Abstract Title Page

Not included in page count.

Title:
Dimensionality of Upper Elementary Mathematics Instruction: Exploring Factors Across Two Observational Instruments

Authors and Affiliations:
David Blazar, Harvard Graduate School of Education
David Braslow, Harvard Graduate School of Education
Charalambos Y. Charalambous, University of Cyprus

Abstract Body

Limit 4 pages single-spaced.

Background / Purpose:

Description of prior research, intellectual context, and the focus of the research.

Over the past several years, research teams have developed observational instruments to measure the quality of teachers' instructional practices. Instruments such as Framework for Teaching (FFT) and the Classroom Assessment Scoring System (CLASS) assess general teaching practices, including student-teacher interactions, behavior management, and instructional pedagogy (Kane, Taylor, Tyler, & Wooten, 2011; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). Other instruments such as the Protocol for Language Arts Teaching Observations (PLATO) and the Mathematical Quality of Instruction (MQI) attend to content-specific practices that are more pertinent for teaching and learning in specific disciplines (Grossman et al, 2012; Hill et al, 2008). Working at the intersection of both types of practices, we attempt to describe instruction using both generic and content-specific measures of teaching practice. As research teams have focused on the measurement properties of individual instruments, the extent to which generic and content-specific instruments capture related constructs is unclear. This research is of value to both researchers and school leaders, who might be interested in enumerating a parsimonious list of teaching practices that brings together generic and content-specific aspects of instruction; such a list could be useful for a number of purposes, including the creation of comprehensive evaluation frameworks and studies of the relationships between different domains of teaching practice and student learning.

To our knowledge, only one study has begun to address this area of inquiry. The Measures of Effective Teaching Project collected data from teachers across six urban school districts on multiple observation instruments including the four listed above. Kane and Staiger (2012) found that items tended to cluster instrument to form up to three principal components. Using the same data and a factor analysis framework, McClellan and colleagues (2013) examined overlap between content-specific and general observation instruments, finding little and as many as twelve factors. At the same time, neither set of authors attempt to explore potential overlap between instruments through more complex factor structures, such as bi-factor models that attempt to account for instrument-specific variation.

In this study we use exploratory and confirmatory factor analysis to examine scores of math instruction generated using two observation instruments, the MQI and CLASS, across a sample of over 300 fourth- or fifth-grade teachers. We attempt to answer the following two research questions: (1) How many unique dimensions of instruction do we measure? (2) To what extent is there overlap in the dimensions of instruction captured by these two instruments?

Data / Participants:

Description of the participants in the study: who, how many, key features, or characteristics.

Our sample consists of fourth- and fifth-grade teachers from four school districts in the 2010-2011 and 2011-2012 school years. Schools were selected into the study based on district referrals and size; the study design required that schools have a minimum of two teachers in each of the sampled grades. Of eligible teachers, 309 (roughly 55%) agreed to participate.

Teachers' mathematics lessons (n=1,362) were captured over a two-year period, with three lessons per teacher recorded each year. Videos were recorded using a three-camera, unmanned unit; site coordinators turned the camera on prior to the lesson and off at its conclusion. Most lessons lasted between 45 and 60 minutes. Teachers were allowed to choose the dates for capture in advance, and were directed to select typical lesson and exclude days on which students were taking a test. Although it is possible that these videotaped lessons are different from teachers' general instruction, teachers did not have any incentive to select lessons strategically as no rewards or sanctions were involved with data collection. In addition, analyses from the Measures of Effective Teaching project indicate that teachers are ranked almost identically when they choose lessons to be observed compared to when lessons are chosen for them (Ho & Kane, 2013).

Trained raters scored these lessons on two established observational instruments: the MQI, which focuses on mathematics-specific practices, and the CLASS, which focuses on general teaching practices. Validity studies have shown that both instruments successfully capture the quality of teachers' instruction, and specific dimensions from each instrument have been shown to relate to student outcomes (Bell, Gitomer, McCaffrey, Hamre, & Pianta, 2012; Blazar, 2014; Hill, Charalambous, & Kraft, 2012). For the MQI, two raters watched each lesson and scored teachers' instruction on 14 items for each seven-and-a-half-minute segment on a scale from Low (1) to High (3). A single item, *Classroom Work is Connected to Math*, is scored as Not True (0) True (1). For the CLASS, one rater watched each lesson and scored teachers' instruction on 12 items for each fifteen-minute segment on a scale from Low (1) to High (7) (see Table 1 for a full list of items). Three items from the MQI (*Major Errors, Language Imprecisions,* and *Lack of Clarity*) and one from the CLASS (*Negative Climate*) have a negative valence and therefore were reversed coded in this analysis. For both instruments, raters had to complete an online training, pass a certification exam, and participate in ongoing calibration sessions.

