Get Ready to Learn Randomized Control Trial

AppleTree Institute conducted a randomized control trial (RCT) to better-understand the impact of *Get Ready to Learn*, the first unit of our *Every Child Ready* instructional model. AppleTree worked with 51 classrooms in 26 community-based early childhood centers in Cuyahoga County, Ohio to evaluate the impact of GRTL curriculum and coaching content on teacher and child outcomes. The applied research project spanned 3 months and yielded four general conclusions about GRTL:

1. Classrooms that used GRTL with at least a moderate degree of fidelity saw positive results for both teachers and children.
2. Teachers had improved emotional support, classroom organization, and instructional strategies.
3. Children had improved social-emotional skills, self-regulation skills, initiative, and attachment to adults and peers.
4. Children learning in moderate-risk centers (based on state child care licensing regulations) experienced greater social-emotional benefits.

Get Ready to Learn Study Overview

Procedures

Get Ready to Learn (GRTL) is a four-week, full-day program for 3- and 4-year-olds, designed to build social-emotional and approaches to learning skills. GRTL is the first unit in the *Every Child Ready* (ECR) instructional model and focuses on acclimating children to the school environment and helping teachers use sustainable, developmentally appropriate classroom management practices and positive behavior supports.

51 classrooms in 26 community-based early childhood centers were randomized into a treatment or control group. Randomization happened at the center level, meaning all classrooms within a center were assigned to the same group. 13 centers were randomized into the treatment group and implemented GRTL. 13 centers were randomized into the control group and continued with business as usual (non-GRLT) practices in their classrooms.

Resources for the 26 GRTL included:

- A printed manual of all GRTL content including lessons and reproducibles.
- All of the materials necessary to implement GRTL.

- Two copies of each GRTL book, both trade and ATI-published.
- Ongoing coaching and professional learning focused on high-quality classroom environment, classroom management, and social-emotional teaching strategies.
- Lots of extra materials

Non-GRTL classrooms received all of the same content and resources after the project ended and all outcome data were collected.

Participants

Participating centers were all licensed child care providers in Cuyahoga County, Ohio. Classrooms were made up of children between 2- and 5-years-old (average age at baseline was 3 years, 6 months). Any classrooms in a center with children in these age ranges were included, with between one and four classrooms per center participating in the study.

All centers had an assigned Step Up To Quality Rating (SUTQ) star rating. Ratings at time of recruitment were used to randomize centers into the treatment or control group using a block design randomization procedure. This is a five-star quality rating and improvement system used by the Ohio Department of Education and the Ohio Department of Job and Family Services to rank childcare centers on multiple aspects of quality. These ratings are based on four general categories: Learning and Development, Staff Qualifications and Professional Development, Administrative and Leadership Practices, and Family and Community Partnerships. Ratings are tied to state licensing and funding. Participation in the SUTQ system is required for all publicly funded early learning programs.

Randomization occurred in 3 blocks, with block 1 including 1-star centers (n = 22), block 2 including 2-star centers (n = 8), and block 3 including 3-, 4-, and 5-star centers (n = 6).

Measures

To determine the effectiveness of GRTL, all 51 participating classrooms were measured on multiple teacher and child outcomes.

Teacher Outcomes. Teacher outcomes were measured using three direct observations and one teacher survey. The Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) and the Quality Indicators (QI; AppleTree Institute, 2011) are observational tools that focus on general classroom quality across multiple domains. The Preschool Observation of Social-Emotional Teaching (POST: Mathis & Hartz, 2019) is an observation of classroom social-emotional practices. Finally, the Social Emotional Learning Classroom

Practices Scale - Teacher Report (SEL-CP-TR; Sutton, 2016) is a teacher survey that captures teachers' perceptions about their own social-emotional teaching practices.

Child Outcomes. Child outcomes focused on broad social-emotional skills and self-regulation. These data were collected through two teacher reports and one direct assessment. The Devereux Early Childhood Assessment for Preschoolers, Second Edition (DECA-P2; LeBuffe & Naglieri, 2012) and the Positive Behavior Rating Scale (PBRS; AppleTree Institute, 2011) are teacher-completed rating scales that focus on children's social-emotional development. The DECA breaks ratings down into two broad subscales: Total Protective Factors (TPF) and Behavioral Concerns. TPF consists of ratings in the areas of Initiative, Self-Regulation, and Attachments/Relationships. The PBRS is reported as a total score and two subscales: Behavioral Self-Regulation and Self and Social Awareness. Both rating scales were completed by teachers for all children in the treatment and control groups pre and post GRTL implementation.

