# Third-Year Results From an Efficacy Study of the Acuity Data System

**Shana Shaw, Ph.D.**[1]
**Jeffrey C. Wayman, Ph.D.** [2]

**November 26, 2012**

[1]**SEDL, Austin TX**
[2]**The University of Texas at Austin**

## ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of various individuals who have provided work, knowledge, and support during the course of this project. First, we wish to thank David Seitter, Linda Bond, Jessalyn Smith, Dan Lewis, Mark Mantel, and Tariq Durrani of CTB/McGraw-Hill for their support and input as we collected data for this study. Our thanks also go to Joe O'Reilly and Cindy Bochna of Mesa Public Schools for their generous support in providing data and answering questions. Thanks also to Ronald Heck and Arie van der Ploeg, who agreed to review a prior draft of this report – their comments were tremendously helpful and we thank them for their time and insight. Finally, we wish to thank Vincent Cho of Boston College for his expertise and collegial input in earlier portions of this study.

This study is funded by CTB/McGraw-Hill. CTB/McGraw-Hill has a number of efficacy studies to evaluate the effectiveness of the Acuity system and we feel fortunate to be trusted to do this research. By agreement, CTB/McGraw-Hill has no influence or control over the results or reporting from this study, making this a fully unbiased study of Acuity. The findings and conclusions from this report are solely the responsibility of the authors.

We thank CTB/McGraw Hill for undertaking a relationship of this sort. While many vendors may be uncomfortable giving up control of the content of the evaluation, this study represents a commitment by CTB/McGraw-Hill to unbiased evaluation and the betterment of education. We commend them for their efforts and we endeavor to provide good work in return.

# INTRODUCTION

This study was conducted in the third year of a project examining the Acuity assessment system in Mesa Public Schools (MPS). The primary goal of this research was to identify the effects of teachers' Acuity use on student achievement. To better understand this relationship, we explored facets of teachers' Acuity use, factors impacting this use, and district support for Acuity use (Wayman, Cho, & Shaw, 2009; Wayman, Shaw, & Cho, 2011). Our previous findings indicate that teachers' use of the system increased dramatically in the second year of Acuity implementation in MPS, though student achievement gains associated with that use were small and inconsistent across grades and subjects (Wayman et al., 2011). Our previous reports also identified actions we believed MPS could take to improve the impact of Acuity use on student achievement, including increasing district support, providing better training for educators in using Acuity, and consistently integrating Acuity into educators' work (Wayman et al., 2009; Wayman et al., 2011). In the present study, we continued our examination of the relationship between teachers' Acuity use and student achievement in MPS, though from a more longitudinal perspective than our previous studies.

## Studies of Acuity Use

Our work joins that of others that explored Acuity implementations in other districts (Konstantopoulos, Miller, van der Ploeg, Li & Traynor, in press; Spradlin, 2012). While independent from one another, these studies have been conducted concurrently, generating a small body of empirical literature on Acuity. Taken together, this work highlights aspects of how Acuity is used in educational settings, and relationships between that use and student achievement.

One was a multi-year, mixed methods study conducted by researchers at the Center for Evaluation and Education Policy (CEEP) (Spradlin, Dickinson, Dingjing, Shi, & Whiteman, 2012). Exploring relationships between Acuity use and achievement in Indiana schools, CEEP findings were similar to those we observed in the second year of our work in MPS (Wayman et al., 2011). Specifically, CEEP researchers found small, statistically significant relationships between teachers' Acuity use and student achievement at both the elementary and junior high level (we detected trends at only the elementary level). When interpreting the practical significance of those associations, CEEP researchers reached conclusions similar to ours; that is, while statistically significant, relationships between Acuity use and achievement did not indicate meaningful achievement gains among students on the whole (Spradlin et al., 2012).

The second study, also conducted in Indiana, used a randomized design to examine the causal impact of Acuity use on math achievement (Konstantopoulos et al., in press). Unlike the CEEP study and our work in Mesa, these authors observed Acuity effects that they considered both statistically and practically significant, particularly in the upper elementary grades. After just one year of implementation, Acuity use resulted in additional math gains among fifth and sixth grade students that ranged from one-fourth to one-third of a standard deviation (Konstantopoulos et al., in press). Those findings seem to be more impactful than trends we observed in MPS, with high levels of Acuity use associated with achievement gains among elementary math students of less than one-tenth of standard deviation (Wayman et al., 2011).

The discrepancies between findings we observed in Mesa and those of Konstantopoulos and colleagues in Indiana are not unique to Acuity initiatives. A number of other researchers have studied similar assessment initiatives in the past five years, with varied conclusions about the impact of the respective initiatives (Carlson, Borman, & Robinson, 2011; Henderson, Petrosino, Guckenburg, & Hamilton, 2007; May & Robinson, 2007; Quint, Sepanik, & Smith, 2008; Slavin, Cheung, Holmes, Madden, & Chamberlain, 2011).

All of these studies were predicated on the same underlying assumption operating in MPS: interim assessments and analytic tools can help teachers improve their capacity to plan and deliver instruction. In the next section, we describe these studies and provide an overview of themes derived from this relatively new line of research.

**Other Literature**

Two recent studies were initiated by researchers at the *Center for Data-Driven Reform in Education* (CDDRE) at Johns Hopkins University (Carlson et al., 2011; Slavin, et al., 2011). Both reported findings from experimental investigations of a benchmark assessment initiative implemented in 59 districts, selected randomly across seven states. CDDRE consultants worked with individual districts to develop and administer quarterly benchmark assessments. They also counseled district personnel on how to use the data generated by the benchmark assessments to identify areas of curricular and programmatic weakness as well as potential interventions to address those weaknesses.

After one year of the CDDRE initiative, Carlson et al. (2011) reported a small, positive, statistically significant impact of the program on math achievement and a positive—though not statistically significant—impact on reading achievement. These authors also reported that, while educators were generally positive about the training they received on using the assessment data, complications in implementing the intervention still arose, including: a lack of compliance among some schools in administering all four benchmark assessments, unanticipated challenges in hosting data review and training sessions for practitioners, and intervention contamination between CDDRE districts and two districts in the control condition where benchmark assessments were also administered. Carlson et al. (2011) reported that these issues were largely resolved after the first year.

After four years of the CDDRE intervention, Slavin et al. (2011) found larger effects on elementary reading and math achievement than Carlson et al. (2011), effects Slavin et al. cited as both practically and statistically significant. The impact of the CDDRE program on middle school achievement—specifically eighth grade reading and math—was less substantial: the researchers observed a positive impact on reading and math achievement in years one and two, but the same effects were not observed in the third and fourth years of the study. When assessing the overall impact of the CDDRE initiative, Slavin et al. concluded that it had a meaningful impact on student achievement, particularly when educators used the benchmark assessment data to select research-based instructional interventions. In other words, achievement improvements resulted from teachers' informed and purposeful use of the assessment data to guide their instruction rather than simply having access to students' assessment results.

Two other recent studies explored benchmark assessment initiatives designed to help educators forecast students' preparedness for annual state tests in Massachusetts (the *Massachusetts Comprehensive Assessment System*, or MCAS). Quint, Sepanik, and Smith (2008) explored the *Formative Assessments of Student Thinking in Reading* (FAST-R) program, which periodically assessed reading comprehension among third and fourth grade students in Boston Public Schools (BPS). Similar to the CDDRE initiative, FAST-R provided consultants (i.e., data coaches) who met with teachers to review student results and suggest interventions based on the data. The second Massachusetts study explored a pilot program where benchmark assessments were administered in junior high math (Henderson et al., 2007).

Both studies used quasi-experimental research methods to identify similar non-participating schools for "matched comparison" samples. Neither study, however, found statistically significant improvements in achievement attributable to their respective benchmark assessment initiatives. Both sets of authors identified factors that may have restricted their ability to assess program impacts. For instance, Henderson et al. (2007) noted that their study was statistically underpowered: by using schools as the unit of analysis, their sample was limited to only 22 treatment and 44 comparison schools. Quint et al. (2008) reported that most data coaches met with teachers too infrequently to provide impactful support. Whether teachers received training on using the benchmark assessment data in the Henderson et al. study is unclear—their report focused on the analytic methods and research findings rather than the initiative. Like Slavin et al, Henderson and colleagues did endorse the notion that improving student achievement via benchmark assessment initiatives relies on how educators use data: "Higher mathematics scores will come not because benchmarks exist but because of how the benchmark assessment data are used by a school's teachers and leaders" (p. 8).

Another experimental study examined Ohio's *Personalized Assessment Reporting System* (PARS) initiative, which was implemented in 51 randomly-selected high schools in 60 districts in Ohio. The PARS initiative provided resources to students, parents, teachers, and administrators to prepare students for the Ohio Graduation Test (OGT) (May & Robinson, 2007). The authors found no impact of the tools for most students who took the OGTs, but did find statistically and practically significant effects for students who retook the OGT after previously unsuccessful attempts. As with the two Massachusetts studies, May and Robinson (2007) identified limitations in their study, including deficient information provided in PARS reports and school- and district-leaders' lukewarm support for the initiative. The authors also reported that, after only one year of implementation, it was difficult to assess the impact PARS reports and tools could have on teaching and learning given a more sustained intervention.

Our prior research covered 2009[1], the first full year of Acuity implementation (Wayman et al., 2009), and 2010, the second year of Acuity implementation (Wayman et al., 2011). After the first full year of Acuity implementation in MPS, our research revealed negligible Acuity use, resulting largely from low levels of district- and school-level support and leadership in using the system (Wayman et al., 2009). Our study of the second year of implementation examined single-year effects of Acuity use on achievement[2], and revealed that increased use was associated with slightly higher achievement on state tests in elementary schools, but found no effects in junior high (Wayman et al., 2011). At that time, we speculated that increased support and familiarity with Acuity could result in larger effects in later years.

Altogether, these studies reveal themes in the implementation and impact of benchmark assessment initiatives. In terms of implementation, several studies echoed our findings that educators are often positive about assessment initiatives when it is clear how these initiatives may improve their practice (Carlson et al. 2011; May & Robinson, 2007; Quint et al., 2008; Wayman et al., 2011). However, many initiatives also faced difficulties, experiencing growing pains during early stages of implementation, or in light of competing initiatives (Carlson et al. 2011; May & Robinson, 2007; Quint et al., 2008). MPS was not impervious to these difficulties: in the early stages of Acuity's rollout, a competing program diverted training resources and

[1] We refer to the 2010-2011 school year as "2011", 2009-2010 as "2010", and 2008-2009 as "2009."
[2] Acuity use was too low in year one of implementation to consider longitudinal relationships in year two.

educator time away from the system (Wayman et al., 2009). The resulting lack of professional development led many teachers to report that Acuity was not user-friendly and was too time-consuming to navigate; a greater focus on system support the following year ameliorated many of these concerns (Wayman et al., 2011).

Despite problems with implementation and support, some authors found that benchmark assessment initiatives had positive, *statistically* significant impacts on achievement. However, only Slavin et al. (2011) identified effects large enough that they considered *educationally* significant, effects they attributed to the sustained nature of the initiative and interventions that were chosen based on the assessment data.

**The Present Study**

The previous review of literature is unclear whether it is reasonable to expect one-year effects of benchmark assessment on student achievement. The present study adds to this research base by examining effects of the Acuity system in its third year of implementation, with MPS district administrators reporting ongoing, increased support for teachers in using the system, particularly in the second and third year. Thus, in conducting this study, we sought to examine effects of two years of sustained Acuity use on student achievement. In doing so, we sought to first establish consistency with our prior reports, identifying correlates of teacher Acuity use and the one-year correlation of Acuity use and student achievement in the 2011 school year. Additionally, we focused on the two-year effects of Acuity use on student achievement.

We addressed three research questions:

1. Which factors were associated with teachers' 2011 Acuity use?
2. Did 2011 teacher Acuity use correlate with 2011 student achievement?
3. Did teacher Acuity use over two consecutive years (2010 and 2011) correlate with 2011 student achievement?

Research Questions 1 and 2 are replications of research questions in our second-year report (Wayman et al., 2011). In exploring these questions, we aim to verify that there are no major changes between 2010 and 2011 regarding factors that are associated with teacher Acuity use and the one-year correlation between Acuity use and student achievement. Research Question 3 is the main question in our study and is designed to investigate two-year effects of teacher Acuity use on student achievement.

Our report consists of four sections. First, we present the methodology for our study. Second, we offer the results of our analyses. Third, we discuss these results and provide recommendations for improvement. Finally, we offer a short conclusion section.

## METHOD

We employed a purely quantitative research design in conducting this study. Our analytic method expanded the scope of our previous studies by examining longitudinal impact of teachers' Acuity use, controlling for their teaching experience and student-level factors that may influence student achievement. Analytic samples were disaggregated by research question, school type (i.e., elementary/junior high) and content area (reading/math). This section outlines our methods and procedures: first, we describe the Acuity system, then the MPS context. Following this, we offer sections outlining our procedures for collecting data, the measures used, and analyses employed.

### About Acuity

Acuity is a software program by CTB/McGraw-Hill that offers numerous functions for accessing data and standards-based instructional content. It is intended to serve educators by assessing student progress on state learning standards and determining students' readiness for state tests (CTB/McGraw-Hill, 2009). Acuity assessments may be offered online or via pencil and paper, the results of which are then accessible via the Acuity data system. These assessments target reading and math in grades 3-8, as well as Algebra. When fully implemented, Acuity offers three predictive assessments to assess learning and predict progress toward state tests, four diagnostic assessments to assess learning, and a variety of reports and tools for working with these data. For instance, Acuity offers roster and individual level reporting, distracter analysis, item banks, and the ability to create and customize assessments for individual students. The research team was trained in Acuity by CTB staff and provided access to MPS Acuity data for the purposes of becoming more familiar with the system.

There are three forms of predictive tests; the material on these tests is based on state standards. Form A is a baseline test that is given at the start of a year. In each grade, this test consists of material from the prior grade and material that is to be taught in the current grade. Form B is a mid-year test that is typically given in late falls. It consists of material already taught, but also contains material yet to be taught. Form C is given in the early spring and is intended to prepare students for the state test by testing all material for the current grade.

Districts are able to determine what levels of functionality are accessible to various roles. For instance, central office administrators usually have access to different functions or data than principals, who have different levels of access than teachers. In this report, we focus on the instructional functions provided to MPS teachers. To evaluate these functions, we examined use logs generated by the Acuity system to track weekly educator use. We categorized instructional functions into five areas: Custom Tests, Instructional Resources, Management, Reports, and Tracking Completion Status (see the *Measures* section for information on how each function was quantified).

### Mesa Public Schools and Acuity Implementation

Mesa Public Schools (MPS) is located in Mesa, Arizona. The city of Mesa is a suburb of Phoenix with an inner city core, and a population of approximately 439,000 (Census Bureau, 2011). For the 2011 school year, MPS enrolled about 65,000 students and employed 3,752 teachers. Approximately 76% of MPS students identify English as their primary language at home. The district's two largest student demographic groups are Non-Latino White (48%) and Latino (41%). Roughly 66% of students are enrolled in free or reduced lunch (Mesa Public

Schools, 2011). MPS has 56 elementary schools (serving grades K-6), 11 junior high schools (serving grades 7-8), and six senior high schools (serving grades 9-12), not including other more specialized schools or academies.

