Office of Accountability
Department of Research, Evaluation, and Assessment

Data Modeling of the Elementary Response to Instruction and Intervention Implementation in Knox County Schools

Technical Report

Clint Sattler
Supervisor, Research and Evaluation
Knox County Schools
Department of Research, Evaluation, and Assessment

December 2015

Overview

In the 2014-2015 school year (SY1415), the state of Tennessee mandated the implementation of their Response to Instruction and Intervention (RTI^2) framework at the elementary level. The state RTI^2 framework was created in order to standardize the disparate practices that were being used across the state to provide enrichment to high performing students, support struggling students, and ultimately determine if a student has a learning disability in basic academic skills (such as reading fluency, basic mathematical calculations, and written expression). Student performance was to be gauged by a nationally normed screening assessment, and any academic intervention provided to at-risk students were to be provided by research-based programs.

The RTI^2 framework promotes individualization to meet student needs through core instruction and additional intervention. The RTI^2 guidelines focus on early interventions utilizing evidence based practices, instructionally appropriate assessment, data based decision making and educator professional development to foster an environment conducive to academic success for all students. The process is tiered in order to meet student needs from special education through general education and to ensure that there is a continuum of academic support for students regardless of their ability. Students in Tier III receive the most intensive academic intervention in an effort to close achievement gaps with students in less intensive tiers.

The Knox County Schools (KCS) implemented a local RTI^2 framework based on the principles of the state framework. The overall goals of both of the frameworks were identical. The local framework provides district specific guidelines around the screening and progress monitoring tools as well as meeting norms for the school based RTI^2 leadership teams.

The effective implementation of appropriate academic interventions is a key tenant of the district's strategic plan, *Excellence for Every Child*. An evaluation of the RTI^2 initiative was therefore requested because of its relative importance to the long term strategy of the district. The district's RTI^2 leadership team defined the following research questions to be answered after the initial year of RTI^2 implementation.

- What were the relevant enrollment and movement patters of the students in the first year of the Knox County Schools' RTI^2 initiative?
- Were the RTI^2 teams making defendable decisions to place students in the appropriate intervention tiers and move the students to appropriate tiers per their progress monitoring data?
- What processes and procedures need to be refined in order to improve the RTI^2 process for SY1516?

The results from this analysis indicated that, in a general sense, the KCS practitioners were implementing the RTI² framework as intended. The Knox County tier enrollment was similar to the state theoretical distributions, there were statistically significant differences between the mean performance of students in the various tiers, and students were generally exiting tiers when their progress monitoring data indicated that they were ready for promotion to a less intensive tier. Although there are a variety of areas in which KCS RTI² can improve, it appears that the process has largely met the goals of the state RTI² initiative.

Methodology

A variety of information was gathered regarding the implementation of RTI² in SY1415. The exploration of both quantitative and qualitative RTI² data was the emphasis for this first year of implementation.

Methodology: Enrollment Statistics

The initial phase of this analysis required an extensive examination of the descriptive statistics associated with the launch of RTI². This was done in order to determine how closely KCS statistics aligned with the theoretical distribution presented in the state RTI² framework and to search for relevant trends in the RTI² enrollment data. The data regarding tier enrollment was extracted from a database maintained by district-level RTI² coaches.

Methodology: Qualitative Data Collection

The secondary phase of the analysis involved collecting qualitative data from school-based practitioners. A representative from the Office of Research, Evaluation and Assessment (REA) attended RTI² meetings and interviewed principals, teachers, education assistants, RTI² coaches and academic coaches regarding their experiences with RTI². The schools that participated in the qualitative data collection were selected on a random basis. Approximately 20% of the elementary schools in the district were visited by the REA representative.

Methodology: Quantitative Analysis of Academic Gains

A third phase of the analysis focused on the outcomes of different groups of students. The analysis compared the RTI² progress monitoring data from students who moved from a more intensive tier to a less intensive tier (i.e. from Tier II to Tier 1), students who moved from a less intensive tier to a more intensive tier (i.e. from Tier II to Tier III), and students who remained in a given tier for the entire academic year. Tier II progress monitoring generally occurred through the STAR Renaissance suites of assessment (STAR Reading, Math and Early Literacy assessments) and Tier III progress monitoring generally occurred through the Aimsweb suite of assessments (R-CBM, M-CAP and M-COMP). Each analysis began with correlative studies to estimate the potential for multicollinearity between variables. The output from the correlative studies was used to determine which variables were to be used in the deeper analyses. This, and all subsequent analyses, only included students who were enrolled in intervention for a minimum of 45 instructional days in the hopes of excluding data from students who were placed in an intervention tier based on an erroneous initial data point.

The initial screening of the data indicated that statistical testing and modeling should focus on the initial SY1415 normal curve equivalent (NCE) of each student on their progress monitoring tool, their final SY1415 NCE on their progress monitoring tool, and a measure of the student's rate of improvement. Although the STAR suite of assessments quantified gains in test scores using student growth percentiles (SGPs), these were deemed too volatile to use in the calculation. STAR calculates SGP based on only two data points (initial and final data points for the SY1415 school year), and these point estimates are prone to the typical errors of testing. The rate of improvement was therefore calculated as the slope of the (ordinary least squares) best-fit line through each student's longitudinal progress monitoring data.

A variety of techniques were used to determine if there were any significant differences in the mean performance between the groups of students. Analysis of variance (ANOVA) was used to determine if there were any significant differences in mean performance between the groups of students when the data were normally distributed and of equal variance. The Brown-Forsyth statistic was used when the data were normally distributed and of un-equal variance. Post-hoc (Tukey) and contrast testing was done to better understand the nature of any statically significant results of the ANOVA and Brown-Forsyth testing. Non-parametric testing was used to detect differences in student performance when the data of interest was not normally distributed. An Independent Samples Median test was done to determine if the medians of the groups of students were the same between the groups of students. Additionally, an Independent Samples Kruskal-Wallis test was done to determine if the distribution of the outcome data was the same between the groups of students. The threshold for statistical significance for all testing was $\alpha=0.05$. Finally, longitudinal trends in the data were graphed to visualize the time dependent structure of the progress monitoring data.

Methodology: Progress Monitoring Sensitivity

The state RTI^2 framework mandated the collection of progress monitoring data at two week intervals. An analysis of the sensitivity of each of the tools to detect changes in student performance in two week intervals was therefore desired.

A rate of change in the outcomes of each assessment was calculated for every administration of a STAR and Aimsweb assessment. The rate of change was calculated as the absolute value of the change in scaled score (for STAR) or the number of correct responses (for Aimsweb) divided by the number of weeks between test administrations. The STAR suite of assessments provides a standard error of measurement (SEM) for each individual assessment, whereas the Aimsweb technical manuals provide a single estimate of the SEM for their assessments. The SEM was divided by the rate of change in order to determine the number of weeks, at that observed growth rates that would be required to equal one SEM. The results were averaged by student to determine the student-mean number of weeks required to equal one SEM. The distribution of the student-mean number of weeks (as determined by the 1st, 2nd and 3rd quartile cuts) was then generated.

Methodology: Predictive Modeling

The final phase of the analysis attempted to use logistic modeling to determine the probability of a success in intervention from observable characteristics. Any student who moved from a more intensive tier to a less intensive tier in SY1415, and remained in the less intensive tier in SY1516 was coded as a "success". The initial logistic regression screening included all of the non-multicollinear data from the progress monitoring tools (initial NCE, final NCE, and rate of improvement) as well as demographic data (gender, BHN membership, ED membership and ELL membership). The variables were added to the model one at a time in order to determine which of the variables provided a statistically significant contribution to the model. Reliable models could not be created for the Tier III Math progress monitoring tools (R-CAP and R-COMP) because of a low number of data points.

Receiver Operator Characteristic (ROC) curves were built from the output of the logistic models. Each probability for a student moving from a less intensive tier to a more intensive tier was plotted against the binary "success" variable. The specificity and sensitivity of the model was used to determine the cut-off probability that simultaneously maximized the number of "successes" that were predicted to be "successes", the number of "non-successes" that were predicted to be "non-successes" and the overall classification rate. These cut-off values were then applied to the output from the logistic regression to determine which students had a probability of success higher than the cut-off but were never promoted to a less intensive tier. The data were aggregated by school to look for outliers.

The percent of RTI² students who were predicted to move to a less intensive tier but did not was calculated for each school. A school was labeled as an outlier if the percent of students who were misclassified as successes exceeded 1.5 interquartile ranges above the 3rd quartile of the school distribution. Schools were only considered outliers if they tested a minimum of 30 students with the progress monitoring tool of interest.

Results: Enrollment Statistics

Population statistics provided by the supervisor of the school psychologists indicated that the percentage of elementary students classified with a specific learning disability (LD) declined sharply in SY1415.

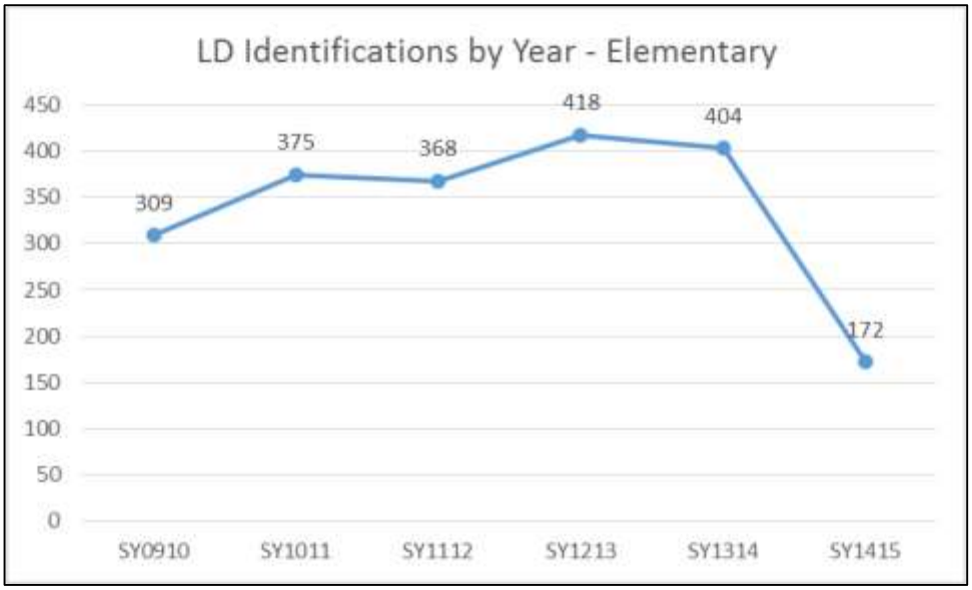

Figure 1: Number of Elementary Students Identified as Learning Disabled by Year

It is important to note that this drop in the number of students identified with a specific learning disability may have been due changes in the data requirements regarding the formal referral process. The state framework recommends eight data points be collected before a student is identified as learning disabled. Because most progress monitoring occurred at two week intervals, there could be a 16 week interval between the collection of a student's initial data point and the first formal recommendation for special education services. This lag could be partly responsible for the drop in the percentage of elementary students identified as learning disabled, rather than the efficacy of the RTI^2 process. Anecdotally, a similar drop in LD identification occurred when the district adopted its first RTI^2 model. Therefore, this data should be monitored longitudinally to determine if the decrease in LD enrollments will persist or if LD enrollment will rise back to historic levels.

