UCDAVIS
SCHOOL OF EDUCATION
CRESS Center—Center for Education and Evaluation Services

Mindful Schools Program Evaluation

Prepared by:

Amy Smith, Alberto Guzman-Alvarez, Theresa Westover, Shani Keller, and Stacey Fuller

July 2012

Table of Contents

Table of Figures

Introduction

In order to examine the effects of Mindful Schools Program, as well the impact of two post-program augmentations, a randomized-controlled study was conducted in three urban public K-5 elementary schools during the 2011-12 school year. Classrooms were randomly assigned to one of three groups: control (no treatment in 2011-12), treatment with classroom follow-up, and treatment with professional development. Teachers assigned to the control group were offered the opportunity to receive the program in the subsequent school year. Specifically, the evaluation was designed to answer the following questions:

1. What is the impact of the Mindful Schools Program on students:'
   a. Behavior?
   b. Attention?
   c. Self-reported Mindfulness?
   d. Transition time?
2. What is the impact of the Mindful Schools Program on teachers:'
   a. Compassion and satisfaction?
   b. Feelings of burnout?
   c. Self-efficacy?
   d. Mindfulness?
3. Is either the professional development or follow-up program associated with sustained impacts on students:'
   a. Behavior?
   b. Attention?
   c. Self-reported Mindfulness?
   d. Transition time?
4. Is either the professional development or follow-up program associated with sustained impacts on teachers:'
   a. Compassion satisfaction?
   b. Feelings of burnout?
   c. Self-efficacy?
   d. Mindfulness?
5. What are teachers' general impressions of the Mindful Schools program?

Methodology

The impact of the Mindful Schools program was examined using an experimental design. Teachers were randomly assigned to one of three groups, two of which received the Mindful Schools program and one which served as the control group. After receiving the standard Mindful Schools program, one treatment group received an additional classroom follow-up component and teachers in the other group received professional development outside of their classroom. This allowed us to compare the effectiveness of these two augmentations in sustaining the effects of the Mindful Schools program. Data were collected prior to the start of the Mindful Schools program, immediately after the completion of the standard program, and at the end of the school year. This allowed us to examine the short-term effects of the program, as well as the long-term effects associated with the two sustainability models.

Participants and Random Assignment

Participants included 47 kindergarten through 5th grade teachers and their students from three urban elementary schools. Teachers were randomly assigned to one of three treatment conditions using a stratified random sampling procedure with teachers stratified by school and grade level. The three treatment conditions include: 1) control (no Mindful Schools program in the 2011-12 school year), 2) Mindful Schools + classroom follow up, and 3) Mindful Schools + teacher professional development (PD). In total, 15 teachers were assigned to the control group, 16 were assigned to the follow-up group, and 16 were assigned to the PD group. Table 1 shows the characteristics of the teachers in each of the three groups.

Table 1. Characteristics of Participating Teachers

	Control (n = 15)	Follow Up (n = 16)	PD (n = 16)	Overall (n = 47)
Grade Taught (%)				
Kindergarten	20.0	18.8	12.5	17.0
1	20.0	12.5	18.8	17.0
2	20.0	12.5	18.8	17.0
3	13.3	18.8	18.8	17.0
4	13.3	18.8	18.8	17.0
5	13.3	18.8	12.5	17.0
Years of Teaching Experience (%)				
0-4	26.7	37.5	50.0	38.4
5-9	40.0	37.5	18.8	31.9
10+	33.3	25.0	31.3	29.8

Overall, 800 students participated in the evaluation.[1] Of these, 227 students were in the control group, 297 were in the follow-up group, and 276 were in the PD group. Table 2 below shows the characteristics of the students in each of the three groups.

---

[1] This number includes students who had complete Kinder Rubric data at Times 1, 2 and 3. Students with missing data from any of the three data collection points were omitted from the evaluation.

Table 2. Characteristics of Participating Students

	Control (n = 227)	Follow Up (n = 297)	PD (n = 276)	Overall (n = 800)
Grade Level (%)				
Kindergarten	26.4	19.5	7.2	17.3
1	20.3	12.5	19.9	17.3
2	22.9	11.8	18.8	17.4
3	8.4	17.5	19.9	15.8
4	7.0	17.2	19.9	15.3
5	15.0	21.5	14.1	17.1
Gender (%)				
Male	53.3	50.7	53.2	52.4
Female	46.7	49.3	46.8	47.6
Age in Months (Mean)	95.6	104.4	104.7	102.0

Statistical Test Methodology

Various aspects of the Mindful Schools program were examined using a statistical test termed Regression as well as a Repeated Measures Analysis of Variance (RM ANOVA). Regression was used for the Kinder Rubric, Teacher Survey, and Student Survey. RM ANOVA was used to examine the Transition Time in the teacher survey.

Regression

Regression is the analysis of the statistical relationships among variables. Regression analysis is utilized to determine the values of parameters for a function that cause the function to best fit a set of data observations. More specifically, Ordinary Least Squares (OLS) Regression is a method of estimation that is used in linear regression that minimizes the errors associated with predicting values for Y. OLS Regression uses a least squares criterion because a simple "least" criterion allows positive and negative deviations from the model to cancel each other out. In OLS regression, certain assumptions must be met, specifically: that error have an expected value of zero, that independent variables are non-random, and that the independent variables are linearly independent. The beta value in the regression analysis is a measure of how strongly each

predictor variable influences the dependent, or criterion variable. Covariates included in regression analyses are sources of information not controlled for in the design of the experiment, but that the researcher believes to affect the dependent variable.

Repeated Measures ANOVA

An ANOVA is a statistical test that compares the means of two or more groups. A repeated measure ANOVA is used when all participants of a random sample are measured at different time periods. As the sample is exposed to each time point, the measurement of the dependent variable is repeated.

This approach is used for several reasons. First, our research hypotheses required repeated measures in the form of the Time factor. Second, in cases where there is a great deal of variation between sample members, error variance estimates from standard ANOVAs are large. Repeated measures of each sample member provide a way of accounting for this variance, thus reducing error variance.

Repeated Measure ANOVA produces an output in which its main statistic is the statistical significance of the Time Effect, and Time*Group interaction. A statistically significant Time Effect explains whether the three means for the three distinct time points are statistically different from each other. A statistically significant Time*Group interaction means that belonging to a group (Control, PD, Follow-Up) had a significant effect on the outcome (Transition Time) over time.

Measures

Kinder Rubric

The teacher ratings used are the Kinder Behavior Rubric, developed by researchers local to the participating schools (Kinder and Associates) and recommended by Mindful Schools staff. Teachers were asked to rate students from 0 – 4 in four areas of behavior: mental, emotional, physical, and social habits.

Table 3. Kinder Behavioral Rubric for teacher rating of students' typical behaviors

Score	Mental Paying Attention	Emotional Self-Calming/ Self-Control	Physical Self-Care/ Participation	Social Shows Care for Others
4	Pays attention all of the time	Demonstrates calmness and self-control	Physically engages in all activities	Shows care and respect for teachers
3	Pays attention most of the time	Demonstrates calmness and self-control most of the time	Physically engages in most activities	Shows care and respect most of the time for teachers and fellow students
2	Pays attention some of the time	Demonstrates calmness and self-control some of the time	Physically engages in some activities	Shows care and respect some of the time for teachers and fellow students
1	Needs continual support to pay attention	Demonstrates little ability to calm or control one's own behavior	Needs continual support to participate in class activities	Needs continual support to show care and respect for teachers and fellow students
0	Made no attempt to pay attention	Made no attempt to calm or control one's own behavior	Made no attempt to participate in class activities	Made no attempt to show care and respect for teachers and fellow students

Attention Network Test for Children (ANT-C)

Assessments of children's attention run the gamut from simple observations to using scalp EEG electrodes to record neural activity. An increasingly popular semi-technical assessment of attention is the Attention Network Test for Children (ANT-C), adapted from the Attention Network Test (ANT) commonly used for adults. The test assesses reaction time to tasks testing three attentional networks – the alerting, orienting, and executive. The data on students' attentional networks can be interpreted with great precision by psychologists and neuroscientists, but the data are also useful to K-12 practitioners, as attentional networks have been linked to behavioral patterns and even attention deficit hyperactivity disorder (ADHD) (Johnson, Robertson, Barry, Mulligan, Daibhis, Daly, Watchorn, Gill, & Bellgove, 2008).

Attentional Networks

Attentional networks are neuronal connections in identified areas of the brain that activate in response to perception of certain types of stimuli. Although many other areas of the brain may be influenced by activation of attentional networks, the identified attentional networks themselves are the source of the brain's response to certain stimuli. Three attentional networks commonly measured are the alerting, orienting, and the executive (conflict).

The alerting network is involved in a heightened and sustained sensitivity to stimuli – the active anticipation of a specific event after a warning cue. The orienting network is activated in response to changes in (usually) visual cues – it is involved in locating and focusing on a stimulus in an unanticipated location.   The executive (conflict) network is considered the most advanced of the three. It is involved in resolving discordant cues – recognizing what is "wrong" about a picture and selecting the most appropriate response. Early development of the executive network has been demonstrated in young infants, who stared longer at representations of incorrect simple addition concepts (1+1 = 1) than at representations of correct addition concepts (1+1 = 2) (Berger, Tzur, & Posner, 2006 as cited in Posner et al., 2006). Higher functioning executive networks in children have been shown to coincide with greater behavioral and emotional regulation among children (Posner et al., 2006). Among infants, higher functioning executive networks have been connected to more frequent self-soothing behaviors when presented with novel or frightening stimuli (Posner et al., 2006).

