showed, the IPD model can be adopted to fit cross-sectional data, and adapted to accommodate drift parameters for multiple groups. When a multiple group IPD model is fit, as was with this study, both drift and DIF can be detected. Additionally, IPD models are estimated with fewer parameters compared to the time-by-group factor interaction models, thus require less computational power.

Whereas the purpose of this paper was to investigate parameter invariance using a multiple group drift model, the paper exposed some real differences in ELL, non-ELL, and exited-ELL students on these measures. This parameter variation across groups allows us to examine the nature of the construct shift not just by treating Latinos as a single group, rather examining the shift by level of language acquisition.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19(6)*, 716-723.

Albano, A.D., & Rodriguez, M.C. (2012, April). *Multilevel modeling of item parameter drift*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, Canada.

Bates, D., Maechler, M., & Bokler, B. (2012). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999999-0. http://CRAN.R-project.org/package=lme4

Benson, P.L., Scales, P.C., Hamilton, S.F., & Sesma, A. (2006). Positive youth development: Theory, research, and applications. In W. Damon & R.M. Lerner (Eds.), *Handbook of Child Psychology: Vol. 1* (6th ed., pp. 894-941). New York: John Wiley & Sons.

Burnham, K.P., & Anderson, D.R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*, 261-304.

Cabrera, J., & Rodriguez, M.C. (2010, May). *Positive Latino youth development: The impact of perceived community support on Latino youths' school plans*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Cabrera, J., & Rodriguez, M.C. (2011, April). *Latino youth's beliefs and attitudes and their influence on after-school activity participation*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Chan, K.Y., Drasgow, F., & Sawin, L.L. (1999). What is the shelf life of a test? The effect of time on the psychometrics of a cognitive ability test battery. *Journal of Applied Psychology, 17*, 610-619.

Contreras, F., Flores-Ragade, A., Lee, Jr., J.M., McGuire, K.M. (2011). *The college completion agenda Research and context brief: Latino edition*. New York, NY: The College Board.

Damon, W. (2004). What is positive youth development? *Annals of the American Academy of Political and Social Science, 591*(1), 13-24.

De Boeck, P., & Wilson, M. (2004). Explanatory item response models: a generalized linear and nonlinear approach. *Statistics for Social Science and Public Policy.* (pp. 213-240). New York, NY. Springer.

Donoghue, J.R., & Isham, S.P. (1998). A comparison of procedures to detect item parameter drift. *Applied Psychological Measurement, 22(1),* 33-51.

Eamon, M. K., & Mulder, C. (2005). Predicting antisocial behavior among Latino young adolescents: An ecological systems analysis. *American Journal of Orthopsychiatry*, 75(1), 117-127.

Fuligni, A.K., & Perreira, K.M. (2009). Immigration and adaptation. In F.A. Villarruel, G. Carlo, J.M. Grau, M. Azmitia, N.J. Cabrera, & T.J. Chahin, (Eds.), *Handbook of U.S. Latino psychology: developmental and community-based perspectives* (pp. 99-113). Los Angeles, CA: SAGE.

Gregory, S.T. (2003). Planning for the increasing number of Latino students. *Planning for Higher Education, 31,* 13-19.

Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement, 20*, 369-377.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, *38*, 79-93.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions of performance. *Sociological Methods & Research, 33*, 188-229.

Lerner, R.M. (2002). *Concepts and theories of human development* (3rd ed.). Mahwah, NJ: Erlbaum.

McLoyd, V. C. (1998). Changing demographics in the American population: Implications for research on minority children and adolescents. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 3- 28). Mahwah, NJ: Lawrence Erlbaum Associates.

Palma, J., Rodriguez, M.C., Cabrera, J.C., Albano, A.D., Vue, Y. Warshawsky, A.J. (2012, April). *Effects of positive assets and socio-economic status on academic performance*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rasch G (1960*). Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen.

Rodriguez, M. C., & Morrobel, D. (2004). A review of Latino youth developmental research and a call for an asset orientation. *Hispanic Journal of Behavioral Sciences,* 26(2), 107-127.

Rupp, A. A., & Zumbo, B.D. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and Psychological Measurement, 66(1)*, 63-84.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6(2)*, 461-464.

Search Institute. (2005). *Developmental Assets Profile technical manual*. Minneapolis, MN: Author.

Sesma, A., Jr., & Roehlkepartain, E. C. (2003). Unique strengths, shared strengths: developmental assets among youth of color. *Search Institute Insights & Evidence*, 1(2). Retrieved from http://www.search-institute.org/research/insightsevidence/november-2003

Smith, R.W. (2004). *The impact of braindump sites on item exposure and item parameter drift.* Paper presented at the Annual Meeting of the American Education Research Association, April 2-6, 2004, San Diego, California.