MPS uses a variety of formal assessments to track student learning. The state test is one example: *Arizona's Instrument to Measure Standards* (AIMS) is a state-mandated, criterion-referenced test, used since the 1990s. The assessment underwent extensive revision and was re-released in its current form in 2005. Since 2005, the only major revision was to the math portion of the AIMS in 2008. At that time, Arizona adopted new content standards in math, to reflect both Common Core and state-specific content standards. Test items based on the new math content standards were developed and field-tested, and the revised portion of the AIMS math test was first administered in 2010[3]. These changes are particularly noteworthy here for their impacts on students' test scores in math after those scores were rescaled in 2010.

For grades 3-8, AIMS is administered for several days, usually in April. Students in grades 10-12 follow a different testing schedule that allows for testing in October, February and April. AIMS comprises several components, including reading, writing, and math. Science is offered to fourth and eighth graders and high school biology students. Examples of other assessments include the *Arizona English Language Learner Assessment* (AZELLA), the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) and various versions of assessments from the TerraNova battery of tests.

The Arizona Department of Education website houses information on district-level performance on federal Adequate Yearly Progress accountability standards since the first year in which those data were publically available (2003). In the intervening years leading up to the acquisition of the Acuity system in 2008, MPS as a district failed to make AYP, mostly due to the reading achievement of subgroups, such as students classified as ELL and in special education. These difficulties led district officials to search for a predictive assessment tool; a search that culminated in the selection of Acuity.

MPS selected Acuity after an extensive process intended to ensure the contribution of many perspectives (Mesa Public Schools, 2009). Some of these strategies included participation and feedback from major user groups, an adoption committee composed of likely users of the system, and a substantial evaluation and RFP process. The district administers Acuity predictive assessments in reading and math three times per school year (in August, October, and January). The tests are administered on paper, with bubble sheets used for responses. Some schools scan on-site but most tests are sent to the district for scanning. Results from the predictive assessments are made available to teachers to help them track student progress toward mastering state standards, and predict students' AIMS performance. District leaders also provide resources and trainings to help teachers use Acuity's instructional resources, both individually and in their professional learning communities (PLCs).

The 2011 school year is the third year of full Acuity implementation. The system has actually been available since midway through the 2008 school year, but a variety of barriers precluded complete, district-wide implementation until the 2009 school year. Training during 2009 focused primarily on the Acuity predictive tests. By the 2010 school year, all teachers had received training in Acuity functions relating to predictive tests, and most had received training on custom tests and instructional resources.

---

[3] Information on the 2010 revisions to AIMS math tests available on the Arizona Department of Education's website http://www.azed.gov/standards-practices/math-standards/.

In 2011, the district hired an Acuity trainer who led two types of trainings: *training by school request* and *training by teacher request*. Principals initiated the trainings by school request, which involved targeted, small group trainings for teachers in one-, two- or four-hour sessions. When possible, the trainings were scheduled during the school day, with principals arranging for substitutes to cover teachers' classes. Trainings by teacher request involved four-hour workshops provided through the district's professional development department. These trainings were leveled according to teachers' familiarity with the system. Initial trainings provided basic information to novice users on navigating the system and accessing assessment reports while other trainings showed mid-level users how to use assessment reports to better understand student achievement, and how to assign instructional resources based on the results from predictive tests. Advanced users received trainings on using Acuity data in concert with information from other data sources.

**Measures**

Data collected for this study included demographic data, weekly use logs from the Acuity system, and student achievement data. A section is provided for each in the following narrative.

**Demographic data.** MPS district personnel provided demographic data for students, teachers, and schools data. Unique identifiers allowed students to be linked with teachers, and teachers to be linked with schools in any given year.

Student demographic data included gender, ethnicity, economic status and grade. Ethnicity was collapsed into three categories: Latino, Non-Latino White, and Other. Economic status was measured by whether students qualified for free and reduced lunch: students were classified as Economically Disadvantaged or Not Economically Disadvantaged. Tables 1-3 describe the student sample by gender, ethnicity, economic status, and school type.

Teacher data included years of experience in education and subject taught (for junior high teachers). Following prior reports (Wayman et al., 2009; Wayman et al., 2011), years of experience was collapsed into a four-level variable: (a) 5 years or less, (b) 6–10 years, (c) 11 – 20 years, and (d) 20 or more years. We also considered including teachers' educational attainment in our analyses, but it was heavily confounded with years of experience. Table 4 describes the teacher sample by school type and years of experience.

School demographic data include state achievement classification (Arizona Learns) and school type (elementary or junior high; high schools are not included because Acuity assessments only range from grades 3-8). Table 5 describes the school and teacher sample by AZ Learns status.

**Acuity use.** Descriptions of teachers' use of the system, in the form of weekly use logs, were developed and provided by Acuity's vendor, CTB/McGraw-Hill. These logs reported actions taken and the date of the action, allowing us to investigate which teachers executed particular actions within the system and how often s/he performed an action. Figure 1 shows Acuity use by week throughout the 2011 school year. Use typically spiked in the weeks after predictive tests were administered. Highest levels of use occurred after administration of Form C, prior to AIMS testing.

As with teacher background data, teachers' unique identification numbers allowed linkage to student achievement data. Since this study focused on whether Acuity helps teachers provide better, more effective instruction, we restricted our focus to analysis of instructional functions of the Acuity system. Accordingly, the terms *use* and *Acuity use* in this report refer only to Acuity's instructional functions.

Use logs enabled us to view Acuity use in two ways: total actions (*prevalence*) and number of weeks used (*consistency*). Prevalence was defined as the total number of actions (relating to instruction) that each teacher performed in 2011. Consistency was defined as the total number of weeks in 2011 that teachers used Acuity for at least one instructional action. Table 6 gives the percent of teachers who used Acuity for one or more instructional actions from 2009 to 2011. Table 7 describes prevalence and consistency of use for these three years.

Instructional functions available in Acuity were tracked throughout the school year and categorized into four areas[4]: (1) Instructional Resources, which allows teachers to assign content to specific students; (2) Management, which allows teachers to manage student information; (3) Reports, which allows teachers to access summaries of data in the form of reports; and (4) Tracking Completion Status, which allows teachers to view and manage the status of individual assignments. Prior reports (Wayman et al., 2009; Wayman et al., 2011) provide more detail on these areas.

**Student achievement.** Student achievement was measured using state test scores (AIMS) for reading and math in grades 4-8. Scores were collected for three school years: 2009, 2010, and 2011. Tables 8 and 9 provide descriptive statistics on AIMS Reading and Math scale scores, respectively. Both portions of the AIMS are vertically equated[5], and Tables 10 and 11 describe differences in annual AIMS scale scores (MPS and statewide).

**Analyses and Analytic Samples**

We used multilevel modeling (MLM) as our primary tool of analysis due to the nested structure of our data: students nested within teachers, who are nested within schools. Unlike traditional methods such as ANOVA or Multiple Regression, MLM accounts for the fact that individual student responses within teachers (or teachers within schools) have some dependence on each other. Thus, MLM gives a more accurate representation of school-to-school variance and provides more statistical power than modeling schools as the unit of analysis (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

In all analyses, statistical significance was assessed at the 0.05 level and 95% confidence intervals were computed for each effect. Full models were compared with null models to determine reductions in level-2 variability due to the inclusion of predictor variables. Model estimates were produced with Restricted Maximum Likelihood (REML) estimation[6] in SPSS Mixed (Version 17). SPSS does not restrict configurations of nested structures in longitudinal datasets; as such, the software program was advantageous for our analyses involving cross-classified data structures (Research Question 3).

In our models, categorical variables were specified with reference categories. At the student level, "male" was specified as the reference category for the gender variable, "other" was the reference category for ethnicity, and "economically advantaged" was the reference category for economic status. At the teacher level, the reference category for teaching experience was "20+ years." At the school level, "elementary" was the reference category for school type and

[4] Custom Tests usage was counted in 2009 but counts were unavailable for 2010 and 2011.

[5] As noted previously, there was a change in the scaling of score on the AIMS math portion in 2010 (see "Mesa Public Schools and Acuity Implementation" section). This change does not affect our analyses—we only use 2009 achievement data as a covariate, and as such, the change does not affect our analyses.

[6] Two estimation procedures are common in MLM: *maximum likelihood* (ML) and *restricted maximum likelihood* (REML). The restricted approach (REML) can lead to better estimates of variance components when fewer than 50 level-two units are included in analyses (Heck et al., 2010; Hox & Maas, 2002).

"excelling" was the reference category for AZ Learns. Teachers included in the present study were "mainstream" teachers responsible for a classroom but are not for special programs. Thus, we did not include teachers whose primary responsibilities included teaching English language learners or students in special education[7].

Acuity variables and AIMS scores were standardized within grades and subjects (i.e., converted to z-scores). Standardizing predictors and outcomes provides a measure of the relative impacts of variables explored in MLMs; in other words, standardizing variables allows researchers to compare the effects of various predictors because those factors have been converted to the same underlying scale[8] (Bloom, Hill, Black, & Lipsey, 2008; Heck, Thomas, & Tabata, 2010). Additionally, descriptive estimates revealed that the prevalence of teachers' Acuity use was more positively skewed than would be expected in normally distributed data, with a portion of frequent users creating a heavy tail in the upper end of the distribution[9]. If not addressed, this type of violation of the normality assumption can lead to biased standard error estimates at both levels of the data structure, which can affect hypothesis test outcomes for predictors in the models. Thus, we addressed the non-normality by applying a square root transformation to the prevalence variable prior to standardization (Garson, 2012; Raudenbush & Bryk, 2002). Appendix A provides more information on these issues.

In the following sections, we outline analyses and features of the analytic samples used to address each research question.

**Research Question 1: Which factors were associated with teachers' 2011 Acuity use?** To answer Research Question 1, we used MLM, with teachers nested within schools. Two dependent variables were of interest: prevalence of Acuity use in 2011 and consistency of Acuity use in 2011. These variables were modeled as a function of years of experience at the teacher level. The teacher-level intercept was allowed to vary randomly, and was modeled as a function of two school-level variables: AZ Learns status and type of school.

Following is an example of the specified models, for teacher *i* in school *j*:

Teachers (Level-1):

$$\text{Prevalence of 2011 Acuity Use}_{ij} = \beta_{0j} + \beta_{1j}(\text{Years of Experience})_{ij} + e_{ij}$$

Schools (Level-2):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{AZ Learns Status})_j + \gamma_{02}(\text{School Type})_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

***Analysis sample.*** Our analysis sample for this research question included only mainstream teachers for whom we were able to obtain complete data on all dependent and independent variables. These delimitations resulted in an analytic sample with 670 teachers from 65 schools, or 82% of the full sample of MPS teachers in our initial dataset. The delimited teacher sample (Table 12) was similar to the full sample (Table 4).

---

[7] The achievement of non-mainstream students is important; however, it carries its own set of correlates and Acuity is designed more for a mainstream environment. We recommend a separate study to examine the relationship between Acuity use and non-mainstream student achievement.

[8] Standardizing predictors in HLMs may reduce the variance components observed across different levels of the data hierarchy (Hox & Maas, 2002). This is of particular concern in HLMs where slopes are modeled as randomly-varying. Our models estimated slopes as fixed across level-2.

[9] This was true even after eliminating 27 teachers from the sample whose Acuity use was more than three standard deviations above average, or "outliers" according to conventional statistical wisdom.

**Research Question 2: Did 2011 teacher Acuity use correlate with 2011 student achievement?** We also used MLM for Research Question 2, with students nested within teachers. School was not included as a third level because we sought to replicate analyses from our second year report (Wayman et al., 2011). Four dependent variables were of interest: students' scale scores on AIMS reading and math tests, separated by elementary and junior high. For each dependent variable, a model was estimated that examined effects due to prevalence of Acuity use and another was estimated that examined effects due to consistency of Acuity use. In all, eight models were estimated.

To isolate changes in 2011 achievement, we controlled for students' prior achievement in 2010. The student-level intercept was allowed to vary randomly[10], modeled as a function of teacher experience and Acuity use. Gender, ethnicity, and economic status were included as covariates, with all student-level covariates fixed across level-2 units. Fixed teacher effects for Acuity use (prevalence or consistency) and teaching experience were included as covariates at the level 2.

Following is an example of the specified models, for student *i* taught by teacher *j*:

Students (Level-1):

$$\text{2011 AIMS Reading Score}_{ij} = \beta_{0j} + \beta_1(\text{Gender})_{ij} + \beta_2(\text{Economically Disadvantaged})_{ij} + \beta_{3j}(\text{Ethnicity})_{ij} + \beta_{4j}(\text{2010 AIMS Reading score})_{ij} + e_{ij}$$

Teachers (Level-2):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Years of Experience})_j + \gamma_{02}(\text{Acuity Use})_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$

***Analysis sample.*** Since our focus was on Acuity's impact on achievement improvements in 2011, we delimited our analytic sample for this research question. In doing so, we included only 2011 mainstream teachers with students in grades 4-8 for whom we were able to obtain valid Acuity use data and years of experience data. Third grade teachers were not included because their students did not take AIMS tests in 2010. Junior high teachers (grades 7 and 8) were linked to a student if they were noted as the student's primary reading or math teacher. Students in grades 4-8 were included if they had valid AIMS data both 2010 and 2011, and valid demographic data for 2011.

These delimitations resulted in an analytic sample of 13,556 students and 566 teachers, representing 67% of the full student sample and 69% of the full teacher sample (Table 13). In the student sample, minor differences were seen from the full MPS student sample: the delimited sample (Table 13) had slightly more economically advantaged and Non-Latino White students than did the full sample (Tables 2-3). Students in the delimited sample (Table 13) also had slightly higher scores on 2010 and 2011 AIMS tests than did the full sample (Table 4). The delimited teacher (Table 13) sample approximated the full sample (Table 4) in terms of teaching experience. Average Acuity use was slightly higher among teachers in the delimited sample

[10]We explored allowing the coefficient for Acuity use to vary randomly among teachers. For most models, there was not significant random variation of this relationship. For clarity, we modeled this effect as fixed.

(Table 13) than in the full teacher sample (Table 5). See Appendix A for more information on missing data for this research question.

**Research Question 3: Did teacher Acuity use over two consecutive years (2010 and 2011) correlate with 2011 student achievement?** To explore two-year effects of Acuity use on student achievement in 2011, we analyzed cross-classified models, nesting students within their 2010 and 2011 teachers (see Appendix A for more information on this choice of analysis). Unlike traditional balanced designs (e.g., repeated-measures ANOVA, hierarchical linear growth models), cross-classified models more accurately reflect the longitudinal data structures found in school settings – contexts where students are assigned to different combinations of teachers over time (Raudenbush & Bryk, 2002). Our models accounted for these complex nesting structures by partitioning the variance in 2011 achievement attributable to teachers into three unique components: achievement variability among students with the same 2010 teachers and different 2011 teachers, achievement variability among students with different 2010 teachers and the same 2011 teachers, and achievement variability among students with the same teachers both years (Heck et al., 2010; Raudenbush & Bryk, 2002). Thus, these models allowed us to estimate cumulative impact of both teachers' Acuity use. As with previous models, we controlled for prior achievement (2009 AIMS scores), student-level demographic factors, and teacher experience.