The ANT-C was developed to assess the alerting, orienting, and conflict attentional networks in children using colorful, child-friendly images (fish), directions ("help feed the fish"), and feedback (vocalized computer response of "woohoo!") in place of the more basic black dashes and arrows used in the adult version of the test. First, students encounter one of four warning cue conditions. On the screen there either flashes a spatial cue (either high or low on the screen), a double-cue (both high and low), a central cue in the middle of the screen, or no cue. When the fish appear, students press a button corresponding to the direction of the central fish to "help feed the fish." The central fish may appear alone (the neutral flanker category) or with a line of fish. Congruent flanker trials comprise all fish, including the central one "swimming" in the same direction. Incongruent flanker trials comprise the central fish swimming in the opposite direction of the others. Student reaction time is recorded in milliseconds from the time when the fish first appear until the time when the student presses a button to "feed the fish."

The ANT-C measures the alerting attention network by subtracting a student's median reaction time for trials where there is no cue from the median reaction time for trials where there is a double cue. This measures students' readiness to react when alerted that fish will appear somewhere on the screen, as compared to their readiness to react without any alert. Using a double cue instead of a single, spatial cue prevents over-estimation of this attention network.

The ANT-C measures the orienting attention network by subtracting a student's median reaction time when they know where on the screen the fish will appear (spatial cue) from the median reaction time when the student focuses on the center of the screen (central cue) then must refocus (re-orient) to another part of the screen to complete the task. The subtraction for orienting indicates how much more time it takes students to orient to the unexpected position than to a known position. The scores were calculated using SAS code provided by previous analyst working with Mindful Schools, with a minor correction to insure scoring was consistent with instrument documentation.

Student Survey

The student survey was based on a 25-item instrument developed by Greco, Baer, and Smith (2011), the Child and Adolescent Mindfulness Measure (CAMM). However, due to the nature of conducting research in public schools and university Institutional Review Board (IRB) requirements for doing so, not all the CAMM items were used in this study. Some items were considered too sensitive and/or open to interpretation to be suitable for this study. Ten of the 25 CAMM items were used, these items were presented in the same order and using the same wording as in the original instrument and all items used the same response set as the original instrument. An additional five items were constructed to specifically address the school setting – e.g., paying attention in class, getting tense when taking tests, focus after recess, etc. Thus, the instrument used in this study varies from the validated CAMM instrument in some potentially important ways.

Review of Greco, et al.'s study (2011) documenting the development and testing of CAMM also indicates that our population was different from their test subjects on several dimensions. First the instrument development test subjects were considerably older (5th-10th grade students) than were the 4th and 5th graders taking the survey in this study. Second, our student sample included a number of English learner students, including students in bi-lingual classrooms, and the students in this study typically come from lower socio-economic families than was true of the CAMM validation and development samples (Greco, et al. 2011). Third, classroom teachers in this study administered the survey, rather than research assistants as in the Greco, et al. study – likely introducing more variation in the administration approach across classrooms.

Possibly due to the differences in the study population and survey content and administration, our student survey results suggest that our instrument does not reflect the validity and reliability reported by Greco, et al. (2011). Several issues were identified with the student survey data. First, there was some evidence that some students did not pay attention to the questions and bubbled questions randomly (e.g., all "c" answers, etc.). These students were flagged and removed from analyses. Second, we found that the items have low internal consistency (determined using Chronbach's alpha). When creating a construct (summing or averaging multiple items to create one scale variable) the items should all be correlated with one another. Typically a Chronbach's alpha of at least .7 to .8 is thought of as an acceptable indicator of the internal consistency of a construct. The alpha for the items in the survey was extremely low, in the range of .3, which is not acceptable to use for analyses or for drawing meaningful conclusions. This may be due, at least in part, to the some of the survey questions being too advanced for the students.

Rather than using all survey items as a single construct, we constructed a variable based on a subset of the items. The items for this construct were selected using two criteria. First, we identified items that were an appropriate reading level for the students. Second, factor analysis was used to identify items that statistically correlate with one another. Both of these methods

yielded the same 5 survey items that we used to create a student Mindfulness construct. The items included:

1. It's hard for me to pay attention to only one thing at a time
2. I have a hard time paying attention in class
3. I get tense when taking tests
4. It is hard to focus on class work before or after recess
5. I get frustrated at school

These five items were averaged to get one student Mindfulness score. Although the internal consistency of these five items was within the acceptable range at all three data points (Time 1 α = .69; Time 2 α = .71; Time 3 α = .78), there may still be some bias in the construct given that four of the items within the construct were at the end of the survey. If many of the initial survey items were in fact too advanced for the students, they may have gotten frustrated with the survey after reading the first several items, causing them to not carefully read and reply to the items used for the construct, which would compromise the validity of the data. Therefore, these analyses should be viewed as exploratory and caution is warranted when basing conclusions upon the findings.

Teacher Survey

An online survey was administered to teachers at each of the three data points. The survey contained items from validated measures designed to assess teachers' compassion satisfaction, burnout, Mindfulness, and self efficacy.[2]

Compassion Satisfaction:

Compassion Satisfaction items were taken largely from the Professional Quality of Life survey (Hudnall & Stamm, 2009), a measure validated with a sample of 1,187 respondents in helping professions. For the compassion satisfaction construct, respondents rated six items such as "I get satisfaction from being able to teach students" on a 5-point scale ranging from "Never" to "Very Often." These five items were then averaged to create one compassion satisfaction construct.

Burnout:

As with the Compassion Satisfaction construct, the Professional Quality of Life Survey (Hudnall Stamm, 2009) was drawn on for questions relating to teacher burnout. Respondents were asked to rate items relating to teacher burnout on the same 5-point scale used for the Compassion Satisfaction construct. The burnout construct consisted of the average of 4 survey items.

Mindfulness:

Mindfulness items were largely taken from the Mindfulness Attention Awareness Scale (MAAS; Brown and Ryan, 2003). According to Brown and Ryan (2003), "Mindfulness" as measured by the Mindfulness Attention Awareness Scale (MAAS), comprises "a receptive state of mind in

---

[2] The internal consistency of each construct at each time point was at the acceptable level (α > .8)

which attention, informed by a sensitive awareness of what is occurring in the present, simply observes what is taking place." Of the 15 items comprising the MAAS, 12 were selected and an additional item was slightly modified for use in this evaluation, creating a 13-item Mindfulness subscale. Respondents rated statements such as, "I rush through activities without being really attentive to them" on a 6-point scale that ranged from "Almost Always" to "Almost Never." The MAAS has been validated with several independent samples of adults.

Self Efficacy:

The survey included items from the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 1998) to measure teachers' efficacy in classroom management and student engagement. Teachers were asked to rate various statements related to how much they believe they can do to impact students on 9-point scale (1 = Nothing; 9 = A great deal).

Finally, the teacher survey included questions asking teachers to report on student transition time to and from recess, as well as questions regarding teachers' implementation of Mindfulness strategies and their perceptions of the Mindful Schools program.

Data Analysis

To examine the short-term impact of the Mindful Schools program, ordinary least squares (OLS) regression was used to examine the differences between teachers and students in the treatment groups compared to those not receiving the Mindful Schools training treatment and a series of covariates. For student-level analyses, covariates included the students' pre-program score on the outcome variable, a variable indicating the students' gender, and the students' age in months. For teacher-level analyses, covariates included the teachers' pre-program score on the outcome variable, a variable representing the teachers' years of teaching experience, and the grade level taught by the teacher. The same model was used to examine the long-term effects of the program, comparing students and teachers in both treatment groups to the students and teachers in the control group. These analyses are described in more detail for each outcome in the results section below. Finally, to compare the effectiveness of being in the PD or the follow-up group, analyses were run using just the subgroup of students in these groups. These analyses had the outcome at Time 3 as a function of a variable indicating whether the student was in the follow-up or PD group, as with the previous models the series of covariates were controlled.

Results

Impact of the Mindful Schools Program on Students

Impact on Student Behavior

The impact of the Mindful Schools program on student behavior was examined using the Kinder Rubric, which was completed by teachers prior to the start of the program (Time 1), immediately following the program (Time 2), and then at the end of the school year (Time 3). Overall, 800

students had complete Kinder Rubric data (scores on all four constructs at all three data points) and were included in analyses. Students' average scores at each of the 3 time points are shown in Table 4 and Figure 3 below.

Table 4. Mean Behavior Ratings of the Control, Professional Development, and Follow Up groups at Times 1, 2, and 3

	Control (n = 227)	PD (n = 297)	Follow Up (n = 276)
Time 1	3.05	2.99	2.90
Time 2	3.19	3.19	3.23
Time 3	3.27	3.29	3.22

Figure 1. Mean Behavior Ratings of the Control, Professional Development, and Follow Up Groups at Times 1, 2, and 3
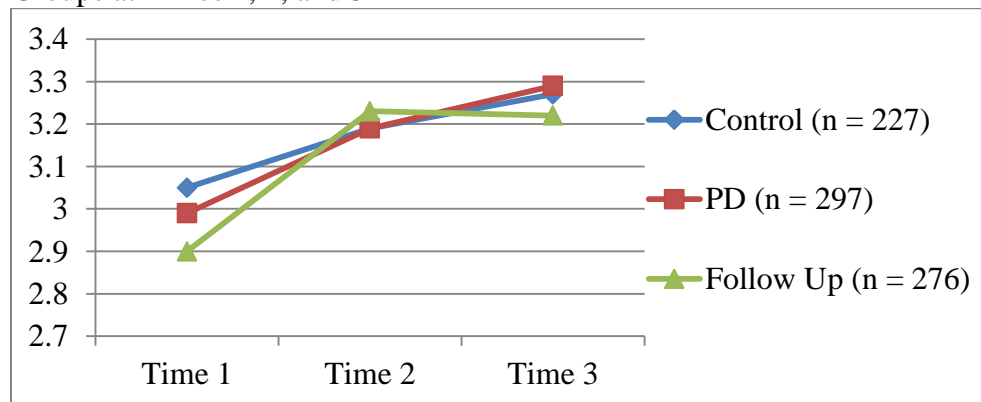
Short-Term Impact of the Mindful Schools Program on Student Behavior
In order to examine the short-term impact of the Mindful Schools program on student behavior, OLS regression was used to model the overall behavior score obtained from the Kinder Rubric at Time 2 as a function of the treatment variable, controlling for students' score at Time 1, gender, and age in months. Behavior scores were standardized (mean = 0, SD = 1), consequently the coefficients can be interpreted as standardized effect sizes. The results indicate that immediately after completion of the Mindful Schools program, there were no statistically significant differences in the behavior of students who received the program and those that did not (β = -.021, p = .752). However, after controlling for students' initial behavior ratings, their age, and gender, there was a marginally statistically significant difference between the groups (β = .104, p = .080), indicating that, on average, students who participated in the Mindful Schools program tended to have behavior ratings that were approximately 0.10 standard deviations higher than students in the control group (see Table 5). The other coefficients in Table 5 similarly demonstrate the relationship between pre-scores, age, and gender with time 2 scores, controlling for the other variables in the equation. In other words, a coefficient of -.26 for males means that, controlling for their initial scores, age, and placement in a control or treatment group, males had

approximately .25 standard deviations lower scores than did female students. The constant can be thought of as the intercept value (or "starting point") for the regression line.