U.S. Census Bureau. (2010). *Statistical Abstract of the United States: 2010.* Retrieved July 23, 2010, from http://www.census.gov/prod/2009pubs/10statab/pop.pdf

Warshawsky, A.J., Rodriguez, M.C., Cabrera, J.C., Palma, J., Albano, A.D., & Vue, Y. (2012, April). *Attitudes toward school and school plans given levels of family alcohol, substance, and physical abuse*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

Zumbo, B.D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing*, *20*(2), 136-147.

Zumbo, B.D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4*, 223-233.

Appendix A: Graphs of Item Difficulty and Drift Estimates Across Grade

This appendix contains graphs of the item difficulty and drift parameters for each of the 6 items on the Positive Identity scale and 7 items on the Suppor scale. Each graphic contains the parameters for the Rasch, IPD, multiple group IPD, and grade-by-group factor models.

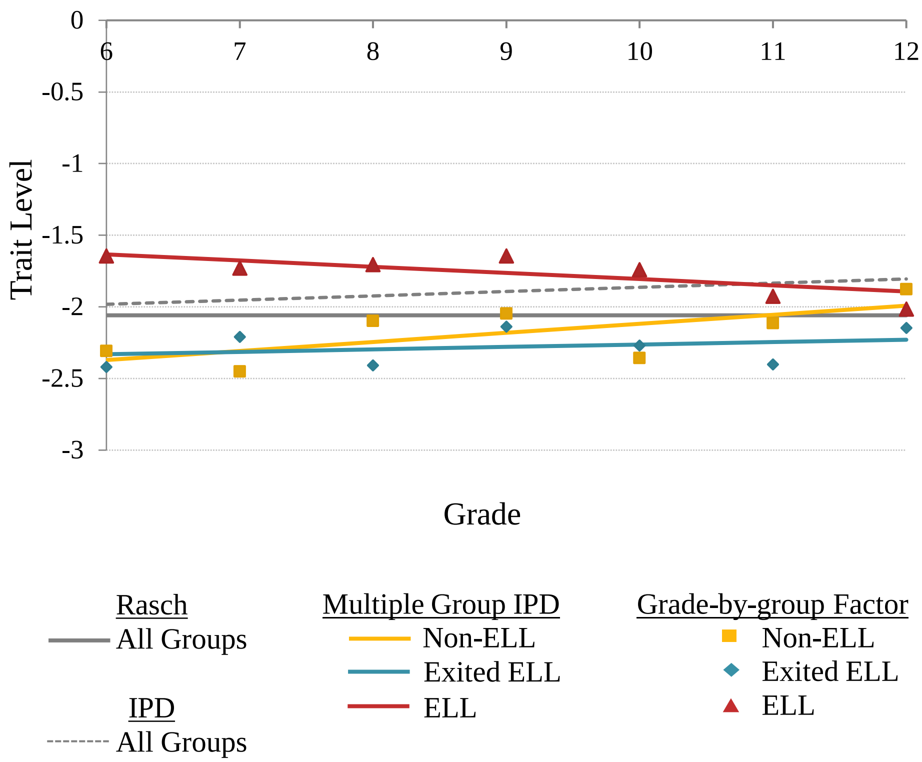

Figure 4. Visualization of the parameter estimates for item 1 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