Intercepts were modeled as randomly varying at the student and teacher levels. Gender, ethnicity, economic status, and prior achievement were modeled as fixed across level-2 units. Fixed teacher effects for Acuity use in 2010 and 2011 were included at the teacher level, along with fixed effects for teaching experience among 2010 and 2011 teachers. We estimated eight models, with different combinations of school type (elementary or junior high), subject (reading or math), and Acuity use (prevalence or consistency). Following is an example of the specified models, for student *i* taught by teacher *j1* in 2010 and teacher *j2* in 2011:

Students (Level-1)[11]:

$$\text{2011 AIMS Reading Score}_{i(j1j2)} = \beta_{0i(j1j2)} + \beta_{1(j1j2)}(\text{Gender})_{i(j1j2)} + \beta_{2(j1j2)}(\text{Economically Disadvantaged})_{i(j1j2)} + \beta_{3(j1j2)}(\text{Ethnicity})_{i(j1j2)} + \beta_{4(j1j2)}(\text{2009 AIMS Reading score})_{i(j1j2)} + e_{\,i(j1j2)}$$

Teachers (Level-2):

$$\beta_{0(j1j2)} = \gamma_{00} + \gamma_{01}(\text{2011 Years of Experience})_{(j1j2)} + \gamma_{02}(\text{2010 Years of Experience})_{(j1j2)} + \gamma_{03}(\text{2011 Acuity Use})_{(j1j2)} + \gamma_{04}(\text{2011 Acuity Use})_{(j1j2)} + u_{j1} + u_{j2}$$

$$\beta_{1(j1j2)} = \gamma_{10}$$

$$\beta_{2(j1j2)} = \gamma_{20}$$

$$\beta_{3(j1j2)} = \gamma_{30}$$

$$\beta_{4(j1j2)} = \gamma_{40}$$

***Analysis sample.*** In addition to delimitations used to specify samples for Research Question 2, the cross-classified models for Research Question 3 required student and teacher data from 2010 and 2011, as well as achievement data from 2009. Since the focus of this question was on the impact of *consecutive* teachers' Acuity use on achievement in 2011, we delimited our analytic sample by mainstream teachers who taught in grades 5-8 in 2011 for whom we were able to obtain Acuity use and teaching experience data from 2010 and 2011. We

[11] For consistency across models, we use traditional factor notation at level-1 and level-2 (i.e., $\beta$ and $\gamma$), rather than notation typically used in cross-classified models (i.e., $\pi$ at level-1 and $\beta$ at level-2).

did not include third and fourth grade teachers because their students did not take AIMS tests in 2009 or 2010. Junior high teachers were linked to a student if they were noted as the student's primary reading or math teacher in 2010 or 2011. Students in grades 5-8 were included in analyses if we were able to obtain AIMS and demographic data for 2011 and 2009 AIMS data.

These delimitations resulted in a sample of 610 teachers: 321 from 2011 and 289 from 2010 (Table 14). The teacher sample represented 39% of the full 2011 teacher sample and 42% of the full 2010 teacher sample[12]. These delimitations also resulted in a sample of 5,581 students in grades 5-8; this represented 27% of the full student sample (see Table 14).

These delimitations required an exploration of possible sample bias. We do not believe these differences are substantial enough to skew our results – Appendix A provides a set of descriptive statistics for these samples, but a general description is given here. The delimited teacher samples for both years approximated the full samples in terms of teaching experience. Teachers' average Acuity use in both 2011 and 2010 was slightly higher among teachers in the sample versus teachers in the full MPS population. The delimited student sample was generally made up of a higher percentage of Non-Latino White and Economically Advantaged students than the full MPS student sample, with more pronounced differences in elementary grades. Students in the delimited sample generally scored higher on 2011 AIMS tests and lower on 2009 AIMS tests than those in the MPS population.

[12] The full 2010 MPS teacher sample was described in our second year report (Wayman et al., 2010).

## RESULTS

In this section, we present the results of our analyses. We first provide a section that describes the proportion of variability associated with level-2 units (i.e., teachers or schools), for each research question. Next, we provide sections that describe relationships between teachers' 2011 Acuity use (Research Question 1), relationships between 2011 Acuity use and student achievement in 2011 (Research Question 2), and relationship between 2011 student achievement and teacher Acuity use over consecutive years (Research Question 3).

### Proportion of Variability Associated with Level-2 Units

For each research question, null models were first estimated, followed by full models that included all variables of interest. Comparison of null and full models allowed estimation of variability due to level-2 units. In the following three sections, we describe the proportion of variability associated with level- units for each research question.

**School-level variability in 2011 teacher Acuity use (Research Question 1).** The proportion of variability in 2011 Acuity use attributable to teachers' school was significantly different than zero in null models for both prevalence and consistency of Acuity use (Table 15). Twenty-three percent of the variability in prevalence and 35% of the variability in consistency of Acuity use was attributable to schools.

The addition of explanatory variables at the teacher- and school-levels did not reduce the school-level variability in prevalence of use; that is, teaching experience and AZ Learns status did not explain variation between schools in the number of instructional actions teachers performed in Acuity. However, the predictors explained roughly six percentage points of school-level variability in consistency of Acuity use, reducing between-school variation in consistency from 35% to 29%[13] (Table 15). For both prevalence and consistency, statistically significant between-school variability remained after the addition of the explanatory variables, implying that a significant amount of variation in average Acuity use at the school-level was not explained by schools' state academic rating or teachers' years of experience.

**Teacher-level variability in 2011 achievement (Research Question 2).** The proportion of variability in 2011 student achievement attributable to teachers was significantly different than zero in all null models, with teacher-level variability ranging from 22% in elementary reading to 46% in junior high math (Table 16). The addition of explanatory variables at the student- and teacher-levels reduced the teacher level variability in each model, ranging from a two percent reduction in elementary math for prevalence and consistency models to 22% percent reduction in teacher-level variability in junior high reading for the consistency model (Table 16).

Statistically significant portions of teacher-level variability remained after adding explanatory variables in all models, except for those exploring junior high reading achievement (Table 16). That is, a significant portion of variation in average class-level performance in elementary reading and math, as well as junior high math, was unexplained by the student- and teacher-level factors.

For this research question, we sought to replicate the two-level models from our second year report (Wayman et al., 2011) and thus did not include school as a third level. Although school level variability was negligible in our previous report (Wayman, 2011), it is possible that

[13] In a previous report, we observed similar patterns, with explanatory variables explaining a larger percentage of school-level variability in consistency of Acuity use (Wayman et al., 2011).

the some of the remaining teacher level variability in the full models for math is due to schools. We discuss these, and other analytic decisions, in Appendix A.

**Two-year teacher-level variability in 2011 student achievement (Research Question 3).** The proportion of variability in 2011 achievement that was attributable to 2011 teachers was significant in each null model estimated, with larger amounts of achievement variability attributable to 2011 teachers than 2010 teachers (Table 17). For instance, 32% of the variability in 2011 junior high math achievement was due to 2011 teachers while less than five percent of the variability in elementary reading and math achievement was attributable to 2010 elementary teachers. The addition of explanatory variables at the student and teacher levels reduced the teacher level variability in each model, most notably a 16% reduction in 2011 junior high math.

In all models, variability in achievement attributable to 2010 teachers was no longer statistically significant after explanatory variables were added. The same was true for junior high math teachers in 2011, though statistically significant teacher-level variability remained after adding explanatory variables in other 2011 models (Table 17)[14].

## Research Question 1: Which Factors Were Associated with Teachers' 2011 Acuity Use?

In this section, we describe correlates of 2011 Acuity use separately for prevalence and consistency. MLM was used to determine associations, with prevalence and consistency of 2011 Acuity use entered as dependent variables. Independent variables included years of experience at the teacher level, and state rating (AZ Learns) and school type (ES/JHS) at the school level. Acuity use was standardized, so differences are discussed in terms of standard deviations.

**Prevalence of Acuity use.** Table 18 shows that teaching experience was a significant predictor of prevalence of Acuity use ($p = 0.01$). Teachers with the most experience (20 or more years) used Acuity less than their colleagues, with the largest difference between teachers with 20 plus years of experience and teachers with 6-10 years of experience (SD = 0.35). Neither AZ Learns status nor school type were significantly associated with prevalence of Acuity use.

**Consistency of Acuity Use.** As with prevalence of Acuity use, teaching experience ($p = 0.02$, see Table 19) was a significant predictor of teachers' consistency of Acuity use. Teachers with the most experience (20 or more years) averaged about a quarter of a standard deviation less than teachers with 6-10 years of experience. School type was also significantly associated with consistency of Acuity use, with elementary school teachers averaging 0.71 standard deviations more than junior high teachers ($p = 0.00$). State rating (AZ Learns) was also statistically significant ($p = 0.03$), with teachers in "Excelling" schools using Acuity more consistently than other teachers, particularly in schools rated as "Performing Plus" or "Highly Performing."

## Research Question 2: Did 2011 Teacher Acuity use Correlate with 2011 Student Achievement?

In this section, we explore the one-year (2011) effects of teacher Acuity use on student achievement growth. For this question, we fit models with students nested within 2011 teachers for both measures of Acuity use (prevalence and consistency), and controlled for prior year achievement. The models also removed effects due to teachers' experience, student demographic

---

[14] Some of the variance initially observed at the teacher-level would likely have been reduced with the addition of a third (i.e., school) level; that is, some variability in achievement outcomes is likely due to the school, rather than the teacher. However, we found that the small increases in the precision of variability estimates resulting from running three-level models (students nested within teachers nested within schools) was outweighed by the computational demands and unclear estimates of predictors entered into those models.

characteristics and prior achievement, thus isolating the unique effect of Acuity use on achievement. We estimated separate models for elementary and junior high students in reading and math. In the following sections, we present four sets of achievement models: elementary reading, elementary math, junior high reading, and junior high math. Acuity use and achievement measures were both standardized, so results are discussed in terms of standard deviation differences.

**Elementary reading.** Controlling for the effects of student- and teacher-level covariates, Table 20 shows that prevalence of Acuity use was marginally associated with elementary reading achievement ($p$ = 0.06). The regression coefficient for prevalence indicates that, when controlling for covariates, a one-standard deviation increase in the square root of total Acuity uses was associated with a 0.02 standard-deviation increase in AIMS elementary reading scale scores (about 0.8 scale points).

Controlling for the effects of student- and teacher-level covariates, Table 21 shows that consistency of Acuity use was significantly associated with elementary reading achievement ($p$ < 0.01). The regression coefficient for consistency indicates that, when controlling for covariates, a one-standard deviation increase in weeks of use was associated with a 0.03 standard-deviation increase in AIMS elementary reading scale scores (about 1.2 scale points).

**Elementary math.** Controlling for the effects of student- and teacher-level covariates, Table 22 shows that prevalence of Acuity use was significantly associated with elementary math achievement ($p$ = 0.05). The regression coefficient for prevalence indicates that, when controlling for covariates, a one-standard deviation increase in the square root of total Acuity uses was associated with a 0.03 standard-deviation increase in AIMS elementary math scale score (about 1.4 scale points).

Controlling for the effects of student- and teacher-level covariates, Table 23 shows that consistency of Acuity use was marginally associated with elementary math achievement ($p$ = 0.09). The regression coefficient for consistency indicates that a one-standard deviation increase in weeks of use was associated with a 0.03 standard-deviation increase in AIMS math scale scores (about 1.4 scale points).

**Junior high reading.** Controlling for the effects of student- and teacher-level covariates, neither prevalence ($p$ = 0.71, Table 24) nor consistency of Acuity use ($p$ = 0.66, Table 25) was significantly associated with junior high reading achievement.

**Junior high math.** Controlling for the effects of student- and teacher-level covariates, neither prevalence ($p$ = 0.59, Table 26) nor consistency of Acuity use ($p$ = 0.92, Table 27) was significantly associated with junior high math achievement.

### Research Question 3: Did Teacher Acuity Use Over Two Consecutive Years (2010 and 2011) Correlate with 2011 Student Achievement?

In this section, we explore the effects of longitudinal Acuity use on student achievement growth. For this question, we again nested students within teachers, but we accounted for 2010 teacher Acuity use and 2011 teacher Acuity use through cross-classified MLMs. In doing so, we controlled for prior (2009) achievement. Beyond the use of cross-classified models, analysis proceeded similar to research question two: We controlled for student and teacher demographic factors, we estimated separate models for elementary and junior high students (for both reading and math), and we examined teacher Acuity use in terms of prevalence and consistency.

In the following sections, we describe four sets of achievement models: elementary reading, elementary math, junior high reading, and junior high math. Acuity measures and achievement measures were both standardized, so results are discussed in terms of standard deviation differences.

**Elementary reading.** Controlling for the effects of student- and teacher-level covariates, Table 28 shows that prevalence of teachers' Acuity use in both 2011 and 2010 was not significantly associated with elementary reading achievement ($p$ = 0.13 and $p$ = 0.71, respectively). When controlling for covariates, a one-standard deviation increase in the square root of total Acuity uses per year was associated with a 0.01 standard-deviation increase in 2011 AIMS elementary reading scale scores (about 0.4 scale points).

Controlling for the effects of student- and teacher-level covariates, Table 29 shows that consistency of Acuity use in 2011 was significantly associated with elementary reading achievement ($p$ = 0.01), but consistency of Acuity use in 2010 was not ($p$ = 0.19). The additive effect of consistency in 2010 and 2011 was 0.03 standard deviations. Thus, when controlling for covariates, a one-standard deviation increase in total weeks of Acuity use per year was associated with a 0.03 standard-deviation increase in AIMS elementary reading scale score (about 1.1 scale points).

**Elementary math.** Controlling for the effects of student- and teacher-level covariates, Table 30 shows that prevalence of Acuity use was not significantly associated with elementary math achievement for either the student's 2010 or 2011 teachers ($p$ = 0.79 and $p$ = 0.49, respectively). Similarly, Table 31 shows that consistency was not significant for either year ($p$ = 0.19 and $p$ = 0.26, respectively).

**Junior high reading.** Controlling for the effects of student- and teacher-level covariates, Table 32 shows that prevalence of Acuity use was not significantly associated with junior high reading achievement for either the student's 2010 or 2011 teachers ($p$ = 0.96 and $p$ = 0.75, respectively). Similarly, Table 33 shows that consistency was not significant for either year ($p$ = 0.62 and $p$ = 0.34, respectively).

**Junior high math.** Controlling for the effects of student- and teacher-level covariates, Table 34 shows that prevalence of Acuity use was not significantly associated with junior high math achievement for either the student's 2010 or 2011 teachers ($p$ = 0.33 and $p$ = 0.65, respectively). Similarly, Table 35 shows that consistency was not significant for either year ($p$ = 0.38 and $p$ = 0.68, respectively).

