

Program Evaluation:

Empowering Education at Denver Green School

Jacob Hyman

University of Denver

Abstract

Social Emotional Learning (SEL) and Mindfulness-Based Techniques (MBTs) have increased in usage and esteem in recent years in public school settings. The effectiveness of these educational techniques for “Tier 3” students with behavior plans, however, has as of yet remained unproven. This program evaluation sought to establish the effectiveness of one MB-SEL program being implemented in a Denver public school with over 500 students from ECE through 8th grade. Behavior observations of students in 3rd through 5th grade with and without behavior plans in classrooms being managed by teachers of varying SEL competencies were implemented over an eight week period. ANOVA indicated a correlation between on-task behaviors and control and experimental groups. Results of Paired T and Wilcoxon Signed-Rank Tests helped identify correlational trends between teacher SEL competency and increased on-task behavior rates, as well as teacher attentiveness. Though the sample being studied is limited, there is a significant correlation between teachers’ SEL integration and increases in on-task behaviors for students with *and* without behavior plans. This bodes well for the current direction of public education programs seeking to implement mindfulness and social-emotional learning programs in elementary schools. Further research should seek to expand sample size to account for greater variability among students’ and teachers’ respective internalization and integration of SEL.

Introduction

Denver Green School (DGS) is a public elementary and middle school located in East Denver. Approximately 550 children between the ages of 4 and 13 (from Early Childhood Education – ECE – through 8th grade) are educated at DGS each year. The student body is made up of an extremely diverse group of children. DGS has students from 28 countries, a wide range of socioeconomic statuses, dozens of different ethnicities and religions, and varying levels of ability (DGS is an Autism Spectrum Disorder magnet school for Denver Public Schools). DGS is also open, accepting, and encouraging of a large number of LGBTQIA students. DGS experiences a turnover rate of approximately 25% per year, the seventh highest in the district, possibly due in part to serving a large proportion of students of low socioeconomic status (SES). As part of the Denver Public School (DPS) system, DGS is mandated to follow all state guidelines regarding students' education and academic performance. DGS staff is comprised of almost entirely White American educators. Three school administrators are Latina American, one teacher is Latino American, and one member of the support staff is Asian American. Several student aides and mentors are also Latina American. Though there is a large African American student population, none of the educational, support, or administrative staff is African American.

DGS is also an "Innovation School." This model allows DGS to incorporate progressive and less traditional educational strategies. One such strategy, in place since the school's founding in 2008, is a focus on experiential learning and sustainability. As connoted by the "Green" in the school's name, there is one acre of active farmland on the school's premises, and students are charged with cultivating and maintaining that land for the benefit of the school's cafeteria program and weekly farm stand. Another strategy being utilized at DGS for the first time this year is the implementation of a Social Emotional Learning (SEL) program. This program,

created to deal with reactive behavior problems that frequently generate a burden in classrooms and schools, is being run by an external non-profit organization, Empowering Education, with the help of a counselor (Cody Wiggs) that has been working with or at DGS for the past 3 years.

SEL programs have been gaining in popularity due to strong initial indications of research outcomes. Empower Education employs a unique brand of Mindfulness Based Social-Emotional Learning (MB-SEL) strategies developed specifically by the nonprofit for its work at DPS. This model combines evidence-based techniques from both mindfulness-based stress reduction (MBSR) and previously researched SEL programs (i.e. CASEL) (CASEL, 2013; Payton, et al., 2008; Schonert-Reichl & Lawlor, 2010). The theoretical basis of Empowering Education and DGS' SEL program is ecological, mindfulness, sustainability, and empowerment based. Leading up to this school year, teachers were given information and training in 22 SEL lessons on which to focus in their classes (one for each week, beginning on 08/31/2015). Lessons include "I-Statements," "Mindful Body," "Mindful Eating," "Self-Compassion," and "Bullying 101," among many others. Teachers have the freedom to integrate a given week's lesson in any way in which they see fit, though all involve specific teacher-led lessons and discussions on the weekly topic. Empowering Education staff will be on hand weekly to assist teachers, but program implementation is primarily in the hands of educators and administrators within the school, in order to create a school culture that is in line with the values and criteria of the SEL program.

Since this year is the first that Empower Education has been officially implementing their brand of SEL at DGS, there has not yet been a program evaluation to establish the efficacy this particular program. Program outcomes are already being assessed, beginning at the end of the 2014 – 2015 school year, by tracking office referral rates, teacher and student pre and post-test

measures, and academic performance. An evaluation of initial data will be underway in early 2016, and it was the intention of this evaluation to enhance and lend specificity to that assessment. The specifics being evaluated by this assessment will focus on the effectiveness of Empowering Education's SEL program in reducing behavioral disruptions in the classroom of third, fourth, and fifth graders that have an Individual Education Plan (IEP) or Behavior Plan (BP). To that end, this evaluation sought to answer several key questions: Is this SEL program effective for children previously identified as having problematic or disruptive behaviors? How do age, nation-of-origin, ethnicity, culture, sexual orientation, and/or religion impact the effectiveness of SEL? How can the implementation of Empowering Education's SEL program be improved? How significantly has teacher implementation in the classroom impacted the effectiveness of Empower Education's SEL curriculum?

Literature Review

The Basis for Social-Emotional Learning

The purpose of Social-Emotional Learning (SEL) programs is the cultivation of five core competencies for students: self-awareness, self-management, social awareness, relationship skills, and responsible decision making (CASEL, 2013). High maintenance behaviors – that is, passive and active behaviors that negatively impact classroom activity of students and their peers – can be effectively targeted by SEL programs, and have been in the past (McCormick, et al., 2015). Some research has shown that these behaviors appear more often in urban neighborhoods and schools with a high proportion of students of low SES (Lewis, et al., 2013). Students of color in urban environments are also identified as higher risk, though whether this is due to SES, race, or environment (or some intersection of the three) remains mostly unclear (2013). This is

particularly difficult to parse out because many educators are of different cultural identities than their students, and biases may exist in the identification of problematic behaviors that result in what may be disproportionate disciplinary referrals and recommendations.