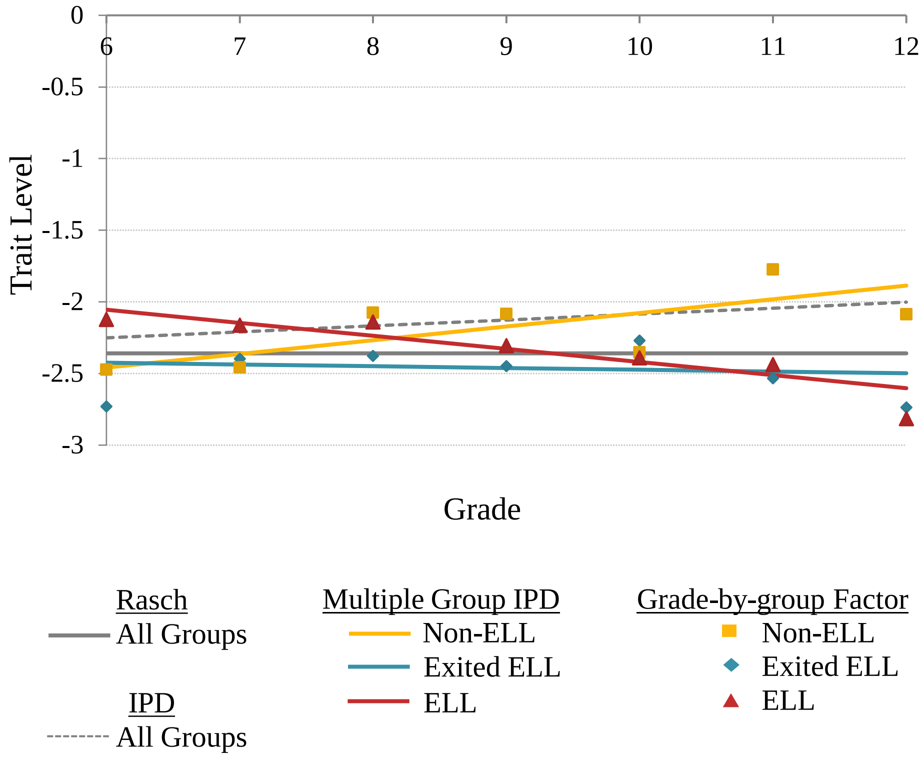

Figure 5. Visualization of the parameter estimates for item 2 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

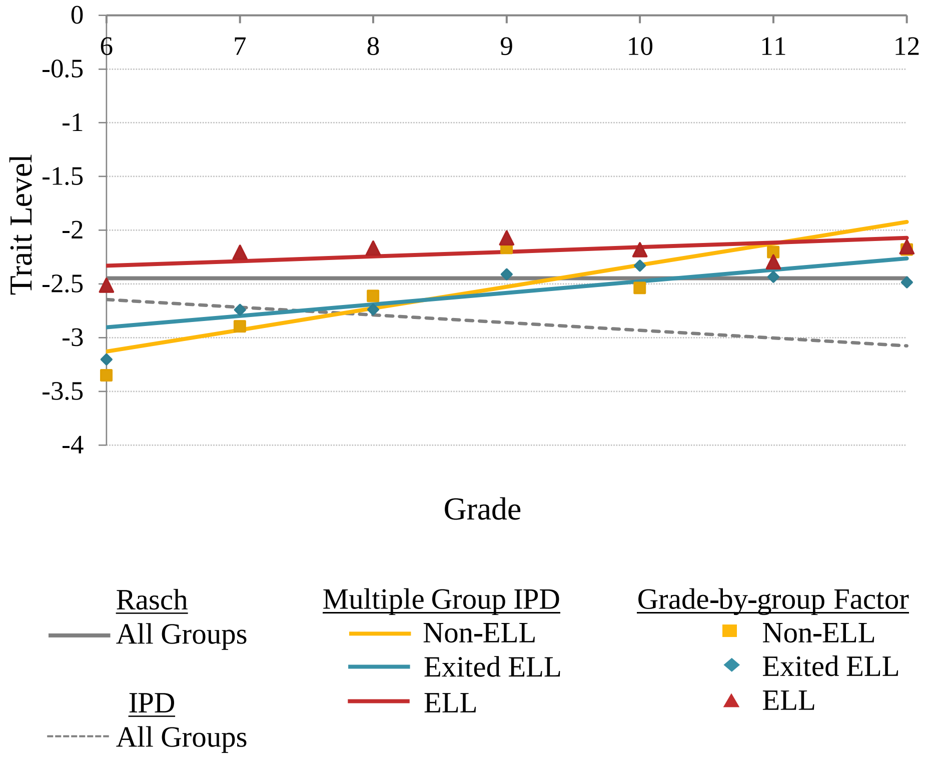

Figure 6. Visualization of the parameter estimates for item 3 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