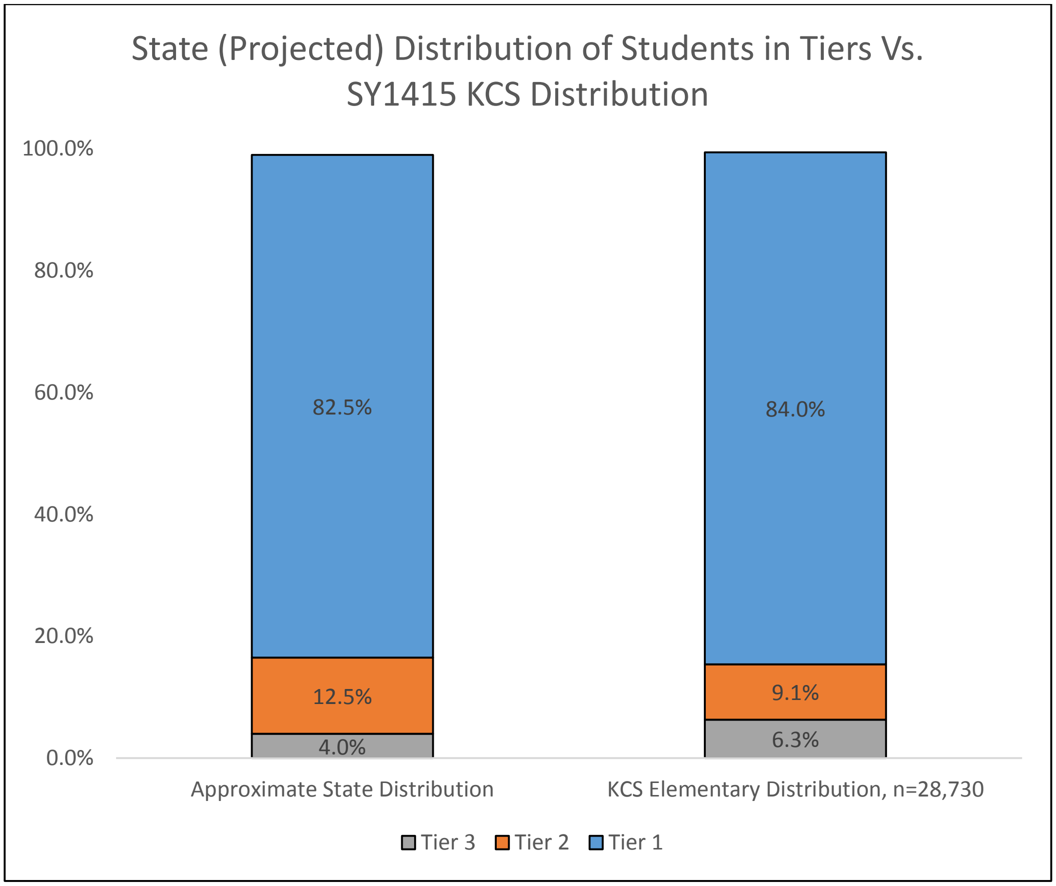

Figure 2: Distribution of Tier Enrollment at the end of SY1415

The distributions of students who ended the year in each tier are presented in the graph above. The approximate state distribution was determined by averaging the upper and lower limits of enrollment in each tier. The SY1415 KCS data suggests that we had a slightly higher percentage of Tier III students when compared to the state distribution (the state approximate distribution indicates that 3-5% of students are expected to be in Tier III). It is possible that this resulted in slightly lower enrollment in Tier II than the state distribution (10-15%). The SY1415 enrollment in Tier I fell within the range of the state theoretical distribution (80-85%).

By far, the students were enrolled in reading/language arts (RLA) intervention at much higher rates than in math. There were very few students who are enrolled in both math and reading intervention in SY1415. Enrollment in intervention in both subjects was very difficult to maintain due to scheduling constraints. The school-based RTI^2 teams generally enrolled a student in reading intervention because reading skills are seen as the gateway to academic success in all of the other subjects. The KCS motto has been "learn to read so you can read to learn". RTI^2 enrollment trends indicated that this motto was put into practice in SY1415.

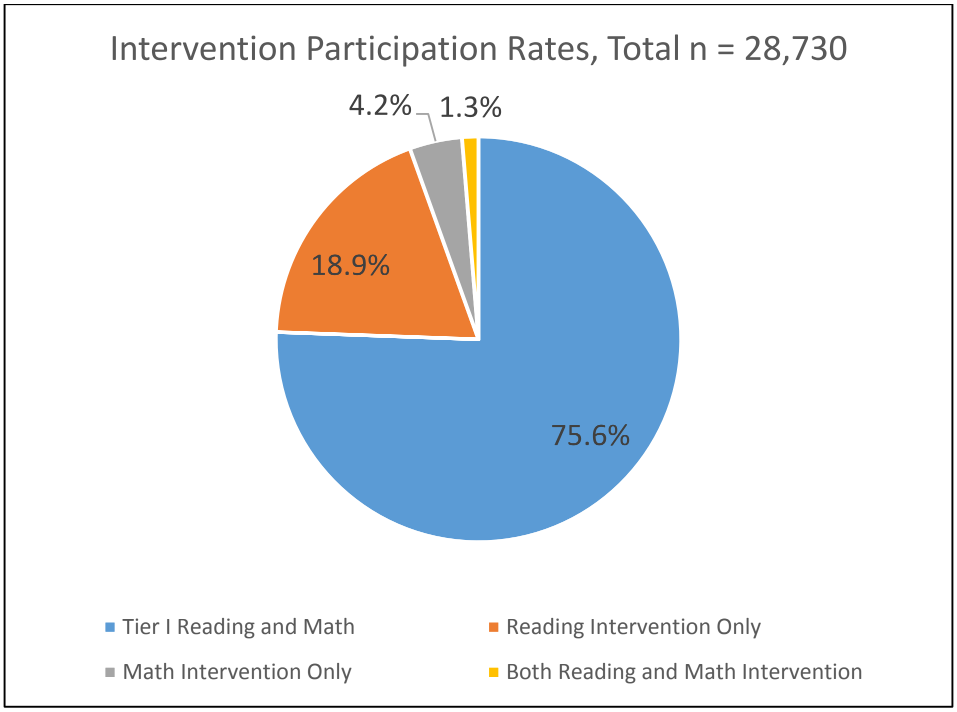

Figure 3: Intervention Enrollment by Subject

The percentage of the student population in each tier varied widely by school. This created challenges around scheduling and resource allocations that were felt more acutely at school with higher percentages of students enrolled in Tiers II and III.

Table 1: Ranges of the % of Population in Each Tier

	Max % in Tier (by School)	Min % in Tier (by School)	School with Max % in Tier	School with Min % in Tier
Tier I	94.5%	63.7%	Northshore Elementary	Green Elementary
Tier II	22.4%	2.6%	Pond Gap Elementary	Northshore Elementary
Tier III	18.4%	1.5%	Green Elementary	Sequoyah Elementary

There was less variation between grade levels than between schools. The percentage of students enrolled in Tiers II and III tended to decrease with increasing grade levels. This provided some evidence that the district RTI² teams believed in early intervention as a form of prevention (as per both the state and district RTI² frameworks), as well as having provided some evidence that previous intervention efforts may have been paying dividends.

Table 2: Percent of Students Ending SY1415 in a Tier (By Grade)

Grade	n	Moved to SpEd	Tier 1	Tier 2	Tier 3
K	4709	0.3%	87.5%	8.0%	4.1%
1	4755	0.5%	79.4%	11.8%	8.3%
2	4941	0.9%	82.5%	8.6%	8.1%
3	4871	1.0%	82.6%	9.5%	6.9%
4	4742	0.4%	84.8%	8.5%	6.3%
5	4712	0.3%	87.4%	8.3%	4.0%

The demographic breakdown of the students enrolled in each tier is presented in the table below. Students who were members of subgroups that the state identified as at-risk (students who are Black, Hispanic or Native American (BHN), English Language Learners (ELL) or Economically Disadvantaged (ED)) were more likely to be enrolled in Tier II or Tier III than students who did not belong to these subgroups. This tier enrollment data tracked with the state trends in academic performance by subgroup.

Table 3: Percent of Students Ending SY1415 in a Tier (By Subgroup)

Subgroup		n	Moved to SpEd	Tier 1	Tier 2	Tier 3
BHN	No	21541	0.56%	87.06%	7.41%	4.97%
	Yes	7189	0.60%	74.85%	14.19%	10.36%
ELL	No	26989	0.57%	84.42%	8.95%	6.06%
	Yes	1741	0.46%	77.60%	11.60%	10.34%
ED	No	13398	0.32%	92.15%	4.83%	2.70%
	Yes	15332	0.78%	76.88%	12.85%	9.48%

The table below provides information as to how students moved through intervention tiers in SY1415. The majority of students remained in the tier in which they were originally enrolled. Approximately 25% of the students who were enrolled in either Tier II or Tier III were promoted to a less intensive tier before the end of SY1415.

Table 4: Movement of SY1415 Students in Intervention Tiers

Intervention Subject	Net Movement	Tier Movement	n	% of Students
RLA	Left District	Left District	147	2.4%
	From a less intensive tier to a more intensive tier	Tier 1 to 2	48	11.1%
		Tier 1 to 3	9	
		Tier 2 to 3	459	
		Tier 2 to 3 to SPED	7	
		Tier 2 to SPED	15	
		Tier 3 to SPED	137	
	No Change	Tier 2	2308	61.8%
		Tier 2 to 1 back to 2	4	
		Tier 2 to 3 back to 2	30	
		Tier 3	1399	
		Tier 3 to 2 back to 3	14	
	From a more intensive tier to a less intensive tier	Tier 2 to 1	1092	24.7%
		Tier 2 to 3 to 1	2	
		Tier 3 to 1	53	
		Tier 3 to 2	287	
		Tier 3 to 2 to 1	68	
Math	Left District	Left District	35	2.1%
	From a less intensive tier to a more intensive tier	Tier 1 to 2	36	10.6%
		Tier 1 to 3	8	
		Tier 2 to 3	100	
		Tier 2 to SPED	1	
		Tier 3 to SPED	34	
	No Change	Tier 2	642	60.1%
		Tier 2 to 1 back to 2	2	
		Tier 2 to 3 back to 2	6	
		Tier 3	369	
		Tier 3 to 2 back to 3	1	
	From a more intensive tier to a less intensive tier	Tier 2 to 1	345	27.2%
		Tier 2 to 3 to 1	2	
		Tier 3 to 1	27	
		Tier 3 to 2	76	
		Tier 3 to 2 to 1	12	

The RTI² enrollment data were re-examined after the first screening assessment was given in SY1516. The intent was to determine the proportion of students who were promoted to a less intensive tier in SY1415 and were subsequently placed back into the more intensive tier at the start of SY1516. Re-enrollment of a student in the more intensive tier could have been a product of summer learning loss, changes in student interest in academic pursuits, or misclassification by RTI² teams at the end of SY1415. The figure below shows the percentage of students who were promoted to a less intensive tier in SY1415 but returned to the more intensive tier in SY1516.