We used these data to create two datasets. The first is a teacher-level dataset with scores for each item on both the MQI and CLASS averaged across segments, lessons, and raters (for the MQI). The second is a segment-level dataset that captures the original scores assigned to each teacher by raters. For the MQI, we averaged scores across raters within a given segment to match the structure of the CLASS. Given that for any individual lesson there are twice as many segments for the MQI than for the CLASS, we assigned CLASS scores of the full fifteen-minute segment to the corresponding seven-and-a-half-minute segments from the MQI.

Analysis:

Description of the methods for collecting and analyzing data.

To answer our research questions, we conducted three sets of analyses. We began by examining pairwise correlations of items across instruments. This allowed us to explore the degree of potential overlap in the dimensions of instruction captured by each instrument. Next, we conducted a set of exploratory factor analyses to identify the maximum number of factors we might expect to see, both within and across instruments. While we conducted these analyses combining data from both instruments, prior research suggests that we would not expect to see much overlap across instruments (McClellan et al, 2013). Therefore, we conducted a set of confirmatory factor analyses to account for sources of variance that had not been addressed in previous analyses. In particular, we utilized a bi-factor model to extract instrument-specific variation, and then tested factor structures that allowed items to cluster across instruments.

Findings / Results:
Description of the main findings with specific details.

(Please insert Table 1 here). In Table 1, we present teacher-level correlations of items across instruments. Here, we find that some items on the MQI and CLASS are highly related. For example, *Analysis and Problem Solving* from CLASS is correlated with multiple items from the MQI (*Multiple Methods, Use Student Productions, Student Explanations, Student Mathematical Questioning and Reasoning,* and *Enacted Task Cognitive Activation*) above 0.3. Three items from the MQI – *Language, Use Student Productions,* and *Student Mathematical Questioning and Reasoning (SMQR)* – are correlated with all items from CLASS, though at lower magnitudes. This suggests that items from the two measures seem to be capturing somewhat similar facets of instruction and that factor structures might include factors with loadings across instruments

(Please insert Table 2 here). At the same time, exploratory factor analyses of teacher-level scores indicate little overlap between instruments. In Table 2, we present eigenvalues and factor loadings for a parsimonious list of factors generated by focusing on the factors with eigenvalues larger than 1 (Kline, 1994). While not shown here, we also conduct separate factor analyses for individual instruments and school years, as well as for segment-level scores, and find similar factor structures across all analyses. Results indicate that four factors are needed to parsimoniously capture the observed variation in instruction, and these factors do not suggest any substantial crossover between instruments. Generally, items appear to load onto only one factor, with the only exception being the MQI item of *Mathematical Language*, which had relatively low loadings on two factors. *Classroom Work is Connected to Math* does not load strongly onto any factor, which may be due either to the unique scaling of this item or to the fact that this item does not reflect content-specific aspects of instruction. We therefore allowed this item to load freely on different generic or the content-specific factors. We label these four factors "Ambitious Mathematics Instruction," "Mathematical Errors", "Classroom Organization", and "Classroom Climate and Support", with the first two from the MQI and the latter two from the CLASS.

(Please insert Table 3 here). In Table 3, we identify other potential model structures, beginning with a model using a single instructional factor and building towards a four-factor model similar to that identified by the exploratory factor analysis results. Models 1 through 5 do bot allow items to load across factors; however, we do explore models with items loading across instruments, which we capture in Model 6. Models 7 through 11 are bi-factor models that extract instrument-specific variation as well as theoretically driven instructional factors. Therefore, all items load onto two factors – one for the instrument on which they are scored and another for a particular instructional domain. Given the nested structure of the data, we run these models at both the teacher- and segment-level. Ideally, we would be able to fit a three-level model with segments nested within lessons, nested within teachers; however, due to non-convergence issues common to bi-factor models, we are only able to show results for a two-level model with segments nested within teachers.

