



SUMMARY MEMO:

Kidznotes RCT Evaluation

December 2020

OVERVIEW & BACKGROUND

Kidznotes is a North Carolina-based orchestral program that uses the Venezuelan El Sistema model to alter the life trajectory of historically underprivileged youth. Founded in 2010, the nonprofit aims to build a network of children, families, and partners in which the passion for music transforms lives and communities in North Carolina's Triangle Region. To this end, Kidznotes partners with Title-1 public elementary schools, where at least 50% of students qualify for free or reduced-price lunch. Research suggests that there are a multitude of benefits associated with music training initiated at an early age¹, including more complex brain development, higher school achievement (e.g., math, reading, SAT scores, graduation rates, etc.), improvements in executive functioning and social emotional learning², and a host of other gauges of long-term social, economic, and emotional well-being³.

Duke University's Social Science Research Institute (SSRI) has partnered with Kidznotes to better understand the effects of youth orchestral training on youth participants and their families. In 2017-18, SSRI's evaluation team, in partnership with Kidznotes, initiated a randomized control trial (RCT) to examine the effects of Kidznotes programming on: (1) youth processes and outcomes in social emotional development, executive function, school engagement, and academic performance; and (2) parental/familial processes and outcomes in social capital and educational

¹ E.g., see Creech, Andrea, Patricia Gonzalez-Moreno, Lisa Lorensino, and Grace Waltman. 2016. *El Sistema and Sistema-inspired programmes: Literature Review of research, evaluation, and critical debates*. Sistema Global. Available 1 August 2016 at <http://sistemaglobal.org/literature-review/>.

² See Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: a relation mediated by executive functions. *Music Perception*, 29, 195-201 and Rickard, N. S., Appelman, P., James, R., Murphy, F., Gill, A., & Bambrick, C. (2013). Orchestrating life skills: The effect of increased school-based music classes on children's social competence and self-esteem. *International Journal of Music Education*, 31(3), 292-309.

³ <https://www.wsj.com/articles/a-musical-fix-for-american-schools-1412954652>

involvement. The study implemented original data collection across the 2018-19 and 2019-20 school years with kindergarteners and first graders – the age of Kidznotes’ youngest participants – as well as their teachers and parents. As data collection associated with the original study design⁴ comes to a close, the evaluation team has turned to continued data management and preliminary analyses, with a focus on comparing outcomes for students who did and did not participate in Kidznotes programming, based on random assignment.

This memo focuses on initial data analysis, with specific emphasis on preliminary findings related to youth outcomes in executive function and social emotional development. Future analyses will also focus on parental/familial outcomes in social capital and educational involvement, as well as youth outcomes in school engagement and academic performance. This memo summarizes information to be shared more fully in a forthcoming report and presents key findings that may be integrated into regular Kidznotes program processes. It starts by addressing driving questions and methods (p. 2), followed by key findings (p. 4), and finally, initial program recommendations (p. 10).

QUESTIONS & METHODS

Research Questions

Evaluation processes addressed the following overarching questions:

1. To what degree, if any, does Kidznotes programming affect youth outcomes in social emotional development and executive function (which are associated with longer-term academic success)?
2. To what degree, if any, does Kidznotes programming affect parental/familial outcomes in the extent to which families become involved in child’s education (educational involvement practice), the extent to which families feel they have the ability to help their child do well in school (educational involvement self-efficacy) and the extent to which families feel part of an educational community (social capital development)?
3. How might these effects vary by youths’ race/ethnicity, gender, income level, and degree of engagement?

Target population/Participants

⁴ Funding from the National Endowment of the Arts has allowed for additional data collection to occur in order to examine the effects of multiple years of participation, which is currently ongoing (see future directions, p. 9 of this memo).

To be eligible for the study, youth participants had to meet the following criteria:

- 1) be an eligible applicant to Kidznotes for the study year;
- 2) enrolled in Kindergarten or 1st grade at a participating Kidznotes partner school; and,
- 3) have no prior participation in Kidznotes.