Best Practices in Social-Emotional Learning Programs

It has been shown in recent years that SEL programs, when properly implemented, can be effective for children of varying cultural identities, abilities, ages, and locations. What constitutes best practices of implementation varies greatly from city to city, from school to school, and even from class to class. Having reviewed much of the recent and pertinent research on this topic, the keys to successful change in the cognitive, social, and behavioral targets of SEL programs lie in three key areas: teacher implementation, cultural competency, and single-component design (Collie, et al, 2011; CPPRG, 2010; Durlak, et al., 2011; Garner, et al., 2014; Lewis, et al., 2013). SEL programs are most effective when teachers implement strategies in the classroom through lessons and overall classroom culture, as opposed to when experts or specialists push-in for weekly or even daily lessons. Programs have also been shown to be most effective for individual students when they are targeted specifically to the norms of that students culture and home life, including race/ethnicity, gender, SES, school environment, and neighborhood/geographic environment. Lastly, studies that followed the implementation of SEL programs within the classroom as a single component of school culture was a key to the effectiveness of SEL programs along the lines of academic, behavioral, and social improvements.

In the past, programs have often been measured through teacher, parent, and student report. The best studies, most notably McCormick, et al. (2015), have utilized evidence-based measures that are less susceptible to bias than self-reporting measures. The most reliable data

collected by measures, in the opinion of this evaluator, come from the use of the Behavioral Observation of Students in Schools (BOSS) observation system, which can utilize blind observers in the assessment of student behavior over the course of SEL implementation. When measuring both control and experimental groups, this outcome measure can be very valid, reliable, and effective, especially when combined with other evidence-based measures (2015).

To date, SEL, and CASEL in particular, has been found to be effective with many different populations of students when tailored to the needs of a student body in a particular environment (i.e. urban vs. rural, or more affluent vs. less affluent). This proves difficult in a school setting, since most student bodies are not made up of a homogenous population of any particular group of students. Particularly in diverse urban areas, such as the one where DGS is located, a school-wide program simply cannot be tailored to the specific cultural needs of every single student. Therefore, a more universal approach is necessary, which *can* negatively impact effectiveness. Such barriers to effectiveness can be addressed by supplementing teacher SEL activities with culturally competent lessons or by addressing the needs of particular students as they arise in the context of SEL practice.

Limitations

Overall, the data analyzed as part of this literature review was of relatively high quality. The evaluator did not address any studies that had small sample sizes, homogenous sample sizes, case studies, or even qualitative research (though the latter may have been useful, had any been available). As such, research of such internal and external validity was hard to come by. Still, the research that has been included in this review represents the best the field has to offer in terms of validity, reliability, and generalizability. While some studies had limitations related to diversity

of sample, many others took diversity and cultural competence into particular account when analyzing data. While some studies used measures that have low validity due to susceptibility to biases, still others took special care to combine subjective and objective measures to best control for biases and moderation effects. And while some studies did not take cultural differences into account in program implementation, others were executed with the express purpose of determining which interventions and programs would be the most effective for individuals of various cultures, abilities, and physical and social locations (Adams, 2013; CASEL, 2013; CPPRG, 2010; Garner, et al., 2014; Lewis, et al., 2013; McCormick, et al., 2015; Payton, et al., 2008; Raimundo, et al., 2013).

Evaluation Methodology

The current evaluation of Empowering Education's Mindfulness Based Social-Emotional Learning program at Denver Green School strove to add depth to assessments already being undertaken by independent evaluators being consulted by Empowering Education. The assessments collected by those evaluators are comprised of a series of surveys given to teachers and students at the start, middle, and end of the 2014-2015 school year (Revak, 2015). Both groups rated students on a Likert scale (1 - none of the time through 4 – all of the time) to measure five specific social-emotional skills: self-control, academic self-efficacy, persistence, social competence, and mastery orientation. Over the course of the school year, both students and teachers reported statistically significant growth in three of the five skill areas: self-control, persistence, and social competence (see Appendix A; Revak 2015).

Two important factors must be considered when reviewing results demonstrated in Figures 1 and 2 (see Appendix A). First, the MB-SEL program at DGS was implemented by a

specialist that was tasked with pushing-in to classrooms weekly. Second, self-report measures like the surveys given to students and teachers have a well-documented and established low level of reliability due to a high potential for bias and subjectivity.

One of the two above factors has been remedied in the 2015 – 2016 implementation of Empowering Education’s self-assessment. A specialist is no longer charged with pushing-in to classrooms in order to teach MB-SEL lessons weekly. Instead, teachers have been trained with executing weekly lessons, and following up on them throughout the course of the week in both the material and atmosphere of the classroom. This approach is more evidence-based, as nearly all of the literature notes that teacher-led, integrated SEL is much more effective than specialist-led lessons. The second factor mentioned above, a low level of reliability of the assessment measures, will still be a significant factor for survey data collected during the current school year.

In an attempt to mitigate this low level of assessment reliability and validity, the current program evaluation sought to add a level of objectivity to the data that is being collected from students and teachers. The program evaluator and a colleague “pushed-in” to third, fourth, and fifth grade classrooms on a weekly basis in order to perform behavioral observations. The observations were executed using the Behavioral Observation of Students in School (BOSS) application. This application (“app”) for the Apple iPad was purchased by Empowering Education a one-time payment of \$29.99.

The observers utilized the BOSS app to assess students’ behavior (see Appendix B) in ten-second intervals for a ten-minute period once weekly. At the end of each observation session, the BOSS application produced a report detailing the trends of that day’s observation. The BOSS application has been used previously in highly valid and reliable research (McCormick, et al.,

2015), and is in line with observation standards that have existed for decades in public schools.

The main advantage of utilizing BOSS instead of paper observations is the ability to process and track data in real time without additional data-entry and analysis.

The sample for this program evaluation is comprised of twelve students at Denver Green School between the ages of 8 and 10 years old (Grades 3 through 5) at the time that the behavior observations began. The mean age of the students is 9 years old. Six of the students are female and six of the students are male. For most of the program evaluation, six of the students are Latino/a, five of the students are African-American, and two of the students are White, though these numbers fluctuated due to participating students leaving the school mid-evaluation. Other students appropriate for either the experimental or control group, respectively, were selected in their stead. The data and subsequent analyses has not controlled for these individuals, but changes in student race/ethnicity may have impacted certain results (i.e. impact of race on observed behavior, teacher attention, etc.)