(Please insert Table 4 here). In Table 4, we present model fit indices for all of these models. Because most models are not nested, we cannot compare them based on formal statistical significance tests. Instead, we rely on criteria from the field and, in particular, on the AIC and BIC indices (Akaike, 1987; Kline, 2011). When we compare models using AIC and BIC indices, we observe that a three- or four-factor structure best fits the data. In the one-level model with scores at the teacher level, the best-fitting model (i.e., those with smallest AIC and BIC values) is Model 9, which is a bi-factor model with three instructional factors: “Ambitious Instruction”, “Mathematical Errors”, and “Classroom Pedagogy”. “Mathematical Errors” consists solely of items from the MQI instrument, while the other two factors include items from both instruments. At the same time, once we extract variation for the MQI instrument, we do not observe substantial variation on the “Ambitious Instruction” factor. This may be because all but four items from the MQI are included in this factor. The second best-fitting model is Model 4, which matches results observed from exploratory factor analyses. Results from the two-level models also indicate that Model 4 has the best fit of the models that do not attempt to correct for instrument-specific variation. Because only one bi-factor, two-level model converged, we are not able to draw conclusions from this set of analyses.

It is important to note that no individual model meets commonly accepted criteria for overall model fit. This likely is due to the fact that there are sources of variation that are not being modeled well (segments, lessons, raters, etc). Other reasons, such as non-normal item score distributions, also play a role. Finally, the purpose of using at most four factors was to be parsimonious, not to capture all of the variation in our data.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

For years, scholars have attended either to generic or to content-specific teaching practices, without systematic attempts to consider both types in tandem. Responding to older (e.g., Brophy) and more contemporary (e.g., Grossman & McDonald, 2008) calls to attend to both types of practices, in this study we explored the benefits that can be accrued by working at the intersection of generic and content-specific practices. Despite their obvious limitations, our results seem to be in favor of integrating the two types of practices: they suggest that, although there still seem to be some more content-specific factors (e.g. “Errors and Imprecision”) and some more generic teaching factors (e.g., “Pedagogy”), other factors combine generic and content-specific aspects of instruction (e.g., “Ambitious instruction”).

This finding has implications for the measurement community. In this study, we found that items correlate both within and across instruments, and that, once we extracted instrument-specific variation – that might be attributed, among other things, to the different scales used in the two instruments – factors comprised of items from both instruments fit the data better than those that do not do so. Of course, as acknowledged above, we are limited in our ability to model the complexity of the data and, therefore, future work may attempt to do so.

In addition, our findings could also inform policy around teacher education and professional development programs. Results highlight a parsimonious list of instructional factors – both general and content-specific – that, in turn, can be used in future work to explore which of these dimensions or combinations thereof matter most to student learning outcomes. Once validation studies link these dimensions to student learning, teacher preparation and professional development programs can target these areas of teacher practice as they prepare pre-service and in-service teachers for the complex work of teaching.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52* (3), 317-332.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., & Pianta, R. C. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2-3), 62-87.

Blazar, D. (2014). *The effect of exposure to high-quality mathematics instruction on student achievement: Exploiting within-school, between-grade, and cross-cohort variation in scores from an observation instrument.* Cambridge, MA: National Center for Teacher Effectiveness, Harvard University.

Brophy, J. (1986). Teaching and learning mathematics: Where research should be going? *Journal for Research in Mathematics Education, 17* (5), 323-346.

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added.* Working Paper No. 16015. Cambridge, MA: National Bureau of Economic Research.

Grossman, P., & McDonald, M. (2008). Back to the future: directions for research in teaching and teacher education. *American Educational Research Journal, 45*, 184-205.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Charlotte, NC: Information Age Publishing.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researchers, 41*(2), 56-64.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel*. Seattle, WA: Measures of Effective Teaching Project, Bill and Melinda Gates Foundation.

Kane, T. J., Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Policy and practice brief.* Seattle, WA: Measures of Effective Teaching Project, Bill and Melinda Gates Foundation.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

McClellan, C., Donoghue, J., & Park, Y. S. (2013). *Commonality and uniqueness in teaching practice observation*. Clowder Consulting.

Pianta, B., Belsky, J. Vandergrift, N. Houts, R. & Morrison, F. (2008) Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*(2), 365-387.

Appendix B. Tables and Figures

Not included in page count.