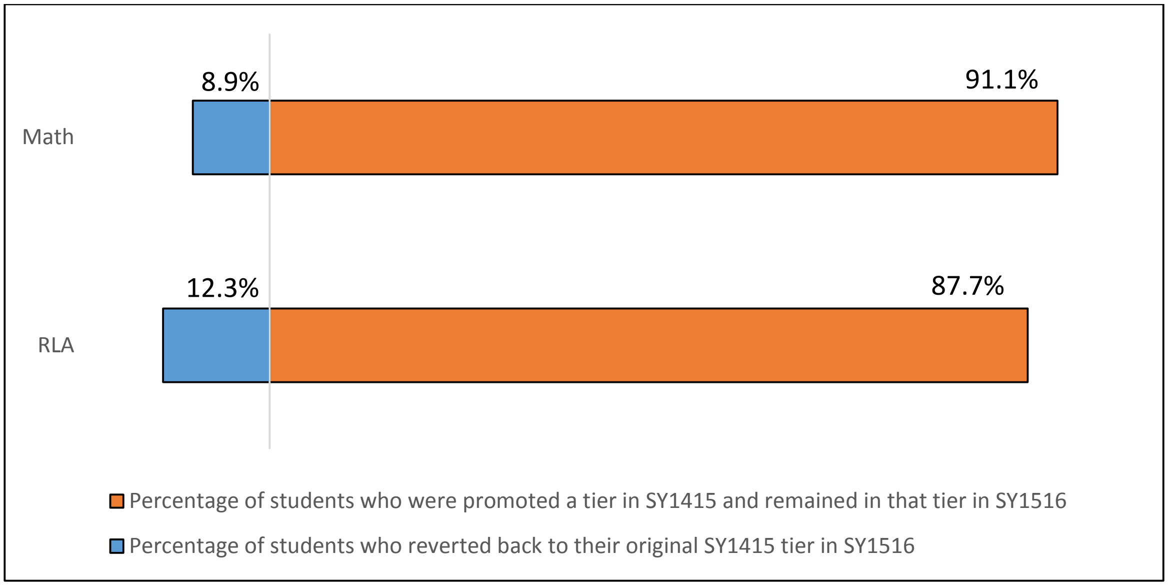

Figure 4: SY1516 Re-enrollment in More Intensive Tiers

The relatively low percentage of students who were re-enrolled in the more intensive tier provided some evidence that RTI² teams were generally not moving students to a less intensive tier without adequate cause.

Results: Qualitative Data Collection

A majority of the staff members who were interviewed during the course of the SY1415 program evaluation of RTI² had a positive view of the RTI² process. The interviewees were generally knowledgeable regarding district and state intervention policies and procedures. The major theme of the feedback from the RTI² implementation in SY1415 can be summed up by a quote from a participant. "Things started out very rough, but the whole process became easier and more valuable as the year progressed".

The teams of school personnel who were responsible for providing intervention services were expected to meet regularly to discuss the progress of the students. Generally, the meetings were originally seen as bureaucratic and of little value to the core mission of RTI². However, as the year progressed, a majority of the teams that we met with felt that these meetings were extremely valuable. Generally, the participants felt that the RTI² team meetings fostered an environment of team decision making, rather than placing the onus for making decisions on the RTI² coach. The participants also indicated that the process brought the focus back to how each individual student was progressing, rather than intervention procedures. Though the time requirements for these meeting were considerable, most of the participants felt that they were worth it. Most of the participants also agreed that RTI^2 teams should be meeting on a monthly basis. More frequent meetings would compete too much with other job functions, and less frequent meetings would lead to lags in making programmatic changes to interventions.

The magnitude of the time commitment required by the school staff to implement the RTI^2 initiative should not be under-estimated. The early RTI^2 team meetings were generally spent trying to organize data and complete forms. As the process matured, RTI^2 teams would organize the data prior to their meetings to ensure the maximum amount of time was available to discuss student progress. Dealing with paperwork became the main focus for many of the schools in the district. Some of the schools spent building-level funds to hire additional help to organize data, complete paperwork, and send parent letters home. Some schools did not have the funds available to hire additional help, and had to resort to using academic coaches and teaching staff to complete the administrative tasks associated with the RTI^2 process. Almost universally, there were complaints that the time requirements for the administration of RTI^2 and the associated paperwork created a sink that would normally have been spent on lesson planning, PLC collaboration, and 1:1 coaching opportunities.

Related to the paperwork burden, the RTI^2 team members who were interviewed noted the difficulty in tracking the tier enrollment data among mobile populations. The district opted to create a database tool to track the students who were in Tier II and III interventions, but the data were managed solely by the RTI^2 coaches. In essence, 12 KCS staff members were responsible for maintaining the list of more than 7,000 elementary students enrolled in Tier II or III. This resulted in severe time lags between any changes in the intervention programming and when the associated data were entered in the database. As a result of this, RTI^2 teams would have to consult the paper copies of student records sent to the school (when a child enrolled) in order to determine which intervention services a student had previously received. Often, student files would be missing paperwork, or the paperwork would be non-standard, which created confusion and lost time. A better system for tracking intervention enrollment was nearly universally requested.

The changes mandated by the state required regular fidelity monitoring of intervention services. The interviewees found value in the fidelity monitoring, even though it was rare for a fidelity check to turn up an issue within an intervention group. The KCS administrators that were interviewed praised the fidelity check process for providing structure around “look-fors” and feedback regarding intervention.

Generally, participants praised the most commonly used RLA intervention programs. The S.P.I.R.E. program (usually used in Tier III RLA) was especially popular among those that provided feedback and should be considered a good investment by the district. The participants appreciated the lesson structure, content and pacing that were a core component of the S.P.I.R.E. program. However, because S.P.I.R.E was generally deployed in Tier III intervention (to aid in reading fluency), it was common for the RTI^2 team members to assume that Tier III RLA intervention was for reading fluency only, and that Tier II RLA intervention was for reading comprehension. In actuality, both the state and district RTI^2 frameworks were adamant that the difference between Tier II and Tier III interventions is related to the intensity of the intervention, not the content. The district RTI^2 team must determine if the current practice is acceptable in terms of the district's RTI^2 strategy, or if the RTI^2 teams need to be re-educated in this point.

Although there was a consensus that the RLA intervention tools were beneficial, opinions were split over the RLA progress monitoring tools (especially Tier II). Some personnel felt that the STAR Renaissance assessment provided information that helped target specific academic weaknesses. The data were especially useful for creating small instructional groups to target specific RLA skills in Tier I instruction. Other teachers disagreed about the utility of the STAR output. Because the Renaissance suite of assessments contains adaptive testing, teachers were unable to see an item analysis of the questions that individual students missed (in contrast to previously available formative tools). Others felt that STAR was an inappropriate tool to monitor progress in reading fluency and comprehension because it is a standards based assessment. Progress in interventions that target fundamental reading skills may not "move the needle" on STAR Renaissance RLA assessments.

Despite the fact that RLA interventions were generally well-regarded by interviewees, the same cannot be said for the math intervention programs. The participants tended to be critical of the lack of a scripted math intervention program, though the nature of math intervention may not lend itself easily to a scripted program. The schools that felt they could not have certified teachers leading math interventions (either because of time requirements, lack of qualified applicants or school-budgets) were especially vocal regarding the need for a high-quality math intervention program.

In general, RTI^2 teams seemed to err on the side of retaining a student in more intensive tiers. Many teams wanted "a few more data points" to ensure trends in progress monitoring data were sustainable, even after the required number of data points had been collected. Additionally, some of the interviewed teams were concerned with the generally low levels of differentiation and personalization in Tier I instruction (when compared to Tiers II and III). Commonly, schools would run Tier II intervention groups during Tier I core extension blocks. In general practice, Tier I core extension blocks were treated similarly to independent study classes. Many of the RTI^2 teams felt that the structure offered in a Tier II intervention would be more beneficial to a student than core extension, so students with strong progress monitoring data were sometimes retained in Tier II. Other roadblocks to tier promotion included concerns over discipline and student motivation. Many of the RTI[2] team members who participated in this study lamented the fact that the academic intervention system was not rolled out simultaneously with an equally strong non-cognitive intervention framework.

Results: Quantitative Analysis of Academic Gains
STAR Reading

The STAR Reading assessment was the most commonly used Tier II RLA progress monitoring tool for grades 3 through 5. The data from the statistical analysis of the progress monitoring data, disaggregated by how the students moved through the intervention tiers, is presented in the table below.

Table 5: Tier II STAR Reading and RLA TCAP

	Population Means (N Counts)			
Population	Initial SR NCE	Final SR NCE	ROI (SS/week)	TCAP RLA NCE
The students moving from a less intensive tier to a more intensive tier (i.e. Tier II to Tier III)	27.95 (329)	23.76 (329)	0.84 (329)	27.71 (176)
The students who remained in the same tier for the entire school year (i.e. Tier II to Tier II)	30.61 (1812)	35.29 (1812)	2.27 (1812)	34.50 (1085)
The students moving from a more intensive tier to a less intensive (i.e. Tier II to Tier I)	32.36 (768)	46.96 (768)	4.53 (768)	42.23 (438)
Levene Statistic	0.077	0.000	0.000	0.110
Sig.	**0.000***	**0.000#**	**0.000#**	**0.000***
r	0.13	0.50	0.47	0.36

* ANOVA, # Brown-Forsythe

The statistical tests indicated that we could reject the null hypothesis that the mean initial STAR Reading NCE, the final STAR Reading NCE, rate of improvement (in scaled score points per week) and TCAP RLA NCE were no different between the groups. Tukey post-hoc testing indicated that we could reject that null hypothesis that any of the means were equal to each other. The results seemed to indicate that there were generally differences in performance on the STAR Reading progress monitoring tool between groups of Tier II students for whom different RTI[2] decisions were being made. The results also helped to corroborate the assertion that RTI[2] teams were using progress monitoring data to inform their decisions regarding tier placement for students. The differences in mean performance (as measured by NCEs) were again manifested on the state assessment (as measured by mean TCAP RLA NCE). It was also noteworthy that the magnitudes of the differences in mean scores could be classified as medium to large (as indicated by the large Pearson's r). It was somewhat troubling that there was already a statistically significant difference in each group's mean STAR Reading NCE prior to the delivery of intervention services, but at least the magnitude of the difference would be classified as small by the Pearson's r.

Longitudinal plots of the mean STAR Reading scaled scores helped to visual time-dependent trends of the data for the groups of students. It was noteworthy that the students who were eventually promoted to Tier I seemed to exhibit higher than expected growth early in the academic year, and thus early in the intervention process. There were at least two possible causes for this. One of the causes could be that the students who responded positively to the intervention did so early in the intervention process (before the end of the first month of intervention). The other possible cause is that the initial screening data that placed the student in intervention could have been a low outlier. This may, when considered with the ANOVA data, provide evidence that students were placed in Tier II instruction when it was not required.