Table 5. Short-Term Effects of the Mindful Schools program participation on Student Behavior

	β
Constant	-.055
Mindful Schools	.104[+]
Pre-Behavior Score	.606***
Age in Months	.001
Male	-.255***

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, [+]$p \leq .1$

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

In addition, the four individual behavior ratings on the Kinder Rubric were analyzed separately to examine whether the Mindful Schools program had a short-term impact on students' emotional, physical, social, and mental behavior. As shown in

Table 6, the students in the treatment group made statistically significantly larger improvements in their mental ($Z^3 = -2.89$, $p^4 = .004$, $r^5 = -0.10$) and physical behavior ($Z = -2.23$, $p = .026$, $r = -0.08$) than did the students in the control group. In contrast, there was no difference in emotional ($Z = -0.25$, $p = .80$, $r = -0.01$) or social ($Z = -1.34$, $p = .165$, $r = -0.05$) behavioral changes of the students in the two groups.

---

[3] Z-scores are standardized values that indicate how much a value deviates from a group mean.

[4] Significance level for Mann-Whitney U analysis

[5] Effect size for Mann-Whitney U analysis, calculated using the following formula: $r = \dfrac{Z}{\sqrt{N}}$
Small effect sizes are indicated at .1, medium at .3, and large at .5 or above.

Table 6. Mean Change from Time 1 to Time 2 on Kinder Rubric Scales

	Treatment (n = 573)	Control (n = 227)
Mental	.28	.09**
Emotional	.27	.23
Physical	.28	.15*
Social	.24	.10

* $p \leq .05$; ** $p \leq .01$; ** $p \leq .001$
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

Long-Term Impact of the Mindful Schools Program on Student Behavior

Overall Long-Term Effects

To examine whether there was any impact of the Mindful Schools program that persisted through the end of the school year, an OLS regression was used to model the students' overall behavior scores at Time 3 as a function of the treatment variable (overall), controlling for students' score at Time 1, gender, and age in months. As with the Time 2 data, behavior scores were standardized (mean = 0, SD = 1) so that coefficients can be interpreted as standardized effect sizes.

The results indicate that, at Time 3, there were no statistically significant differences in the behavior of students who received the Mindful Schools program and those that did not (β = -017, p = .829). However, after controlling for students' initial behavior ratings, their age, and gender, there was a marginally statistically significant difference between the groups (β = .117, p = .061), indicating that on average, students who participated in the Mindful Schools program had behavior ratings that were approximately 0.12 standard deviations higher than students in the control group at the end of the school year (see **Error! Reference source not found.**). This suggests that the effects that were observed immediately after the completion of the program persist.

Table 7. Long-Term Effects of the Mindful Schools program participation on Student Behavior

	β
Constant	.741***
Mindful Schools	.117[+]
Pre-Behavior Score	.579***
Age in Months	-.007***
Male	-.258***

$*p \leq .05$; $**p \leq .01$; $*** p \leq .001$, $+p \leq .1$

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

In addition, the four individual behavior ratings on the Kinder Rubric were analyzed separately to examine whether the Mindful Schools program had a long-term impact on students' emotional, physical, social, and mental behavior. As shown in Table 8, the students in the treatment group made statistically significantly larger improvements in their mental ($Z = -1.95$, $p = .05$, $r = -0.07$) and social behavior ($Z = -2.56$, $p = .01$, $r = -0.01$) than the students in the control group. In contrast, there was no difference in emotional ($Z = -0.29$, $p = .77$, $r = -0.01$) or physical ($Z = -0.71$, $p = .48$, $r = -0.01$) behavioral changes of the students in the two groups.

Table 8. Mean Change from Time 1 to Time 3 on Kinder Rubric Scales

	Treatment (n = 573)	Control (n = 227)
Mental	.39	.26*
Emotional	.29	.29
Physical	.31	.24
Social	.26	.10**

$* p \leq .05$; $** p \leq .01$; $** p \leq .001$

NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

Comparison of the Effectiveness of Follow-Up and Professional Development

To compare the behavior scores of students in the follow-up and professional development groups, regression was used to model the students' overall behavior scores at Time 3 as a function of the follow up and PD variables, controlling for students' score at Time 1, gender, and age in months. Behavior scores were standardized (mean = 0, SD = 1) so coefficients can be interpreted in standard deviation units. As with the overall results, there were no differences in the behavior ratings of the students in the three groups prior to controlling for students' initial behavior ratings, age, and gender. However, after controlling for these variables, the effect of being in the professional development group was statistically significant (β = .134, p = .05). In contrast, the behavior ratings of the students in the follow-up group did not differ significantly from the ratings of students in the control group (β = .098, p = .17; see Table 9). Although this suggests that effect of the Mindful Schools program only persisted for students who were in the professional development group, it does not necessarily indicate that the students in the professional development group had behavior ratings that were significantly higher than those in the follow-up group. The regression model compares the ratings of students in the professional development group to the ratings of those in the control group and compares the ratings of the students in the follow-up group to those in the control group, but does not directly compare students in the two treatment conditions to each other.

Table 9. Effect of Mindful Schools program plus follow up or Professional Development on Student Behavior

	β
Constant	.740***
Mindful Schools + PD	.134*
Mindful Schools + Follow Up	.098
Pre-Behavior Score	.578***
Age in Months	-.007***
Male	-.259***

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, $^+ p \leq .1$

Note: The reference group includes students in the control group. The Mindful Schools + PD coefficient represents the difference in the behavior ratings of the PD and control groups, after controlling for the other variables in the model. The Mindful Schools + Follow Up coefficient represents the difference in the behavior ratings of the follow up and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

To examine differences between students in the two treatment conditions, an additional regression model was run, including only treatment students. The results indicate that there was only a marginally statistically significant difference in the behavior ratings of the students in the follow up and professional development groups ($\beta = .132$, $p = .06$), suggesting that, on average, students in the professional development group had behavior ratings that were approximately 0.13 standard deviations higher than those in the follow-up group.

As shown in Table 10, the average Time 3 behavior rating of students in the follow up group, represented by the constant in the regression model, was .70, indicating that, on average, these students had behavior ratings that were .70 standard deviations higher than the average student in the overall sample, after controlling for the students' initial behavior ratings, age, and gender. Furthermore, although the effect was only marginally statistically significant ($p = .06$), students in PD group received a rating that was .13 standard deviations higher than those in the follow up group, for an average rating of .83 after controlling for the students' initial behavior ratings, age, and gender.

Although these results do not provide conclusive evidence that one sustainability option is superior to the other, they do suggest that the professional development model may be more effective for sustaining the initial effects of the Mindful Schools program.

Table 10. Comparison of the Effects of the Professional Development and Follow Up Sustainability Models

	β
Constant	.700***
PD	.132[+]
Pre-Behavior Score	.599***
Age in Months	-.006***
Male	-.261***

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, [+]$p \leq .1$

Note: The reference group includes the students in the follow up group. The PD coefficient represents the difference in the behavior ratings of the PD and follow up groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

In addition, the four individual behavior ratings on the Kinder Rubric were analyzed separately to examine whether the follow-up or professional development program was a more effective model for sustaining the effects of the Mindful Schools program on student behavior. As shown in 1, there were no statistically significant differences in the students' behavioral change on any of the four individual constructs from Time 1 to Time 3.[6]

Table 11. Comparison of Students in the Follow-Up and Professional Development Groups Change in Behavior from Time 1 to Time 3

	Follow-Up (n = 276)	PD (n = 297)
Mental	.36	.43
Emotional	.30	.27
Physical	.30	.31
Social	.33	.20

*$p \le .05$; ** $p \le .01$; ** $p \le .001$
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

---

[6] Mental Z = 1.67, p = .096; Emotional Z = -0.16, p = .870; Physical Z = -0.31, p = .760, Social Z = -.013, p = .895

Impact on Students' Attention (Ant-C)

The impact of the Mindful Schools on students' attention was measured using data from the Ant-C. Overall, 146 students had complete Ant-C data (complete data at all three data points) and were included in analyses. Students' average scores at each of the 3 time points are shown in

Table 12. Mean Ant-C Score of Students in the Control, Professional Development, and Follow Up groups at Times 1, 2, and 3

	Control (n = 32)	PD (n = 54)	Follow Up (n = 60)
Alertness			
Time 1	85.2	89.6	71.5
Time 2	91.9	85.2	95.9
Time 3	127.2	104.3	94.3
Orientation			
Time 1	19.4	41.9	35.4
Time 2	37.9	25.1	26.0
Time 3	28.9	35.4	24.6
Conflict			
Time 1	117.6	95.7	84.6
Time 2	71.1	73.9	58.9
Time 3	53.3	73.3	68.3

Short-Term Impact of the Mindful Schools Program on Student Attention

In order to examine the short-term impact of the Mindful Schools program on student attention, OLS regression was used to model each sub-score obtained from the Ant-C Time 2 as a function of the treatment variable, controlling for students' score at Time 1, gender, and age in months. Ant-C scores were standardized (mean = 0, SD = 1), consequently the coefficients can be interpreted as standardized effect sizes. The results indicate that immediately after completion of the Mindful Schools program, there were no statistically significant differences in the attention behavior of students in who received the program and those that did not, on any of the three constructs assessed by the Ant-C, after controlling for their initial Ant-C scores, their age, and

their gender (See Table 13; Alertness β = -.007, *p* = .972; Orientation β = -.158, *p* = .446; Conflict β = .018, *p* = .928).