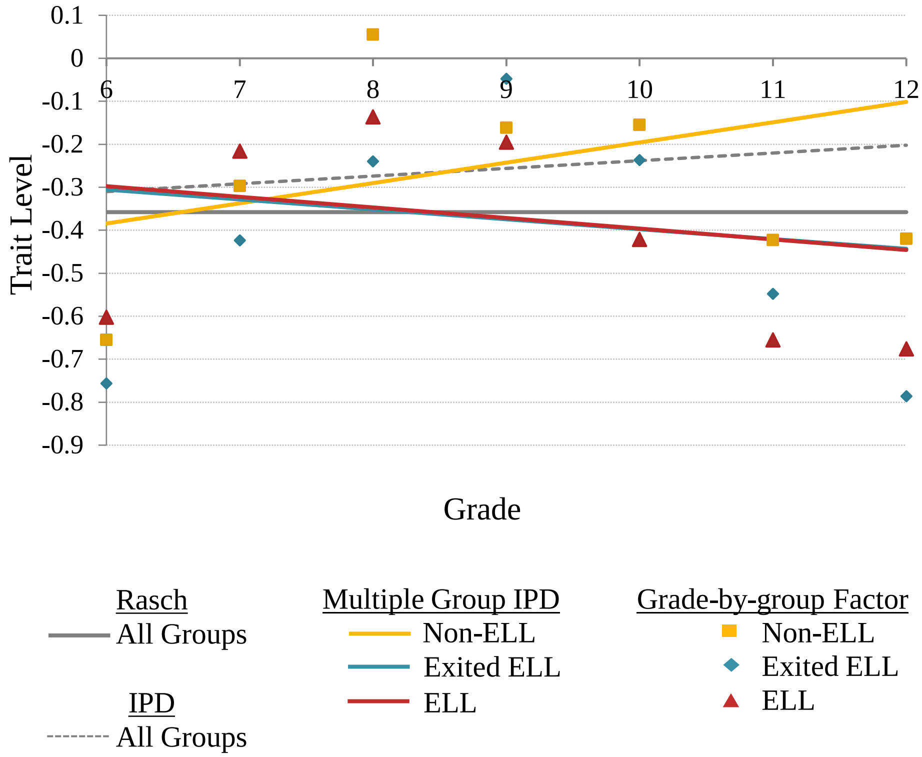

Figure 7. Visualization of the parameter estimates for item 4 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

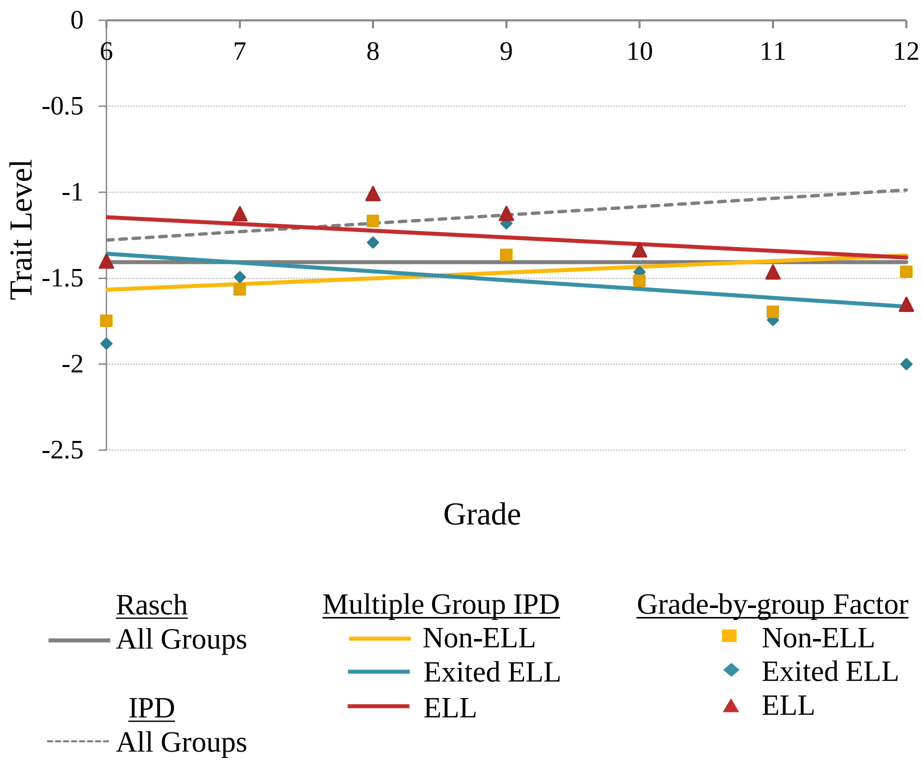

Figure 8. Visualization of the parameter estimates for item 5 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