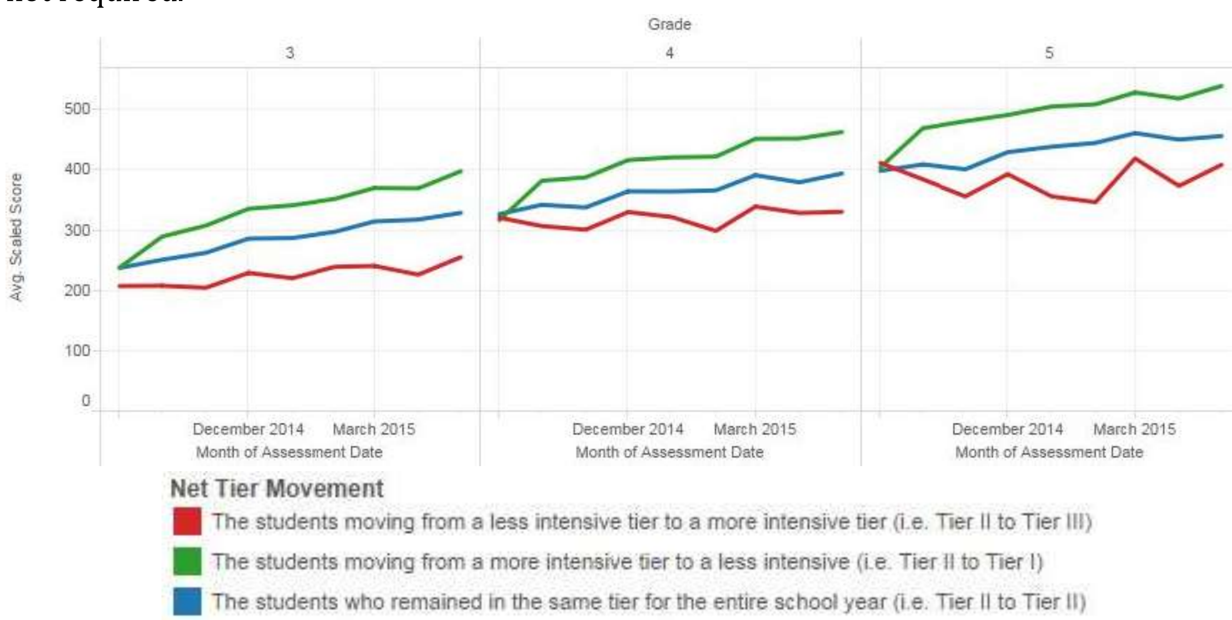

Figure 5: Longitudinal Trends in STAR Reading Progress Monitoring Data

STAR Early Literacy

The STAR Early Literacy assessment was the most commonly used Tier II RLA progress monitoring tool for grades K through 2. The data from the statistical analysis of the progress monitoring data, disaggregated by how the students moved through the intervention tiers, is presented in the table below.

Table 6: Tier II STAR Early Literacy

Population	Initial SEL NCE	Population Means (N Counts) Final SEL NCE	ROI (SS/week)
The students moving from a less intensive tier to a more intensive tier (i.e. Tier II to Tier III)	32.09 (65)	42.07 (65)	3.66 (65)
The students who remained in the same tier for the entire school year (i.e. Tier II to Tier II)	34.56 (425)	53.05 (425)	4.36 (425)
The students moving from a more intensive tier to a less intensive (i.e. Tier II to Tier I)	33.92 (296)	64.00 (296)	6.65 (296)
Levene Statistic	0.756	0.021	0.000
Sig.	0.246*	**0.000#**	**0.000#**
r	NA	0.37	0.36

* ANOVA, # Brown-Forsythe

The statistical tests indicated that we failed to reject the null hypothesis that the mean initial STAR Early Literacy NCE was no different between the groups. The statistical tests indicated that we could reject the null hypothesis that the mean final STAR Early Literacy NCE, and rate of improvement (in scaled score points per week) were no different between the groups. Tukey post-hoc testing indicated that we could reject that null hypothesis that any of the means were equal to each other. The results seemed to indicate that there were generally real differences in performance on the STAR Early Literacy progress monitoring tool between the groups of Tier II students for whom different RTI² decisions were being made. The results helped to corroborate the assertion that the RTI² teams were using progress monitoring data to inform decisions regarding tier placement for students. The difference in mean performance (as measured by NCEs) could not be compared to the performance on state assessments since the early grades do not participate in state testing. The magnitude of the differences between the groups was not as large as those reported with the STAR Reading results, though the Pearson's r was still large enough to classify the difference as a medium effect. The smaller effect size may have been a by-product of the test instrument itself. The reader may notice that the mean final STAR Early Literacy NCEs appeared to be higher than those in STAR Reading. STAR Renaissance provided percentile rankings for early literacy test takers that appeared to be artificially high from March of 2014 until the end of the year. Renaissance has since found the error and re-normed their end-of-the year early literacy rankings, but students who were being progress monitored on the STAR Early Literacy assessment easily could have been misclassified as making significant progress because of Renaissance's oversight. Most of the interviewed RTI² teams reported that they used early literacy percentile ranks as relative measure, rather than an absolute measure, since the rankings did not adhere to their classroom observations of the same students' performances. That practice would have helped to minimize the risk of misclassifying a student using the STAR Early Literacy rankings.

The longitudinal plots of the mean STAR Early Literacy scaled scores showed some interesting trends. The relationship between the mean Kindergarten scaled score for the students for whom different RTI² placement decisions were being made was erratic early in SY1415. However, there were clear separations between the mean scaled scores from October onward. This information may be useful in helping to inform district policy about when to start intervention for Kindergarten students. The first grade plot of the mean scaled score also showed an interesting trend. The grade-to-grade transition in the mean STAR Reading scaled score was continuous for the students who remained in Tier II for all of SY1415 (the blue line in Fig. 5). However, there was a discontinuity between these students when considering the Kindergarten to first grade data (the blue line in Fig. 6). This discontinuity may have been related to 1st grade students' lack of computer literacy skills.

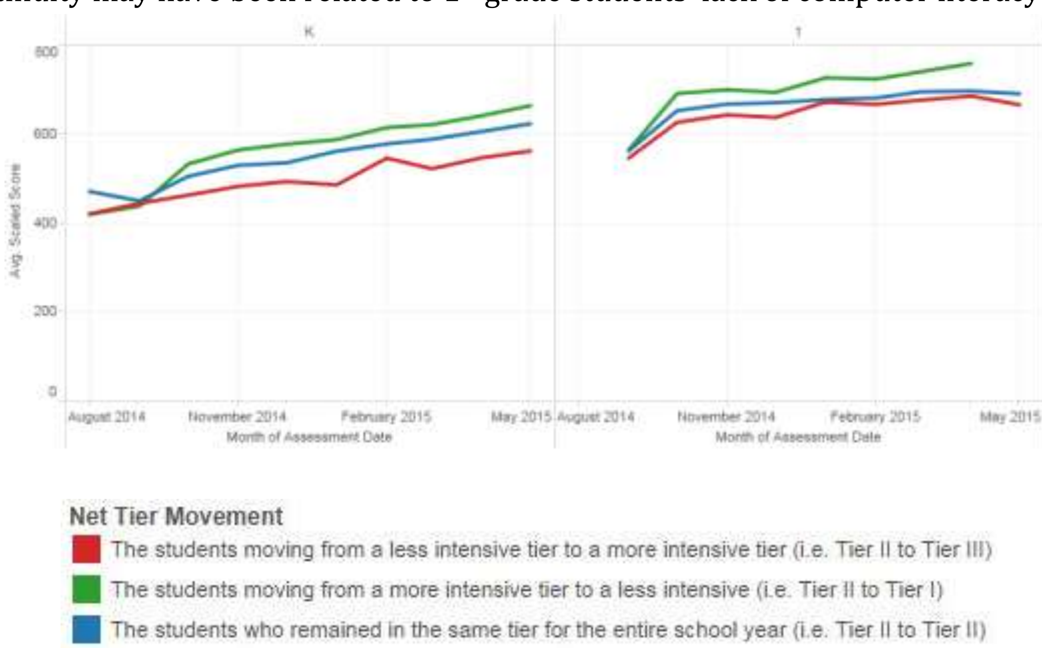

Figure 6: Longitudinal Trends in STAR Early Literacy Progress Monitoring Data

Aimsweb R-CBM

The Tier III RLA services were difficult to evaluate quantitatively. There were multiple progress monitoring tools that were being utilized to track student progress in Tier III. Commonly, the Aimsweb Reading Curriculum-Based Measurement assessments (R-CBM) were used to progress monitor students in reading fluency, while the Aimsweb Reading Maze assessments were used to progress monitor reading comprehension (in Tier III). The students were progress monitored on grade level, on their instructional level, or sometimes on both levels. This fact added to the confusion as to what data were being used to make RTI[2] decisions. This analysis only considered R-CBM data collected on-grade level because this was the most abundant of the available data. Unlike the Tier II STAR data, the Aimsweb data were not normally distributed, so all of the statistical testing utilized non-parametric testing.

Table 7: Tier III Aimsweb R-CBM and RLA TCAP

	Population Means - Population Medians (N Counts)			
Population	Initial R-CBM NCE	Final R-CBM NCE	ROI (WRC/week)	TCAP RLA NCE
The students moving from a less intensive tier to a more intensive tier (i.e. Tier III to Tier SpEd)	33.04 - 31.60 (43)	26.37 - 29.30 (43)	0.29 - 0.24 (43)	20.13 - 18.00 (68)
The students who remained in the same tier for the entire school year (i.e. Tier III to Tier III)	25.39 - 27.3 (963)	26.07 - 29.40 (963)	0.62 - 0.58 (963)	22.15 - 22.00 (542)
The students moving from a more intensive tier to a less intensive tier (i.e. Tier III to Tier II)	35.38 - 36.20 (182)	40.58 - 41.75 (182)	1.21 - 1.11 (182)	30.00 - 30.00 (97)
Probability that the Medians are the same across all categories	**0.000***	**0.000***	**0.000***	**0.000***
Probability that the distribution of scores are the same across all categories	**0.000#**	**0.000#**	**0.000#**	**0.000#**

* Independent Samples Median Test, # Independent Samples Kruskal-Wallis Test

The results of the Independent Sample Median Tests indicated that we could reject the null hypothesis that the medians were no different for all the groups of students. The results of the Independent Samples Kruskal-Wallis Tests indicated that we could reject the null hypothesis that the distribution of NCEs and rates of improvement were the same for all the groups of students. There are no post-hoc testing procedures for these tests that are analogous to the ANOVA post hoc procedures to tell us if there are differences between each individual category. Therefore, we ran contrasts manually and adjusted our alpha value to limit our total type I error rate to 5% (adjusted $\alpha=0.025$). The results of the contrast tests are contained below.

- The contrast testing indicated that we could reject the null hypothesis that the distributions of initial R-CBM NCEs were no different when comparing the students who moved to a more intensive intervention and students who stayed in Tier III all year ($p=0.001$). Visual inspection of boxplots of the data indicated that the students who were formally referred to special education generally had higher initial R-CBM NCEs. We failed to reject the null hypothesis that the median initial NCE was no different between the same students ($p=0.058$). The contrast testing between the students who eventually moved to Tier II and the students who remained in Tier III indicated that we could reject the null hypothesis that both the medians and the distribution of initial R-CBM NCEs were the same between the groups of students ($p=0.000$, both tests). Visual inspection of the plots indicated that the group that eventually moved to Tier II intervention had the higher scores.