Table 13. Short-Term Effects of the Mindful Schools Program Participation on Student Attention

	Alertness	Orientation	Conflict
Constant	.998	-1.212	-.203
Treatment	-.007	-.158	.018
Pre-ANT-C Score	.066	-.024	.270
Male	-.030	.276	.040
Age in Months	-.009	.010	.002

*p ≤.05; **p ≤.01; *** p ≤ .001, +p ≤.1

Note: The reference group includes students in the control group. The treatment coefficient represents the difference in the respected ANT-C scores of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, ANT-C score at Time 2, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Long-Term Impact of the Mindful Schools Program on Student Attention

Overall Long-Term Effects

To examine whether the Mindful Schools program had an impact on student attention that persisted through the end of the school year, an OLS regression was used to model the students' Ant-C scores at Time 3 as a function of the treatment variable (overall), controlling for students' score at Time 1, gender, and age in months. As with the Time 2 data, Ant-C scores were standardized (mean = 0, SD = 1) so that coefficients can be interpreted as standardized effect sizes.

The results indicate that, at Time 3, there were no statistically significant differences in the Orientation or Conflict of students in who received the Mindful Schools program and those that did not, after controlling for students' initial behavior ratings, their age, and gender (see Table 14; Orientation β = .026 *p* = .900; Conflict β = .326 *p* = .120). However, the results indicate that at Time 3, students who were in the treatment group had Alertness scores that were approximately .36 standard deviations lower than those in the control group (β = -.363, *p* = .048).

Table 14. Long-Term Effects of the Mindful Schools Program Participation on Student Attention

	Alertness	Orientation	Conflict
Constant	1.739*	.138	1.007
Treatment	-.363*	.026	.326
Pre-ANT-C Score	.180*	-.030	.133
Male	-.294	.018	-.030
Age in Months	-.011	-.001	-.011

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$,

Note: The reference group includes students in the control group. The treatment coefficient represents the difference in the respected ANT-C scores of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Pre-ANT-C, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Comparison of the Effectiveness of Follow-Up and Professional Development

To compare the attention scores of students in the follow-up and professional development groups, regression was used to model the students' overall Ant-C scores at Time 3 as a function of the follow up and PD variables, controlling for students' scores at Time 1, gender, and age in months. Ant-C scores were standardized (mean = 0, SD = 1) so coefficients can be interpreted in standard deviation units.

Results indicate that, compared to the control group, there was no statistically significant effect of being in the professional development or follow-up group on students Orientation (PD β = .125, p = .590; follow up β = -.070, p = .757) or Conflict scores (PD β = .320, p = .169; follow up β = .254, p = .267) after controlling for the other variables in the model. In contrast, students in the Follow Up group had Alertness scores that were approximately .41 standard deviations lower than the scores of students in the control group (β = -.411, p = .041). There was no statistically significant difference in the Alertness scores of students in the professional development and control groups (β = -.310, p = .129).

Table 15. Effect of Mindful Schools Program Plus Follow-up or Professional Development on Student Attention

	Alertness	Orientation	Conflict
Constant	1.753*	.135	1.124
Mindful Schools + PD	-.310	.125	.320
Mindful Schools + Follow Up	-.411*	-.070	.254
Pre-ANT-C Score	.176*	.066	.098
Male	-.294[+]	.014	-.025
Age in Months	-.011	-.001	-.011

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, [+]$p \leq .1$

Note: The reference group includes students in the control group. The Mindful Schools + PD coefficient represents the difference in the behavior ratings of the PD and control groups, after controlling for the other variables in the model. The Mindful Schools + Follow Up coefficient represents the difference in the behavior ratings of the follow up and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

These results suggest that placement in the follow-up program is negatively associated with student alertness, compared to the students who did not receive the Mindful Schools program (i.e., control group); however, it does not directly compare students in the two treatment conditions to each other. To examine differences between the Alertness scores of students in the two treatment conditions, an additional regression model was run, including only treatment students. The results indicate that there was no statistically significant difference in the Alertness scores of the students in the follow up and professional development groups ($\beta = .092$, $p = .578$; see Table 16). These results suggest that there is no difference in the impact of the two sustainability models on students' Alertness scores.

Table 16. Comparison of the Effects of the Professional Development and Follow Up Sustainability Models on Student Alertness

	β
Constant	.976
PD	.092
Pre-Behavior Score	.220**
Age in Months	-.463**
Male	-.007

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, $^+ p \leq .1$

Note: The reference group includes the students in the follow up group. The PD coefficient represents the difference in the behavior ratings of the PD and follow up groups, after controlling for the other variables in the model. The outcome variable, Time 3 Ant-C Alertness score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Impact on Students' Self-Reported Mindfulness

Students' Mindfulness was examined using data from the student survey, which was administered to all fourth and fifth grade students. Overall 307 students had data from at least one time point, but only 182 had complete data from all three time points and were included in the analyses. Table 17 and Figure 4 below show the average Mindfulness scores of students at Times 1, 2, and 3. [7]

Table 17. Mean Mindfulness Scores of Students in the Control, Professional Development, and Follow Up Groups at Times 1, 2, and 3

	Control (n = 43)	PD (n = 73)	Follow Up (n = 66)
Time 1	3.15	3.22	3.36
Time 2	3.23	3.11	3.11
Time 3	3.32	3.15	3.27

Figure 2. Mean Mindfulness Scores of Students in the Control, Professional Development, and Follow Up Groups at Times 1, 2, and 3

---

[7] Note that Questions 5,7,11,12,14,15 of the student survey were reversed scored

Short-Term Impact on Students' Self-Reported Mindfulness

In order to examine the short-term impact of the Mindful Schools program on students' self-reported Mindfulness, OLS regression was used to model the Mindfulness construct at Time 2 as a function of the treatment variable (overall), controlling for students' score at Time 1, gender, and age in months. Mindfulness scores were standardized (mean = 0, SD = 1) so that coefficients can be interpreted as standardized effect sizes. The results indicate that, at Time 2, there were no statistically significant differences in the self-reported Mindfulness of students who participated in the Mindful Schools program and those who did not (β = -.126, p = .470). There continued to be no statistically significant difference, even after controlling for students' Mindfulness scores prior to the intervention, their age, and their gender (β = -.208, p = .134).

Table 18. Effect of the Mindful Schools Program Participation on Students' Self-Reported Mindfulness After Completion of the Program

	β
Constant	.139
Mindful Schools	-.208
Pre-Mindfulness Score	.620***
Age in Months	.000
Male	.126

*$p \leq$ .05; **$p \leq$ .01; *** $p \leq$ .001, $^+p \leq .1$

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the Mindfulness scores of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 2 overall Mindfulness score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.
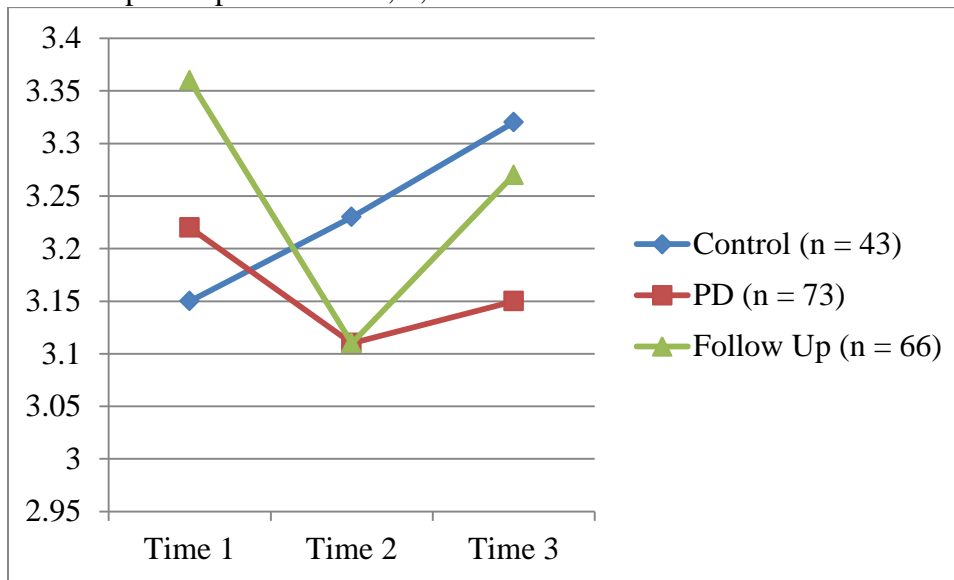
Overall Long-Term Effects

The long-term effects of the Mindful Schools program on students' self-reported Mindfulness were examined using OLS regression to model the Mindfulness construct at Time 3 as a function of the treatment variable (overall), controlling for students' score at Time 1, gender, and age in months. Mindfulness scores were standardized (mean = 0, SD = 1) so that coefficients can be interpreted as standardized effect sizes.