## DISCUSSION

Over the past three years, our research team has explored the Acuity implementation in Mesa Public Schools (Wayman et al., 2009; Wayman et al, 2011; the present study). These studies have helped us provide a description of how Acuity is used; they also have allowed us to explore the relationship between Acuity use and student achievement. The present study represents a step beyond the first two studies in that it enabled us to examine, over consecutive years, the relationship between Acuity use and student achievement.

In this section, we interpret findings from the present study in light of previous research and our previous studies. In doing so, we first offer a section that summarizes Acuity use over the three years of this project. This is followed by a section discussing implications and next steps for MPS. Finally, we offer a brief section discussing the effects of Acuity use at the junior high level.

### Three Years of Acuity Use in Mesa Public Schools

Across the three years of this project, we studied Acuity use[15] in two primary ways: (1) describing the nature and correlates of teacher Acuity use, and (2) exploring the relationship between teacher Acuity use and student achievement. In the following narrative, we provide a section for each.

**Descriptions of teacher Acuity use.** Teachers' use of Acuity increased dramatically between the first and second years of full implementation (Wayman et al., 2011). The present study shows these increases were maintained – but did not increase – during the third year. Teacher use typically clustered in the weeks following Acuity assessments, but the timing of this use changed during the three years. In 2009, teachers used Acuity more following Form B, with the least use following Form C and preceding AIMS (Wayman et al., 2009). Data from the present study showed that teachers used Acuity considerably more following Form C (and preceding AIMS), with Form A drawing the least attention.

Across the three years of the study, Acuity use was consistently and significantly higher in elementary schools than junior high schools. Teaching experience may matter slightly, as experience was significantly associated with Acuity use in the present study. However, this variable was not significant in either of our prior studies (Wayman et al., 2009; Wayman et al., 2011).

Overall, we believe observed changes in Acuity use may be largely due to intensified district-level support and training on Acuity. Data from the first two years of this project show increases in both the amount and breadth of Acuity training (Wayman et al., 2009; Wayman et al., 2011) and our informal discussions with MPS staff during the current study suggest this trend is continuing.

**Relationship between Acuity use and student achievement.** In our second-year study (Wayman et al., 2011) and in the present study, we demonstrated one-year, "snapshot" relationships between teacher Acuity use and student achievement. We found positive, significant associations between Acuity use and elementary reading and math achievement in both years, but were not sure these relationships were educationally meaningful. In neither year were we able to demonstrate one-year effects in junior high school.

---

[15] We studied Acuity use in terms of prevalence and consistency. While there were occasional differences between these types of use in our statistical models, we find it clearer and equally meaningful to refer to Acuity use generally in interpreting these results.

The present study was unique in that we considered the effects of consecutive years of teachers' Acuity use. However, results from our cross-classified models did not demonstrate significant effects of two years of Acuity use on reading and math achievement.

**Explaining the Relationship between Acuity Use and Student Achievement**

Over the three years of our study, we have been unable to demonstrate broad, clear effects of Acuity use on student achievement. While we do not believe these results are evidence that Acuity is ineffective, the fact that these results are not stronger does warrant further exploration. In this section, we explore two possible interpretations of these results: (1) that Acuity effects are manifested in different ways than total use, and (2) Acuity use should be more tightly coupled to practice.

**Acuity effects are manifested in different ways than total use.** In this project, we measured Acuity effects in terms of total use over the course of a school year, measuring prevalence (total uses during the year) and consistency (total weeks used during the year). Use of these measures carries an implicit hypothesis that more uses and more frequent use can lead to improved student achievement – and that it can be seen in two school years (the duration of our study). However, it is possible that this implicit hypothesis is flawed. A number of alternative hypotheses are possible.

First, it is possible that Acuity effects take more than two years to appear. The present study showed negligible effects after two years, but prior studies have described the difficulties that districts encounter in adopting a data-based initiative such as Acuity (Wayman et al., 2009; Wayman et al., 2011; Spradlin, 2012). Thus, it is possible that it takes more than two years to embed Acuity into school culture, make changes in educational practice, and see these changes manifest in improved achievement.

Second, it is possible that Acuity effects will accumulate over time and result in significant, impactful effects over the course of a student's career rather than just over two years. Bloom et al. (2008) noted that if students make small yearly gains, these gains might prove significant as students advance. Most of the effects in the present study were smaller than the ones Bloom et al. (2008) used to make this point (e.g., 0.1 standard deviations vs. our effects of 0.05 and smaller), but when paired with other effective initiatives, it is possible that even the small effects found in the present study might prove significant over the duration of a student's education.

Third, it is possible that merely measuring system use will not demonstrate achievement effects. Thus, future research might consider ways to measure practice along with system use. For instance, it is possible that teachers who use Acuity alone might realize different achievement effects than those who use it collaboratively. Or, perhaps measures could be constructed that demonstrate different classroom actions taken after similar Acuity uses.

Finally, it is possible that the small effects found here are somehow unique to the MPS context. Konstantopoulos et al. (2011) found larger effects of Acuity use on student achievement in Indiana, so perhaps there are contextual factors at work in either context that caused differential results.

**Acuity use should be more tightly coupled with practice.** Goertz, Olah, & Riggan (2010) surmised that many data initiatives are successful in increasing levels of teacher data use, but it is more difficult to attain associated increases in student achievement. While the authors were speaking of data use more generally than our present focus on data system use, the point applies here: our data suggest that MPS has been successful in increasing Acuity use but has been less successful in realizing associated student achievement gains.

With this in mind, one potential area for growth is more clearly pairing Acuity use with practice-based solutions. In this section, we offer examples of the sorts of solutions that research suggests MPS or other districts could explore to realize student achievement gains built on increased Acuity use.

First, we believe it is necessary for vendors and districts to focus training opportunities on helping teachers make decisions and change practice based on data. This is in contrast with research demonstrating that training is often focused on manipulating the data system (Jimerson & Wayman, 2011; Means et al., 2010). In fact, we advocate little or no training specific to a data system. Instead, we advocate training that helps teachers link specific data analyses to specific pedagogical solutions. In doing so, these sessions can expose teachers to system functions that will help them access data they need, but in focusing on practice rather than the system, will help them build a repertoire of instructional solutions related to data.

Second, it is possible that vendors and districts might look for ways to better understand the sense that teachers make of their systems. Cho & Wayman (in press) studied teacher sensemaking and how it related to data system use, concluding:

> Some district leaders or vendors might be surprised when a powerful, easy to use tool fails to make the impact on teachers' data use that they had envisioned. Whereas a natural response might be to doubt the tool or its design, a better response might be to investigate what sense teachers have made of those artifacts.

In fact, we observed this phenomenon at work during the first full year of Acuity implementation (Wayman et al., 2009). We observed teachers interpreting terminology on the Acuity site differently than Acuity designers or MPS leaders had intended. This led to some confusion and frustration in interpreting data.

Third, districts might attend to the district-based messages that are sent regarding data systems. Recent research is starting to describe the impact that district-based actions and statements hold for data use (Cho & Wayman, in press; Honig & Venkateswaran, 2012). This is particularly important in light of the prior two recommendations: these studies suggest that districts might considerably influence the ways systems are used by implementing activities and supporting messages that promote system use in terms of practice and sensemaking.

Fourth, districts could promote collaboration around Acuity. Collaboration is prominently forwarded in research as effective method for helping teachers construct practice-based responses to the information they get from data and data systems (Anderson, Leithwood, & Strauss, 2010; Datnow, Park, & Wohlstetter, 2007; Lachat & Smith, 2005; Wayman, Cho, Jimerson, & Spikes, 2012; Wayman & Stringfield, 2006; Young, 2006). Mesa Public Schools has experimented with activities that promote collaboration (e.g., the "principal homework assignment," Wayman et al., 2011), but it is unclear how prevalent activities such as this are in current MPS data use.

Finally, it is possible that greater achievement gains will be seen by focusing on practice through programmatic issues rather than at the individual teacher level. Slavin et al. (2011) attributed the gains seen in their study to selection of programs based on assessment results. The authors contended, "it is these programs that lead to achievement gains, not the consultation or benchmarks in themselves" (Slavin et al., 2011, p.22). Certainly, other studies are necessary to bear this finding out, but it is at least an avenue for districts to consider.

**Acuity Effects in Junior High**

In both the present study and our previous report (Wayman et al., 2011), Acuity use was shown to be largely unrelated to junior high (i.e., 7th and 8th grade) student achievement. While these trends are discouraging, other studies have also failed to find significant impacts of benchmark assessment initiatives on achievement during in the junior high grades (Henderson, et. al., 2007; Konstantopoulos et al., in press; Slavin et al., 2011).

The reasons for this are yet unclear and a variety of explanations are possible. For instance, Slavin et al. (2011) offered the explanation that middle school educators in their study failed to use benchmark assessment data to select high-quality instructional interventions. As another example, we noted in our first report (Wayman et al., 2009) that junior high teachers cited large class sizes as a barrier for finding time to use Acuity. It is also possible that junior high teachers focus their Acuity use on a much narrower group of students than do elementary teachers: in earlier reports, we found that some MPS junior high teachers described their use of Acuity as restricted to situations where they provided one-on-one tutoring for students who were behind grade-level in reading or math (Wayman et al., 2009; Wayman et al., 2011). Or, perhaps junior high teachers have less unique effect on individual student learning, given that they are only one of many teachers a student may learn from in a given year.

Given the lack of clarity about junior high effects, we recommend that MPS staff explore junior high Acuity use. Such an exploration would involve district officials (or outside researchers) visiting junior high schools, talking to teachers, administrators, and staff about how Acuity is used, and listening to their opinions about its utility. Armed with this knowledge, we are hopeful that MPS officials can make effective recommendations to help junior high teachers make better use of this vast resource.

## CONCLUSION

Acuity is a powerful data system that offers teachers access to voluminous tools and data that help them know more about their students. To date, MPS personnel have done well in increasing use of the system. However, it may be time to move to a new phase in Acuity support. Should MPS move in this direction, it likely will require increased resources for supporting Acuity. We believe this will be money well spent, because this support focuses on issues of practice and educational improvement. As Mike Flicek (2010) once pointed out, most good practices around the use of data are “just good education.” Accordingly, we believe the recommendations in this report should serve MPS educators for years to come.

## REFERENCES


Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9*(3), 292-327.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1, 289-328.*

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement (2011). *Educational Evaluation and Policy Analysis*, *33*(3), 378–398.

Census Bureau (2011). Mesa, Arizona. US Census Bureau. Retrieved January 14, 2012, http://quickfacts.census.gov/qfd/states/04/0446000.html.

Cho, V. & Wayman, J. C (in press). Districts' efforts for data use and computer data systems: The role of sensemaking in system use and implementation. *Teachers College Record.*

Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis, 25*, 375–395.

CTB/McGraw-Hill. (2009). *Acuity Formative Assessments from CTB/McGraw-Hill*. Retrieved November 28, 2010, from http://www.acuityforschool.com.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students.* Los Angeles: University of Southern California, Rossier School of Education, Center on Educational Governance.

Flicek, M. (2010, May). Discussant remarks for the symposium *The Data-Informed District.* 2010 Annual Meeting of the American Educational Research Association, Denver CO.

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal 38*(4), 915-945.

Garson, G. D. (2012). *Testing Statistical Assumptions (2012 Edition)*. Statistical Associates Publishing: Asheboro, NC.

Goertz, M. E., Olah, L. N., Riggan, M. (2010) *From testing to teaching: The use of interim assessments in classroom instruction.* Philadelphia PA: The Consortium for Policy Research in Education.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and Longitudinal Modeling with IBM SPSS.* New York, London: Routledge Academic.

Hox, J. J. & Maas, C. J. M. (2002). Sample sizes for multilevel modeling. In J. Blasius, J. J. Hox, E. De Leeuw, & P. Schmidt (Eds.), *Social Science Methodology in the New Millennium: Proceedings of the Fifth International Conference on Logic and Methodology* (2nd expanded Ed.).

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007 No. 039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Honig, M. I., & Venkateswaran, N. (2012). School–central office relationships in evidence use: Understanding evidence use as a systems problem. *American Journal of Education*, *118*(2), 199–222.

Jimerson, J. B. & Wayman, J. C. (2011, November). *Data-related professional learning: What teachers need vs. what teachers get.* Paper presented at the 2011 Annual Convention of the University Council for Educational Administration, Pittsburgh PA.

Konstantopoulos, S., Miller, S., van der Ploeg, A., Li, C. H., & Traynor, A. (in press). The impact of Indiana's system of benchmark assessments on mathematics achievement. *Educational Evaluation and Policy Analysis*.

Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed At Risk, 10*(3), 333–349.

May, H., & Robinson, M.A. (2007). *A randomized evaluation of Ohio's Personalized Assessment Reporting System (PARS).* Madison, WI: Consortium for Policy Research in Education.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2010). *Implementing data-informed decision-making in schools—Teacher access, supports and use.* Washington, D.C.: U.S. Department of Education Office of Planning, Evaluation and Policy Development. Retrieved February 27, 2012 from www.ed.gov/about/offices/list/opepd/ppss/reports.html.

Mesa Public Schools. (2009, January). Acuity in Mesa Public Schools. MPS educational technology and training. Retrieved January 21, 2009, from http://www2.mpsaz.org/edtech/acuity_turnleaf.

Quint, J., Sepanik, S., & Smith, J. (2008). *Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) program in Boston elementary schools.* New York: MDRC.

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

Slavin, R. E., Holmes, G., Madden, N. A., Chamberlain, A., Cheung, A., & Borman, G. D. (2010). *Effects of a data-driven district-level reform model,* Working Paper. Baltimore, MD: Center for Data-Driven Reform, Johns Hopkins University.

Snijders. T., & Bosker, R. (1999) *Multilevel Analysis.* London: Sage.

Spradlin, T. E., Dickinson, S. L., Dingjing, L., Shi, D., & Whiteman, R. S. (2012). *Acuity Assessment Program Indiana Efficacy Study: 2008-2011 Data Analysis (Final Edition)*. Bloomington, IN: Center for Evaluation & Education Policy.

Wayman, J. C., Cho, V., Jimerson, J. B., & Spikes, D. D. (2012). District-wide effects on data use in the classroom. *Education Policy Analysis Archives, (20)* 25. Retrieved November 26, 2012 from http://epaa.asu.edu/ojs/article/view/979.

Wayman, J. C., Cho, V., & Shaw, S. M. (2009). *First-year results from an efficacy study of the Acuity data system.* Austin: The University of Texas.

Wayman, J. C. & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education 112*(4), 549-571.

Wayman, J. C., Shaw, S. M. & Cho, V., & (2011). *Second-year results from an efficacy study of the Acuity data system.* Austin: The University of Texas.

Young, V.M. (2006) Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education 112*(4), 521-548.