Families who met these criteria were invited to enroll in the study at the time of program application. Families were informed that their decision to participate (or not) in the study would have no bearing on whether they are accepted into Kidznotes. Following program application, families who consented to participate in the study were identified and randomly assigned to be accepted into Kidznotes (treatment) or not (control). This selection process took advantage of the fact that the number of program applicants typically exceeds program capacity and ensured that the treatment and control groups were as similar as possible in terms of interest and need for Kidznotes programming. Following random assignment, families were informed of their acceptance (or not) to Kidznotes and were reminded of their enrollment in the RCT. The evaluation team then communicated with partner schools in order to locate youth participants within teacher classrooms, and teachers with study-enrolled youth were invited to participate in the study (reporting on youth outcomes).

Data were collected and pooled across two cohorts (i.e. 2018-19 and 2019-20 academic years) in order to match Kidznotes' operating capacity (e.g. approximately 50 "Mozart" students per year across Durham and Wake counties), while also reaching a target sample size of 200 participants (100 control and 100 treatment participants). Across two cohorts, 218 youth and their families were enrolled in the study, with 138 participants in Cohort 1 and 80 participants in Cohort 2. Fourteen participants were removed prior to the start of data collection because they were no longer eligible to participate in the study (e.g. students were no longer enrolled in the partner school indicated on the Kidznotes application, students were not enrolled in the grade indicated on the Kidznotes application, etc.), resulting in 204 participants who were eligible for data collection. Additionally, 17 participants were removed prior to conducting analyses because they were assigned to the treatment group but participated in less than 20% of Kidznotes classes. Of the resulting 187 participants included in the analysis, 45% received the treatment (i.e. participated in Kidznotes) while 55% were part of the control group (i.e. did not participate in Kidznotes). This subset of participants that are included in the analyses closely matches the demographic breakdown of the initial sample that was enrolled at the beginning of the study, with participants more likely to be enrolled in

Kindergarten (61%), attend a Durham school (61%), identify as female (55%), Black (44%), or Hispanic (35%), and qualify for free and reduced lunch (67%). See Table 1 for breakdown by demographic.

Table 1. Similar demographic breakdown for participants enrolled in the study and those included in the analysis

Table 2. Demographic breakdown by treatment and control for students in the analysis (n = 187)

number of students		treatm ent	control
Grade	K	27%	34%
	1	18%	20%
Gender	Male	21%	24%
	Female	24%	31%
District	Durham	27%	34%
	Wake	19%	20%

		Total enrolled in study	Total eligible for analysis	Total included in EF analysis	Total included in SEL analysis
Total N		218	187	110	106
Condition	Treatme nt	50%	45%	45%	42%
	Control	50%	55%	55%	58%
Grade	K	62%	61%	62%	61%
	1	38%	39%	38%	39%
Gender	Male	45%	45%	45%	42%
	Female	55%	55%	55%	58%
District	Durham	61%	61%	66%	59%
	Wake	39%	39%	34%	41%
Race	Hispanic	35%	35%	44%	36%
	Black	45%	44%	40%	45%
	White	16%	17%	20%	17%
Socioeconomic Status	Low SES	66%	67%	74%	70%
	High SES	34%	33%	26%	30%

Materials and Methods

Data, which focus primarily on youth and parental outcomes, were collected from four sources (see appendix A for a more detailed description of each measure):

- *Youth assessments (Executive Functioning tasks)*. Youth participants completed three computer-based tasks used to study executive functioning. These tasks examine three executive functioning constructs, including working memory (Backwards Digit Span task), inhibitory control (Go/No-Go task), and flexible thinking/planning (Tower of London task). Data collection occurred in-person with a research team member and took place in the students' elementary schools in Durham and Wake County.
- *Teacher assessments (Social Emotional Learning surveys)*. Teachers were asked to assess study-enrolled youth participants' social emotional learning via three widely used developmental surveys (Dimensions of Mastery Questionnaire, Child Behavior Scale, and Devereux Student Strengths Assessment (Dessa-mini)). Data collection occurred with teachers outside of school time, with compensation.
- *Parent survey (surveys assessing parental/familial outcomes)*. Parents were asked to complete an online survey assessing outcomes in educational involvement practice, educational involvement self-efficacy, and social capital development. In addition, parents were asked to report on select youth-focused outcomes, examining social emotional development. Data collection occurred with parents outside of school time, with compensation.
- *Kidznotes administrative data (e.g. program attendance and engagement)*. Kidznotes teaching artists were asked to report treatment participants' attendance and engagement in the program. These data were collected once, at the end of the academic year (May-July), and are used in analyses to control for the extent to which treatment participants actually participated in Kidznotes (i.e. just because students were assigned to the treatment group does not mean they regularly participated in the program).