At Time 3, there were no statistically significant differences in the self-reported Mindfulness of students who participated in the Mindful Schools program and those who did not (β = -.119, p = .497). There continued to be no statistically significant difference, even after controlling for students' Mindfulness scores prior to the intervention, their age, and their gender (β = -.196, p = .161)

Table 19. Effect of the Mindful Schools Program Participation on Students' Self-Reported Mindfulness at the End of the School Year (Time 3)

	B
Constant	.451
Mindful Schools	-.196
Pre-Mindfulness Score	.607***
Age in Months	-.003
Male	.158

*$p \le .05$; **$p \le .01$; *** $p \le .001$, $^+ p \le .1$

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the Mindfulness scores of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Mindfulness score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Comparison of the Effectiveness of Follow-Up and Professional Development

In order to examine the effectiveness of the follow-up and professional development augmentations, a regression model was used to compare the self-reported Mindfulness scores of students in the follow-up and professional development group at Time 3. Specifically, OLS regression was used to model the students' overall Mindfulness scores at Time 3 as a function of the follow-up and PD variables, controlling for students' score at Time 1, gender, and age in months. Behavior scores were standardized (mean = 0, SD = 1). As with the overall results, there were no differences in the Mindfulness ratings of the students in the three groups prior to controlling for students' initial ratings, age, and gender (follow up β = -.056, p = .774; PD β = -.176, p = .364). Furthermore, there continued to be no difference in the groups even after controlling for students' initial Mindfulness, age, and gender (follow up β = -.187, p = .235; PD β = -.205, p = .187;

Table 20).

Table 20. Effect of Mindful Schools Program plus Follow Up or Professional Development on Students' Self-Reported Mindfulness

	β
Constant	.437
Mindful Schools + PD	-.205
Mindful Schools + Follow Up	-.187
Pre-Behavior Score	.606***
Age in Months	.157
Male	-.003

*$p \leq.05$; **$p \leq.01$; *** $p \leq .001,$ $^{+}p \leq .1$
Note: The reference group includes students in the control group. The Mindful Schools + PD coefficient represents the difference in the Mindfulness ratings of the PD and control groups, after controlling for the other variables in the model. The Mindful Schools + Follow Up coefficient represents the difference in the Mindfulness ratings of the follow up and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 Mindfulness score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Impact on Student Transition Time

In order to measure classroom transition time, which was considered a concrete measure of classroom management, participating teachers responded to items in the Teacher Survey that asked the number of minutes that their class spent in transition to and from recess during the school day. Overall, 36 teachers had complete data from all three time points and were included in the analyses. Respondents selected from a drop-down menu of whole minutes ranging from "less that 1 minute," which was recoded as 0 minutes, to "more than 10 minutes," which was recoded as 11 minutes. Table 21 and Figure 5 below show the average transition times of teachers at Times 1, 2, and 3. Items were summed across transition times, creating both a "transition to" and a "transition from" score. Both of these composite scores were then summed to create a total transition score.

Table 21. Mean of Transition Times by Treatment Group

Transition (min.)	Treatment	Time 1 (n = 11)	Time 2 (n = 14)	Time 3 (n = 11)
Mean Transition Times to Recess	Control	6.26	6.25	6.42
	PD	5.81	6.35	5.18
	Follow-Up	5.75	5.81	5.31
Mean Transition Times from Recess	Control	10.2	9.50	7. 92
	PD	10.2	8.14	6.37
	Follow-Up	9.18	7.90	6.31
Mean Total Transition Time	Control	16.4	15.7	14.3
	PD	16.0	14.5	11.5
	Follow-Up	14.9	13.7	11.6

Figure 3. Mean Total Transition Scores (Teachers Self-Report) in the Control, Professional Development, and Follow Up Groups at Times 1, 2, and 3
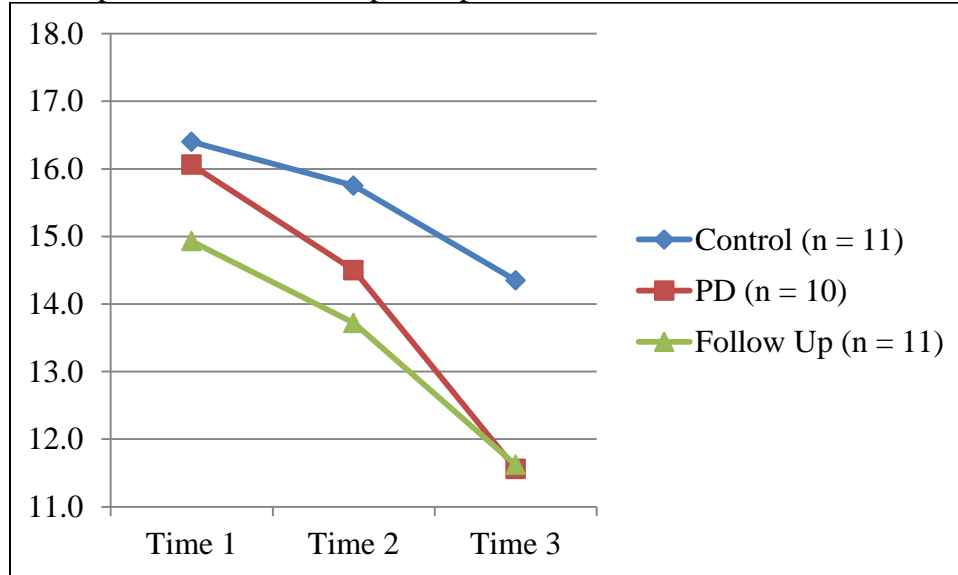
In order to examine the effectiveness of the Mindful Schools program on teachers self-reported transition times were examined using a Repeated Measures ANOVA, comparing the means of all three treatment groups, at Time 1, 2, and 3. There were no statistically significant difference in the Transition to Recess (F= 1.35[8], $p = .253$, $\eta^2 = 0.04$[9]), but there were statically significant results for the Time main effect, meaning there were statistically significant changes in all three time points for, Transition from Recess (F = 10.48, $p = .000$, $\eta^2 = 0.35$) as well as the Total Transition Time (F = 32.0, $p = .004$, $\eta^2 = 0.30$). Although the Time main effect was significant, the time*group main effect was not for all three transition scores[10][11]. As teachers' determination of start and end of a transition time is somewhat subjective, the analysis should be taken as exploratory.[12]

---

[8] The F-test is a ratio of between group and within group variability.

[9] Effect size for ANOVA

[10] Transition to Recess (F = .302, $p = .875$) Transition from Recess (F = .167, $p = .954$) Total Transition (F = .101, $p = .101$, $p = .982$)

[11] Although the Teacher Survey also asked about transition time to lunch, those times were not included in the report analysis. Analyses were conducted using the transition to recess, and lunch, but yielded identical results as presented therefore only results from the transition to recess are presented.

[12] To test whether there was any non-response bias, T-test were conducted to compare teachers who had missing data for Time 2 and/or time 3 to their Time 1 score. There were no statistically significant differences between the teachers who had missing data and those that did not.

Impact of the Mindful Schools Program on Teachers

The impact of the Mindful Schools program on teacher behavior, attitudes, and opinions was examined using the Teacher Survey, which was completed by teachers prior to the start of the program (Time 1), immediately following the program (Time 2), and at the end of the school year (Time 3). Overall, 39 teachers had complete Teacher Survey data (scores on all four constructs at all three data points) and were included in analyses. Teachers' average scores at each of the 3 time points are shown in Table 22 below.

Table 22. Mean Burnout, Compassion Satisfaction, Mindfulness, and Self Efficacy Scores of Teachers at Times 1, 2, and 3

	Control (n = 11)	PD (n = 15)	Follow Up (n = 13)
Burnout			
Time 1	2.84	3.06	3.03
Time 2	2.75	3.20	3.17
Time 3	2.75	3.20	3.17
Compassion Satisfaction			
Time 1	4.57	4.28	4.34
Time 2	4.43	4.11	4.32
Time 3	4.30	4.03	4.22
Mindfulness			
Time 1	4.32	3.91	4.09
Time 2	4.22	4.04	4.24
Time 3	4.18	4.00	4.27
Efficacy			
Time 1	7.60	6.90	7.22
Time 2	7.82	7.08	7.40
Time 3	7.62	7.04	7.69

Short-Term Impact of the Mindful Schools Program on Teachers Self-Reported Behavior and Attitudes

In order to examine the short-term impact of the Mindful Schools program on the participating teachers, OLS regression was used to model the scores obtained from the Teacher Survey at Time 2 as a function of the treatment variable, controlling for teachers' score at Time 1, grade taught, and years taught. Scores on each construct were standardized (mean = 0, SD = 1), consequently the coefficients can be interpreted as standardized effect sizes. The results indicate that immediately after completion of the Mindful Schools program, there were no statistically significant differences in the Burnout, Compassion Satisfaction, or Self Efficacy of teachers who received the program and those that did not, after controlling for grade taught and years teaching (see Table 22; Burnout β = .279, p = .177; Compassion Satisfaction β = .067, p = .770; Self Efficacy β = -.101, p = .658). However, there was a marginally statistically significant effect on teachers' Mindfulness (β = .326, p = .054), suggesting that on average, teachers who participated in the Mindful Schools program tended to have mindfulness ratings that were approximately 0.33 standard deviations higher than teachers in the control group at Time 2.