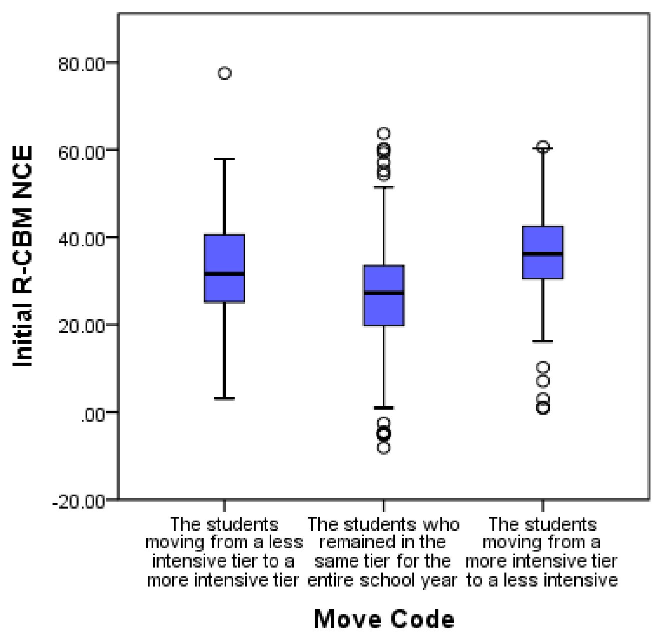

Figure 7: Boxplot of Tier III R-CBM Initial NCEs

- The contrast testing indicated that we failed to reject the null hypothesis that the medians and distribution of final R-CBM NCEs were no different between students who were referred to special education and students who remained in Tier III all year (p=0.957 and 0.832, respectively). It is important to reiterate that students may have been formally referred to special education due to reading comprehension rather than reading fluency, so it is difficult to draw any conclusions from this result. The contrast testing between students who eventually moved to Tier II and students who remained in Tier III indicated that we could reject the null hypothesis that both the medians and the distribution of final R-CBM NCEs were the same between the groups of students (p=0.000, both tests). A visual inspection of the plots indicated that the group that eventually moved to Tier II intervention had the higher scores. A visual comparison of the score distributions in Figures 7 and 8 suggests that the rate of change in R-CBM scores was a major driver for Tier III placement decisions.

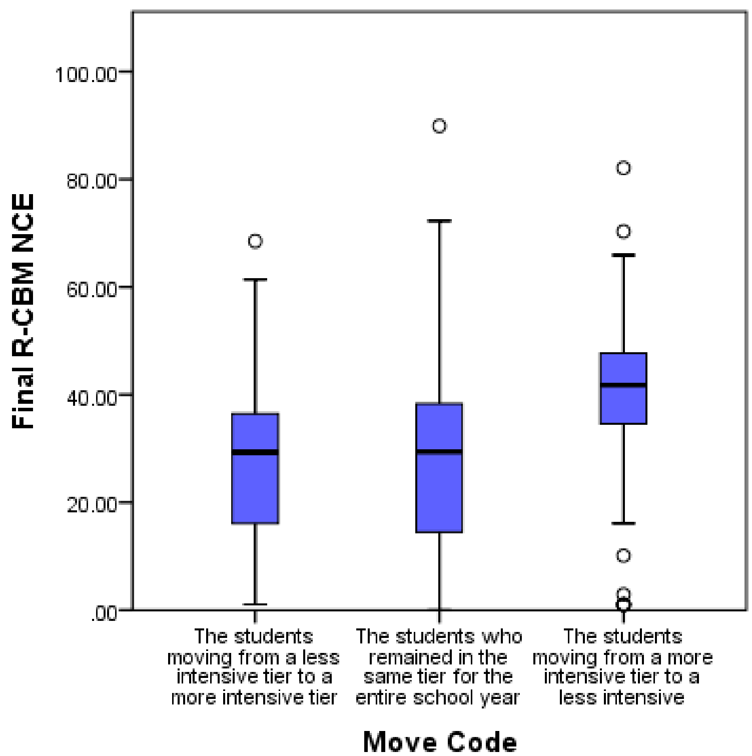

Figure 8: Boxplot of Tier III R-CBM Final NCEs

- The contrast testing indicated that we could reject the null hypothesis that the medians and distributions of the rates of improvement were no different between the students who were referred to special education and the students who remained in Tier III (p=0.002 and p=0.000 respectively). A visual inspection of the boxplots of the ROI data indicated that the students who remained in Tier III had generally higher rates of improvement than the students who were referred to special education. The contrast testing between students who eventually moved to Tier II and students who remained in Tier III indicated that we could reject the null hypothesis that both the medians and the distribution of the rates of improvement were the same between the groups of students (p=0.000, both tests). A visual inspection of the plots indicated that the group that eventually moved to Tier II intervention had the higher scores.

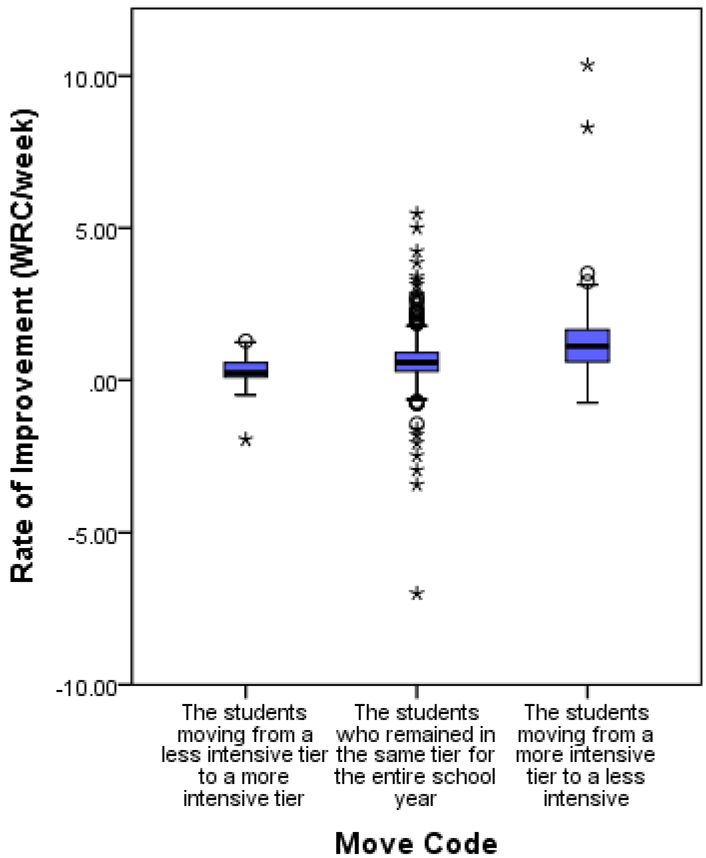

Figure 9: Boxplot of Tier III R-CBM ROI

- The contrast testing indicated that we failed to reject the null hypothesis that the medians and the distribution of TCAP RLA NCE were no different between the students who were referred to special education and those that remained in Tier III (p=0.283 and p=0.175 respectively). We could reject the null hypothesis that both the medians and the distribution of TCAP RLA NCEs were no different between the students who eventually moved to Tier II and those that remained in Tier III (p=0.000 for both tests).

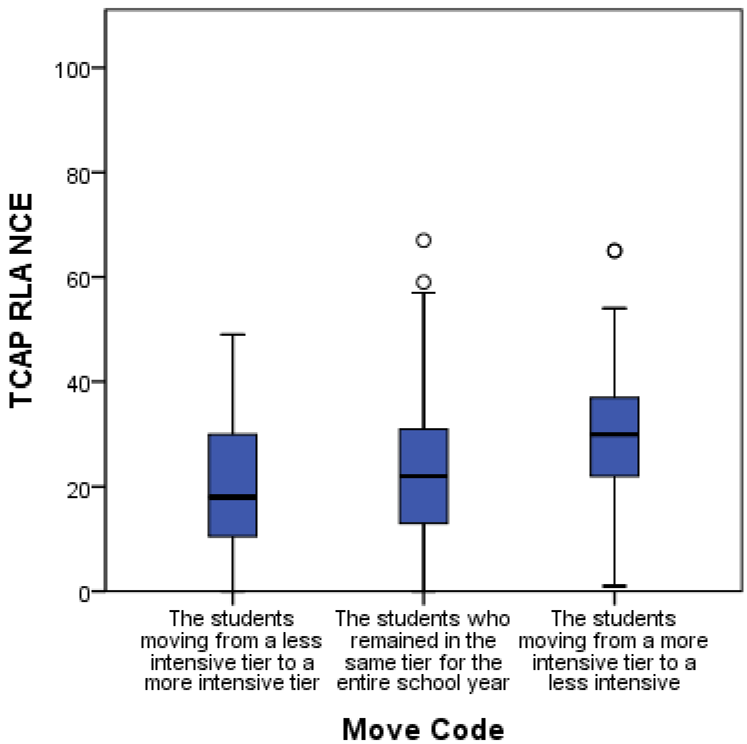

Figure 10: Boxplot of Tier III TCAP RLA NCEs

The longitudinal trends of the mean number of words read correct between the groups allowed us to visualize the differences in performance between the groups of students. In most instances, the differences in mean performance were manifested early in the intervention process, and in most cases, the differences in mean performance were evident before Tier III services were being provided.

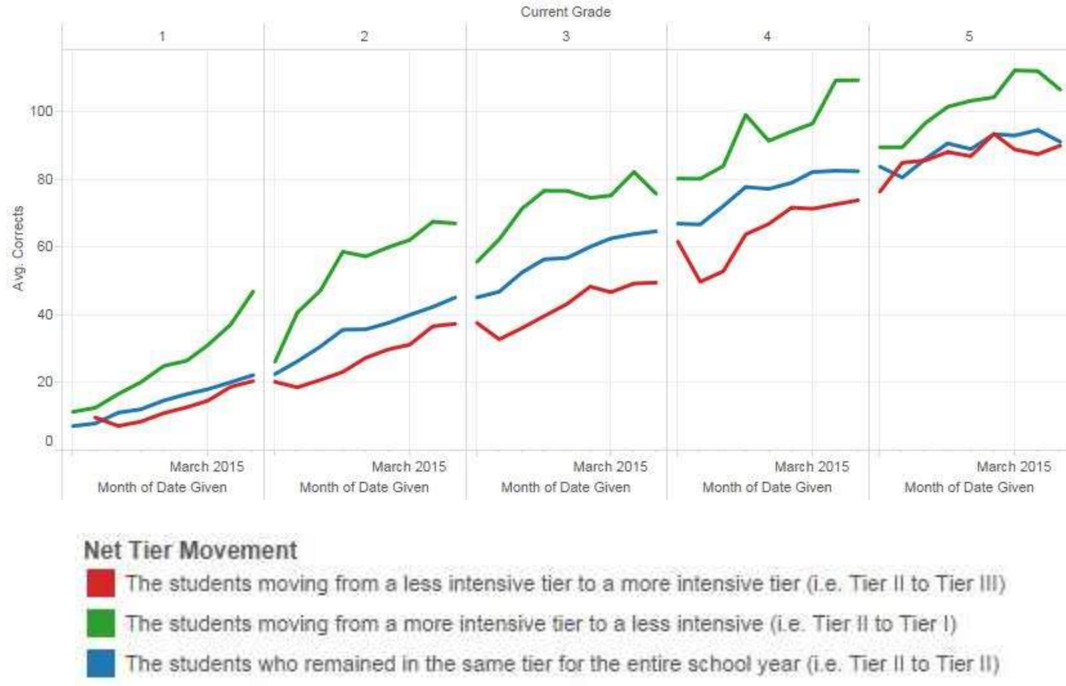

Figure 11: Longitudinal Trends in R-CBM Progress Monitoring Data

It is difficult to come to any real conclusions from the results of the Tier III R-CBM data. There are at least two reasons why the data may have shown conflicting trends. One reason why we may have seen conflicting results using R-CBM data was that the grade level R-CBM results may not have been the data that was used to decide if a child should be formally referred to special education. It is possible that there were more compelling R-CBM data collected off-grade level for students who were formally referred to special education. Additionally, the results of R-CBM testing is related to reading fluency. Students may have been referred to special education because of deficits in reading comprehension (which would be evident in the R-Maze data). The need to have a consistent basis for student performance in this analysis (by monitoring only grade-level R-CBM data), a general dearth of R-Maze data and the quantity of parallel testing events (both R-CBM and R-Maze and grade level and instructional level R-CBM) could have led to some data inconsistencies. However, there were clear trends that differentiated the performance between the students who were eventually moved to Tier II and those that remained in Tier III.