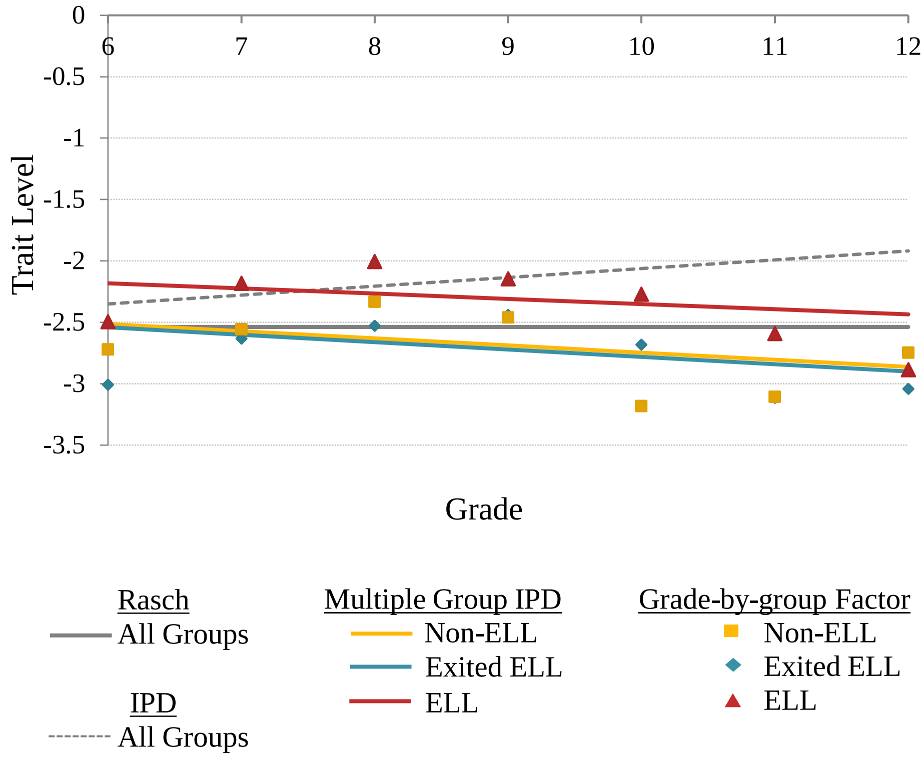

Figure 9. Visualization of the parameter estimates for item 6 on the Positive Identity scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

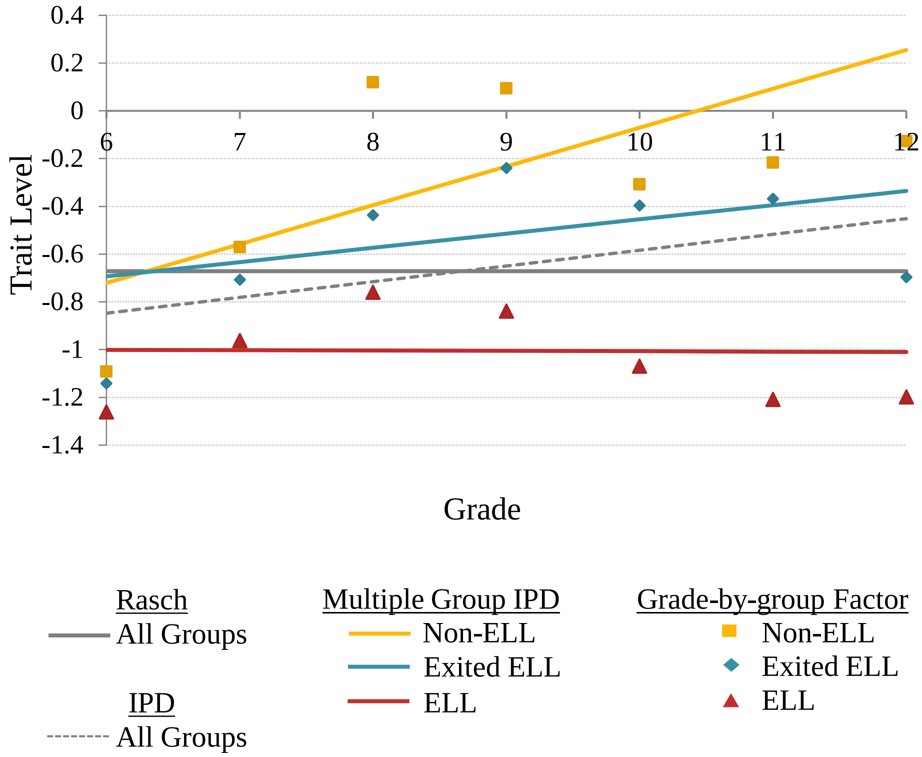

Figure 10. Visualization of the parameter estimates for item 1 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