Table 23. Short-Term Effects of the Mindful Schools Program on Teachers

	Burnout	Compassion Satisfaction	Mindfulness	Self Efficacy
Constant	-.208	.018	-.044	.061
Treatment	.279	.067	.326[+]	-.101
Pre-Survey Score	.797***	.822***	.932***	.738***
Grade 3 - 5	.315	-.209	-.438**	-.176
6+ Years Experience	-.164	.059	.100	.191

*$p \leq$.05; **$p \leq$.01; *** $p \leq$ .001, [+]$p \leq$.1*

Long-Term Impact of the Mindful Schools Program on Teacher Behavior and Attitudes

Overall Long-Term Effects

To examine whether there was any impact of the Mindful Schools program that persisted through the end of the school year, an OLS regression was used to model the teachers' overall scores at Time 3 as a function of the treatment variable (overall), controlling for teachers' scores at Time 1, grade level taught, and years of teaching experience. As with the Time 2 data, scores were standardized (mean= 0, SD = 1) so that coefficients can be interpreted as standardized effect sizes.

The results indicate that, at Time 3, there were no statistically significant differences in the responses of teachers who received the Mindful Schools program and those that did not for all four constructs (see Table 24; Burnout $\beta = -.005$, $p = .985$; Compassion Satisfaction $\beta = .181$, $p = .499$; Mindfulness $\beta = .314$, $p = .192$; Efficacy $\beta = .193$, $p = .478$). These results indicate that any initial effect of the program on teachers' Mindfulness was not sustained.

Table 24. Long-Term Effect of the Mindful Schools Program Participation on Teachers (Time 3)

	Burnout	Compassion Satisfaction	Mindfulness	Efficacy
Constant	.082	-.217	.035	-.187
Treatment	-.005	.181	.314	.193
Pre-Survey Score	.586***	.743***	.815***	.765***
Grade 3 - 5	.363	-.184	-.608**	-.060
6+ Years Experience	-.433+	.131	-.016	.056

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, $^+ p \leq .1$

Comparison of the Effectiveness of Follow-Up and Professional Development

To compare the scores of teachers in the follow-up and professional development groups, regression was used to model the survey scores at Time 3 as a function of the follow up and PD variables, controlling for teachers' scores at Time 1, grade, and years taught. Survey construct scores were standardized (mean = 0, SD = 1) so coefficients can be interpreted in standard deviation units. There were no differences in the scores of the teachers in the three groups prior to controlling for teachers' initial scores, grade, and years taught (see Table 25; Burnout $\beta = .034$, $p = .904$; Compassion Satisfaction $\beta = .070$, $p = .813$; Mindfulness $\beta = .242$, $p = .369$; Self Efficacy $\beta = .006$, $p = .983$).

Table 25. Effect of Mindful Schools Program plus Follow Up or Professional Development on Teachers

	Burnout	Compassion Satisfaction	Mindfulness	Self Efficacy
Constant	.087	-.233	.024	-.216
Mindful Schools + PD	.034	.070	.242	.006
Mindful Schools + Follow Up	-.050	.302	.392	.388
Pre-Survey Score	.585***	.733***	.806***	.734***
Grade 3 - 5	.358	-.168	-.596**	-.043
6+ Years Experience	-.438	.150	-.004	.104

*$p \leq .05$; **$p \leq .01$; *** $p \leq .001$, $^{+}p \leq .1$

Teachers' Perceptions of the Mindful Schools Program

Teachers' General Attitudes toward Mindfulness

In the teacher survey, teachers were asked several questions regarding their attitudes toward Mindfulness and their perceptions of the program. After completion of the program, teachers were asked to describe their attitude toward Mindfulness on a 4-point scale (1 = "very unfavorable" to 4 = "very favorable"). As shown in Figure 6 below, teachers in the professional development group tended to have the most positive attitudes toward Mindfulness, followed by teachers in the follow-up group.[13] However, from the pre-survey to the end of the school year, there was no change in the proportion of teachers who reported "very favorable" attitudes toward Mindfulness in either the professional development ($p = 1.00$) or follow-up ($p = .453$) groups.

---

[13] Differences in the proportions of teachers reporting attitudes were very unfavorable, slightly unfavorable, slightly favorable, very favorable, or "don't know/unsure" were not statistically significant across the three groups ($\chi^2[8] = 14.66$, $p = .066$).

Figure 4. Percentage of Teachers Who Reported that Their Attitude Toward Mindfulness was "Slightly" or "Very" Favorable at the End of the School Year
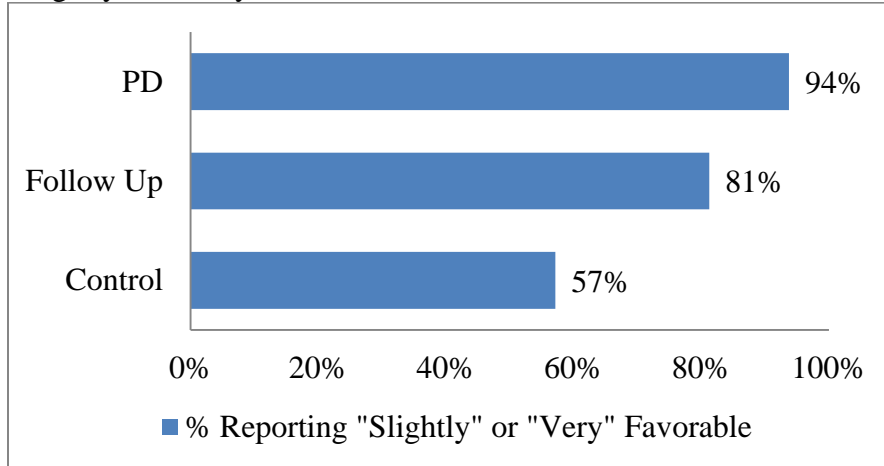
Similarly, teachers were asked to report how useful they believe that it is to teach Mindfulness to students and teachers on a 4-point scale (1 = "not at all useful" to 4 = "highly useful"). The majority of teachers in each of the three groups reported that they feel it is important to teach Mindfulness to students and teachers. The largest proportion of professional development teachers felt that it is useful to teach Mindfulness to students and teachers, with 100% of the teachers reporting that teaching Mindfulness is "moderately" or "highly" useful (Figure 7).[14] However, as with teachers' general attitudes toward Mindfulness, from the pre-survey to the end of the school year, there was no change in the proportion of teachers who reported that they felt it is highly useful to teach Mindfulness in either the professional development group ($p = 1.00$) or the follow-up group ($p = 1.00$).

---

[14] Differences in the proportions of teachers reporting that their attitudes were very unfavorable, slightly unfavorable, slightly favorable, very favorable, or "don't know/unsure" were statistically significant across the three groups ($\chi^2[6] = 15.01, p = .020$).

Figure 5. Percentage of Teachers Reporting That They Feel it is "Moderately" or "Highly" Useful to Teach Mindfulness to Students and Teachers
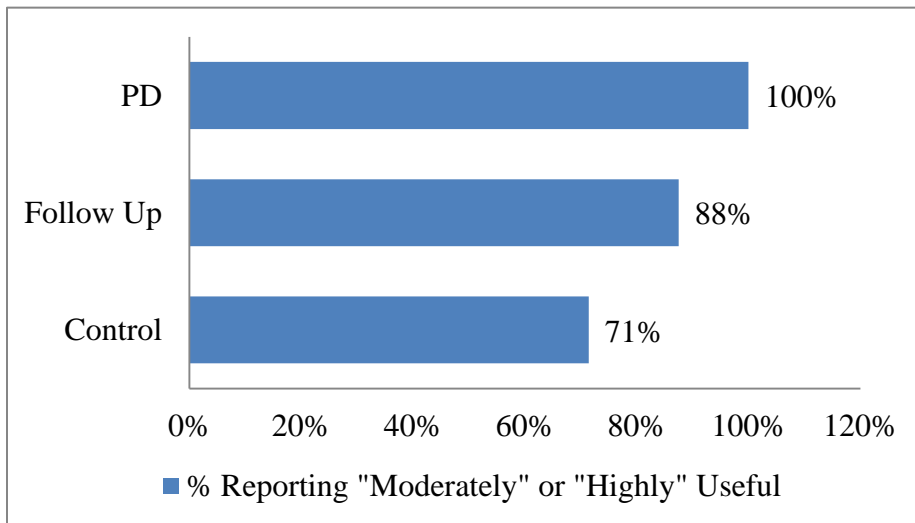
When asked to explain why they did or did not believe that it is useful to teach Mindfulness to students and teachers, teachers generally responded positively. The most common reason teachers cited for why it is useful to teach Mindfulness was that it helps both teachers and students improve their awareness and to stay grounded in the moment. For example, one teacher wrote that "for many of our students, Mindfulness is essential for having our students 'leave their lives at home' and focus on being at school so they can learn and enjoy school." Similarly, another teacher reported that "for children, it helps them become aware of their ability to control their thoughts and bodies, allowing them space to pay attention to their learning" and that Mindfulness was a "tool for [teachers] to have in order to help us remain present when we work with our students."

Likewise, several teachers reported that Mindfulness helps both students and teachers deal with stress. As one teacher reported, Mindfulness "calms everyone down, regains the group's focus, gives moments of peace that we usually don't get in the classroom" and another reported that, for teachers, it "help[s] manage the high stress and emotions that come with the work we do."

In addition, several teachers reported that it integrates well into the classroom and into students' lives. For example, one teacher reported that Mindfulness "has become a regular part of the day" and that the "students talk about Mindfulness often…and ask to do Mindfulness before tests."

Although most responses were generally positive, a few teachers did cite some shortcomings of the program. Two teachers indicated it was difficult to find time to fit Mindfulness into their lessons, saying that "it is highly useful; however it takes time off our packed curriculum." Similarly, one teacher said "I have a really hard time trading learning time for Mindfulness time, even though I know it can be powerful." Another teacher also reported that the "techniques and

methods used to teach Mindfulness were not appropriate for each grade level" and that "the exercises felt repetitive and boring to some of the children."