Table 1

Item Correlations

	CLASS Items											
MQI Items	*Negative Climate*	*Behavior Management*	*Productivity*	*Student Engagement*	*Positive Climate*	*Teacher Sensitivity*	*Respect for Student Perspectives*	*Instructional Learning Formats*	*Content Understanding*	*Analysis and Problem Solving*	*Quality of Feedback*	*Instructional Dialogue*
Linking	0.095	0.112~	0.102~	0.043	-0.006	0.1~	0.087	0.108~	0.047	0.204***	0.051	0.16**
Explanations	0.138*	0.183**	0.169**	0.114*	0.036	0.214***	0.11~	0.148*	0.194***	0.277***	0.147*	0.212***
Multiple Methods	0.072	-0.016	-0.004	-0.051	-0.062	0.046	0.124*	0.007	-0.063	0.31***	0.026	0.15**
Generalizations	0.101~	0.144*	0.167**	0.044	-0.019	0.091	-0.008	0.101~	0.178**	0.059	0.121*	0.073
Language	0.2***	0.285***	0.217***	0.211***	0.142*	0.216***	0.112~	0.229***	0.276***	0.213***	0.112~	0.227***
Remediation	0.048	0.082	0.108~	-0.008	-0.02	0.134*	0.049	0.027	0.094	0.176**	0.101~	0.11~
Use Student Productions	0.133*	0.154**	0.157**	0.141*	0.106~	0.255***	0.272***	0.196***	0.165**	0.375***	0.271***	0.332***
Student Explanations	0.089	0.12*	0.131*	0.112~	0.079	0.252***	0.255***	0.163**	0.146*	0.375***	0.258***	0.336***
SMQR	0.111~	0.168**	0.134*	0.139*	0.114~	0.24***	0.237***	0.16**	0.14*	0.325***	0.166**	0.269***
ETCA	0.138*	0.167**	0.184**	0.123*	0.09	0.257***	0.271***	0.176**	0.156**	0.302***	0.253***	0.267***
Major Errors	-0.022	-0.082	-0.105~	0.03	-0.031	0.009	0.007	0.003	-0.084	0.052	-0.048	0.039
Language Imprecisions	-0.085	-0.091	-0.056	0.009	-0.044	-0.075	0.06	0.012	-0.061	-0.065	0.035	0.039
Lack of Clarity	-0.007	-0.068	-0.078	0.018	0.023	0.007	0.021	0.034	-0.04	0.028	-0.011	0.018
CWCM	0.125*	0.176**	0.204***	0.027	0.092	0.161**	0.051	0.063	0.142*	0.174**	0.256***	0.161**

Table 2
Exploratory Factor Analyses Loadings

	Factor 1	Factor 2	Factor 3	Factor 4
Eigenvalues	5.779	4.755	2.240	1.717
MQI				
Linking	0.011	0.540	0.101	0.148
Explanations	0.047	0.789	0.188	0.161
Multiple Methods	0.009	0.611	-0.101	0.043
Generalizations	-0.013	0.355	0.214	0.069
Language	0.139	0.301	0.277	0.146
Remediation	-0.021	0.619	0.098	0.231
Use Student Productions	0.181	0.879	0.018	0.054
Student Explanations	0.184	0.812	-0.032	-0.002
SMQR	0.160	0.680	0.034	-0.087
ETCA	0.147	0.803	0.075	0.122
Major Errors	-0.005	0.145	-0.066	0.783
Language Imprecisions	-0.008	0.142	-0.090	0.510
Lack of Clarity	0.012	0.106	-0.036	0.813
CWCM	0.078	0.247	0.183	-0.080
Class				
Negative Climate	0.361	0.047	0.586	-0.011
Behavior Management	0.318	0.079	0.776	-0.061
Productivity	0.314	0.084	0.765	-0.078
Student Engagement	0.718	-0.013	0.280	0.060
Positive Climate	0.758	-0.063	0.238	0.008
Teacher Sensitivity	0.775	0.126	0.318	-0.003
Respect for Student Perspectives	0.815	0.134	-0.087	0.002
Instructional Learning Formats	0.821	0.052	0.173	0.043
Content Understanding	0.703	0.055	0.367	-0.047
Analysis and Problem Solving	0.728	0.315	-0.056	-0.042
Quality of Feedback	0.727	0.124	0.219	-0.043
Instructional Dialogue	0.851	0.202	0.085	0.008