Data from the first three sources (i.e. student, teacher, and parent data) were collected twice within the school year—once as a pre-program assessment (collected in October) and once as post-program assessment (collected in April)⁵. In addition to these data sources, the evaluation team is in the process of accessing school administrative data (e.g. attendance, disciplinary action), which will be integrated into future analyses.

⁵ As noted below, student data collection did not occur for Cohort 2 Time 2 (April 2020), due to coronavirus-related school closures.

Measures were taken to ensure that data were collected for as many participants as possible (including multiple onsite visits for youth, follow-up emails and phone calls to teachers and parents, and monetary incentives for teacher and parent participation). In addition to the typical challenges of data collection (e.g. student absences, opt-outs, etc.), circumstances related to the coronavirus pandemic limited Cohort 2 youth assessment data collection in Spring 2020; for this reason, current analyses using youth assessment data (i.e. executive functioning tasks) include only Cohort 1.

Because data were collected across three different data sources (youth, parent, and teacher) at two different time points, we have different response rates for each type of information. Table S1 summarizes these response rates across key demographics. As noted in the table, we received the highest response rate among youth participants at the beginning of the year (100%). Barring unforeseen coronavirus-related school closures, we would have expected to have the highest completion rate from this data source. Teacher and parent response rates ranged from 64-70% and 70-79% respectively. While these response rates are relatively high, the numbers drop when considering the response rates for participants for whom we have both pre and post data. Of the 187 participants eligible for analyses, we have pre and post data for 57% of the participants from teacher surveys, 59% from youth tasks (mostly attributable to coronavirus-related school closures for Cohort 2), and 66% for parent surveys.

Preliminary data analysis used descriptive statistics, paired and independent *t*-tests, and multiple regression analyses to focus on youth outcomes in executive function and social emotional development. Future analyses will examine parental/familial outcomes in social capital and educational involvement, as well as youth outcomes in school engagement and academic performance.

Current findings can be categorized into three main areas. First, given the relatively novel use of the executive functioning tasks with youth of this age, we examined measurement-related outcomes within the executive functioning tasks to ensure that the outcomes were valid and measuring change across time. Next, we examined whether there were differences in outcome variables between the treatment and control groups. Last, we examined whether there were demographic-related differences in outcome measures. Analyses focusing on executive functioning include data from Cohort 1 only, while analyses focusing on social emotional development include data for the overall sample (Cohort 1 and 2 combined). As noted above, treatment participants who participated in less than 20% of Kidznotes classes were excluded from analysis.

KEY FINDINGS

Key findings present select information based on preliminary analyses. They are divided into three categories: (1) measurement related findings and comparison to established findings in education literature, (2) effects of program participation on youth outcomes (social emotional development and executive functioning) and (3) demographic-related differences in effects of program participation on youth outcome measures. Results focus on findings that are based on statistically significant (or approaching significance) analysis results.

I. Measurement-related findings for all study participants, suggesting that the data we collected are measuring change over time

Finding 1: Overall, youth study participants show improvement in executive functioning measures and social emotional development from the beginning of the year to the end of the year; this provides evidence of the validity of outcome measures used.

Given the relatively novel use of the executive functioning tasks with youth of this age, we first examined measurement-related outcomes in order to ensure that the outcomes were valid and measuring change across time. This serves as a check that will allow us to later examine whether there are differences between the treatment and control group.