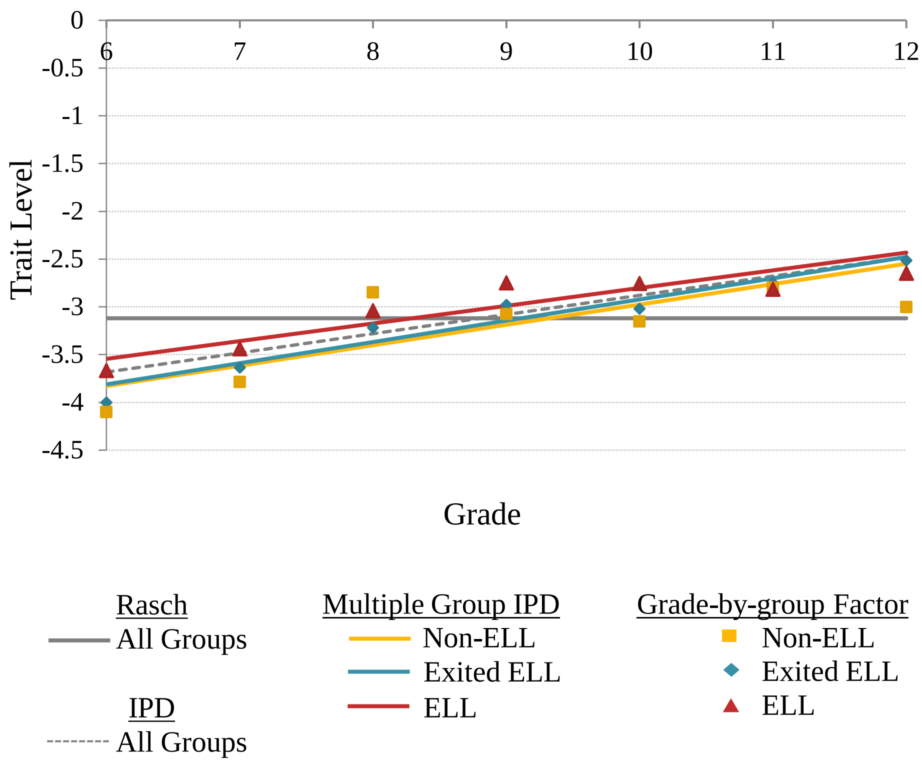

Figure 11. Visualization of the parameter estimates for item 2 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

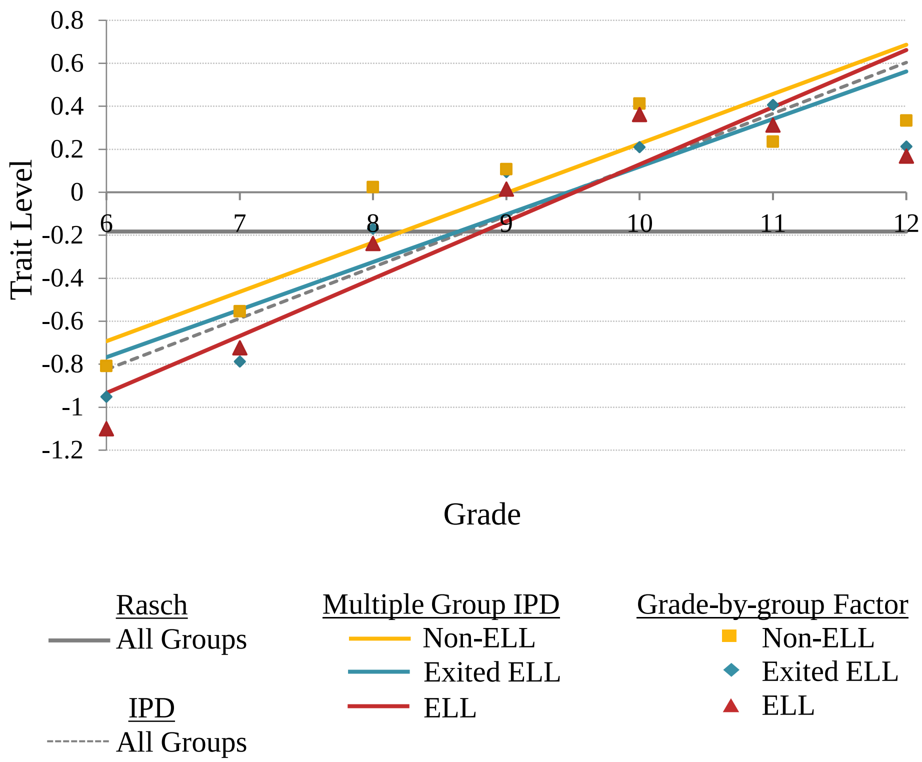

Figure 12. Visualization of the parameter estimates for item 3 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