The teachers in the classrooms being observed were given SEL Ratings of between 1 and 4 on a weekly basis. Five teachers are female and one is male. Five teachers are Caucasian and one is Hispanic and white. The teachers had spent an average of just over 6 years in the classroom, with a standard deviation of 12. Initial SEL Ratings indicated that two teachers were assigned a rating of 4, one was assigned a rating of 3, two were assigned a rating of 2, and one was assigned a rating of 1. These numbers remained consistent over time. When considered by grade, it is clear that the mean of the teacher's SEL Ratings varied markedly by grade level, with the 3rd grade teachers having the highest rating of SEL integration and one 5th grade teacher having the lowest. There is a standard deviation of 0 for 3rd grade teachers, and 0.5 for 4th and 5th grade teachers.

Data Analyses

As mentioned above, data for the quantitative analysis being implemented for this program evaluation were collected via the Behavioral Observation of Students in Schools (BOSS) app for Apple devices (Pearson Education, Inc., 2013). Initially, the observations were to be carried out by one social work intern at Denver Green School, visiting six classrooms at various times throughout the week. After two weeks of pilot observations, scheduling as well as variability of classroom activities and student behaviors rendered one-person observations inconsistent at best. To help prevent threats to internal validity, a second social work intern at Denver Green School agreed to coordinate with the first in order to carry out observations.

As such, observation sessions were implemented by two social work interns at Denver Green School on Wednesday afternoons in two third, fourth, and fifth grade classrooms for ten minutes each. Observations were conducted simultaneously in each classroom in order to achieve the highest level of consistency between the two students being observed. One observer recorded observations students in the experimental group (defined as students that have previously established behavior plans at DGS, referred to as “Tier 1” or “Tier 2” students), and the other observes students in the control group (defined as any student without a previously established behavior plan at DGS, or “Tier 3” students). After each observation period, the BOSS app automatically totaled data by momentary and partial scores, which is manually executed by the observers every ten seconds for the duration of the ten minute period (see Appendix B). Reports from each observation period are then sent to the program evaluator for manual data entry into a statistical analysis program.

Students being observed were assigned to a control or experimental group prior to the first pilot observation period. Data are entered in to SOFA Statistics (Version 1.4.6) in order of student ID, which is assigned using student's grade level (3, 4, or 5), classroom (a or b) and trial group (c or e). A student in one particular third grade classroom in the experimental group, for example, would be given the Student ID 3a-e. The total number and percentages of each observation period for the following behaviors are then entered for each student: looking at teacher (LookTr), participating (Partic), on task (OnTask), movement (Mvmnt), verbal disruption (VbDspt), physical disruption (PhDspt), distracted by an object (Object), and being out of his or her seat (NoSeat). Each of the above eight behaviors can be observed only once per ten-second interval, but could also be co-observed with the following behaviors: attention from teacher (AttnTr) and attention from peer(s) (AttnPr). See Appendix B for more details regarding specific observation coding criteria.

Prior to beginning observations, teachers were rated by one observer and one objective third party regarding their level of successful integration of Empower Education's Mindfulness-Based Social-Emotional Learning curriculum into the classroom environment. Teachers were rated on a Likert scale from 1 to 4 (1 = No Integration; 2 = Minimal Integration; 3 = Some Integration; 4 = Thorough Integration), and scores were averaged together. There were no other items on this survey. The ratings were averaged and entered as "Teacher SEL Rating." Teachers were rated on a weekly basis.

All data collected has been cleaned by observing, upon data entry, whether any outliers exist; this was difficult to determine, however, because students behaved differently from one period of observation to another. Data has been further cleaned by disregarding the first two weeks of observation, in which the evaluator observed control and experimental students during

back-to-back observation sessions. Data was tested for associations using Paired-T and ANOVA tests to seek differences between control and experimental groups along the lines of behavioral data (i.e. overall on-task, off-task, and disruption percentages). Correlations between the changes in on-task behavior percentage, off-task behavior percentage, and/or disruptive behavior percentage and experimental group, race, grade, gender, and/or Teacher SEL Rating were sought using a Paired T-Test. Wilcoxon Signed-Rank Tests were also performed to assess the impact of Teacher SEL Rating on Teacher Attentiveness to students, as well as students' on-task behaviors.

Results

Details about sample of students and teachers within the auspices of this program evaluation can be seen in Appendix C in Tables 1 and 2. It is notable that whereas students' gender and race were roughly evenly selected as part of the sample, teacher race and gender was nearly homogenous. There was only one teacher of color (Hispanic) and one male teacher. Teacher race and gender did not significantly impact on or off task behaviors as measured during observations.

Results of an ANOVA revealed a statistically significant correlation between on-task behavior and group selection ($p = .01018$). This means that members of the control group were more likely to display on-task behaviors than their counterparts on the experimental group (see Appendix E for statistical outputs from SOFA, 1.4.6). A paired T-test revealed a significant correlation between race and displays of on-task behavior ($p < 0.001$ ($1.349e-34$)), with African-American students displaying the least, Latino/a students displaying the second least, and white students displaying the most on task behaviors. This result cannot be generalized, however, due to potential teacher and observer bias.

Results of a second paired T-test revealed a significant positive correlation between Teacher SEL Rating and number of on-task behaviors ($p < 0.001$ (1.290e-33)). This, when coupled with a Wilcoxon Signed-Rank Test indicating a statistically significant correlation between Teacher SEL Rating and prevalence of on-task behaviors ($p < 0.001$ (0.000)), indicate that teachers' level of classroom integration Empower Education's Mindfulness-Based SEL curriculum has an impact on students' behavior, regardless of control or experimental group.

Discussion & Conclusion

Denver Green School and Empowering Education have invested considerable time and resources in developing and implementing an effective Mindfulness-Based Social-Emotional Learning program. For three years, a representative of Empowering Education worked exclusively at DGS in an effort to develop the program. Last year, that same individual spent the year pushing-in to classrooms to begin its implementation. This produced mixed results, as analyses of teacher and student surveys revealed some statistically significant reported improvements. Methods of data collection, however, were not very reliable.

Now, in their fifth year at DGS, Empowering Education has trained teachers and is implementing their MB-SEL program in the most evidence-based way possible. The literature clearly demonstrates that teacher-led, in-class SEL programs – particularly in urban environments – are the most effective programs. Data is still mixed, however, on SEL's effectiveness with students that demonstrate frequent and moderate-to-severe behavioral disruptions (i.e. Tier 3 students). This program evaluation seeks to establish whether or not Empowering Education's teacher-led MB-SEL is effective with such students in third, fourth, and fifth grades at DGS. The evaluation will both supplement and enhance the data already being

collected by Empowering Education by adding a level of reliability to their assessment that is unattainable using their current methods of evaluation.