The Preschool Self-Regulation Assessment (PSRA; Smith-Donald, Raver, Hayes, & Richardson, 2007) is a direct assessment of children's self-regulation skills. The PSRA is broken down into three subscales: Behavioral Control, Emotional Control, and Compliance. These are important aspects of children's school readiness and approaches to learning skills (Blair, 2002). PSRA direct assessments were administered by trained external contractors for all children in the treatment and control groups pre and post GRTL implementation.

More detailed descriptions of all teacher and child measures are available in Appendix A.

Get Ready to Learn Implementation Fidelity Rubric. As part of the study design we created an implementation fidelity rubric used to rate treatment group classrooms on their fidelity to the GRTL program. Fidelity was broken out into three distinct areas: High Quality Environment, Rules and Routines, and Application of Coaching and Content. Ratings were assigned in each of these areas, as well as an overall fidelity score. Multiple data sources were used to rate classrooms, including coaching notes and observational data.

Center Moderate Risk Non-Compliance Reports. As part of the SUTQ rating systems, centers receive multiple onsite licensing visits throughout the year. During these inspections licensing personnel complete a paperwork review and conduct observations in classrooms. These inspection reports are publically available through the Ohio Department of Jobs and Family Services. For the purposes of this study, we examined the number of low, moderate, and serious risk non-compliances for each center, using these as an additional proxy for quality. Low

risk non-compliances include things such as appropriate completion of incident reports, up-to-date medical statements included in employee record charts, and complete child enrollment forms. Moderate risk non-compliances include things like programs being out of teacher:child ratio, timely completion of staff background checks, and incidences of children being left unsupervised while in the building. Serious risk non-compliances include things such as an administrator refusing access to licensing personnel, medication being administered to the wrong child, or children being left unsupervised while outside of the building.

We determined low-risk non-compliances to be of low enough risk as to not be associated directly with child outcomes. Only one center had a serious risk non-compliance. There was high variability in the number of moderate-risk non-compliances across centers, and these types of violations were more clearly associated with child outcomes, so this variable was used as a control in our analyses.

Results and Discussion

GRTL is designed to improve teacher quality and children's social-emotional skills, so we focused on these two areas to understand how GRTL was associated with those important early pieces of early learning. Importantly, we also rated teachers' fidelity to the GRTL program. That is, did GRTL classrooms in this project actually do GRTL? Results emerged in three categories.

Treatment and Control Comparison

We began by comparing the treatment group (GRTL implementers) and the control group (business as usual implementers) on all outcomes. A key assumption of comparisons in a randomized control trial is establishing baseline equivalence between the two groups. This process ensures that the two groups start out the same on all variables of interest. This essentially answers the question, did randomization work. In the case of GRTL, the two groups were not equivalent on all measures, meaning that control and treatment classrooms did not start out in the same place in terms of quality. Baseline equivalence was established for DECA Total Protective Factors, DECA Behavioral Control, PBRS total score, CLASS Emotional Support, and CLASS Instructional Support. There was not baseline equivalence for the PSRA, CLASS Classroom Organization, and any domains of the QI, or POST. This makes it difficult to interpret comparisons between the two groups on the non-equivalent measures.

To estimate the overall impact of the GRTL RCT, an intent-to-treat (ITT) analysis was utilized as the initial analytic approach. The ITT analysis is the recommended approach for

RCTs because it provides unbiased estimates of the effect of participating in the treatment of interest, in this case GRTL. This is achieved through comparing mean differences in outcomes between the entire assigned treatment group and the entire assigned control group. Analyses were conducted in a multilevel framework to account for nesting in the data at the classroom and site levels. Models also included multiple control variables: baseline measures for all outcome variables, randomization block (block 1, 2, or 3), and the number of moderate risk violations at each site. ITT impact models were run to test for main effects of the GRTL program on measures where baseline equivalence had been established. ITT models are not recommended when there are not equivalent groups.

Results from Impact Models.