- Youth participants show improvement across all three executive functioning tasks, exhibiting better working memory, response inhibition, and flexible thinking at the end of the year compared to the beginning of the year (see Table S2 for pre and post means).
- Youth participants also showed improvement in social emotional learning, with higher scores reported by teachers in the post-program assessment than in the pre-program assessment.
- However, while youth participants show improvement in these areas, teachers also reported that youth have increased negative reactions to failures to meet goals at the end of the year compared to the beginning of the year. Although speculative, this may be related to the increased difficulty of material covered at the end of the year compared to the beginning of the year or general attitudinal change over the year.

Finding 2: We observed expected differences by grade and gender (first grade participants outperformed kindergarten participants, and female participants outperform male participants) across a number of outcome measures, providing further evidence of the validity of outcome measures.

We also examined whether there are differences in our outcome measures based on demographic differences that are well-established in the research

literature. For example, we would expect to see older students (i.e. first graders) outperform younger students (i.e. kindergarteners)⁶. Additionally, there is evidence to suggest that there are gender-related differences in select outcome measures, with females outperforming males⁷. When examining our outcome measures, we find differences:

- By grade (see Table S3a)
 - o First graders outperformed kindergarteners in all three executive functioning tasks, with better performance in the working memory, response inhibition, and flexible thinking/planning tasks at the beginning of the year and end of the year.
 - o Kindergarteners' performance on these assessments at the end of the year reached levels of performance similar to first graders at the beginning of the year.
 - o Teachers reported that first graders experienced fewer negative reactions to failed outcomes compared to kindergarteners, particularly in the beginning of the year assessment (but not end of year assessment).
- By gender (see Table S3b)
 - o Female participants outperformed male participants in two out of the three executive functioning tasks. Female participants were more accurate in the working memory task at the beginning of the year; however, this difference was no longer statistically significant at the end of the year. Female participants were also more accurate than males in the response inhibition task at both the beginning and end of the year.
 - o Teachers reported greater social emotional development for female participants compared to male participants across pre-program and post-program assessments.
 - o Teachers reported females experienced fewer negative reactions to failed outcomes compared to males, particularly in the end of the year assessment.

II. Examining differences between treatment (Kidznotes) and control (non-Kidznotes) participants in key outcome measures

⁶ Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, 8(2), 71-82.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, 81(6), 1641-1660.

⁷ Romer, N., Ravitch, N. K., Tom, K., Merrell, K. W., & Wesley, K. L. (2011). Gender differences in positive social-emotional functioning. *Psychology in the Schools*, 48(10), 958-970.

Finding 3: When examining change across the year (i.e. controlling for baseline scores), regression analyses suggest limited evidence of statistically significant differences for treatment and control participants.

After establishing that we have confidence in our outcome measures as a whole (see section I), we next turned to our primary research question—are there differences in outcome measures based on whether students participated in Kidznotes (treatment) or not (control). Using regression analyses in which we control for students' performance at the beginning of the year (see Tables S4 & S5), we found:

- There were no significant differences ($p < .05$) between treatment and control for **youth outcomes** in working memory, inhibitory control, flexible thinking, or social emotional development.
- There were also no significant differences ($p < .05$) between treatment and control for **parental outcomes** in educational involvement practice, educational involvement self-efficacy, or social capital development.
- For variables that were approaching statistical significance ($p < .10$), regression analyses controlling for baseline performance suggested that control participants outperform treatment participants in select executive function measures, with control participants exhibiting *higher* accuracy in the working memory task (control = 36%; treatment = 28%) and *longer* reaction times, suggesting a more thoughtful monitoring process in the inhibitory control task (control = 124.9 milliseconds; treatment = 80.6 milliseconds).
- However, ignoring statistical significance ($p > .10$), we find that treatment students do outperform control students in a few areas (based on higher means for treatment than control) and this directionality remains when additionally controlling for gender, grade, and district. Again, note that these differences are not statistically significant. We observed:
 - o Higher accuracy in the working memory task when including partials⁸ (treatment = 71%, control = 69%)
 - o Greater pleasure experienced when mastering concepts/skills (treatment = 4.2, control = 4.1, on a 1-5 scale where higher numbers are better)
 - o Fewer peer relationship problems (treatment = 1.1, control = 1.2; on a 1-3 scale where lower numbers are better)

⁸ In backwards digit span task, participants are asked to remember a series of digits and enter them in backwards order (e.g. if presented 294, participants enter 492). For the purposes of our analyses, we examined accuracy based on exact matches, as well as partials. These partials are defined as responses in which participants remembered the correct digits that were presented, but did not enter them in backwards order (e.g. if presented 385, participants may have entered 538).