STAR Math

The STAR Math assessment was the most commonly used Tier II Math progress monitoring tool for grades 3 through 5. The data from the statistical analysis of the progress monitoring data, disaggregated by how students moved through the intervention tiers, is presented in the table below.

Table 8: Tier II STAR Math and Math TCAP

Population	Initial SM NCE	Population Means (N Counts) Final SM NCE	ROI (SS/week)	TCAP Math NCE
The students moving from a less intensive tier to a more intensive tier (i.e. Tier II to Tier III)	31.64 (75)	25.33 (75)	0.87 (75)	21.56 (57)
The students who remained in the same tier for the entire school year (i.e. Tier II to Tier II)	34.57 (596)	38.08 (596)	2.30 (596)	30.12 (384)
The students moving from a more intensive tier to a less intensive (i.e. Tier II to Tier I)	34.78 (328)	46.64 (328)	3.25 (327)	37.81 (199)
Levene Statistic	0.209	0.006	0.024	0.306
Sig.	**0.034***	**0.000#**	**0.000***	**0.000***
r	0.08	0.39	0.33	0.12

* ANOVA, # Brown-Forsythe

The statistical testing indicated that we could reject the null hypothesis that the mean initial STAR Math NCE, the final STAR Math NCE, rate of improvement (in scaled score points per week) and TCAP Math NCE were no different between the groups. Tukey post-hoc testing indicated that we could reject that null hypothesis that any of the means were equal to each other except the initial mean STAR Math NCE. In that post-hoc test, we failed to reject the null hypothesis that the initial mean STAR Math NCE was no different between the students who moved to Tier III and the students who remained in Tier II all year. The results seemed to indicate that there were real differences in performance on the STAR Math progress monitoring tool between groups of Tier II students for whom different RTI² decisions were being made. The results also helped to corroborate the assertion that the RTI² teams were using progress monitoring data to inform decision regarding tier placement for students. The effect size indicated that the difference in initial mean STAR Math NCE was small. There were medium effects when considering the differences in mean final STAR Math NCE and mean rates of improvement. However, the effect size was small again when analyzing the differences in mean TCAP NCE.

Longitudinal plots of the mean STAR Math scaled scores helped to visualize the progress of each group of students. Similarly to the results from the STAR Reading analysis, the students who were eventually promoted to Tier I seemed to exhibit higher than expected growth early in the academic year, and thus early in the intervention process (this is especially evident in grades 3 and 4, and less so in grade 5).

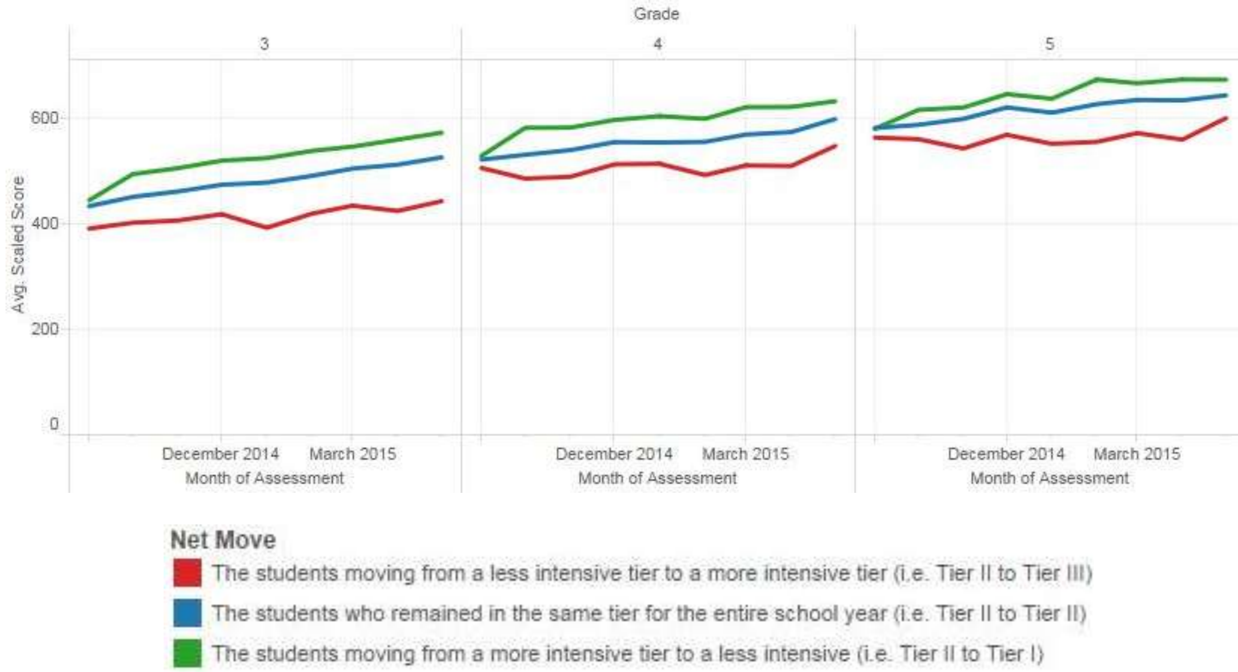

Figure 12: Longitudinal Trends in STAR Math Progress Monitoring Data

Aimsweb M-CAP and M-COMP

The Tier III Math data were even more difficult to analyze than the Tier III RLA data. Progress monitoring was accomplished with two different tools depending on the student's area of need. A majority of students were progress monitored using the Aimsweb Mathematics Concepts and Application assessment (M-CAP), which focused on conceptual understanding and procedural fluency. However, a sizable portion of the Tier III students were progress monitored using the Aimsweb Mathematics Computation assessment (M-COMP) which focused more on mastery of math skills. As with the Tier III reading progress monitoring tools, some students were assessed both on grade level and on their instructional level. On-grade level data were the only data used in this analysis in order to make the comparisons of student performance. There were not enough students who were eventually referred to special education and who were progress monitored using M-CAP and M-COMP to allow inclusion of this group in the study. Non-parametric tests were used for the analysis because the data were not normally distributed.

Table 9: Tier III Aimsweb M-CAP and Math TCAP

Population	Population Means - Population Medians (N Counts) Initial M-CAP NCE	Final M-CAP NCE	ROI (Corr./week)	TCAP Math NCE
The students moving from a less intensive tier to a more intensive tier (i.e. Tier II to Tier III)	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data
The students who remained in the same tier for the entire school year (i.e. Tier III to Tier III)	27.67 - 27.62 (67)	32.84 - 34.39 (67)	0.21 - 0.17 (67)	28.44 - 29.00 (48)
The students moving from a more intensive tier to a less intensive (i.e. Tier III to Tier II)	32.98 - 32.38 (33)	47.92 - 47.33 (33)	0.47 - 0.39 (33)	32.67 - 34.00 (24)
Probability that the Medians are the same across all categories	0.395*	**0.002***	**0.000***	0.278*
Probability that the distribution of scores are the same across all categories	0.152#	**0.000#**	**0.000#**	0.124#

* Independent Samples Median Test, # Independent Samples Kruskal-Wallis Test

We failed to reject the null hypothesis that the distributions and the median of the initial M-CAP NCEs and TCAP Math NCEs were no different between the students who were promoted to Tier II and the students who remained in Tier III. However, we could reject the null hypothesis that there was no difference in the distribution and median of the final M-CAP NCEs and rate of improvement. A visual inspection indicated that the students who were promoted to Tier II generally had the higher scores. A visual inspection of the TCAP data exhibited a directionally higher median and smaller interquartile range for the students who were promoted to Tier II. We failed to reject the null hypothesis that the mean TCAP Math NCE is the same between the groups of students. However, this may (at least partly) be a by-product of the low n-counts for this assessment.

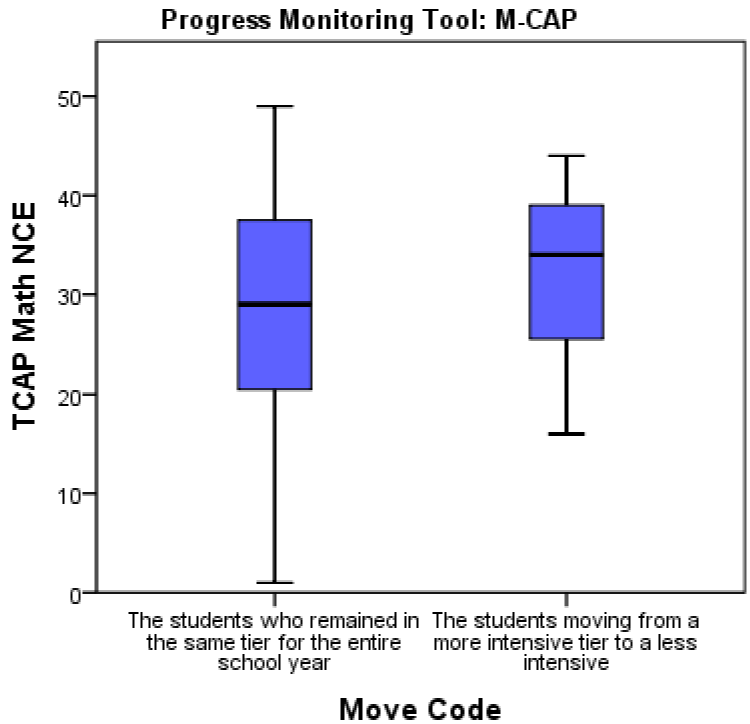

Figure 13: Boxplot of Tier III Math TCAP NCE

Longitudinal trends in M-CAP scores could not be easily interpreted because of the low number of data points.

Table 10: Tier III Aimsweb M-COMP and Math TCAP

Population	Population Means - Population Medians (N Counts) Initial M-COMP NCE	Final M-COMP NCE	ROI (Corr./week)	TCAP Math NCE
The students moving from a less intensive tier to a more intensive tier (i.e. Tier II to Tier III)	Insufficient Data	Insufficient Data	Insufficient Data	Insufficient Data
The students who remained in the same tier for the entire school year (i.e. Tier III to Tier III)	17.14 - 1.00 (237)	31.29 - 35.01 (237)	0.46 - 0.38 (237)	19.04 - 20.00 (157)
The students moving from a more intensive tier to a less intensive (i.e. Tier III to Tier II)	28.62 - 18.34 (55)	49.55 - 46.47 (55)	1.04 - 0.77 (55)	27.07 - 30.00 (45)
Probability that the Medians are the same across all categories	0.639*	**0.031***	**0.015***	**0.000***
Probability that the distribution of scores are the same across all categories	0.106#	**0.000#**	**0.000#**	**0.000#**

* Independent Samples Median Test, # Independent Samples Kruskal-Wallis Test

We failed to reject the null hypothesis that the distributions and the median of the initial M-COMP NCEs were no different between the students who were promoted to Tier II and the students who remained in Tier III. However, we could reject the null hypothesis that there was no difference in the distribution and median of the final M-COMP NCEs as well as the rate of improvement and TCAP Math NCEs. A visual inspection indicated that the students who were promoted to Tier II generally had the higher scores.

The separation in mean final M-Comp scores is evident in the longitudinal graphs of the data. The statistically significant difference in mean initial NCE is also evident.