As a precursor to the ITT impact models, power analyses were run for all study variables to determine minimum detectable effect sizes (MDES). In studies such as this one it is common to consider MDES as a marker for significance, as opposed to p-values alone. Based on sample size and other components of the model (variance at each nesting level, variance accounted for by covariates, etc.), MDES's showed that for all classroom-level variables we would need effect sizes of at least d = .9 to obtain significant results, and for all child-level variables we would need effect sizes of at least d = .5 to obtain significant results. Effect sizes in developmental and educational research are often low, in the range of d = .2 to d = .5. Due to the high MDES results from the power analysis, we expected some non-significant results in terms of p-value, but we also expected these results to have acceptable effect sizes based on standards in the field.

Results from the impact models suggest that children in the treatment group have higher overall PBRS scores at outcome than children in the control group (β= .53, p = .12, d = .61), and higher PBRS self and social awareness: β= .67, p = .04, d = .72). Children in the treatment group are also seen to have higher levels of DECA behavioral concerns at outcome than children in the control group (β= 10.86, p = .12, d = .40). No results emerged for DECA Total Protective Factors or CLASS Emotional Support and Instructional Support.

Children in the treatment group had higher social-emotional ratings and higher teacher-reported behavioral concerns than the control group.

GRTL coaching and professional learning focused heavily on supporting broad social-emotional learning in the classroom, so it is likely that teachers were more aware of and focused on promoting these skills, accounting for the increased social-emotional ratings for

treatment group children. Regarding the greater level of behavioral concerns, it may be that teachers who receive more professional development around these areas become more aware of behavioral concerns in the children in their classrooms. This is a phenomenon that has been reported in similar interventions, with teacher-report tools focused on concerning behaviors increasing for treatment group participants.

These minimal significant findings were unexpected. In an effort to better understand why the associations between GRTL and teacher and child outcomes we explored the question of fidelity to the GRTL program. Did classrooms in the GRTL treatment group actually use GRTL?

Understanding the Treatment Group

In an effort to better understand the results and the effect of GRTL, the analytic approach was shifted to a sensitivity analysis with the goal of uncovering trends and effects within the treatment group. An implementation fidelity rubric was created using multiple data sources. All classrooms were rated on their implementation of the GRTL program in three distinct categories; High Quality Environment, Rules and Routines, and Application of Coaching and Content. Classrooms were also given an overall fidelity score and a fidelity group assignment. 13 classrooms were rated as low fidelity and 12 as moderate fidelity. No classrooms implemented with high fidelity. Of note, high-fidelity implementation is what we see in the majority of our full ECR implementation classrooms in D.C.

Impact models were re-run within the treatment group to examine the effect of fidelity on classroom and child outcomes. The same control variables were included in the impact models, and fidelity group (low or moderate) was added as a predictor to the models. Analyses were conducted on all teacher and child outcomes in a multilevel framework to account for nesting in the data at the classroom and site levels.

Results from Treatment on the Treated Analyses.

Results suggest that classrooms in the moderate fidelity group have higher outcomes in CLASS Classroom Organization, QI Support for Diverse Learners, QI Direct Instruction, QI Independent Learning and Guided Practice, QI General Process Quality, QI General Structural Quality, and POST ratio of Total Praise:Redirections and POST Management and Classroom Climate. Children in moderate fidelity classrooms had higher social-emotional scores on PBRS. Classrooms in the moderate fidelity group are higher than the low fidelity group by .611 points in

CLASS Classroom Organization (p = .084), 1.03 points in Support for Diverse Learners (p = 0.03), .71 points in Direct Instruction (p = .051), 1.05 points in Independent Learning and Guided Practice (p = .002), .93 points in General Process Quality (p = 0.03), 1.14 points in General Structural Quality (p = 0.00), 1.06 points in Total Praise:Redirections (p = .078), and .66 points in Management and Classroom Climate (p = .028).

In addition to overall fidelity to GRTL, fidelity to Rules and Routines was particularly impactful for outcomes. Classrooms that more closely adhered to rules and routines by setting appropriate expectations for children and providing them with predictable routines, self-soothing strategies, and lots of positive behavior reinforcement had higher outcome scores on CLASS Classroom Organization (β= 1.29, p = .00) and CLASS Emotional Support (β= .863, p = .026). Children in these classrooms had higher scores on DECA Initiative (β= , p = .), overall social-emotional ratings on PBRS (β= .57, p = .011), and the PBRS Behavioral Regulation(β= .48, p = .074) and Self and Social Awareness (β= .71, p = .001) subscales.