Figure 1. Acuity Use by Week, 2011 Academic Year, Instructional Acuity Use Functions

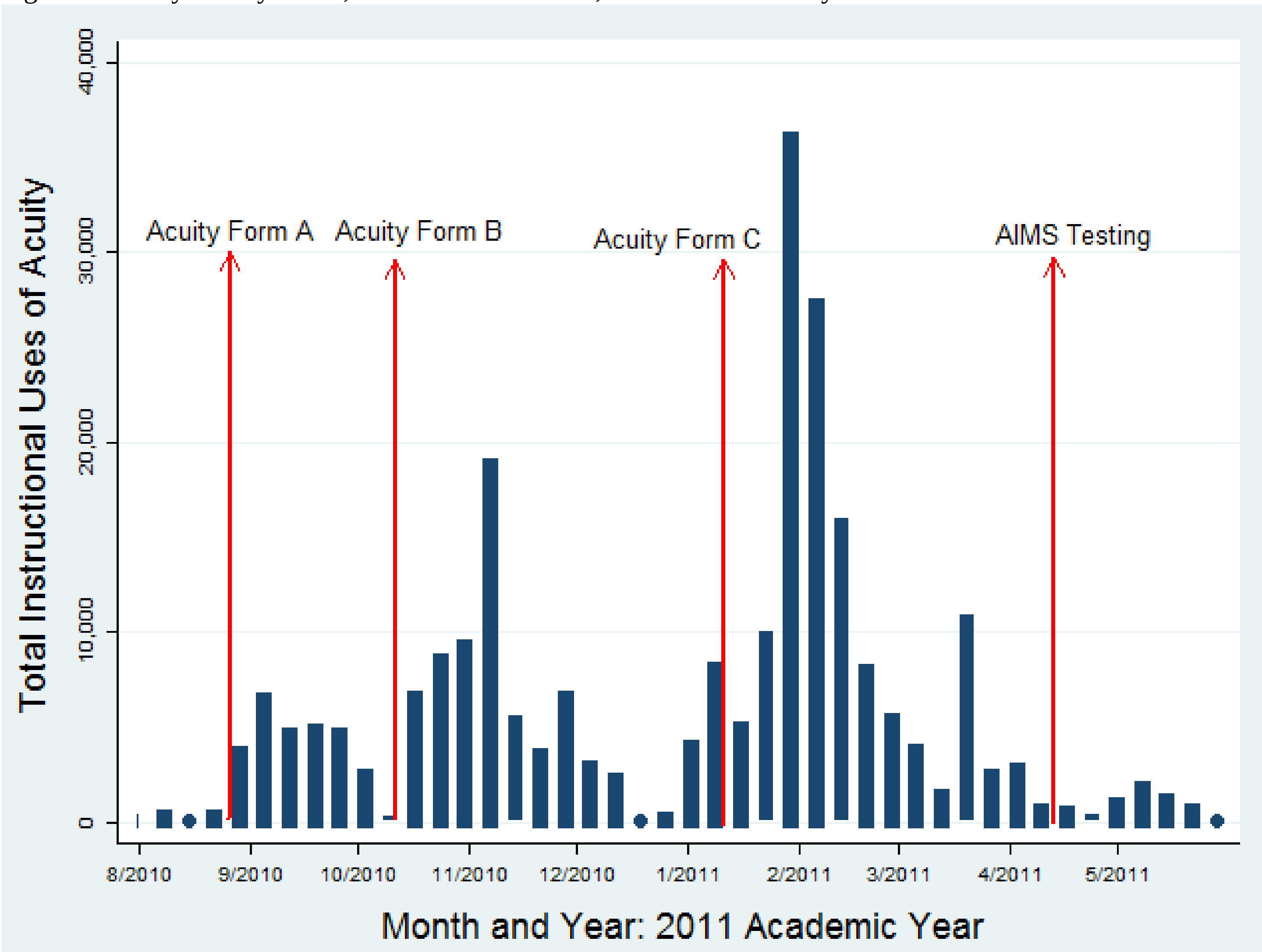

Table 1

*Number and Percent of Students in 2011, by Gender and School Level*

| Gender | ES *n (%)* | JHS *n (%)* | Total *n* (%) |
|---|---|---|---|
| Female | 7063 (49%) | 4836 (49%) | 11899 (49%) |
| Male | 7371 (51%) | 4949 (51%) | 12320 (51%) |
| Total | 14434 | 9785 | 24219 |

Table 2

*Number and Percent of Students in 2011, by Ethnicity and School Level*

| Ethnicity | ES *n (%)* | JHS *n (%)* | Total *n* (%) |
|---|---|---|---|
| Non-Latino White | 6746 (47%) | 4897 (50%) | 11643 (48%) |
| Latino | 5913 (41%) | 3680 (38%) | 9598 (40%) |
| Other | 1775 (12%) | 1208 (12%) | 2983 (12%) |
| Total | 14434 | 9785 | 24219 |

Table 3

*Number and Percent of Students in 2011, by Economic Status and School Level*

| Economic Status | ES *n (%)* | JHS *n (%)* | Total *n* (%) |
|---|---|---|---|
| Economically Advantaged | 5392 (37%) | 4098 (42%) | 9490 (39%) |
| Economically Disadvantaged | 9042 (63%) | 5687 (58%) | 14729 (61%) |
| Total | 14434 | 9785 | 24219 |

Table 4

*Number and Percent of Teachers in 2011, by Teaching Experience and School Level*

| Experience Category | ES *n (%)* | JHS *n (%)* | Total *n* (%) |
|---|---|---|---|
| 0 - 5 years | 110 (18%) | 39 (20%) | 149 (18%) |
| 6 - 10 years | 138 (22%) | 51 (27%) | 189 (23%) |
| 11 - 19 years | 187 (30%) | 60 (31%) | 247 (30%) |
| 20 or more years | 190 (30%) | 41 (22%) | 231 (28%) |
| Total | 625 | 191 | 816 |

Table 5

*Number and Percent of Schools and Teachers in 2011, by State Achievement Classification (AZ Learns)*

| AZ Learns Classification | Schools *n (%)* | Teachers *n (%)* |
|---|---|---|
| Performing | 3 (5%) | 41 (5%) |
| Performing Plus | 33 (51%) | 359 (44%) |
| Highly Performing | 16 (25%) | 228 (28%) |
| Excelling | 13 (20%) | 188 (23%) |
| Total | 65 | 816 |

Table 6

*Number and Percent of Teachers Who Used Acuity, 2009 to 2011*

| | 2009 *n (%)* | 2010 *n (%)* | 2011 *n (%)* |
|---|---|---|---|
| No | 230 (29.6%) | 52 (7.5%) | 68 (9.2%) |
| Yes | 546 (70.4%) | 643 (92.5%) | 670 (90.8%) |
| Total | 776 | 695 | 738 |

Table 7

*Average Teacher Acuity Use, 2010 and 2011*

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| Prevalence of Use | | | | |
| 2009 | 39.7 | 46.4 | 1 | 565 |
| 2010 | 127.5 | 150.9 | 1 | 738 |
| 2011 | 159.3 | 107.0 | 1 | 748 |
| Consistency of Use | | | | |
| 2009 | 4.9 | 3.8 | 1 | 24 |
| 2010 | 10.1 | 5.7 | 1 | 30 |
| 2011 | 8.8 | 5.8 | 1 | 29 |

*n* (2009) = 546; *n* (2010) = 643; *n*(2011) = 670

*Note*. Prevalence of Acuity Use is the number of instructional actions performed.

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used.

Table 8
*Average Student Achievement in AIMS Reading Tests, 2009 to 2011*

| | | 2009 | 2010 | 2011 |
|---|---|---|---|---|
| Elementary | Mean (*SD*) | 492.81 (43.14) | 508.15 (42.24) | 508.81 (43.79) |
| | *N* | 9410 | 9954 | 11941 |
| 4th grade | Mean (*SD*) | 477.07 (45.04) | 494.96 (46.78) | 490.08 (46.70) |
| | *N* | 2990 | 3183 | 3931 |
| 5th grade | Mean (*SD*) | 491.16 (43.06) | 507.91 (38.62) | 510.46 (39.67) |
| | *N* | 3175 | 3347 | 3977 |
| 6th grade | Mean (*SD*) | 507.2 (41.12) | 522.1 (36.32) | 525.42 (37.13) |
| | *N* | 3245 | 3424 | 4033 |
| Junior High | Mean (*SD*) | 528.69 (43.72) | 545.82 (48.30) | 537.43 (45.02) |
| | *N* | 8745 | 9306 | 8391 |
| 7th grade | Mean (*SD*) | 521.15 (42.69) | 540.59 (43.69) | 538.92 (41.67) |
| | *N* | 4056 | 4410 | 4129 |
| 8th grade | Mean (*SD*) | 534.77 (44.60) | 547.3 (48.94) | 535.98 (48.01) |
| | *N* | 4689 | 4896 | 4262 |

*n* (2009) = 18155; *n* (2010) = 19260; *n*(2011) = 20332

Table 9
*Average Student Achievement in AIMS Math Tests, 2009 to 2011*

| Grade | | 2009 | 2010 | 2011 |
|---|---|---|---|---|
| Elementary | Mean (*SD*) | 501.49 (50.33) | 406.50 (45.62) | 409.98 (49.83) |
| | *N* | 9411 | 9953 | 11941 |
| 4th grade | Mean (*SD*) | 467.03 (46.34) | 393. (41.53) | 393.79 (51.33) |
| | *N* | 2989 | 3182 | 3931 |
| 5th grade | Mean (*SD*) | 505.83 (50.58) | 402.94 (44.04) | 406.19 (48.61) |
| | *N* | 3177 | 3346 | 3977 |
| 6th grade | Mean (*SD*) | 524.29 (51.38) | 423.92 (45.20) | 429.49 (51.07) |
| | *N* | 3245 | 3425 | 4033 |
| Junior High | Mean (*SD*) | 568.23 (57.3) | 452.17 (47.32) | 442.70 (46.89) |
| | *N* | | 9304 | 8389 |
| 7th grade | Mean (*SD*) | 553.47 (57.19) | 444.22 (51.17) | 439.69 (51.29) |
| | *N* | 4057 | 4408 | 4128 |
| 8th grade | Mean (*SD*) | 581.73 (54.87) | 455.14 (44.16) | 445.61 (46.57) |
| | *N* | 4688 | 4896 | 4261 |

*n* (2009) = 18156; *n* (2010) = 19257; *n*(2011) = 20330

Table 10

*MPS versus Statewide Scale Score Increases in AIMS Reading, 2009 to 2011*

| | MPS | | Statewide (AZ) | |
|---|---|---|---|---|
| Grades | 2009 to 2010 | 2010 to 2011 | 2009 to 2010 | 2010 to 2011 |
| $4^{th}$ to $5^{th}$ | 31 | 16 | 19 | 17 |
| $5^{th}$ to $6^{th}$ | 31 | 18 | 13 | 17 |
| $6^{th}$ to $7^{th}$ | 33 | 17 | 24 | 14 |
| $7^{th}$ to $8^{th}$ | 26 | -5 | 12 | -4 |

*Note*. AIMS scale scores are vertically equated. Statewide (AZ) AIMS scores were obtained from technical reports (2009-2011) posted on the Arizona Department of Education website (http://www.azed.gov).

Table 11

*MPS versus Statewide Scale Score Increases in AIMS Math, 2010 to 2011*

| Grades | MPS | Statewide (AZ) |
|---|---|---|
| $4^{th}$ to $5^{th}$ | 13 | 15 |
| $5^{th}$ to $6^{th}$ | 27 | 25 |
| $6^{th}$ to $7^{th}$ | 16 | 8 |
| $7^{th}$ to $8^{th}$ | 1 | 13 |

*Note*. AIMS scale scores are vertically equated. Statewide (AZ) AIMS scores were obtained from technical reports (2009-2011) posted on the Arizona Department of Education website (http://www.azed.gov).

Table 12

*Descriptive Statistics for Teacher Analytic Sample, Research Question One*

| Covariates | *n* | % |
|---|---|---|
| Years of Experience | | |
| 0 to 5 years | 131 | 19.6 |
| 6 to 10 years | 162 | 24.2 |
| 11 to 19 years | 205 | 30.6 |
| 20+ years | 199 | 29.7 |
| Elementary School | 525 | 78.4 |
| 2011 AZ Learns status | | |
| Performing | 32 | 4.8 |
| Performing Plus | 323 | 48.2 |
| Highly Performing | 186 | 27.8 |
| Excelling | 156 | 23.3 |
| | Acuity Use (Mean / SD) | |
| Prevalence of Acuity Use | 159.3 | 150.9 |
| Consistency of Acuity Use | 8.8 | 5.8 |

*Note*. *n*(teachers) = 670; *n*(schools) = 65

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

Table 13

*Descriptive Statistics for Student- and Teacher-Level Analytic Sample, Research Question Two*

| Student-Level Variables (*n* / %) | Elementary (*n* = 8312) | | JHS Reading (*n* = 5244) | | JHS Math (*n* = 5020) | |
|---|---|---|---|---|---|---|
| Gender | | | | | | |
| Female | 4327 | 52.1% | 2667 | 50.9% | 2558 | 51.0% |
| Male | 3985 | 47.9% | 2557 | 48.8% | 2462 | 49.0% |
| Ethnicity | | | | | | |
| Non-Latino White | 4339 | 52.2% | 2905 | 55.4% | 2801 | 55.8% |
| Latino | 3033 | 36.5% | 1761 | 33.6% | 1671 | 33.3% |
| Other | 940 | 11.3% | 578 | 11.0% | 548 | 10.9% |
| Economically Disadvantaged | 4585 | 55.2% | 2666 | 50.8% | 2512 | 50.0% |
| | AIMS Achievement (Mean / SD) | | | | | |
| Prior Reading Score (2010) | 494.4 | 43.6 | 546.4 | 42.1 | -- | -- |
| Prior Math Score (2010) | 393.7 | 43.9 | -- | -- | 444.9 | 45.4 |
| Reading Score (2011) | 512.2 | 40.2 | 540.5 | 44.1 | -- | -- |
| Math Score (2011) | 416.7 | 47.6 | -- | -- | 453.6 | 41.8 |
| Teacher-Level Variables (*n* / %) | Elementary (*n* = 423) | | JHS Reading (*n* = 60) | | JHS Math (*n* = 83) | |
| Years of Experience | | | | | | |
| 0 to 5 years | 75 | 18% | 19 | 32% | 12 | 15% |
| 6 to 10 years | 92 | 22% | 14 | 23% | 27 | 33% |
| 11 to 19 years | 124 | 29% | 19 | 32% | 26 | 31% |
| 20+ years | 132 | 31% | 8 | 13% | 18 | 22% |
| | Acuity Use Variables (Mean / SD) | | | | | |
| Prevalence of Acuity Use | 170.9 | 179.2 | 102.4 | 102.4 | 138.8 | 138.8 |
| Consistency of Acuity Use | 9.9 | 6.0 | 4.0 | 4.0 | 7.6 | 7.6 |

*Note*. Cells with '--' are empty because variable(s) were not included in respective analyses.