There are several strengths to this evaluation approach. The behavioral data that were collected as part of the evaluation are objective. There was no need for qualitative interpretation by the observer/evaluator, nor was there an impetus for students to change their behavior intentionally based on observation. Further, the BOSS application streamlines the behavioral observation and data entry process, leading to real-time analyses that saved resources – particularly time and personnel – for all parties involved in the evaluation. Finally, the behavioral data collected through observations will yet be an important complementary tool in understanding assessments of the MB-SEL program that are already underway.

While there were many strengths to this evaluation, there are also several weaknesses. While the BOSS application has a high level of reliability and objectivity, the evaluators are only two individuals and were not free from biases. They spend time in classrooms as part of their roles as social workers at DGS, and may have had previous relationships with and knowledge of some of the students that are to be observed. A potential remedy for this weakness would be to hire non-invested third parties to conduct behavior observations for future program evaluations. Past studies utilizing the BOSS application have been blind, in the sense that observers did not have prior experience with students being observed (McCormick, et al., 2015). Unfortunately, neither DGS nor Empowering Education has the resources to support such an evaluation.

Further, the limited sample size of twelve students – while representative of gender and race/ethnicity – means that this evaluation has a low level of internal validity. This implies that results cannot be generalizable beyond DGS, and are not generalizable to the rest of the student body at DGS. Future evaluations should involve a wider range of students in a wider array of

grades. Generalizability will be contingent on a representative sample of a diverse student body. What's more, while steps are being taken to parse out mediators in the final statistical analyses, it is virtually impossible to account for all external or moderating factors that impact students in and out of the academic environment. Home life, family conflict, trauma history, cultural background, and teacher relationship are all shown to be significant factors in the effectiveness of SEL, and the current evaluation will not be able to control for such effects.

Time was another obstacle to successful evaluation implementation. As implemented here, the observers could not make time to expand the convenience sample. Scheduling time for consistently weekly observations amidst scheduled sessions and inevitable crises that arise throughout the week proved difficult. In the future, this can be mitigated at least in part by communicating clearly with Empowering Education and DGS administrators about the purpose and evolving trends that arise from frequent and consistent behavioral observations. Future evaluations should account for the amount of time it takes to observe a large number of students, and schedule accordingly.

Weaknesses and limitations notwithstanding, this program evaluation shed valuable light on the effectiveness of Empowering Education's MB-SEL curriculum for students with behavior plans (i.e. Tier 3 students), as well as those without such plans (i.e. Tier 1 and Tier 2 students). The literature supports the results of this evaluation regarding Tier 1 and Tier 2 students. Multiple analyses (two Paired T-Tests and a Wilcoxon Signed-Rank Test) revealed statistically significant improvements in classroom behavior over time in such students when teachers adhered well to SEL curricula. What the literature has been unable to indicate previously, though, is effectiveness of SEL curricula for Tier 3 students. This program evaluation has identified a trend that Tier 3 students' classroom behavior may improve alongside that of their

Tier 1 and Tier 2 classmates when teachers integrate specific SEL lessons and attitudes into the classroom environment. This represents a potentially significant and impressive step forward for SEL implementation in public schools.

The data also bear out the assertions in most previously conducted research that teacher integration of and adherence to SEL curricula is *paramount* to successful implementation of SEL programs like the one evaluated here. As mentioned above, the link between (admittedly subjective) Teacher SEL Rating and improvement in student behavior was confirmed as statistically significant by multiple statistical analyses. Future research should seek to use validated Teacher SEL Rating measures in order to confirm the significant findings of this program evaluation.

Given how much time, energy, and money has been invested in the partnership between DGS and Empowering Education, this program evaluation has significant implications for both the program and the agency. Both parties should continue to partner for years to come, carefully training and *retraining* teachers, between and *during* school years to increase adherence to the evidence-based MB-SEL curriculum. It is also recommended that professionals, such as Mr. Wiggs, continue to push-in to classrooms to complement the hard work being done by teachers to create an atmosphere of social-emotional learning in a diverse and challenging public school environment. Future assessments can build on the knowledge gleaned from this evaluation by expanding to cover a larger, more representative sample of the student body. The most important implication for Empowering Education, DGS, and school social workers moving forward will be to find a way to carefully consider the impact of students' intersecting cultures on the effectiveness of SEL programs. The literature is clear that culturally competent, teacher-led SEL is the best possible evidence-based approach to implementing this evolving educational

technique. It will be the job of programs, schools, and social workers to adapt SEL – much like Empowering Education is trying to do by incorporating mindfulness – to the needs of diverse student bodies like that at DGS, throughout Denver, and across the United States.

References

- Adams, D.A. (2013). The application of social-emotional learning principles to a special education environment. *Korean Educational Development Institute Journal of Educational Policy, Special Issue (2013)*, 103 – 118.
- Bar On, R., Tranel, D., Denburg, N. L., & Bechara, A. (2003). Exploring the neurological substrate of emotional and social intelligence. *Brain*, 126, 1790–1800.
- CASEL—Collaborative for Academic, Social, and Emotional Learning. (2013). The 2013 CASEL Guide: effective social and emotional learning programs—Preschool and Elementary School Edition. Chicago, IL: Author.
- Collie, R.J., Shapka, J.D., & Perry, N.E. (2011). Predicting teacher commitment: the impact of social climate and social-emotional learning. *Psychology in the Schools*, 48(10), 1034 – 1048.
- Conduct Problems Prevention Research Group (CPPRG). (2010). The effects of a multiyear universal social-emotional learning program: the role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, 78(2), 156 – 168.
- Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D., & Schellinger, K.B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal intentions. *Child Development*, 82(1), 405 – 423.
- Garner, P.W., Mahatmya, D., Brown, E.L., & Vesely, C.K. (2014). Promoting desirable outcomes among culturally and ethnically diverse children in social emotional learning programs: a multilevel heuristic model. *Educational Psychology Review*, 26, 165 – 189.
- Lewis, K.M., Schure, M.B., Bavarian, N., DuBois, D.L., Day, J., Ji, P., Silverthorn, N., Acock, A., Vuchinich, S., & Flay, B.R. (2013). Problem behavior and Urban, low-income youth:

- a randomized controlled trial of Positive Action in Chicago. *American Journal of Preventative Medicine*, 44(6), 622 – 630.
- McCormick, M.P., O'Connor, E.E., Cappella, E., & McClowry, S.G. (2015). Getting a good start in school: effects of *INSIGHTS* on children with high maintenance temperaments. *Early Childhood Research Quarterly*, 30, 128 – 139.
- Payton, J., Weissberg, R.P., Durlak, J.A., Dymnicki, A.B., Taylor, R.D., Schellinger, K.B., & Pachan, M. (2008). The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Pearson Education, Inc. (2013). Users guide. *Behavioral Observation of Students in Schools (BOSS)*. Bloomington, MN.
- Raimundo, R., Marques-Pinto, A., & Lima, M.L. (2013). The effects of a social-emotional learning program on elementary school children: the role of pupils' characteristics. *Psychology in the Schools*, 50(2), 165 – 180.
- Revak, M. (2015). Combined report: measuring the effectiveness of the *Empowering Education* curriculum. *Empowering Education Internal Report*.
- Riggs, N.R., Greenberg, M.T., Kusché, C.A., & Pentz, M.A. (2006). The meditational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: effects of the PATHS curriculum. *Prevention Science*, 7(1), 91 – 102.
- Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education program on preand early adolescents' wellbeing and social and emotional competence. *Mindfulness*, 1, 137–151.

Statistics performed using Sofa Statistics version 1.4.6. Paton-Simpson & Associates Ltd:

Auckland, New Zealand.

Appendices

A: Figures.....23

B: Observation and Data Collection Measures.....24

C: Tables.....27

D: Logic Model.....28

E: Output of Quantitative Analyses (SOFA, 1.4.6).....30

Appendix A: Figures

Figure 1: Skill Gains as Measured by Teacher Survey – Beginning to End of Year (2014 – 2015)

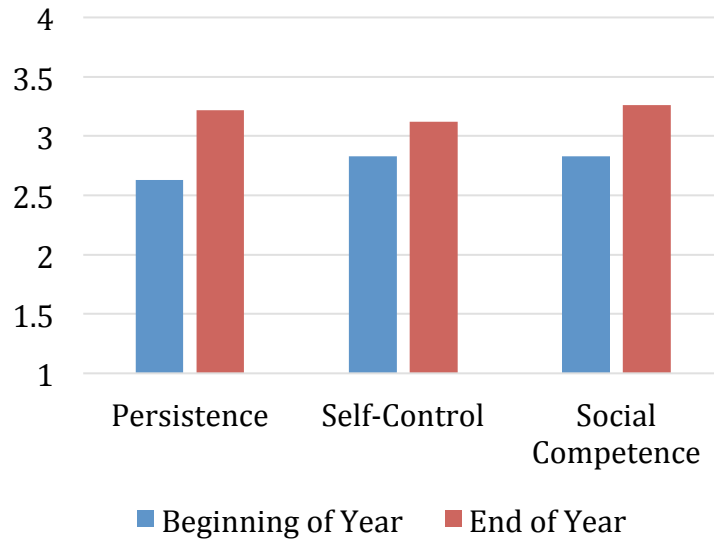
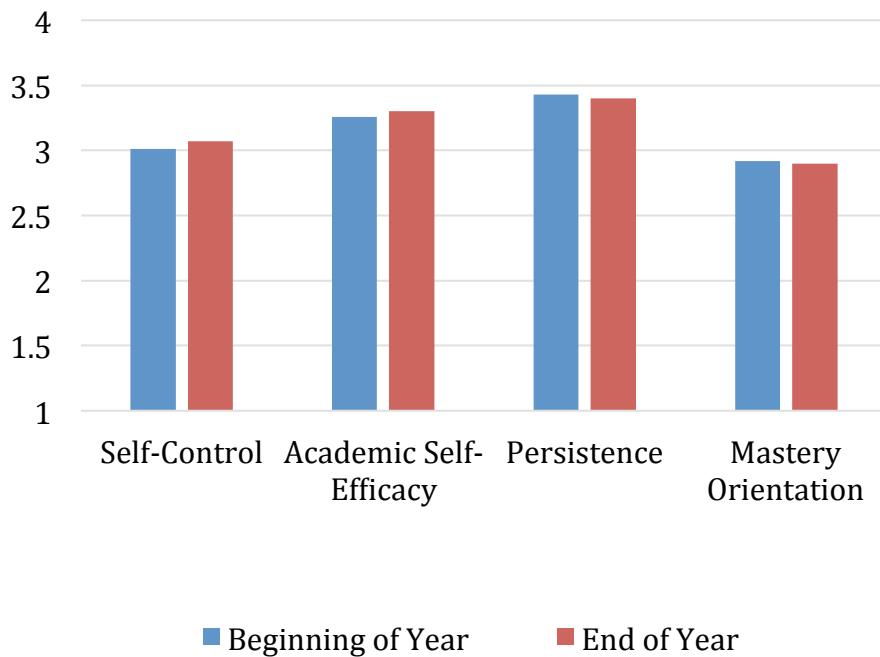


Figure 2: Skills Gains as Measured by Student Survey – Beginning to End of Year (2014 – 2015)



Appendix B: Observation and Data Collection Measures (Pearson Education, Inc., 2013)

Coding Academic Engagement

The SEL template divides academic engagement into three subcategories: on task (OnTask), looking at the teacher (LookTr), or participating (Partic). In either case, the student is considered to be on-task. Each of these behaviors is recorded as a momentary time sample. At the beginning of each cued interval, the observer looks at the targeted student: determines whether the student is on task, and, if so, whether the on-task behavior constitutes an active or passive form of engagement as defined below. At the start of each interval, the BOSS makes a soft “click” sound and vibrates lightly, cueing the user to record the momentary behavior evident by touching either OnTask, LookTr, or Partic. Touching the button turns the button blue. If the user feels they made an error, retouching the button within the interval returns the button to a white background. The user can only record either OnTask, LookTr, or Partic at the start of the interval. These buttons work as either one or the other. If neither OnTask, LookTr, nor Partic is present at the start of the interval, neither button should be touched. The “clicking” or tone sound can be turned off by muting the sound on the device. The occurrence of the behavior at that moment is recorded by tapping the appropriate button on the observation screen.

Coding Nonengagement

When a student is not engaged in academic behavior, five possible categories of off-task behavior are coded. These behaviors are recorded by means of a partial interval observation method: if any of the three behaviors occurs at any point during the interval, tap the appropriate button on the observation screen. Multiple occurrences of the same behavior within a single interval are noted only once.