When teachers implement GRTL with at least a moderate degree of fidelity they are likely to see positive results for their children and themselves.

In order to see the hypothesized benefits of the GRTL program, teachers must use the GRTL resources and implement the best practices in social-emotional teaching that were the focus of GRTL professional development. By examining fidelity to this model, we were able to determine that non-implementers, or those with low fidelity, did not see meaningful changes in outcomes over the course of the program, while partial-implementers. Of note, no participating classrooms implemented with high fidelity. This is not unusual during the first year of a new program in a new setting. Increases in implementation fidelity often occur after users have time to understand the program and incorporate a program into their daily practice. This is typical of or D.C. partners as well, with new partners taking at least two years to begin to see meaningful benefits from the full Every Child Ready curriculum. Importantly, these findings confirm that even if teachers do not implement the program with high fidelity, they will still see positive outcomes with partial implementation.

In understanding what may lead to a classroom implementing with moderate fidelity, these classrooms had teachers who were willing to try something new and were open to using the resources provided and working on the strategies they learned through coaching. These findings are important, as they will guide implementation science and planning for future partnerships.

As a final step we explored how these varying fidelity classrooms differed from the control group classrooms on teacher and child outcomes. A clear pattern emerged, described below.

Exploratory Comparison of the Different Fidelity Groups and the Control Group

In an effort to further understand the impact of fidelity on outcomes, a final exploratory analyses was conducted. Similar to the models in the treatment on the treated analyses, new models were run comparing the low fidelity group, the moderate fidelity group, and the control group. The control group sites were all assigned fidelity scores of zero for the purpose of this analysis. Comparisons between these three groups were conducted through the lens of overall fidelity, as well as the three fidelity categories.

Children in classrooms that implemented GRTL with moderate fidelity had better outcomes in multiple areas compared to the low fidelity and control groups. Overall, low fidelity and control were very similar, while moderate fidelity outperformed both. This reinforces the conclusion that to see the positive benefits of GRTL, it should be used in settings where teachers are able to implement with at least moderate fidelity.

Results of Exploratory Analysis. Results emerged in similar patterns for the three groups (low fidelity, moderate fidelity, control), with some variations based on category of fidelity. Classrooms implementing with moderate fidelity in the area of Rules and Routines had higher QI General Process Quality (β= .54, p = .039) and Independent Learning and Guided Practice scores than teachers in low-fidelity and control classrooms (β= .71, p = .008). The same pattern was true for child outcomes. Children in moderate-fidelity classrooms in the areas of Rules and Routines had higher total social-emotional PBRS ratings compared to children in low-fidelity classrooms and those in control classrooms (β= .30, p = .05). Children in moderate-fidelity classrooms in the areas of Rules and Routines and Application of Coaching and Content had higher DECA Attachment ratings (β= 11.25, p = .031 and β= 13.23, p = .023, respectively). Children in moderate fidelity classrooms in the areas of Rules and Routines and High Quality Environment had higher PBRS Self and Social Awareness (β= .44, p = .003 and β= .29, p = .025, respectively) compared to children in low-fidelity classrooms and those in control classrooms.

To further understand these relationships, interactions were tested between treatment and center-level moderate risk violations. The purpose of the interaction models is to determine if the effect of GRTL varies based on the number of moderate risk violations a site has. An

interaction term was created and included as a predictor in the initial impact models described above. The treatment * risk interaction was significant for CLASS Instructional Support, total social-emotional PBRS ratings, PBRS Behavior Regulation, DECA Self-Regulation, and DECA Behavioral Concerns, such that sites with higher instances of moderate risk violations benefited more on all five outcomes compared to sites with fewer moderate risk violations.

Moderate-Fidelity Implementers See Stronger Results than Both Low Fidelity Implementers and Control Group Classrooms.

As discussed above, at least a moderate degree of fidelity to the GRTL model is needed to yield positive outcomes. These exploratory analyses further confirmed this conclusion, as the moderate-fidelity group outperformed both the low-fidelity group and the business-as-usual control group. Although we are unable to confirm extensive causal results from our overall impact models, which necessitates grouping low and moderate fidelity classrooms in the same group, these findings suggest that when teachers use GRTL, it works better than business as usual.