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

Table 14
*Descriptive Statistics for Student- and Teacher-Level Analytic Sample, Research Question Three*

| Student-Level Variables (*n* / %) | Elementary (*n* = 3360) | | JHS Reading (*n* = 2221) | | JHS Math (*n* = 2070) | |
|---|---|---|---|---|---|---|
| Gender | | | | | | |
| Female | 1726 | 51.4% | 1160 | 52.2% | 1085 | 52.4% |
| Male | 1634 | 48.6% | 1061 | 47.8% | 985 | 47.6% |
| Ethnicity | | | | | | |
| Non-Latino White | 1935 | 57.6% | 1283 | 57.8% | 1161 | 56.1% |
| Latino | 1097 | 32.6% | 725 | 32.6% | 707 | 34.2% |
| Other | 328 | 9.8% | 213 | 9.6% | 202 | 9.8% |
| Economically Disadvantaged | 1672 | 49.8% | 1061 | 47.8% | 1081 | 52.2% |
| | AIMS Achievement (Mean / SD) | | | | | |
| Prior Reading Score (2009) | 488.8 | 42.6 | 527.1 | 40.4 | -- | -- |
| Prior Math Score (2009) | 492.6 | 52.7 | -- | -- | 555.3 | 52.1 |
| Reading Score (2011) | 525.4 | 35.7 | 550.3 | 46.7 | -- | -- |
| Math Score (2011) | 426.4 | 46.3 | -- | -- | 454.5 | 42.2 |

| Teacher-Level Variables (*n* / %) | Elementary (*n* = 489) | | JHS Reading (*n* = 48) | | JHS Math (*n* = 73) | |
|---|---|---|---|---|---|---|
| Years of Experience (2010) | *n*(2011) = 250 | | *n*(2011) = 29 | | *n*(2011) = 42 | |
| 0 to 5 years | 43 | 17.2% | 13 | 44.8% | 6 | 14.3% |
| 6 to 10 years | 50 | 20.0% | 5 | 17.2% | 13 | 31.0% |
| 11 to 19 years | 80 | 32.0% | 8 | 27.6% | 15 | 35.7% |
| 20+ years | 78 | 31.2% | 3 | 10.3% | 8 | 19.0% |
| Years of Experience (2011) | *n*(2011) = 239 | | *n*(2011) = 24 | | *n*(2011) = 26 | |
| 0 to 5 years | 38 | 15.9% | 4 | 16.7% | 6 | 23.1% |
| 6 to 10 years | 45 | 18.8% | 6 | 25.0% | 2 | 7.7% |
| 11 to 19 years | 72 | 30.1% | 11 | 45.8% | 7 | 26.9% |
| 20+ years | 84 | 35.1% | 4 | 16.7% | 11 | 42.3% |

*(Table 14 cont'd)*

| | Acuity Use Variables (Mean / SD) | | | | | |
|---|---|---|---|---|---|---|
| | Elementary (*n* = 489) | | JHS Reading (*n* = 48) | | JHS Math (*n* = 73) | |
| Prevalence of Acuity Use (2010) | 132.46 | 109.26 | 81.84 | 89.86 | 135.91 | 115.63 |
| Prevalence of Acuity Use (2011) | 162.58 | 164.39 | 82.63 | 78.42 | 136.98 | 133.94 |
| Consistency of Acuity Use (2010) | 10.99 | 5.89 | 6.12 | 3.23 | 9.93 | 5.74 |
| Consistency of Acuity Use (2011) | 9.83 | 5.95 | 3.81 | 4.18 | 7.54 | 5.52 |

*Note*. Cells with '--' are empty because variable(s) were not included in respective analyses.
*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2010 or 2011.
*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2010 or 2011.

Table 15

*School-Level Variability in Acuity Use, Research Question 1*

| | *Null Model* | | *Full Model* | | |
|---|---|---|---|---|---|
| Actions | % | *p*-value | % | *p*-value | PVE |
| Prevalence | 23.3 | 0.00 | 23.3 | 0.00 | 0.0% |
| Consistency | 35.4 | 0.00 | 28.5 | 0.00 | 27.6% |

*Note*. *n*(teachers) = 670; *n*(schools) = 65

*Note.* The *%* column represents the percentage of total variance at the school-level.

*Note*. The *PVE* column represents the *proportion of variance explained* at the teacher-level by the introduction of predictors at the teacher- and school-levels.

Table 16

*Teacher-Level Variability in AIMS Reading and Math Test Scores, Research Question 2*

| | Null Model | | Full Model (Prevalence) | | | Full Model (Consistency) | | |
|---|---|---|---|---|---|---|---|---|
| | % | *p*-value | % | *p*-value | PVE | % | *p*-value | PVE |
| Elementary | | | | | | | | |
| Reading | 22.1 | 0.00 | 8.7 | 0.00 | 83.3% | 8.6 | 0.00 | 83.7% |
| Math | 23.7 | 0.00 | 22.0 | 0.00 | 64.7% | 22.0 | 0.00 | 64.7% |
| Junior High | | | | | | | | |
| Reading | 25.0 | 0.00 | 3.3 | 0.06 | 93.8% | 3.4 | 0.05 | 93.5% |
| Math | 46.0 | 0.00 | 16.4 | 0.00 | 89.1% | 16.5 | 0.00 | 89.1% |

*Note*. *n*(students) = 8312; *n*(teachers) = 423

*Note.* The *%* column represents the percentage of total variance at the teacher-level.

*Note*. The *PVE* column represents the *proportion of variance explained* at the teacher-level by the introduction of predictors at the student- and teacher-levels.

Table 17

*Teacher-Level Variability in AIMS Reading and Math Test Scores, Research Question 3*

| | Null Model | | Full Model: Prevalence | | | Full Model: Consistency | | |
|---|---|---|---|---|---|---|---|---|
| | % | *p*-value | % | *p*-value | PVE | % | *p*-value | PVE |
| Elementary | | | | | | | | |
| Reading (2010) | 4.6 | 0.00 | 0.5 | 0.55 | 94.6% | 0.4 | 0.62 | 95.9% |
| Reading (2011) | 9.4 | 0.00 | 5.0 | 0.00 | 73.1% | 4.9 | 0.00 | 74.0% |
| Math (2010) | 4.4 | 0.00 | 1.7 | 0.09 | 81.8% | 1.7 | 0.10 | 82.4% |
| Math (2011) | 10.8 | 0.00 | 11.6 | 0.00 | 49.6% | 11.5 | 0.00 | 50.0% |
| Junior High | | | | | | | | |
| Reading (2010) | 12.2 | 0.04 | 0.2 | 0.86 | 98.9% | 0.3 | 0.81 | 98.5% |
| Reading (2011) | 19.6 | 0.01 | 7.7 | 0.02 | 77.3% | 7.7 | 0.02 | 77.3% |
| Math (2010) | 18.5 | 0.01 | 5.9 | 0.11 | 81.6% | 5.5 | 0.14 | 82.6% |
| Math (2011) | 31.9 | 0.00 | 15.4 | 0.05 | 85.2% | 10.1 | 0.05 | 84.0% |

*Note*. Elementary: *n*(students) = 3,360; *n*(2011 teachers) = 250; *n*(2010 teachers) = 239
JHS Reading: *n*(students) = 2,221; *n*(2011 teachers) = 29; *n*(2010 teachers) = 24
JHS Math: *n*(students) = 2,070; *n*(2011 teachers) = 42; *n*(2010 teachers) = 26

*Note.* The *%* column represents the percentage of total variance at the teacher-level.

*Note*. The *PVE* column represents the *proportion of variance explained* at the teacher-level by the introduction of predictors at the student- and teacher-levels.

Table 18
*Regression Model for Prevalence of Acuity Use*

| | Teacher-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Years of Experience | | | | 0.01 |
| 0 - 5 years | 0.10 | -0.11 | 0.31 | |
| 6 - 10 years | 0.35 | 0.15 | 0.55 | |
| 11 - 19 years | 0.18 | 0.00 | 0.36 | |
| | School-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| AZ Learns | | | | 0.78 |
| Performing | -0.21 | -0.91 | 0.49 | |
| Performing Plus | -0.33 | -1.06 | 0.39 | |
| Highly Performing | -0.20 | -0.93 | 0.53 | |
| Elementary | 0.19 | -0.15 | 0.54 | 0.27 |

*Note*. *n*(teachers) = 670; *n*(schools) = 65
*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.
*Note.* Reference category for Years of Experience is *20+*; AZ Learns is *Excelling*.

Table 19
*Regression Model for Consistency of Acuity Use*

| | Teacher-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Years of Experience | | | | 0.02 |
| 0 - 5 years | 0.16 | -0.03 | 0.35 | |
| 6 - 10 years | 0.26 | 0.08 | 0.45 | |
| 11 - 19 years | 0.22 | 0.05 | 0.39 | |
| | School-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| AZ Learns | | | | 0.03 |
| Performing | -0.71 | -1.43 | 0.00 | |
| Performing Plus | -0.99 | -1.73 | -0.25 | |
| Highly Performing | -0.51 | -1.26 | 0.24 | |
| Elementary | 0.71 | 0.37 | 1.06 | 0.00 |

*Note*. *n*(teachers) = 670; *n*(schools) = 65
*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.
*Note.* Reference category for Years of Experience is *20+*; AZ Learns is *Excelling*.

Table 20

*Regression Model Associating Elementary AIMS Reading with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.07 | 0.04 | 0.10 | 0.00 |
| Ethnicity | | | | 0.01 |
| Non-Latino White | 0.05 | 0.00 | 0.11 | |
| Latino | -0.01 | -0.06 | 0.05 | |
| Economically Advantaged | 0.07 | 0.04 | 0.11 | 0.00 |
| Prior Year Reading Score | 0.71 | 0.70 | 0.73 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use | 0.02 | 0.00 | 0.05 | 0.06 |
| Years of Experience | | | | 0.60 |
| 0 - 5 years | -0.05 | -0.12 | 0.02 | |
| 6 - 10 years | -0.02 | -0.09 | 0.05 | |
| 11 - 19 years | -0.01 | -0.07 | 0.06 | |

*Note*. *n*(students) = 8312; *n*(teachers) = 423

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 21

*Regression Model Associating Elementary AIMS Reading with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.07 | 0.04 | 0.10 | 0.00 |
| Ethnicity | | | | 0.01 |
| Non-Latino White | 0.05 | 0.00 | 0.11 | |
| Latino | -0.01 | -0.06 | 0.05 | |
| Economically Advantaged | 0.07 | 0.04 | 0.11 | 0.00 |
| Prior Year Reading Score | 0.71 | 0.70 | 0.73 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use | 0.04 | 0.01 | 0.06 | 0.00 |
| Years of Experience | | | | 0.50 |
| 0 - 5 years | -0.06 | -0.13 | 0.02 | |
| 6 - 10 years | -0.02 | -0.09 | 0.05 | |
| 11 - 19 years | -0.01 | -0.08 | 0.05 | |

*Note*. *n*(students) = 8312; *n*(teachers) = 423

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 22

*Regression Model Associating Elementary AIMS Math with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.03 | 0.00 | 0.05 | 0.06 |
| Ethnicity | | | | 0.00 |
| Non-Latino White | 0.07 | 0.02 | 0.12 | |
| Latino | 0.02 | -0.03 | 0.07 | |
| Economically Advantaged | 0.07 | 0.03 | 0.10 | 0.00 |
| Prior Year Math Score | 0.76 | 0.74 | 0.77 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use | 0.03 | 0.00 | 0.06 | 0.05 |
| Years of Experience | | | | 0.79 |
| 0 - 5 years | 0.02 | -0.08 | 0.11 | |
| 6 - 10 years | 0.05 | -0.04 | 0.14 | |
| 11 - 19 years | 0.01 | -0.07 | 0.10 | |

*Note*. *n*(students) = 8312; *n*(teachers) = 423

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 23

*Regression Model Associating Elementary AIMS Math with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.03 | 0.00 | 0.05 | 0.06 |
| Ethnicity | | | | 0.00 |
| Non-Latino White | 0.07 | 0.02 | 0.12 | |
| Latino | 0.02 | -0.03 | 0.07 | |
| Economically Advantaged | 0.07 | 0.03 | 0.10 | 0.00 |
| Prior Year Math Score | 0.76 | 0.74 | 0.78 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use | 0.03 | 0.00 | 0.06 | 0.09 |
| Years of Experience | | | | 0.72 |
| 0 - 5 years | 0.01 | -0.08 | 0.11 | |
| 6 - 10 years | 0.05 | -0.04 | 0.14 | |
| 11 - 19 years | 0.01 | -0.07 | 0.09 | |

*Note*. *n*(students) = 8312; *n*(teachers) = 423

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 24

*Regression Model Associating Junior High AIMS Reading with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.17 | 0.11 | 0.24 | 0.00 |
| Ethnicity | | | | 0.00 |
| Non-Latino White | 0.09 | -0.03 | 0.21 | |
| Latino | -0.06 | -0.18 | 0.07 | |
| Economically Advantaged | 0.09 | 0.01 | 0.17 | 0.02 |
| Prior Year Reading Score | 0.71 | 0.67 | 0.75 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use | 0.01 | -0.06 | 0.08 | 0.71 |
| Years of Experience | | | | 0.21 |
| 0 - 5 years | 0.13 | -0.11 | 0.37 | |
| 6 - 10 years | -0.08 | -0.23 | 0.08 | |
| 11 - 19 years | -0.11 | -0.29 | 0.08 | |

*Note*. *n*(students) = 5244; *n*(teachers) = 60

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 25

*Regression Model Associating Junior High AIMS Reading with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.18 | 0.11 | 0.24 | 0.00 |
| Ethnicity | | | | 0.00 |
| Non-Latino White | 0.09 | -0.03 | 0.21 | |
| Latino | -0.06 | -0.18 | 0.07 | |
| Economically Advantaged | 0.09 | 0.01 | 0.17 | 0.03 |
| Prior Year Reading Score | 0.71 | 0.67 | 0.75 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use | -0.01 | -0.07 | 0.04 | 0.66 |
| Years of Experience | | | | 0.19 |
| 0 - 5 years | 0.14 | -0.10 | 0.38 | |
| 6 - 10 years | -0.07 | -0.23 | 0.08 | |
| 11 - 19 years | -0.11 | -0.30 | 0.08 | |

*Note*. *n*(students) = 5244; *n*(teachers) = 60

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 26

*Regression Model Associating Junior High AIMS Math with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | -0.01 | -0.07 | 0.05 | 0.69 |
| Ethnicity | | | | 0.17 |
| Non-Latino White | 0.02 | -0.08 | 0.11 | |
| Latino | -0.06 | -0.16 | 0.05 | |
| Economically Advantaged | 0.05 | -0.02 | 0.12 | 0.21 |
| Prior Year Math Score | 0.73 | 0.69 | 0.77 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use | 0.02 | -0.06 | 0.10 | 0.59 |
| Years of Experience | | | | 0.55 |
| 0 - 5 years | -0.19 | -0.48 | 0.11 | |
| 6 - 10 years | -0.08 | -0.32 | 0.16 | |
| 11 - 19 years | -0.02 | -0.26 | 0.22 | |

*Note*. *n*(students) = 5020; *n*(teachers) = 83

*Note*. Prevalence of Acuity Use is the number of instructional actions performed in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 27