Recording Behaviors – Partial

Movement (Mvmt)

Movements (Mvmt) are defined as any instance of motor activity that are not directly associated with an assigned academic task. Examples of Mvmt include:

- Engaging in any out-of-seat behavior (defined as buttocks not in contact with the seat)
- Turning around in seat, oriented away from the classroom instruction
- Fidgeting in seat (i.e., engaging in repetitive motor movements for at least 3 consecutive seconds) while not on task

Mvmt should not be scored if the student is:

- Passing paper to a student as instructed by the teacher
- Coloring on an assigned worksheet as instructed (OnTask)
- Laughing at a joke told by another student (Verbal Disruption)
- Swinging feet or fidgeting while working on assigned material (OnTask)

Verbal Disruption (VbDsrp)

Verbal Disruptions (VbDsrp) are defined as any audible verbalizations that are not permitted and/or are not related to an assigned academic task. Examples of VbDsrp include:

- Making any audible sound, such as whistling, humming, forced burping
- Talking to another student about issues unrelated to an assigned academic task
- Talking to another student about an assigned academic task when such talk is prohibited by the teacher
- Making unauthorized comments or remarks

- Calling out answers to academic problems when the teacher has not specifically asked for an answer or permitted such behavior

VbDsrp should not be scored if the student is:

- Laughing at a joke told by the teacher
- Talking to another student about the assigned academic work during a cooperative learning group (OnTask)
- Calling out the answer to a problem when the teacher has permitted such behavior during instruction (Partic)

Physical Disruption (PhDsrp)

Physical Disruptions (PhDsrp) are defined as any physical actions that are not permitted and/or are not related to an assigned academic task. Examples of PhDsrp include:

- Throwing items in the classroom
- Making intentional or distracting physical contact with a peer or adult
- Physically touching another student when not related to an academic task.
- Using physical actions to create audible distractions

PhDsrp should not be scored if the student is:

- Fidgeting in seat (i.e., engaging in repetitive motor movements for at least 3 consecutive seconds) while not on task (Mvmt)
- Bending or reaching, such as picking up a pencil on the floor (Object)

Object (Object)

Object related behaviors (Object) are defined as any physical actions involving objects that are not permitted and/or related to an assigned academic task. Examples of Object include:

- Aimlessly flipping the pages of a book
- Manipulating objects not related to the academic task (e.g., playing with a paper clip, throwing paper, twirling a pencil, folding paper)
- Bending or reaching, such as picking up a pencil on the floor
- Drawing or writing not related to an assigned academic activity

Object should not be scored if the student is:

- Fidgeting in seat (i.e., engaging in repetitive motor movements for at least 3 consecutive seconds) while not on task (Mvmt)
- Passing paper to a student as instructed by the teacher (Partic)
- Coloring on an assigned worksheet as instructed (OnTask)

Passive (Pssve)

Passive behaviors (Pssve) are defined as those times when a student is passively not attending to an assigned academic activity for a period of at least 3 consecutive seconds. Included are those times when a student is quietly waiting after the completion of an assigned task, but is not engaged in an activity authorized by the teacher. Examples of Pssve behavior include:

- Sitting quietly in an unassigned activity
- Looking around the room
- Staring out the window
- Passively listening to other students talk about issues unrelated to the assigned academic activity

It is important to note that the student must be passively off-task for 3 consecutive seconds within an interval to be scored. Should the interval end before the full 3 second period occurs, Pssve is not scored for that interval, and a new consecutive 3 second period is required for the

next interval. For instance, suppose a student begins to stare out the window during the third interval of observation. The observer counts only 2 seconds before the fourth interval begins. The student continues to stare out the window for over 3 seconds in this interval. In this case, only the fourth interval should be scored for Pssve. If the student stops staring out the window after 2 seconds of the fourth interval, than Pssve should not be scored for either interval. In addition, Pssve should not be scored if the student is:

- Quietly reading an assigned book (OnTask)
- Passively listening to other students talk about the assigned work in a cooperative learning group (OnTask)

No Seat (NoSeat)

No seat (NoSeat) behaviors are defined as those times when a student leaves his or her seat for a period of at least 3 seconds without the permission of the teacher or other adult helper. Examples of NoSeat include:

- This includes getting up to throw out paper, sharpen pencils, get water, or use the restroom without prior acknowledgement from a teacher or other helping adult
- Standing up from seat for more than 3 seconds for any off-task reason
- Standing up and walking to speak with a peer during non-group work time

NoSeat should not be scored if the student is:

- Standing, but still engaged in any of the 3 on-task behaviors
- Going to use the restroom, sink, or leaves the room for any reason *with permission* of teacher or other helping adult

Recording Behaviors – Momentary

Teacher Attention (AttnTr)

Teacher Attention (AttnTr) is defined as those times when the student is actively engaged with, or by, the teacher or other adult helper. Examples of AttnTr include:

- Speaking to the teacher
- Receiving *direct* feedback (positive or negative) from the teacher
- Receiving one-on-one help from the teacher or teacher assistant
- Interacting in any way with teacher or other helping adult
- Talking to a teacher or other helping adult about the assigned material
- Receiving attention from a teacher or other helping adult during group work

Peer Attention (AttnPr)

Peer Attention (AttnPr) is defined as those times when the student is engaged in any way with, or by, a peer or peers. Examples of AttnPr include:

- Doing group work with a peer or peers
- Discussing classroom activities, as directed by a teacher, with a peer
- Getting attention from peers for disrupting the class
- Giving attention to peers for disrupting the class
- Attempting to get the attention of a peer
- Redirecting a peer for seeking attention

Appendix C: Tables

Table 1: Sample Characteristics

Grade Level	African-American	Latino/a	White	Female	Male
3 rd	0	1	3	2	2
4 th	3	1	0	2	2
5 th	2	2	0	2	2
TOTAL	5	4	3	6	6

Table 2: Mean Teacher SEL Rating (by grade)

	3 rd Grade	4 th Grade	5 th Grade
Mean Teacher SEL Rating	4	2.5	1.5
Standard Deviation	0	0.5	0.5