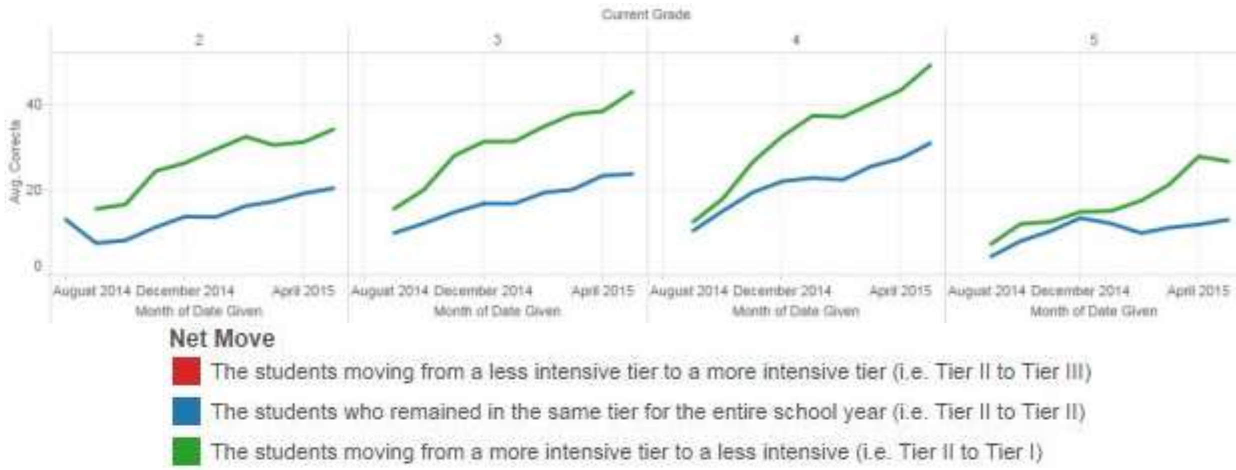

Figure 14: Longitudinal Trends in M-COMP Progress Monitoring Data

Relativity of Tier Placement

Non parametric testing was utilized to compare the medians and distribution of the final progress monitoring NCEs for the students who ended the SY1415 school year enrolled in a tier. The testing indicated that we can reject the null hypothesis that the median and distributions of initial and final NCEs were no different between schools in RLA. The outcome of the analysis of Math scores is less consistent.

Table 11: Final NCE from Progress Monitoring Tool

Final SY1415 Tier	Progress Monitoring Tool	Probability that the median NCE is no different between schools*	Probability that the distribution of NCEs is no different between schools#
II	SR	**0.000**	**0.000**
	SEL	**0.000**	**0.000**
	SM	0.142	**0.001**
III	R-CBM	**0.018**	**0.01**
	M-CAP	0.295	0.147
	M-COMP	**0.015**	**0.013**

* Independent Samples Median Test, # Independent Samples Kruskal-Wallis Test

These results are probably not very surprising, considering the number of elementary schools implementing RTI[2] and the nature of non-parametric testing. Box plots of the data provided visual evidence of the school-to-school variation. An example boxplot from the STAR Reading data generated by the students who ended the SY1415 academic year in Tier II is available below. Generally, the schools that have been high performing on state assessments and had a proportionally smaller population of economically disadvantaged students tended to have higher performing students who remain in Tier II (when compared to other schools). Although the state and district framework indicate that relative enrollment in intervention tiers is acceptable, we must monitor the effect that relative tier enrollment has on students as they matriculate to middle schools. Students who may have come to thrive in intervention may find that there are not enough seats in middle school interventions available to accommodate them. The boxplots for all of the other progress monitoring tools show very similar characteristics.

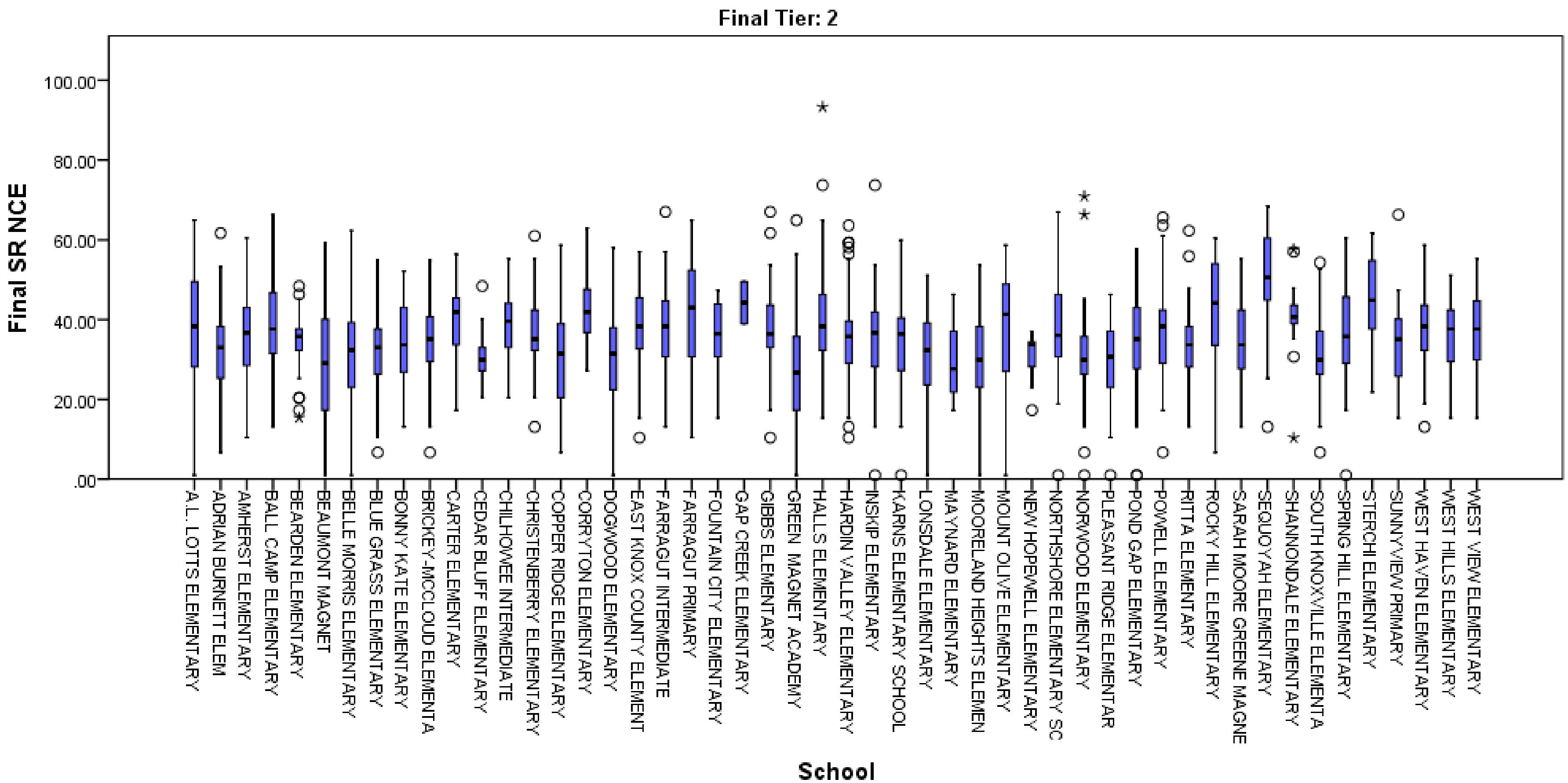

Figure 15: Example Boxplot for STAR Reading NCE by School

Progress Monitoring Sensitivity

The data below provided some evidence that progress monitoring at two week intervals may have been too frequent to actually measure true academic gains given the uncertainty in the test instruments. Personnel at the state Department of Education have been notified of this finding in the hopes that the psychrometric properties of the approved progress monitoring tools will be used to inform the decision making on the required frequency of RTI[2] data point collection.

Table 12: Distribution Statistics of Mean Weeks Required to Meet One Standard Error of Measurement for Each KCS Progress Monitoring Tool

	STAR Renaissance			Aimsweb		
	SR	SEL	SM	R- CBM	M-CAP	M-COMP
1st Quartile	3.11	3.84	3.55	2.41	3.42	2.46
Median	4.68	6.37	5.67	3.66	4.27	3.55
3rd Quartile	7.92	11.62	9.33	5.75	6.01	4.98

Results: Predictive Modeling

The predictive models provided some insight into what data were generally being used by RTI[2] teams to promote a student to a less intensive tier. It was encouraging that the two variables that contribute significantly to every prediction model were the rate of improvement and the most recent test result (in NCE). It was also interesting that the initial progress monitoring NCE was a significant predictor in the Tier III R-CBM model. This provided some evidence that the students who were eventually promoted to Tier II were outperforming the students who remained in Tier III before they even entered intervention.

Tier II STAR Reading

We could create a statistically significant model to provide a probability that a student would move from Tier II to Tier I in RLA given their STAR Reading rate of improvement and their most recent STAR Reading NCE.

$$P_{Tier\ II\ to\ Tier\ I,SR} = \frac{1}{1 + exp^{-(-4.903+0.226*ROI+0.07*NCE_{Recent})}}$$

The model statistics, parameters and the classification table are provided below.

Table 13: Tier II STAR Reading Model Statistics

Chi-square	Sig.	-2 Log likelihood	Nagelkerke R^2
680.075	0.000	2440.747	0.315

Table 14: Tier II STAR Reading Model Parameters

	B	S.E.	Sig.	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Rate of Imp.	0.226	0.028	0.000	1.254	1.187	1.324
Most Recent SR NCE	0.07	0.006	0.000	1.073	1.06	1.085
Constant	-4.903	0.221	0.000	0.007	NA	NA

The results of the logistic regression illustrated the relatively equal consideration that the RTI² teams were giving to the rate of improvement and final NCE when deciding to move a student to Tier I (see Exp(B) values in Table 14).

Table 15: Tier II STAR Reading Classification Table

	Actual	# Correctly Predicted	% Correctly Classified
Moved to less intensive tier	652	190	29.1%
Remained in Original tier	2307	2197	95.2%
		Total	80.7%

The cutoff value of 0.52 minimized the misclassification of the students. The application of the model seemed to result in a conservative probability of a student successfully moving to a less intensive intervention tier (300 students were predicted to move to a less intensive tier from the data whereas 652 actually moved).

The following schools were considered outliers in terms of the percentage students who were predicted to be moved to a less intensive tier but did not: A.L. Lotts Elementary, Farragut Primary, Rocky Hill Elementary, Mount Olive Elementary and Seqouyah Elementary.

Tier II STAR Early Literacy

We could create a statistically significant model to provide a probability that a student would move from Tier II to Tier I in RLA given their STAR Early Literacy rate of improvement, their most recent STAR Early Literacy NCE and their ED status (1 means a student is ED, 0 means a student is not ED).

$$P_{Tier\ II\ to\ Tier\ I,SEL} = \frac{1}{1 + exp^{-(-3.961+0.106*ROI+0.04*NCE_{Recent}+0.64*ED)}}$$

Table 16: Tier II STAR Early Literacy Model Statistics

Chi-square	Sig.	-2 Log likelihood	Nagelkerke R^2
133.645	0.000	840.736	0.22

Table 17: Tier II STAR Early Literacy Model Parameters

					95% C.I. for Exp(B)	
	B	S.E.	Sig.	Exp(B)	Lower	Upper
Rate of Imp.	0.106	0.03	0.000	1.112	1.048	1.179
Most Recent SEL NCE	0.04	0.006	0.000	1.04	1.028	1.052
ED	0.64	0.176	0.000	1.897	1.344	2.678
Constant	-3.961	0.355	0.000	0.019	NA	NA

It was interesting that membership in the ED subgroup actually increased the probability of a student moving from Tier II to Tier I. The high Exp(B) value may have indicated that ED students were responding better to intervention services than non-ED students. This finding provided some evidence of the effectiveness of early intervention in SY1415. This hypothesis assumes that the ED students would have been obtaining exposure to the foundation RLA skills in intervention that they had not previously seen prior to enrollment in elementary schools. This may explain why the ED students responded so well to the Tier II RLA intervention being monitored by the STAR Early Literacy assessment.