Finally, several teachers reported that Mindfulness really depends on the attitude and buy-in of the teacher. As one teacher stated, "I feel that it really depends on the person whether they believe Mindfulness is for them or not. As for myself, I am very much in tune with my body, mind and soul…Mindfulness is for me, but it may not be for others who aren't in tune with themselves."

Professional Development and Follow-Up Teachers Use of Mindfulness

In addition, the teachers in the professional development and follow-up groups were asked an additional set of questions on the post-survey regarding their experiences with the program. These questions related to their use of Mindfulness strategies in the classroom, the benefits of the program, aspects they would change, and their feeling about using Mindfulness in the future.

Use of Mindfulness in the Classroom

First, the teachers were asked to report how often, since January, they practiced Mindfulness in their class. The majority of teachers in both groups reported that they practiced Mindfulness "often" (see Figure 8).

Figure 6. Teachers' Report of the Frequency they Practiced Mindfulness in their Class since January 2012
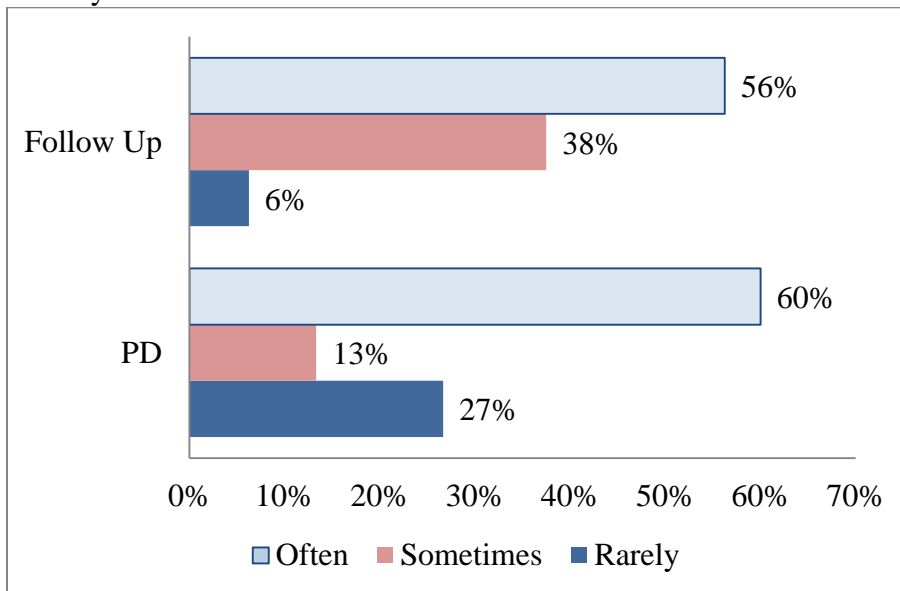
There was no statistically significant difference in the frequency with which the two groups reported practicing Mindfulness (Mann-Whitney U = 113.5, Z = -.290, *p* = .800).

Similarly, the teachers were asked to report on when they practiced Mindfulness in their classes.[15] Across both groups, the most common time that Mindfulness was practiced was after lunch, followed by in the morning (see Figure 9). Several teachers reported that they used Mindfulness strategies when students returned from recess or lunch because it "helps us to get focused again," and because it is "good for getting students to settle down." One teacher reported that he/she used Mindfulness "every day when they come in, after lunch…it is a routine they value and enjoy. It gets them calm and ready to do their best."

Figure 7. Percentage of Teachers Reporting Practicing Mindfulness in their Class During Different Times of the Day
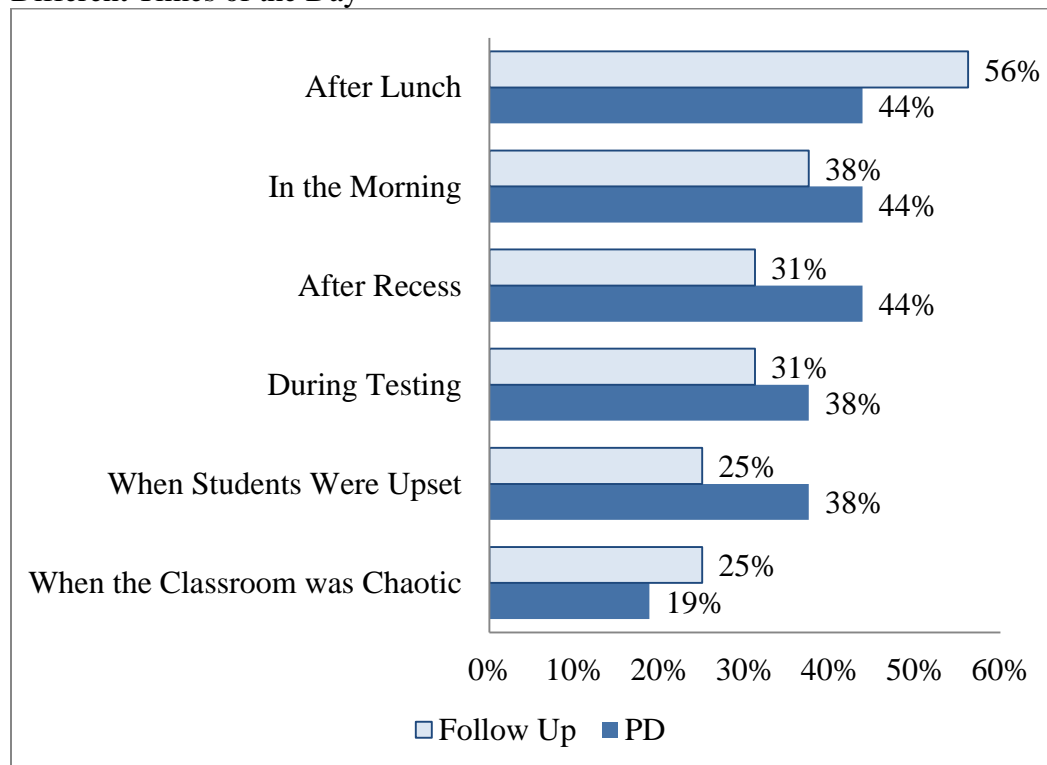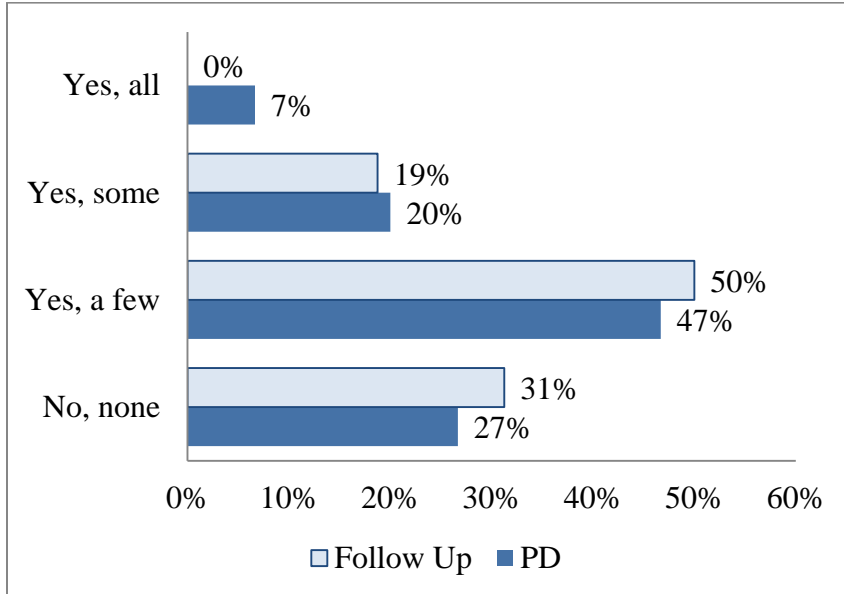

There were no statistically significant differences in the proportions of teachers in each group that reported practiced Mindfulness at any of the times.

Finally, teachers were asked whether or not they noticed their students using Mindfulness strategies on their own. In both groups, approximately 20% of the teachers reported that they noticed at least some of their students using Mindfulness strategies on their own and another 50% reported that they noticed a few students using the strategies (Figure 10). However, approximately 30% of the teachers reported that they did not notice any students using the strategies on their own.

---

[15] Teachers were asked to report whether or not they practiced Mindfulness during six different times of the day: a) in the morning, b) after recess, c) after lunch, d) when students were upset, e) when the classroom was chaotic, and f) during testing.

Figure 8. Percentage of Teachers who Noticed "All," "Some," and "A Few" Students Practicing Mindfulness on their Own

There was no statistically significant difference in the proportions of teachers who noticed students using Mindfulness strategies on their own across the two groups (Mann-Whitney U = 107.5, Z = 243.5, p = .593).

When asked to report on when they noticed students using Mindfulness strategies on their own, most teachers reported that they noticed students using the strategies when students were in stressful or difficult situations. For example, one teacher reported that, "A student recited a poem for the school-wide Oratorical Fest. He forgot the next line, took a deep breath. And another. And continued the poem without any hesitation. Afterwards he exclaimed to me, 'I almost forgot but then I used Mindfulness!'" Similarly, other teachers reported that they saw students use Mindfulness strategies when they got upset or when the classroom became chaotic. Finally, one teacher said that during test taking, "some students reported using [Mindfulness] to help them focus."

Benefits of the Mindful Schools Program
The teachers in the professional development and follow-up groups were asked to respond to open-ended questions regarding what they felt was the greatest benefit to their students and to themselves, as teachers. Teachers cited many benefits of the program for students, including improving their self-awareness, self-control, behavior, and awareness of the others. In addition, several teachers mentioned that the Mindfulness helped students learn to be calm and focused. As one teacher said, "There are so many things they have gained from Mindful Schools - being able to be more self aware, to think of others, to be quiet and sit still for several minutes, to be able to calm themselves down, to be able to find courage and strength, to be able to share this with their families."