Table 3
Confirmatory Factor Analysis Model Organization

	Single Factor - No Cross Loadings						Bifactor - Items Load onto their Respective Instruments, Plus Other Factors				
Items	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
MQI											
Linking	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Explanations	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Multiple Methods	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Generalizations	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Language	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Remediation	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Use Student Productions	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Student Explanations	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
SMQR	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
ETCA	Instruction	MQI	Content	Ambitious	Ambitious	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Major Errors	Instruction	MQI	Content	Errors	Errors	Errors	Instruction	Content	Errors	Content	Errors
Language Imprecisions	Instruction	MQI	Content	Errors	Errors	Errors	Instruction	Content	Errors	Content	Errors
Lack of Clarity	Instruction	MQI	Content	Errors	Errors	Errors	Instruction	Content	Errors	Content	Errors
CWCM	Instruction	MQI	Pedagogy	Ambitious	Organization	Organization	Instruction	Pedagogy	Pedagogy	Organization	Organization
CLASS											
Negative Climate	Instruction	CLASS	Pedagogy	Organization	Organization	Organization	Instruction	Pedagogy	Pedagogy	Organization	Organization
Behavior Management	Instruction	CLASS	Pedagogy	Organization	Organization	Organization	Instruction	Pedagogy	Pedagogy	Organization	Organization
Productivity	Instruction	CLASS	Pedagogy	Organization	Organization	Organization	Instruction	Pedagogy	Pedagogy	Organization	Organization
Student Engagement	Instruction	CLASS	Pedagogy	Support	Support	Support	Instruction	Pedagogy	Pedagogy	Climate	Climate
Positive Climate	Instruction	CLASS	Pedagogy	Support	Support	Support	Instruction	Pedagogy	Pedagogy	Climate	Climate
Teacher Sensitivity	Instruction	CLASS	Pedagogy	Support	Support	Support	Instruction	Pedagogy	Pedagogy	Climate	Climate
Respect for Student Perspectives	Instruction	CLASS	Pedagogy	Support	Support	Support	Instruction	Pedagogy	Pedagogy	Climate	Climate
Instructional Learning Formats	Instruction	CLASS	Content	Support	Support	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Content Understanding	Instruction	CLASS	Content	Support	Support	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Analysis and Problem Solving	Instruction	CLASS	Content	Support	Support	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Quality of Feedback	Instruction	CLASS	Content	Support	Support	Ambitious	Instruction	Content	Ambitious	Content	Ambitious
Instructional Dialogue	Instruction	CLASS	Content	Support	Support	Ambitious	Instruction	Content	Ambitious	Content	Ambitious

Table 4

Model Fit Indices for Confirmatory Factor Analysis Models

Fit Indices	Fit Criterion	Null Model	Single Factor - No Cross Loadings						Bifactor - Items Load onto their Respective Instruments, Plus Other Factors				
			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
<u>One-Level Models</u>													
Akaike (AIC)	Smallest Value	-543.417	-2871.42	-4355.4	-3276.02	**-4961.7**	-4956.34	-3578.06	-4737.62	-5041.38	**-5183.24**	-5038.67	-5173.51
Bayesian (BIC)	Smallest Value	-341.63	-2568.74	-4048.84	-2969.46	**-4635.73**	-4630.38	-3252.09	-4330.17	-4630.05	**-4764.14**	-4619.57	-4742.78
Sample-Size Adjusted BIC	Smallest Value	-506.599	-2816.2	-4299.47	-3220.08	**-4902.22**	-4896.86	-3518.58	-4663.28	-4966.33	**-5106.77**	-4962.2	-5094.92
Chi-Square Test of Model Fit	>.05	0	0	0	0	0	0	0	0	0	0	0	0
RMSEA (Root Mean Square Error Of Approximation)	<.05	0.202	0.158	0.113	0.148	0.09	0.091	0.148	0.103	0.092	0.084	0.091	0.086
CFI	>.95	0	0.433	0.712	0.509	0.819	0.817	0.513	0.784	0.828	0.857	0.833	0.85
SRMR (Standardized Root Mean Square Residual)	<.1	0.301	0.165	**0.092**	0.184	**0.071**	**0.073**	0.146	**0.07**	**0.083**	**0.07**	**0.069**	**0.066**
<u>Two-Level Models (Segments Nested Within Teacher)</u>													
Akaike (AIC)	Smallest Value	318885.2	292344.2	284937.3	290766.1	**282835.8**	283162.6	288716.9	**282400.1**				
Bayesian (BIC)	Smallest Value	319711.72	293517.3	286123.7	291952.6	**284088.9**	284415.7	289970	**283913.2**				
Sample-Size Adjusted BIC	Smallest Value	319317.68	292958	285558.1	291386.9	**283491.5**	283818.3	289372.6	**283191.8**				

Note: Two-level models 8-11 did not converge. Nor did any model with three levels (segements nested within lesson, within teacher). Models that meet fit criterion are bolded.