IV. Examining patterns in gain among Kidznotes participants by demographic.

Finding 4: When examining specific demographic groups, participation in Kidznotes was associated with *greater gain* for certain groups (e.g., first graders, males, low SES, Raleigh students).

While we may not see an overarching effect of Kidznotes programming, it is possible that Kidznotes may have positive effects for certain demographic groups. We therefore examined whether the effect of participating in Kidznotes varies by demographics (e.g. grade, gender, race, etc; See Table S6 for statistics related to these regression analyses). This included multiple forms of analysis; results below draw on all of these, with a focus on results that are statistically significant.⁹ Controlling for baseline performance, we observed differences in outcome measures based on:

Grade. Analyses evidenced benefits of Kidznotes participation for first graders, relative to kindergarten students.

- Regression analyses revealed that **first graders** participating in Kidznotes were reported to experience greater *pleasure in mastering concepts/skills* (first grade: treatment (4.3) > control (4.0), on a scale of 1-5, $p < .10$). However, this gain was not observed for kindergarteners.
- In fact, for kindergarteners, participating in Kidznotes was associated with *lower end-of-year social emotional learning scores* (treatment (51.9) < control (55.8), $p < .05$), a difference that was not observed for first graders.
- Similarly, for treatment students, **first graders** were reported to have higher *social emotional development* scores (first grade = 57.0, kindergarten = 53.2, $p < .10$). However, there were no statistically significant differences by grade for control students
- For control students, **first graders** were reported to experience *less pleasure associated with meeting goals* than kindergarteners (first grade = 4.0, kindergarten = 4.3, $p < .10$). However, for treatment participants there were no statistically significant differences by grade, suggesting that perhaps participating in Kidznotes reduces these differences by grade.

Gender: Analyses suggest benefits of Kidznotes participation for males, but not for females.

- For **males**, participating in Kidznotes was associated with higher accuracy in the *working memory task* (when including partials)

⁹ Analyses include: XXX

(treatment (76%) > control (63%), $p < .10$). However, there were no such differences for females.

- Similarly, within Kidznotes participants, **males** had higher accuracy in *working memory* than female participants (male = 78%, female = 66%, $p < .05$). However, for control participants, there were no statistically significant differences by gender.
- Together these suggest that males who participated in Kidznotes show the greatest gains in working memory.

SES. Analyses suggest benefits of Kidznotes participation for low-SES students, but not for high-SES students; in fact, in certain cases, we saw lesser gain for high-SES Kidznotes participants than for their non-participant counterparts.

- For treatment participants, **lower-SES students** (i.e. those who qualify for free and reduced lunch) were reported to have fewer *negative reactions to failed outcomes* than higher-SES students (lower-SES = 2.5, higher-SES = 2.9, $p < .10$). However, for control participants, there were no statistically significant differences between lower-SES and higher-SES students.
- For control participants, **lower-SES students** had lower accuracy on *working memory* compared to higher-SES students (lower-SES = 34%, higher-SES = 50%, $p < .05$). However, there were no statistically significant differences by SES for treatment students, suggesting that perhaps Kidznotes helps eliminate these differences by SES.
- For **higher-SES students** (i.e. those who do not qualify for free and reduced lunch), participating in Kidznotes was associated with lower accuracy in the *working memory* task (treatment (25%) < control (49%) $p < .05$) and greater *negative reactions to failure to meet goals* (treatment (2.9) > control (2.5), $p < .10$). However, for lower-SES students (i.e. do qualify for free and reduced lunch), there were no statistically significant differences between treatment and control in either of these outcomes.