*Regression Model Associating Junior High AIMS Math with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | -0.01 | -0.07 | 0.05 | 0.68 |
| Ethnicity | | | | 0.18 |
| Non-Latino White | 0.02 | -0.08 | 0.11 | |
| Latino | -0.05 | -0.16 | 0.05 | |
| Economically Advantaged | 0.05 | -0.02 | 0.12 | 0.21 |
| Prior Year Math Score | 0.73 | 0.69 | 0.77 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use | 0.00 | -0.08 | 0.08 | 0.92 |
| Years of Experience | | | | 0.56 |
| 0 - 5 years | -0.18 | -0.47 | 0.12 | |
| 6 - 10 years | -0.07 | -0.31 | 0.17 | |
| 11 - 19 years | -0.01 | -0.24 | 0.23 | |

*Note*. *n*(students) = 5020; *n*(teachers) = 83

*Note*. Consistency of Acuity Use is the number of weeks Acuity was used in 2011.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 28
*Cross-Classified Model, Elementary AIMS Reading with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.08 | 0.04 | 0.13 | 0.00 |
| Ethnicity | | | | 0.49 |
| Non-Latino White | 0.02 | -0.06 | 0.10 | |
| Latino | -0.01 | -0.09 | 0.07 | |
| Economically Advantaged | 0.15 | 0.10 | 0.20 | 0.00 |
| 2009 Reading Score | 0.67 | 0.65 | 0.70 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use (2010) | -0.01 | -0.03 | 0.02 | 0.71 |
| Prevalence of Acuity Use (2011) | 0.02 | -0.01 | 0.05 | 0.13 |
| Years of Experience (2010) | | | | 0.10 |
| 0 - 5 years | 0.09 | 0.01 | 0.17 | |
| 6 - 10 years | 0.02 | -0.05 | 0.10 | |
| 11 - 19 years | 0.06 | -0.01 | 0.12 | |
| Years of Experience (2011) | | | | 0.91 |
| 0 - 5 years | -0.02 | -0.11 | 0.07 | |
| 6 - 10 years | 0.02 | -0.07 | 0.11 | |
| 11 - 19 years | 0.00 | -0.08 | 0.07 | |

*Note*. *n*(students) = 3,360; *n*(2011 teachers) = 250; *n*(2010 teachers) = 239
*Note*. Prevalence of Acuity Use is the number of instructional actions.
*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 29

*Cross-Classified Model, Elementary AIMS Reading with Consistency of Acuity Uses*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.08 | 0.04 | 0.13 | 0.00 |
| Ethnicity | | | | 0.48 |
| Non-Latino White | 0.02 | -0.06 | 0.10 | |
| Latino | -0.01 | -0.09 | 0.07 | |
| Economically Advantaged | 0.15 | 0.10 | 0.20 | 0.00 |
| 2009 Reading Score | 0.67 | 0.65 | 0.70 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use (2010) | -0.02 | -0.05 | 0.01 | 0.19 |
| Consistency of Acuity Use (2011) | 0.05 | 0.01 | 0.08 | 0.01 |
| Years of Experience (2010) | | | | 0.06 |
| 0 - 5 years | 0.10 | 0.02 | 0.18 | |
| 6 - 10 years | 0.03 | -0.04 | 0.11 | |
| 11 - 19 years | 0.06 | 0.00 | 0.13 | |
| Years of Experience (2011) | | | | 0.87 |
| 0 - 5 years | -0.03 | -0.12 | 0.07 | |
| 6 - 10 years | 0.02 | -0.07 | 0.10 | |
| 11 - 19 years | -0.01 | -0.08 | 0.07 | |

*Note*. *n*(students) = 3,360; *n*(2011 teachers) = 250; *n*(2010 teachers) = 239

*Note.* Consistency of Acuity Use is the number of weeks Acuity was used.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 30

*Cross-Classified Model, Elementary AIMS Math with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.03 | -0.01 | 0.07 | 0.11 |
| Ethnicity | | | | 0.01 |
| Non-Latino White | 0.07 | 0.00 | 0.14 | |
| Latino | 0.00 | -0.07 | 0.07 | |
| Economically Advantaged | 0.08 | 0.03 | 0.12 | 0.00 |
| 2009 Math Score | 0.69 | 0.67 | 0.72 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use (2010) | 0.00 | -0.02 | 0.03 | 0.79 |
| Prevalence of Acuity Use (2011) | 0.01 | -0.02 | 0.04 | 0.49 |
| Years of Experience (2010) | | | | 0.26 |
| 0 - 5 years | -0.09 | -0.19 | 0.01 | |
| 6 - 10 years | 0.06 | -0.04 | 0.15 | |
| 11 - 19 years | 0.02 | -0.06 | 0.10 | |
| Years of Experience (2011) | | | | 0.07 |
| 0 - 5 years | 0.02 | -0.06 | 0.10 | |
| 6 - 10 years | -0.06 | -0.13 | 0.02 | |
| 11 - 19 years | -0.03 | -0.09 | 0.03 | |

*Note*. *n*(students) = 3,360; *n*(2011 teachers) = 250; *n*(2010 teachers) = 239

*Note*. Prevalence of Acuity Use is the number of instructional actions performed.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 31

*Cross-Classified Model, Elementary AIMS Math with Consistency of Acuity Uses*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.03 | -0.01 | 0.07 | 0.11 |
| Ethnicity | | | | 0.01 |
| Non-Latino White | 0.07 | 0.00 | 0.14 | |
| Latino | 0.00 | -0.07 | 0.07 | |
| Economically Advantaged | 0.08 | 0.03 | 0.12 | 0.00 |
| 2009 Math Score | 0.69 | 0.67 | 0.72 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use (2010) | 0.02 | -0.01 | 0.05 | 0.19 |
| Consistency of Acuity Use (2011) | 0.02 | -0.02 | 0.06 | 0.26 |
| Years of Experience (2010) | | | | 0.22 |
| 0 - 5 years | -0.09 | -0.19 | 0.01 | |
| 6 - 10 years | 0.06 | -0.04 | 0.15 | |
| 11 - 19 years | 0.02 | -0.07 | 0.10 | |
| Years of Experience (2011) | | | | 0.07 |
| 0 - 5 years | 0.02 | -0.06 | 0.09 | |
| 6 - 10 years | -0.07 | -0.14 | 0.01 | |
| 11 - 19 years | -0.03 | -0.09 | 0.03 | |

*Note*. *n*(students) = 3,360; *n*(2011 teachers) = 250; *n*(2010 teachers) = 239

*Note.* Consistency of Acuity Use is the number of weeks Acuity was used.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 32

*Cross-Classified Model, Junior High AIMS Reading with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.21 | 0.13 | 0.28 | 0.00 |
| Ethnicity | | | | 0.11 |
| Non-Latino White | 0.06 | -0.07 | 0.20 | |
| Latino | -0.04 | -0.18 | 0.10 | |
| Economically Advantaged | 0.14 | 0.05 | 0.23 | 0.00 |
| 2009 Reading Score | 0.56 | 0.52 | 0.61 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use (2010) | 0.00 | -0.07 | 0.06 | 0.96 |
| Prevalence of Acuity Use (2011) | -0.01 | -0.11 | 0.08 | 0.75 |
| Years of Experience (2010) | | | | 0.12 |
| 0 - 5 years | 0.15 | -0.08 | 0.38 | |
| 6 - 10 years | 0.23 | 0.02 | 0.44 | |
| 11 - 19 years | 0.21 | 0.01 | 0.41 | |
| Years of Experience (2011) | | | | 0.36 |
| 0 - 5 years | -0.22 | -0.54 | 0.10 | |
| 6 - 10 years | -0.31 | -0.68 | 0.05 | |
| 11 - 19 years | -0.19 | -0.52 | 0.14 | |

*Note*. *n*(students) = 2,221; *n*(2011 teachers) = 29; *n*(2010 teachers) = 24

*Note*. Prevalence of Acuity Use is the number of instructional actions performed.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 33

*Cross-Classified Model, Junior High AIMS Reading with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | 0.21 | 0.13 | 0.28 | 0.00 |
| Ethnicity | | | | 0.12 |
| Non-Latino White | 0.06 | -0.07 | 0.20 | |
| Latino | -0.04 | -0.18 | 0.10 | |
| Economically Advantaged | 0.14 | 0.05 | 0.23 | 0.00 |
| 2009 Reading Score | 0.56 | 0.52 | 0.61 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use (2010) | -0.01 | -0.07 | 0.04 | 0.62 |
| Consistency of Acuity Use (2011) | -0.04 | -0.14 | 0.05 | 0.34 |
| Years of Experience (2010) | | | | 0.14 |
| 0 - 5 years | 0.15 | -0.08 | 0.37 | |
| 6 - 10 years | 0.22 | 0.00 | 0.43 | |
| 11 - 19 years | 0.20 | 0.01 | 0.39 | |
| Years of Experience (2011) | | | | 0.32 |
| 0 - 5 years | -0.21 | -0.52 | 0.11 | |
| 6 - 10 years | -0.32 | -0.68 | 0.03 | |
| 11 - 19 years | -0.18 | -0.51 | 0.14 | |

*Note*. *n*(students) = 2,221; *n*(2011 teachers) = 29; *n*(2010 teachers) = 24

*Note.* Consistency of Acuity Use is the number of weeks Acuity was used.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 34

*Cross-Classified Model, Junior High AIMS Math with Prevalence of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | -0.05 | -0.12 | 0.02 | 0.15 |
| Ethnicity | | | | 0.12 |
| Non-Latino White | 0.07 | -0.04 | 0.18 | |
| Latino | -0.02 | -0.13 | 0.10 | |
| Economically Advantaged | 0.11 | 0.03 | 0.19 | 0.01 |
| 2009 Math Score | 0.60 | 0.56 | 0.65 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Prevalence of Acuity Use (2010) | -0.04 | -0.13 | 0.05 | 0.33 |
| Prevalence of Acuity Use (2011) | 0.02 | -0.06 | 0.10 | 0.65 |
| Years of Experience (2010) | | | | 0.25 |
| 0 - 5 years | 0.07 | -0.14 | 0.28 | |
| 6 - 10 years | -0.22 | -0.53 | 0.10 | |
| 11 - 19 years | -0.12 | -0.33 | 0.08 | |
| Years of Experience (2011) | | | | 0.11 |
| 0 - 5 years | -0.35 | -0.64 | -0.06 | |
| 6 - 10 years | -0.20 | -0.43 | 0.04 | |
| 11 - 19 years | -0.19 | -0.41 | 0.03 | |

*Note*. *n*(students) = 2,070; *n*(2011 teachers) = 42; *n*(2010 teachers) = 26

*Note*. Prevalence of Acuity Use is the number of instructional actions performed.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

Table 35

*Cross-Classified Model, Junior High AIMS Math with Consistency of Acuity Use*

| | Student-Level Variables | | | |
|---|---|---|---|---|
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Gender | -0.05 | -0.12 | 0.02 | 0.15 |
| Ethnicity | | | | 0.12 |
| Non-Latino White | 0.07 | -0.04 | 0.18 | |
| Latino | -0.02 | -0.13 | 0.10 | |
| Economically Advantaged | 0.11 | 0.03 | 0.19 | 0.01 |
| 2009 Math Score | 0.60 | 0.56 | 0.65 | 0.00 |
| | Teacher-Level Variables | | | |
| Factor | SD Difference | 95% Confidence Interval | | *p*-value |
| Consistency of Acuity Use (2010) | -0.04 | -0.13 | 0.05 | 0.38 |
| Consistency of Acuity Use (2011) | 0.02 | -0.06 | 0.10 | 0.68 |
| Years of Experience (2010) | | | | 0.27 |
| 0 - 5 years | 0.07 | -0.14 | 0.28 | |
| 6 - 10 years | -0.19 | -0.51 | 0.12 | |
| 11 - 19 years | -0.12 | -0.33 | 0.09 | |
| Years of Experience (2011) | | | | 0.13 |
| 0 - 5 years | -0.35 | -0.64 | -0.05 | |
| 6 - 10 years | -0.20 | -0.44 | 0.04 | |
| 11 - 19 years | -0.19 | -0.41 | 0.04 | |

*Note*. *n*(students) = 2,070; *n*(2011 teachers) = 42; *n*(2010 teachers) = 26

*Note.* Consistency of Acuity Use is the number of weeks Acuity was used.

*Note.* Reference category for Gender is *Male*; Ethnicity is *Other*; Years of Experience is *20+*.

## APPENDIX A
## Technical Appendix of Data Management and Analysis Descriptions

This appendix describes the data and analyses used in the present study. The appendix is divided into three parts. The first part describes our management of data sets. The second part describes data analyses, including analytic considerations we encountered in choosing modeling approaches. The third part describes post-hoc analyses conducted to ensure that no hidden relationships pertaining to Acuity use remained in our data.

### Data Management

Before we could analyze the multilevel models, a substantial amount of data management was needed to construct the analysis data sets. First, the district (MPS) produced separate files for students and teachers, each containing information on factors included in the models. Using the district files, our research team merged them into seven analysis data sets. In analyses addressing Research Question 1, one file contained teacher-level data (e.g., teacher experience, Acuity use) by data on MPS schools (e.g., school-level). The remaining files contained data for the second and third research questions: three contained students assigned to MPS teachers and schools in 2011 who also had achievement data for two consecutive years (2011 and 2010). These three were organized by school-level and tested content area (i.e., Elementary math/reading, Junior high reading, and Junior high math). For the cross-classified MLMs used to address Research Question 3, the remaining three files contained three consecutive years of student achievement by students' teachers and schools, which were organized in the same manner as the files for Research Question 2.

**Preparation of student-level data files.** The primary variables of interest in the student-level files consisted of student test scores from the *Arizona Instrument to Measure Standards* (AIMS) tests in math and Reading/English Language Arts (ELA) for students in grades three through eight in 2008-09 to 2010-11. Achievement outcomes were normalized within grade, year, and subject (i.e., transformed to z-scores). Other student-level variable management included:

- Categorical variables were recoded to produce meaningful reference categories for analyses (e.g., gender, race/ethnicity; see Table 36).
- Students in special education programs were excluded from analytic data sets.
- Student data files were separated by school-level (e.g., ES and JHS) and within junior high files, by subject (e.g., JHS Math JHS English).

**Preparation of teacher-level data files.** Students were assigned to teachers in the data files provided by MPS. The data set linked each student to their math teacher and to their reading/ELA teacher. As a result, students were nested within teachers.
Other teacher-level variable management included:

- Use logs: Information on teachers' Acuity use was captured in use logs, the full version of which was sent to our research team in spring 2012. We then linked teachers' use log data with their demographic data.
- In all analyses, we did not include Acuity users who were identified as non-teachers (users with IDs coded as 9999999).
- As with student-level factors, categorical variables were recoded to produce meaningful reference categories for analyses (e.g., years of experience).