Table 18: Tier II STAR Early Literacy Classification Table

	Actual	# Correctly Predicted	% Correctly Classified
Moved to less intensive tier	243	76	31.3%
Did not move to a less intensive tier	546	509	93.2%
		Total	74.1%

The cutoff value that minimized the misclassification of students was 0.52. Again, the model provided a conservative estimate of student movement to Tier I (243 students were promoted to a less intensive tier versus the 133 predicted by the data). There were no outliers when considering the percentage of students who were predicted to be promoted to Tier I but remained in Tier II (with a minimum n count of 30 students).

Tier III Aimsweb R-CBM

We could create a statistically significant model to provide a probability that a student would move from Tier III to Tier II in RLA given their R-CBM rate of improvement, their most recent R-CBM NCE and their initial R-CBM NCE.

$$P_{Tier\ III\ to\ Tier\ II,RCBM} = \frac{1}{1 + exp^{-(-4.965+0.552*ROI+0.029*NCE_{Recent}+0.055*NCE_{Initial})}}$$

Table 19: Tier III Aimsweb R-CBM Model Statistics

Chi-square	Sig.	-2 Log likelihood	Nagelkerke R^2
159.15	0.000	809.008	0.225

Table 20: Tier III Aimsweb R-CBM Model Parameters

					95% C.I. for Exp(B)	
	B	S.E.	Sig.	Exp(B)	Lower	Upper
Rate of Imp.	0.552	0.149	0.000	1.736	1.297	2.324
Most Recent RCBM NCE	0.029	0.01	0.003	1.029	1.01	1.049
Initial RCBM NCE	0.055	0.011	0.000	1.056	1.034	1.08
Constant	-4.965	0.354	0.000	0.007	NA	NA

The relatively high value of Exp(B) for the rate of improvement, coupled with the information gleaned from the boxplots in Figures 7 and 8, provided strong evidence that the RTI[2] teams were focused on rates of improvement when they used R-CBM data to move a student to Tier II.

Table 21: Tier III Aimsweb R-CBM Classification Table

	Actual	# Correctly Predicted	% Correctly Classified
Moved to less intensive tier	168	34	20.2%
Did not move to a less intensive tier	1020	993	97.4%
		Total	86.4%

The cutoff value that minimized the misclassification of students was 0.42. It was less likely that this model would correctly classify a student who moved to a less intensive tier when compared to the Tier II models. This may be due to the fact that Tier III students were commonly progressed monitored in both fluency and comprehension, so one set of outcome data may not have been sufficient to truly capture student performance. It is also possible that some of the students who were moved to Tier II had strong off-grade level R-CBM data (and only a few on-grade level data points).

The following schools were considered outliers in terms of the percentage students who were predicted to be moved to a less intensive tier but did not: Bearden Elementary, Blue Grass Elementary, Rocky Hill Elementary and West View Elementary.

Tier II STAR Math

We could create a statistically significant model to provide a probability that a student would move from Tier II to Tier I in Math given their STAR Math rate of improvement, their most recent STAR Math NCE, their ELL membership (1 means a student is ELL) and BHN membership (1 means a student is BHN).

$$P_{Tier\ II\ to\ Tier\ I,SM} = \frac{1}{1 + exp^{-(-3.4+0.169*ROI+0.042*NCE_{Recent}+0.57*ELL-0.35*BHN)}}$$

Table 22: Tier II STAR Math Model Statistics

Chi-square	Sig.	-2 Log likelihood	Nagelkerke R^2
25.357	0.001	1117.123	0.168

Table 23: Tier II STAR Math Model Parameters

	B	S.E.	Sig.	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Rate of Imp.	0.169	0.048	0	1.184	1.078	1.3
Most Recent SM NCE	0.042	0.007	0	1.043	1.029	1.056
ELL	0.57	0.283	0.044	1.767	1.014	3.079
BHN	-0.35	0.155	0.024	0.705	0.52	0.955
Constant	-3.4	0.357	0	0.033	NA	NA

There were some curious results with this prediction model. It was troubling that membership in the BHN subgroup decreased a student's probability of advancing to Tier I. It was also interesting that ELL membership increased a student's probability of advancing to a less intensive tier.

Table 24: Tier II STAR Math Classification Table

	Actual	# Correctly Predicted	% Correctly Classified
Moved to less intensive tier	304	121	39.8%
Did not move to a less intensive tier	725	634	87.4%
		Total	73.4%

The cutoff that minimized the misclassification of students was 0.42. There were an insufficient number of students at each school to report outliers.

Conclusions & Considerations

The analysis of the Knox County School's implementation of the response to instruction and intervention allowed us to answer the research questions that were posed in the introduction of this study.

What were the relevant enrollment and movement patters of the students in the first year of the Knox County Schools' RTI^2 initiative?

The Knox County enrollment statistics closely matched the theoretical enrollment distributions presented in the state RTI^2 framework. It is possible that KCS was retaining more students in Tier III intervention as suggested by the theoretical distribution and some of the statistical modeling. However, the same evidence indicated that this population would most likely be rather small.

The state RTI^2 framework was designed to address individual student needs. It was clear from the KCS data that the district RTI^2 teams felt that foundational reading skills were the area of greatest need for our elementary students. There were more than four times as many students receiving intervention services in reading instruction when compared to math.

The leadership of the district RTI^2 team were, and continue to be, proponents of early intervention. The grade level enrollment figures indicated that this philosophy was put into practice since the percentage of students enrolled solely in Tier I instruction increased with grade level.

The number of students placed in Tiers II and III were highly variable across the schools of Knox County. This level of variation led to some unique challenges in scheduling and resource allocation. The variation may eventually lead to additional challenges as the intervention students matriculate to the middle schools.

Were the RTI^2 teams making defendable decisions to place students in the appropriate intervention tiers and move the students to appropriate tiers per their progress monitoring data?

There were many indicators that, in general, the RTI^2 teams were making defendable decisions regarding intervention enrollment and movement between intervention tiers.

In SY1415, KCS promoted about 25% of intervention students from a more intensive tier to a less intensive tier. When intervention services for SY1516 were launched, only about 12%

of the students who were promoted to a less intensive tier in SY1415 returned to the more intensive tier. This provided some evidence that RTI^2 teams were generally not recommending moves to less intensive tiers before students were prepared for it.

The data from both subject areas exhibited characteristics of data collected from populations with significant differences in end-of-the-year Tier II performance (measured by the progress monitoring tools and between the students who moved to a less intensive tier, moved to a more intensive tier or remained in the same tier through the course of the year). The longitudinal trends in the mean Tier II data seemed to suggest that the growth of the students who eventually graduated to Tier I occurred relatively early in the process. This may have been due to the students' response to intervention, but it may also have been that these students were misclassified as at-risk students due to erroneous data from their first screening (i.e. low scores due to lack of student motivation, uncertainty of how to use the assessment tool, students having a "bad day", etc.). The data were not collected in a manner that could determine a causal relationship between intervention and gains in student performance.

The longitudinal differences in the R-CBM results exhibited characteristics of data collected from populations with significant differences in Tier III performance (on progress monitoring tools and between students who moved to a less intensive tier, moved to a more intensive tier or remained in the same tier through the course of the year). However, in contrast to the Tier II trends, there was some evidence that the differences in performance was evident before any SY1415 intervention services were administered. This may indicate that the students who were eventually moved to Tier II from Tier III may have belonged in Tier II from the start of SY1415. This information, coupled with the percentage of the elementary student population that ended the year in Tier III intervention provided, some evidence that we may have over-identified students for Tier III services.

It was evident that the RTI^2 teams were making decisions based on the data from the progress monitoring tools since predictive models built on the progress monitoring data were statistically significant. It was also evident that teacher input helped to drive decision making, since the prediction model created from progress monitoring data did not account for all of the variation in student placement. However, it is important to note that the RTI^2 teams must understand the limitations of the tools that they have been given to monitor student progress, and continue to use their professional judgment when making any decisions regarding intervention. The current progress monitoring tools that KCS is utilizing are likely not sensitive enough to really capture student growth at two week intervals. A better understanding of the uncertainty in the RTI^2 data should aid in decision making and help to ensure that we are not over-testing the students in Tiers I through III. Minimizing the time spent testing will help to maximize the amount of time each student is receiving leveled instruction.

Similarly, there is a delicate balance between conservative decision-making and leaving a student in a more intensive tier longer than their progress really warrants. Generally, the RTI^2 teams in the district would err on the side of students remaining in more intensive tiers. The district RTI^2 team must decide if this is a strategy that KCS wants to continue, or provide some more structured criteria for tier enrollment. One shortcoming of the state and district RTI^2 frameworks is that each framework is proscriptive about when a student should be enrolled in Tier II or Tier III, but relatively vague about when a student should move to a less intensive tier.

What processes and procedures need to be refined in order to improve the RTI^2 process for SY1516?

The largest hurdle to the implementation of the RTI^2 framework was the time requirements placed on the staff to implement the RTI^2 framework as intended. Most of the teachers who were interviewed as part of this study felt that monthly RTI^2 team meetings were productive and enlightening. However, our skilled educators (from teachers to academic coaches) were using planning time to compile attendance records, test results and complete paperwork for monthly RTI^2 team meetings. This does not seem like a wise use of time for our specialized staff members when a centralized clerical position or specialized software could accomplish many of the same tasks. Some schools had the budget to allow them to hire extra staff to deal with the increased workload from RTI^2, or had such small numbers of students in intervention that no additional support was needed. The schools that seemed most likely to feel the resource crunch from RTI^2 were the schools that received little to no Title 1 funds, but had a relatively large population of students in Tiers II and III. These schools should probably be the first to be targeted if the district is looking for a place to prioritize RTI^2 support for administrative tasks.

The district must also ensure that they have proper tools for implementing RTI^2. The staff that were interviewed as part of this study were generally happy with the intervention programs offered in RLA but were less positive regarding the intervention programs available in Mathematics. The scripted nature of S.P.I.R.E. and Voyager allowed resource-strapped schools to still feel that the interventions were implemented with a great deal of fidelity, but there is currently no similar math intervention program available at the district level. In addition, the district currently has no large scale intervention to target student non-cognitive skills. For example, a lack of motivation was commonly cited as a rationale for the poor performance of students on the progress monitoring tools. However, continued enrollment in an intervention that targets foundational Math and Reading skills is unlikely to solve motivation issues. Knox County should investigate a holistic intervention model that fully integrates the non-cognitive characteristics of students as a compliment to the academic RTI^2 framework.

Finally, the number of data points required for a formal referral to special education is still largely viewed as a “wait to fail” model. Although it is understandable that the state requires quantitative evidence supporting the placement of students in special education, the RTI^2 teams want to make sure that students truly receive the services that they need without delay. The district must promote practices that meet both of these needs simultaneously in order to ensure that we are meeting individual student needs in a timely manner.