Appendix D: Logic Model

Inputs	Underlying Assumptions	Program Components	Outcomes
<p>CLIENTS <i>Clients served by the Program.</i></p>	<p>THEORY OF THE PROBLEM <i>Why does the problem occur? What are its underlying causes? How do the desired assets improve clients' condition?</i></p>	<p>ACTIVITIES <i>Activities within the Program that address the problem(s) or asset(s).</i></p>	<p>SHORT-TERM OUTCOMES <i>Activities are expected to lead to the following changes in _____ (shorter amount of time):</i></p>
<p>Children at Denver Green School (DGS) (~550 students, ECE – Grade 8)</p> <p>Students in grades 3 – 5 having undergone Functional Behavior Assessment (FAB) or with Behavior Plans (BPs) in place</p>	<p>Behavior problems may result from a variety of individual and cultural factors that are beyond the scope of control in an academic environment.</p> <p>Behavior problems/reactive behaviors in school-aged children manifest themselves frequently in school, and are often detrimental to those children's academic progress, as well as that of other students.</p>	<p>Mindfulness-Based Social Emotional Learning (MB-SEL) lessons inside the classroom (weekly with daily reinforcement)</p> <p>Behavior observation (10 minutes at 12 second intervals) utilizing the Behavioral Observation of Students in Schools (BOSS) application</p> <p>Individual counseling (15-30 minutes)</p> <p>Group counseling (20-30 minutes)</p>	<p>Identification of emotional and behavioral responses</p> <p>Identification of connection between internal states (emotions) and external actions (behaviors)</p> <p>Cessation of disruptive and/or unsafe reactive behaviors.</p> <p>Knowledge of potential emotional/behavioral regulation skills.</p> <p>Improved familial and peer interpersonal communication skills</p> <p>Teachers' patience</p> <p>Teachers' ability to regulate and motivate classroom</p>
<p>PROBLEMS & ASSETS <i>The Program aims to address this problem(s) and develop this asset(s).</i></p>	<p>THEORY OF CHANGE <i>How can the problem state be changed / improved? How are the assets developed? How does change occur?</i></p>	<p>RESOURCES <i>The resources required/available to accomplish the activities listed above.</i></p>	<p>LONG-TERM OUTCOMES <i>Activities are expected to lead to the following changes in _____ (longer amount of time):</i></p>
<p>Problems:</p> <ul style="list-style-type: none"> - Behavior disruptions in classroom - Emotion and behavior dysregulation 	<p>Through weekly MB-SEL lessons and daily reinforcement, students can gain more insight into emotions and learn to prevent problematic and disruptive behaviors</p>	<p>20 school teachers (ECE through 8th grade) (15 minutes per week)</p> <p>2 school lead partners</p> <p>1 school Psychologist</p>	<p>More effective behavior regulation</p> <p>Reduction in disruptive behaviors</p> <p>Increased awareness of</p>

<ul style="list-style-type: none"> - Inappropriate identification and expression of feeling <p>Assets:</p> <ul style="list-style-type: none"> - Emotion and behavior regulation skills - Patience and self-compassion - Self/Other Awareness <p>External factors:</p> <ul style="list-style-type: none"> - Parental engagement - Home environment - Prior knowledge of MB-SEL material 	<p>before they come to fruition.</p>	<p>3 school Social Workers</p> <p>1 SEL counselor (as needed)</p> <p>Empowering Education curriculum</p>	<p>self and others</p> <p>Understanding of connection between emotions and behavior, resulting in healthy emotional and behavioral regulation skills</p> <p>Improved academic performance</p> <p>Demonstration of social, emotional, and mindfulness skills in home setting</p> <p>Teachers' connection to and understanding of students</p>
---	--------------------------------------	--	--

Appendix E: Output of Quantitative Analyses (SOFA, 1.4.6)

Results of ANOVA test of average Total On Task (%) for groups from "Control" to "Experimental"

Source	Sum of Squares	df	Mean Sum of Squares	F	p ¹
Between	4405.975	1	4405.975	6.920	0.01018
Within	52211.000	82	636.720		

O'Brien's test for homogeneity of variance: 0.2304 ²

Group	N	Mean	CI 95% ³	Standard Deviation ⁴	Min	Max	Kurtosis ⁵	Skew ⁶	p abnormal ⁷
C	42	70.04	63.055 - 77.024	23.093	8.33	100.0	-0.300	-0.670	0.1684
E	42	55.555	47.327 - 63.783	27.205	2.08	100.0	-0.743	0.105	0.5044

¹ If p is small, e.g. less than 0.01, or 0.001, you can assume the result is statistically significant i.e. there is a difference between at least two groups. Note: a statistically significant difference may not necessarily be of any practical significance.

² If the value is small, e.g. less than 0.01, or 0.001, you can assume there is a difference in variance.

³ There is a 95% chance the population mean is within the confidence interval calculated for this sample. Don't forget, of course, that the population mean could lie well outside the interval bounds. Note - many statisticians argue about the best wording for this conclusion.

⁴ Standard Deviation measures the spread of values.

⁵ Kurtosis measures the peakedness or flatness of values. Between -2 and 2 means kurtosis is unlikely to be a problem. Between -1 and 1 means kurtosis is quite unlikely to be a problem.

Results of Paired Samples t-test of "Race/Ethnicity" vs "Total On Task Behaviors"