Table 36.
*Coding of Dichotomous and Categorical Variables*

| | *Level of Measurement* | *Reference Category* | *Additional Categories* |
|---|---|---|---|
| *Student-level variables* | | | |
| Gender | Dichotomous | Male | Female |
| Ethnicity | Categorical | Other | Non-Latino White; |
| Economic Disadvantage | Dichotomous | Yes | No |
| *Teacher-level variables* | | | |
| Years of experience | Categorical | 20+ years | 0-5 years; 6-10 years; 11-19 years |
| *School-level variables* | | | |
| School-Level | Dichotomous | Junior High | Elementary |
| AZ Learns | Categorical | Excelling | Performing; Performing Plus; Highly Performing |

**Continuous variables.** We standardized the continuous predictors and outcomes to provide measures of effect size; that is, a measure of the impact of predictor variables relative to other variables entered into the models. One implication of this transformation is that statistically significant effects ($p < 0.05$) in raw metrics could be less likely to be significant in standardized analyses (Hox & Maas, 2002). While we did detect this effect in our analyses, it was negligible (see Tables 37 and 38). Thus, we determined that the interpretability advantage associated with standardized predictors outweighed the disadvantage of slightly higher p-values (i.e., slight reduction in observed statistical significance). The standardizations did not affect the nature – or direction – of the observed relationships.

Table 37.
*Illustration: Notable Predictor Coefficient Differences by Analytic Samples (ES Reading Analyses)*

| Parameter | Reading: Year 2 Models | Reading: CC Models | Interpretation |
|---|---|---|---|
| *Student-Level:* | | | |
| 2011 achievement | -0.10 | -0.19 | Average starting reading achievement was 0.04 SDs lower, and ending reading achievement was 0.09 SDs lower among students in the CC versus the Year 2 samples. |
| 2010/2009 achievement | 0.71 | 0.67 | |
| White vs. Other | 0.05 | 0.03 | The reading achievement gap between *White* and *Other* students was 0.02 SDs higher in the Year 2 sample, implying that reading achievement differences between *White* and *Other* students were more pronounced in Year 2 sample. |
| Economically Advantaged | 0.07 | 0.15 | The reading achievement gap between *Economically Advantaged* and *Disadvantaged* students was 0.08 SDs higher in the CC sample, implying that reading achievement differences between *Economically Advantaged* and *Disadvantaged* students were more pronounced in the CC sample. |
| *Teacher-Level: Experience* | | | |
| 0 to 5 years | -0.05 | -0.02 | Reading achievement was 0.03 SDs lower among students with new teachers in the Year 2 sample, while students in the CC sample with new teachers scored only 0.02 SDs less than their peers with veteran teachers. Reading achievement gaps associated with new teachers were more pronounced in the Year 2 sample. |
| 6 to 10 years | -0.02 | 0.02 | Reading achievement was 0.02 SDs lower among students with mid-career teachers (6-10 years) in the Year 2 sample. Students with mid-career teachers scored 0.02 SDs higher than students with veteran teachers in the CC sample. Reading achievement associated with mid-career teachers was lower than that associated with veteran teachers in the Year 2 sample, while the reverse was true in the CC sample. |

Table 38.
*Illustration: Notable Predictor Coefficient Differences by Analytic Samples (ES Math Analyses)*

| Parameter | Math Year 2 Models | Math CC Models | Interpretation |
|---|---|---|---|
| *Student-Level:* | | | |
| 2011 achievement | -0.12 | 0.08 | Average starting math achievement was 0.07 SDs lower in the CC samples, but ending math achievement was 0.20 SDs higher, implying that students' growth trajectory was steeper in CC versus Year 2 samples. |
| 2010/2009 achievement | 0.76 | 0.69 | |
| Latino vs. Other | 0.02 | 0.00 | The math achievement gap between *Latino* and *Other* students was 0.02 SDs higher in the Year 2 sample, implying that math achievement differences between *Latino* and *Other* students were more pronounced in the Year 2 sample. |
| *Teacher-Level: Experience* | | | |
| 0 to 5 years | 0.02 | -0.09 | Math achievement was 0.09 SDs lower among students with new teachers (versus veteran teachers) in the CC sample, while students in the Year 2 sample with new teachers scored 0.02 SDs higher than their peers with veteran teachers. Math achievement associated with new teachers was lower than that associated with veteran teachers in the CC sample, while the reverse was true in the Year 2 sample. |

## Data Analyses

**Initial diagnostic analyses: Explore descriptives and variable distributions.** Our first step in analyses was to examine means, frequencies, ranges, and kurtosis/skewness estimates for predictors and outcomes. We noted that one of the measures of teachers' Acuity use – prevalence – displayed considerable elements of non-normality. Specifically, we explored kurtosis and skewness statistics, and according to conventional wisdom, the prevalence distribution violated a rule of thumb that these statistics be less than or equal to two. Prevalence had a kurtosis statistic that indicated a leptokurtic distribution (11.4) and skewness statistic that indicated a positively-skewed distribution (2.8).

We note an issue associated with lack of normality in predictor variables; namely, complications that may arise from potentially biased standard errors. This issue is of particular concern in multilevel modeling, given that distributional assumptions are made about standard errors at each level in the model. Specifically, nonnormality of errors at level-1 will not bias estimation of level-2 effects, but it will introduce bias into standard errors at both levels, and therefore, into hypothesis tests and confidence intervals. In other words, lack of normality may bias standard errors at both levels of the data structure (Hox & Maas, 2002; Raudenbush & Bryk, 2002). To explore whether these potential impacts on observed standard errors in our models, we ran some initial multilevel models with prevalence and consistency of use factors. Results indicated counter-intuitive divergent findings between the two Acuity use factors; namely, different confidence intervals and p-values in prevalence models as compared to the consistency models. Because we do not believe that one conceptualization of teachers' use over another should dramatically impact observed relationships between Acuity use and achievement, we hypothesized that non-normality in the prevalence factor was a more viable explanation.

To address the normality issue, we explored possible outliers and eliminated 27 teachers from analyses whose prevalence of Acuity use was above three standard deviations above average Acuity use (above 740 total uses). Despite the elimination of these outliers, we still noted a positively skewed distribution in the prevalence variable. As such, we then explored extant literature and concluded that a square-root transformation of the positively skewed variable (prevalence of Acuity use) could mitigate the lack of normality in the factor's distribution (Raudenbush & Bryk, 2002). After transforming the factor, statistics representing distributional properties (i.e., kurtosis and skewness) were in normal range (kurtosis= 1.62 and skewness= 1.13). In the substantive analyses, described more fully below, we also observed compatibility between the two Acuity use factors. We should note a disadvantage associated with the transformation; namely, that it rendered predictor coefficients (i.e., beta coefficients) difficult to interpret in real world terms, because they underwent the square-root transformation *and* were standardized into z-scores.

**Substantive analyses: Multilevel models.** In planning our analyses, we had a number of analytic decisions. This section describes these, in terms of prior years' analyses, cross-classified models, and missing data.

***Replication of prior year analyses.*** In planning for the present study, we made some initial analysis decisions: specifically, to replicate analyses from year two in order to determine whether effects we observed in the second year would be consistent with those we observed in the third.

***Cross-classified models.*** We also sought to extend those analyses by including an additional year of Acuity use and achievement data in the cross-classified models. We chose cross-classified modeling over traditional growth models for two reasons:

- Growth models assume students are nested in same level-2 unit over time, and this was not true in our data (students cross-classified across 2010 and 2011 teachers).
- If we used growth models, teachers' Acuity use would have to have been modeled at student level, because it is a "time invariant" factor. We felt it more correct to model Acuity use at the source of that use—the teacher-level.

Of course, any modeling decision brings trade-offs, and we wish to note a few that arise from our decision to use cross-classified modeling. First, cross-classified MLMs model relationships between predictors and outcomes as linear even when that may not be the case. Given the aim of our substantive analyses was to examine relationships between Acuity use and achievement on tests that are vertically-equated across grades, we felt that the linearity assumption was justifiably secure. However, this does not mean that achievement was, in reality, perfectly linear across grades, and readers should be aware that cross-classified models do not adjust for sources of nonlinearity.

Another important limitation of cross-classified models is the use of 2009 achievement as a covariate for student achievement gains across 2010 and 2011. We note that the impact of 2009 achievement is likely much greater in 2010 than in 2011. This implies a "back-loading" of the covariate, which could have muted effects associated with 2010 teachers. Additionally, in general, including covariates for prior achievement likely reduced the available proportion of test variability in the following year. For our purposes, this is acceptable, given that we wish to isolate teacher-level impacts on achievement from those outside of the current teachers' control, such as achievement demonstrated before students entered their current classrooms.

***Consequences of missing data.*** We employed *listwise deletion* of missing data (i.e., complete deletion of a case that has missing data on any variable included in the analyses). In our data files, this means that students were included only if we had complete achievement, demographic, and teacher-level data for them. These restrictions created opportunities for missing data, particularly in data file used to address the third research question, where the student sample reduced because it was restricted to 5-8 grades. Listwise deletion is predicated on the assumption that data are missing completely at random (MCAR), which means that underlying reasons for missing data bear no systematic relationship with the outcomes under investigation. In reality, likely that our data were only partially MCAR, but additionally missing at random (MAR), where missingness is assumed to be independent of unobserved missing values but potentially dependent on other variables in model. These dependencies may be direct or indirectly correlated with unobserved values. In contexts where MAR dominates missingness, it is appropriate to use more principled methods such as multiple imputation (MI) or full information maximum likelihood (ML) estimation (Heck et al., 2010). In the present study, however, analyses demonstrate that our delimited samples were not greatly deviant from the population (see Table 39). Consequently, we feel comfortable that our data are mostly MCAR and the main impact of missing data on our analyses is a loss of statistical power.

Table 39
*Descriptive Statistics by Analytic Sample, Research Questions 2 and 3 (RQs 2 and 3)*

| | Elementary | | Junior High Reading | | Junior High Math | |
|---|---|---|---|---|---|---|
| | RQ 2 | RQ 3 | RQ 2 | RQ 3 | RQ 2 | RQ 3 |
| *Student-Level Factors* | (*n*= 8312) | (*n*=3360) | (*n*= 5244) | (*n*= 2221) | (*n*= 5020) | (*n*= 2070) |
| Average Reading Achievement | | | | | | |
| 2009 | -- | 488.1 | -- | 527.1 | -- | -- |
| 2010 | 494.7 | 505.3 | 546.4 | 547.3 | -- | -- |
| 2011 | 515.4 | 525.2 | 540.4 | 550.3 | -- | -- |
| Average Math Achievement | | | | | | |
| 2009 | -- | 492.3 | -- | -- | -- | 555.3 |
| 2010 | 393.9 | 402.1 | -- | -- | 444.9 | 445.7 |
| 2011 | 416.9 | 426.4 | -- | -- | 453.6 | 454.5 |
| Gender (*n*, %) | | | | | | |
| Female | 4322 (52%) | 1781 (53%) | 2674 (51%) | 1155 (52%) | 2661 (53%) | 1097 (53%) |
| Male | 3990 (48%) | 1579 (47%) | 2570 (49%) | 1066 (48%) | 2359 (47%) | 973 (47%) |
| Ethnicity (*n*, %) | | | | | | |
| Non-Latino White | 4405 (53%) | 1982 (59%) | 2937 (56%) | 1377 (62%) | 2811 (56%) | 1180 (57%) |
| Latino | 2992 (36%) | 1008 (30%) | 1678 (32%) | 644 (29%) | 1606 (32%) | 662 (32%) |
| Other | 914 (11%) | 370 (11%) | 629 (12%) | 200 (9%) | 602 (12%) | 228 (11%) |
| Economically Disadvantaged (*n*, %) | 3574 (43%) | 1478 (44%) | 2517 (48%) | 955 (43%) | 2460 (49%) | 994 (48%) |

*(Table 15 cont'd)*

| | Elementary | | Junior High Reading | | Junior High Math | |
|---|---|---|---|---|---|---|
| *Teacher-Level Factors* | RQ 2 | RQ 3 | RQ 2 | RQ 2 | RQ 3 | RQ 2 |
| 2011 Years of Experience (*n*, %) | (*n*= 423) | (*n*= 250) | (*n*= 60) | (*n*=29) | (*n*= 83) | (*n*= 42) |
| 0 to 5 years | 80 (19%) | 40 (16%) | 15 (25%) | 12 (40%) | 10 (12%) | 5 (12%) |
| 6 to 10 years | 86 (20%) | 48 (19%) | 19 (31%) | 5 (18%) | 30 (36%) | 15 (36%) |
| 11 to 19 years | 128 (30%) | 83 (33%) | 7 (11%) | 10 (34%) | 23 (28%) | 12 (28%) |
| 20+ years | 132 (31%) | 80 (32%) | 20 (34%) | 2 (8%) | 20 (24%) | 11 (26%) |
| 2010 Years of Experience (*n*, %) | | (*n*= 239) | | (*n*= 24) | | (*n*= 26) |
| 0 to 5 years | -- | 38 (16%) | -- | 5 (21%) | -- | 7 (25%) |
| 6 to 10 years | -- | 41 (17%) | -- | 5 (19%) | -- | 2 (9%) |
| 11 to 19 years | -- | 72 (30%) | -- | 11 (45%) | -- | 7 (28%) |
| 20+ years | -- | 88 (37%) | -- | 4 (15%) | -- | 10 (39%) |
| Average Acuity Use | | | | | | |
| Prevalence (2011) | 169.2 | 173.4 | 102.4 | 82.6 | 138.8 | 137.0 |
| Prevalence (2010) | -- | 134.0 | -- | 81.8 | -- | 135.9 |
| Consistency (2011) | 9.8 | 9.9 | 4.0 | 3.8 | 7.6 | 7.5 |
| Consistency (2010) | -- | 11.0 | -- | 6.1 | -- | 9.9 |

*Note*. Cells with '--' are empty because variable(s) were not included in respective analyses.

**Post-Hoc Analyses**

We conducted additional post hoc analyses to ensure that no significant patterns, or relationships, between Acuity use and achievement went unexplored. We first explored interactions between Acuity use and achievement. Specifically, to assess whether Acuity use had differential impacts depending on students' achievement level, we broke achievement and Acuity use into quartiles, and explored interactions between prior achievement and Acuity use as well as current achievement and use, controlling for background characteristics. We noted nothing statistically significant except for an interaction between Acuity use and the second and third achievement quartiles in junior high math.

We also explored possible descriptive explanations for different Acuity impacts in elementary versus junior high schools. In terms of different Acuity functions accessed by teachers, we noted a couple of interesting, but likely unimportant trends:

- While 33% of elementary teachers accessed Longitudinal Reports only 17% of junior high teachers did so.
- A larger proportion of junior high teachers accessed the Previewing Instructional Resources feature (66% versus 58% of elementary teachers).
- We also explored differences in Acuity functions accessed by teachers within school-levels, and observed the following:
    - No significant differences among elementary teachers, regardless of grade taught.
    - Statistically significant differences in functions used among junior high teachers.
    - Math teachers accessed Acuity's instructional resources – not just Reports – more frequently than reading teachers. We note that this could be due to larger student to teacher ratios in reading: in 2011, average reading teachers were linked to 101 junior high students while average math teachers were linked to 78. Larger class sizes in junior high made using Acuity resources time-prohibitive for some teachers (Wayman et al., 2009).