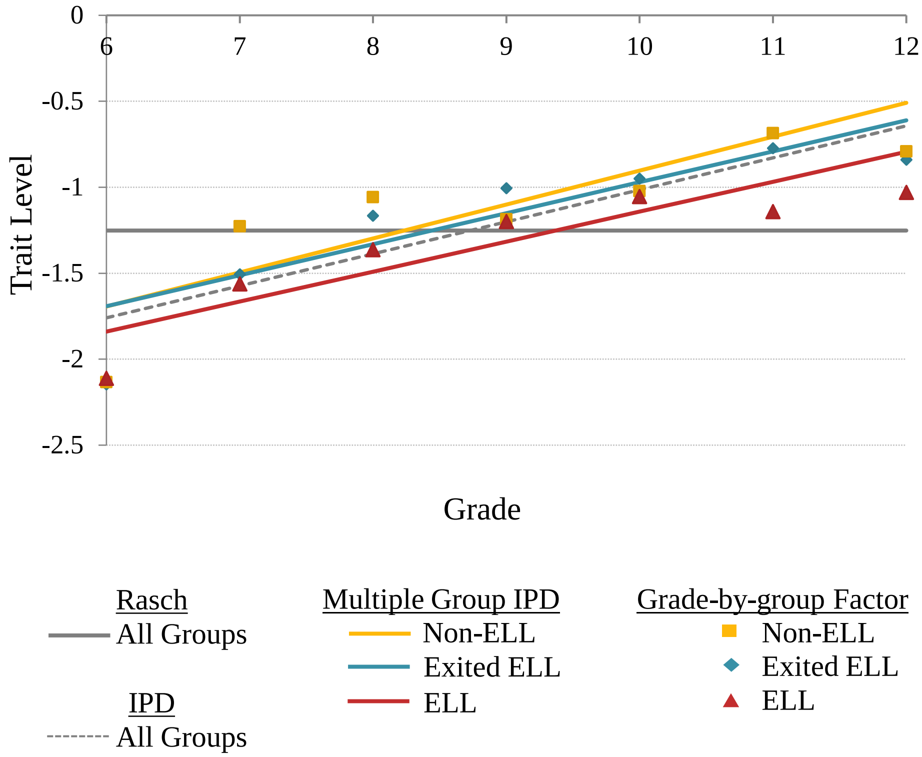

Figure 13. Visualization of the parameter estimates for item 4 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

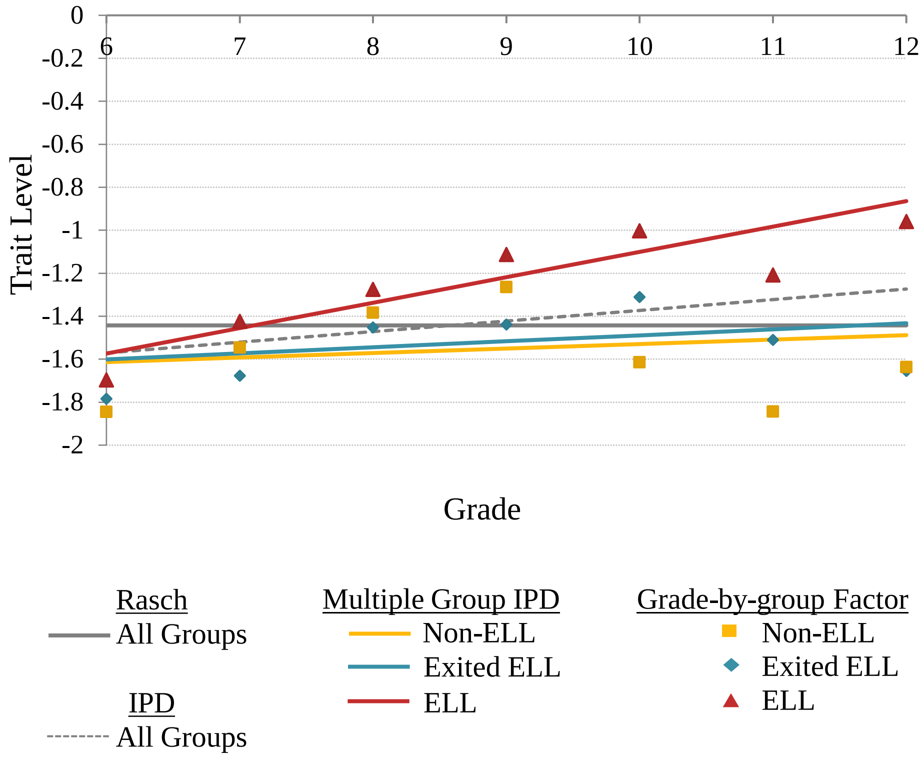

Figure 14. Visualization of the parameter estimates for item 5 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