p value: < 0.001 (1.349e-34)¹

t statistic: -20.75

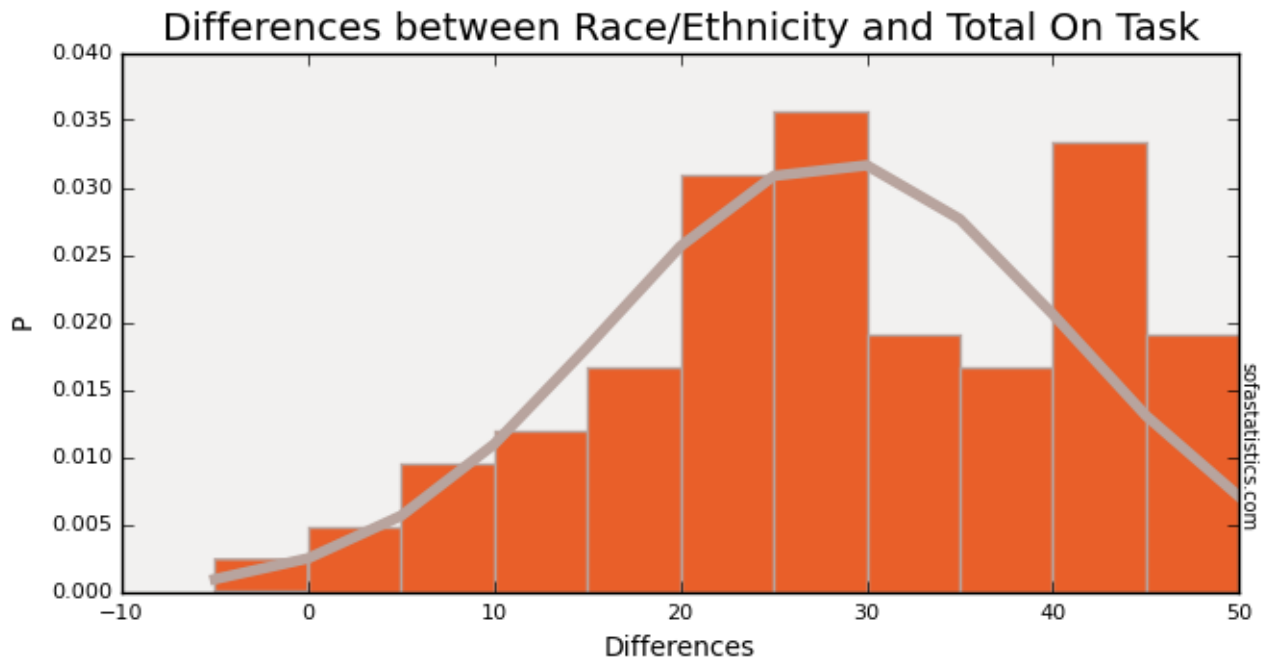
Degrees of Freedom (df): 83

Group	N	Mean	CI 95% ²	Standard Deviation ³	Min	Max
Race/Ethnicity	84	1.857	1.684 - 2.030	0.809	1.0	3.0
Total On Task	84	30.143	27.462 - 32.824	12.536	1.0	48.0

¹ If p is small, e.g. less than 0.01, or 0.001, you can assume the result is statistically significant i.e. there is a difference between at least two groups. Note: a statistically significant difference may not necessarily be of any practical significance.

² There is a 95% chance the population mean is within the confidence interval calculated for this sample. Don't forget, of course, that the population mean could lie well outside the interval bounds. Note - many statisticians argue about the best wording for this conclusion.

³ Standard Deviation measures the spread of values.



*Results of Paired Samples t-test of "Teacher Sel Rating" vs "Total On Task Behaviors"*p value: < 0.001 (1.290e-33)¹

t statistic: -20.086

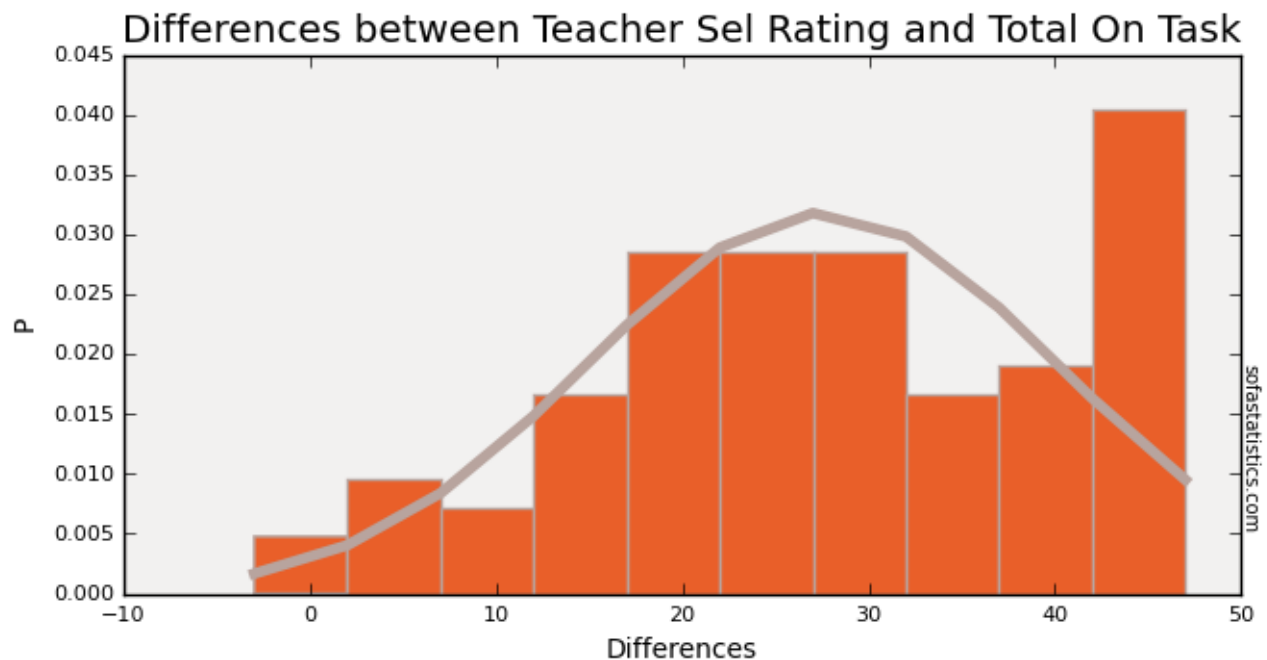
Degrees of Freedom (df): 83

Group	N	Mean	CI 95% ²	Standard Deviation ³	Min	Max
Teacher Sel Rating	84	2.643	2.400 - 2.886	1.137	1.0	4.0
Total On Task	84	30.143	27.462 - 32.824	12.536	1.0	48.0

¹ If p is small, e.g. less than 0.01, or 0.001, you can assume the result is statistically significant i.e. there is a difference between at least two groups. Note: a statistically significant difference may not necessarily be of any practical significance.

² There is a 95% chance the population mean is within the confidence interval calculated for this sample. Don't forget, of course, that the population mean could lie well outside the interval bounds. Note - many statisticians argue about the best wording for this conclusion.

³ Standard Deviation measures the spread of values.



Results of Wilcoxon Signed Ranks Test of "Teacher Sel Rating" vs "Total On Task (%)"

Two-tailed p value: < 0.001 (0.000) ¹

Wilcoxon Signed Ranks statistic: 1.0 ²

Variable	N	Median	Min	Max
Teacher Sel Rating	84	2.5	1.0	4.0
Total On Task (%)	84	64.58	2.08	100.0

¹ If p is small, e.g. less than 0.01, or 0.001, you can assume the result is statistically significant i.e. there is a difference between at least two groups. Note: a statistically significant difference may not necessarily be of any practical significance.

² Different statistics applications will show different results here depending on the reporting approach taken.