Conclusions

Taken together, these findings suggest that GRTL has positive impacts for sites and classrooms that implement the program with at least a moderate degree of fidelity, and is especially effective for those who implement well in the area of Rules and Regulations. Additionally, sites with higher instances of moderate risk violations (which is likely an indicator of center quality) benefit more from GRTL than programs with fewer instances of moderate risk violations. All children saw benefits, but these benefits were greater for children in more high-risk settings.

We analyzed how the low and moderate groups performed. Classrooms, where the GRTL model was implemented with moderate fidelity, had higher teacher quality and child social-emotional skills than classrooms that implemented with low fidelity. This was particularly true for classrooms that closely adhered to the rules and routines piece of fidelity. This included setting appropriate expectations for children and providing them with predictable routines and lots of positive behavior reinforcement.

Appendix A: Descriptions of Teacher and Child Measures

The Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008): The CLASS is an observational tool for pre-K classrooms was completed at baseline and outcome. Scores range from 1 to 7, with 7 representing the highest quality. Scores are aggregated into three broad domains: Emotional Support, Classroom Organization, and Instructional Support. Each domain is made up of three to four dimensions which are averaged to create the aggregate domain scores. CLASS observers must meet rigorous reliability standards prior to observing.

The Quality Indicators (QI; AppleTree Institute, 2011): The QI observational tool was completed at baseline and outcome. Scores on the QI range from 1 to 6, with 6 representing the highest quality. Scores are aggregated into five broad domains: Support for Diverse Learners, Direct Instruction, Independent Learning, General Process Quality, General Structural Quality. Each domain is made up of four indicators which are averaged to create the aggregate domain scores. QI observers must meet rigorous reliability standards prior to observing.

Preschool Observation of Social-Emotional Teaching (POST; Mathis & Hartz, 2019): The POST is a broad measure of social-emotional teaching behaviors that includes a 10 minute observation and frequency count and a 20 minute observation and subsequent classroom climate likert-scale rating. The frequency count includes teacher behaviors such as praise, redirections, effective commands, and ineffective commands, which are tallied and totaled during a 10 minute window. The Classroom Management and Climate rating scale uses a 5-point likert scale to assess 8 components of classroom climate such as whether teachers obtain/maintain children's attention and if teachers follow the classroom schedule.

The Social Emotional Learning Classroom Practices Scale - Teacher Report (SEL-CP-TR; Sutton, 2016) is a 24-item survey designed to gauge the presence of classroom practices that support children's social-emotional learning. The teacher respondent indicates how often they use the practice or strategy described in each item on a 5-point scale ranging from 1 (Never) to 5 (Almost Always). Example items include, *I teach active listening practices (e.g., have child summarize what they just heard from me or another child)* and *My child and I discuss different strategies for resolving conflicts*.

The Positive Behavior Rating Scale (PBRS; AppleTree Institute, 2011) is a teacher-completed 10-item survey designed to measure a children's social-emotional development. The teacher indicates how often a child exhibits the strength described in each item on a 5-point scale ranging from 1 (Never) to 5 (Always).

The Devereux Early Childhood Assessment for Preschoolers, Second Edition (DECA-P2; LeBuffe & Naglieri, 2012) is a teacher-completed 38-item survey designed to measure how often a child exhibits positive or negative behaviors in the previous four weeks on a 5-point scale ranging from 0 (Never) to 4 (Very Frequently). Each item is attributed to one of four scales: Initiative, Self-Regulation, Attachments, and Behavioral Concerns. The first three

domains are also measured in a composite Total Protective Factors scale. The DECA-P2 manual provides nationally normed t-scores and percentile ranks associated with the sum of the raw scores of the items on each scale.

Preschool Self-Regulation Assessment (PSRA; Smith-Donald, Raver, Hayes, & Richardson, 2007) is used to measure underlying executive function skills such as self regulation, inhibitory control, and working memory. The PSRA yields three subscales: Behavioral Control, Emotional Control, and Compliance. The assessment is administered in a one-to-one setting with a trained assessor. Administration includes multiple engaging tasks, such as sorting toys, walking on a pretend balance beam, and building a tower. In this study we used a modified version of the PSRA, excluding any items that involved food (Snack Delay and Tongue Task), as many early childhood programs have policies around using food for non-nutritional practices.