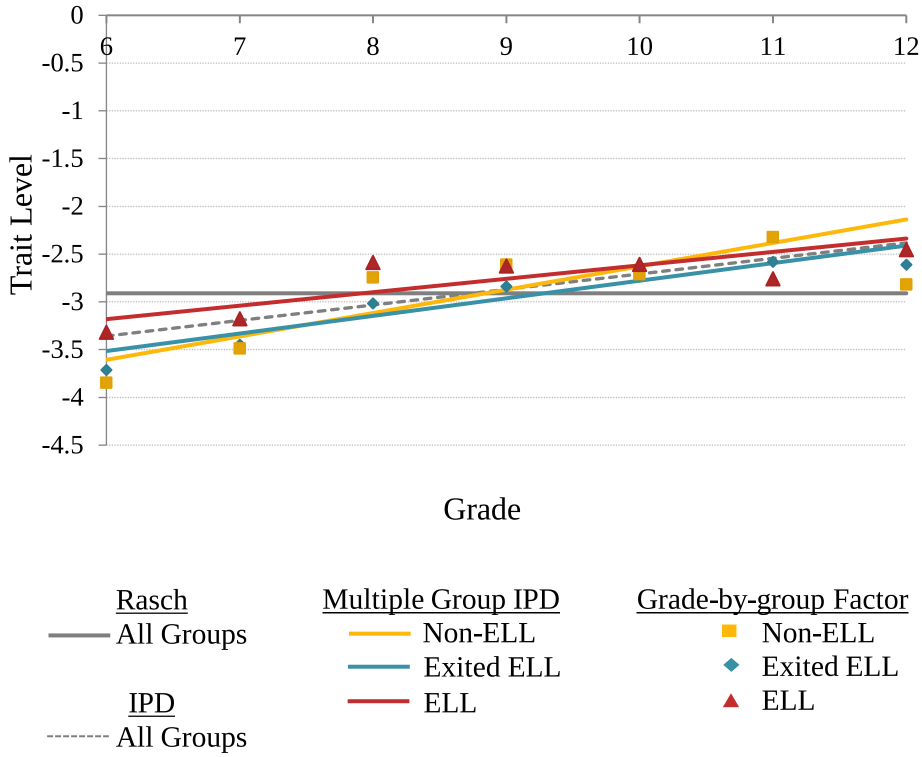

Figure 15. Visualization of the parameter estimates for item 6 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.

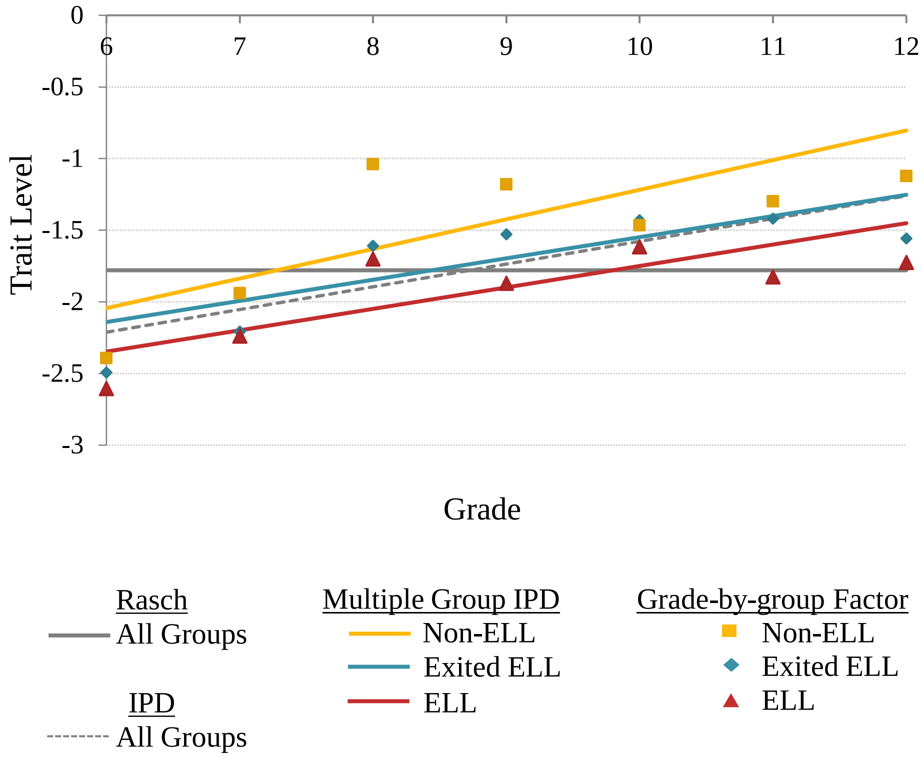

Figure 16. Visualization of the parameter estimates for item 7 on the Support scale across grade level for the Rasch, IPD, multiple groups IPD, and grade-by-group factor models.