District. There is evidence that Kidznotes has select benefits for peer relationships for Raleigh, but not Durham participants

- For **Raleigh** students, participating in Kidznotes was associated with fewer peer relationship problems (control (1.28) > treatment (1.15), $p < .05$), where lower numbers are better). However, there were no statistically significant differences between treatment and control for Durham students.

Race/ethnicity: Analyses suggest that participating in Kidznotes may help eliminate ethnicity-based differences in peer relationships, provide mixed results for White students, and fewer benefits to Black students; in fact, in certain cases, we saw lesser gain among Black Kidznotes participants than for their non-participant Black counterparts.

- For **White** students, we found that participating in Kidznotes can provide **both benefits and drawbacks**. White students participating in Kidznotes outperformed non-White students in Kidznotes in *working memory* (white = 6.4 trials completed, non-white = 4.1 trials completed $p < .05$), a difference that was not observed for control students. However, we also found that participating in Kidznotes was associated with greater *negative reactions to failed outcomes* for White students (treatment (2.8) > control (2.2) $p < .10$), which was not true for non-White students.
- For **Black** students, we found that participating in Kidznotes had **negative impacts** on response inhibition. For Black students, participating in Kidznotes was associated with lower accuracy in the *inhibitory control task* (treatment (81%) < control (87%) $p < .10$). However, for non-Black students the opposite was true—participating in Kidznotes was associated with higher accuracy in the inhibitory control (treatment (88%) > control (84%) $p < .10$). Similarly, within Kidznotes participants, Black students had lower accuracy in the *inhibitory control task* compared to non-Black students (Black = 81% accurate, non-Black = 88% accurate $p < .05$). However, for control participants, there were no statistically significant differences by race.
- Kidznotes may **help eliminate/prevent ethnicity-based differences** amongst Hispanic and non-Hispanic students. For control students, Hispanic students were reported to have fewer *peer relationship problems* than non-Hispanic students (Hispanic = 1.1, non-Hispanic = 1.2, $p < .05$). However for treatment participants there were no statistically significant differences by Hispanic ethnicity.

Finding 5: Among Kidznotes participants, we find demographic-related differences in overall Kidznotes engagement (higher engagement among first grade, female, non-African American/Black students). Regression analyses suggest higher engagement in Kidznotes is associated with gains in executive functioning and social emotional development. However, these effects are no longer

significant when taking into account a student's grade, gender, and race.

Lastly, we examined whether there were any demographic-related differences in treatment participants' engagement in Kidznotes programming and whether engagement in Kidznotes programming is related to executive functioning and social emotional development (e.g. does being more engaged lead to better outcomes in executive functioning?). We found:

- Kidznotes teaching artists reported that **first graders** (compared to kindergarteners) and **female** (compared to male) students were **more engaged** in Kidznotes programming, while **Black** participants (compared to non-Black) were reported to have **lower engagement**.
- Regression analyses suggest that higher engagement in Kidznotes is associated with better working memory and higher social emotional development scores.
- However, these effects of engagement are no longer statistically significant when we also take into account differences between demographics (e.g. students' grade, gender, and race (Black vs non-Black)).

RECOMMENDATIONS & NEXT STEPS

Based on these results from the RCT, we offer selected program recommendations. These will be more fully developed in a forthcoming full report. Selected initial recommendations include:

- o Kidznotes may increase recruitment efforts targeted toward demographics who showed the greatest benefits of Kidznotes programming. In addition, or as an alternative, they may consider exploring mechanisms for enhancing effect for students who demonstrate lesser benefit.
- o Consider potential ways to reach demographic groups who show lower engagement in Kidznotes programming, including Kindergarteners and Black students

As the research team continues to work through analyses, areas of focus will include:

- Gaining access to student administrative data (e.g. academic data from the North Carolina Education Research Data Center) in order to examine effect of participation in Kidznotes in academic data and the

role executive functioning plays in contributing to this relationship (e.g. Does participation in Kidznotes lead to increased executive functioning which then leads to better academic performance)?

- Additional data collection with students and parents (Time 3) in order to examine benefits of Kidznotes participation over multiple years.
- Employing additional analytical strategies (bootstrapping/missing data analyses) in order to test the robustness of these and other effects.