Teachers also reported that they, themselves, benefited from the program in similar ways. Many teachers reported that the program helped them to control their emotions, stay calm, stay in the present moment, and regain focus. For example, one teacher said, "I have been able to use Mindfulness to calm anxiety, to help control my temper. To allow myself a quiet space in my mind (or working towards being able to have a quiet space)" and another reported that, "For me, the aspect I benefited from most was learning to look at each situation as it was in the present moment. I learned to look at events without judgment and to take a step back to think about certain situations before acting upon them."

Several teachers also reported that the greatest benefit was that they learned new strategies for keeping the class focused and calm. Finally, two teachers reported that they felt the greatest benefit was having a strategy that they shared with their students. One teacher said, "My students and I use Mindfulness daily, and because we both went through the same 8 week course we have a shared vocabulary around Mindfulness." Similarly, another teacher reported that the greatest benefit was "sharing a common knowledge and experience with the students of a strategy that we can use."

Benefits of the Follow-Up and Professional Development Augmentations
In addition, teachers in both the follow-up and professional development groups were asked whether they found the training that they received in addition to the standard Mindful schools program to be valuable. For the most part, teachers in both the follow-up and professional development groups responded positively, reporting that the additional training was beneficial. The most frequently cited benefit among teachers in the professional development was that the professional development allowed them to focus on their own Mindfulness. For example, one teacher reported that "being able to focus myself and my own self care was a great gift."

Although several teachers in the follow-up group reported that the additional training was redundant, saying that "the lessons were the same" and that "no new material or exercises were learned," many teachers reported that the continued training helped to reinforce what students learned during the regular Mindful Schools program. Teachers reported that it gave "students some extra Mindfulness practice" and that there was "some value in reinforcing what they learned." Many teachers felt that this extra practice led to students being more confident and being comfortable leading Mindfulness exercises. For example, one teacher reported that the "students are now able to be Mindful teachers and they enjoy taking the lead." Similarly, another teacher said that the students are "truly Mindfulness teachers and lead the sessions."

Suggestions for Changes to the Program
In addition, the teachers were asked for suggestions for ways to improve the program. The most commonly cited way in which teachers suggested the program could be changed was to make it longer or more frequent. Teachers suggested making the program "last the entire school year," "making each session longer," and having the "teachers responsible come more often in the beginning." Despite the large number of teachers who felt the length or frequency of the program

should be increased, two teachers did suggest that the program could be made shorter and one suggested that the program should be implemented outside of class time.

Another common suggestion for improving the program was to ensure that the Mindful Schools staffs were adequately prepared to work in the classroom. One teacher reported that "the actual interactions I had with Mindful schools staffs were often very strained, and I felt that there was both a lack of understanding of classroom management and a lack of communication skills with teachers." Similarly, another said, "Our initial teacher didn't have much experience with classroom management. It worked out fine with my well-behaved class, but I could see it being a potential problem. He wasn't very patient...even with my well-behaved students." The skills and preparation seemed to vary from staff member to staff member. As one teacher reported "The teacher makes a big difference. Ms. Eileen was very real, down to earth, calm, and confident. That made a big difference in my classroom in comparison to the first teacher."

Use of Mindfulness in the Future
Teachers were asked whether they will continue to use Mindfulness in the classroom next year. Over half (58.1%) of the teachers reported that they would "definitely" use Mindfulness and no teacher reported that they would not use Mindfulness in the classroom. There were no statistically significant differences in teachers' responses across the two groups (Figure 11).

Figure 9. Percentage of Teachers Reporting That They Will Use Mindfulness in Their Classroom Next Year
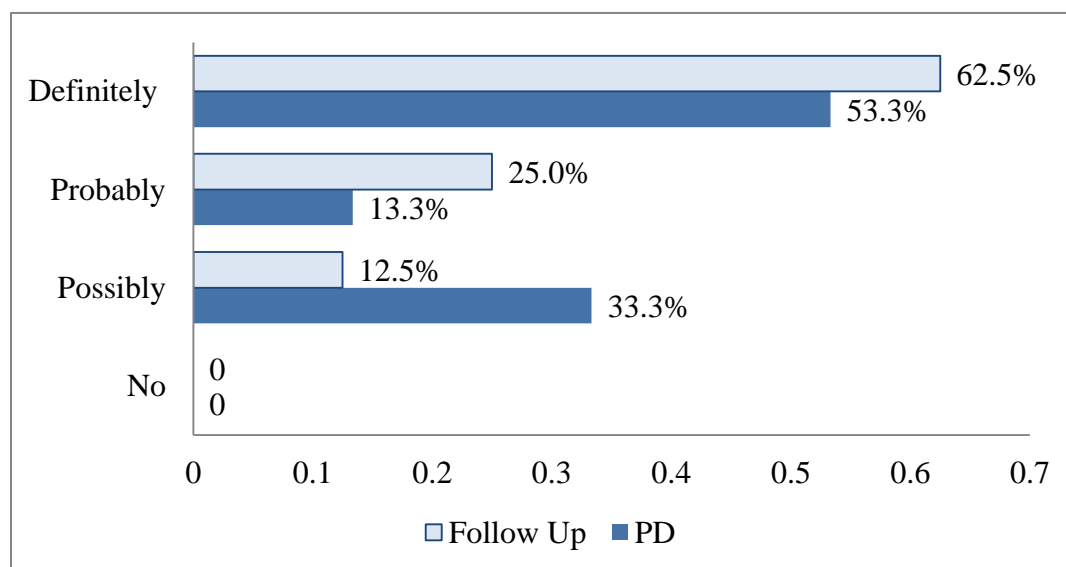
There was no statistically significant difference in the proportions of teachers who reported that they are planning to use Mindfulness in the classroom next year across the two groups (Mann-Whitney U = 101.00, Z = -0.847, p = .470).

In addition, teachers were asked what types of Mindfulness support they would like next year. Among teachers in the professional development group, the largest proportion of teachers reported that they would like sample scripts (44%), followed by additional training (38%; (Figure 12). Among teachers in the follow-up group, the largest proportion of teachers reported that they would like sample scripts (50%) and a Mindfulness calendar (50%).

Figure 10. Percentage of Teachers Reporting They Would like Mindfulness Support in the Following Areas:
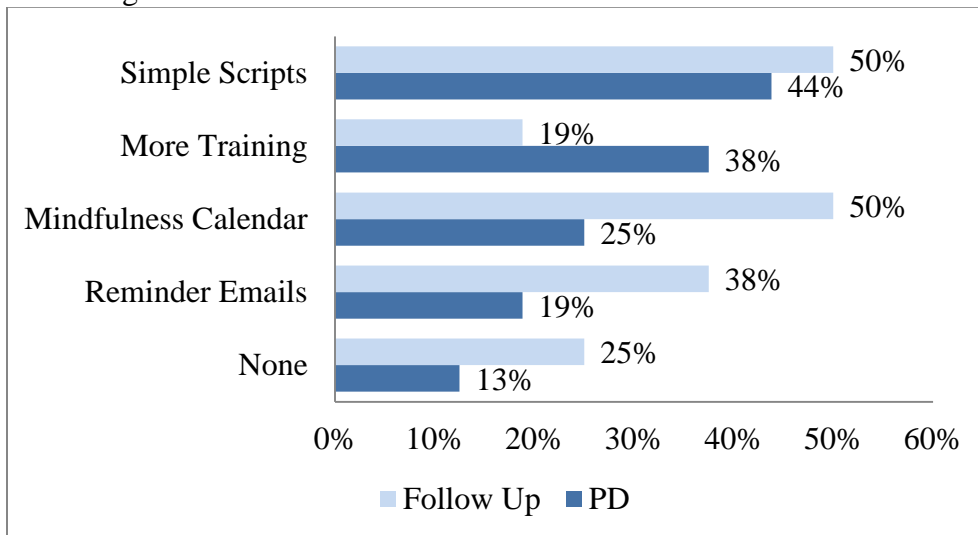
There were no statistically significant differences in the proportions of teachers who reported wanting each type of support across the two groups

Finally, the teachers were asked whether or not they would like to see the Mindful Schools Program return to their school. In each group, 93% of teachers reported that they would like to see the program return. Several teachers reported that they would like to see the program return in order to receive ongoing coaching and training for teachers and students. One teacher reported that "Teachers and students will need a refresher by next year" and another said that "I think ongoing coaching for classes would be extremely helpful in continuing to grow a culture of Mindful Schools on our campus." Similarly, several teachers reported that it would be beneficial for the program to return in order to have an impact on teachers and students who were not exposed to the program this year.

Conclusions and Recommendations

In sum, the evaluation of the Mindful Schools program sought to examine the short and long-term effects of the program, as well as to compare the effectiveness of two program augmentations designed to sustain the effects of the program. The results of the evaluation suggest that the program had a small effect on student behavior (as rated by their teachers) immediately following the standard program. The students in the treatment group had behavior

ratings that were, on average, one tenth of a standard deviation higher than the behavior ratings of students in the control group. This impact of the program on student behavior persisted through the end of the school year. Students who participated in the Mindful Schools program continued to have behavior ratings that were approximately 0.10 standard deviations higher than those in the control group several months after the end of the standard Mindful Schools program. Furthermore, the findings suggest that the professional development component may have been somewhat more effective for sustaining the impact of the standard program on student behavior. Students taught by teachers who received the professional development had behavior ratings that were approximately 0.13 standard deviations higher than those in the follow-up group, which is a small effect size.

In addition, the evaluation results suggest that the program had a short-term effect on the Mindfulness of the teachers, immediately following the standard Mindful Schools program. However, these effects did not persist through the end of the school year.

In contrast, the program was associated with a negative impact on students' Alertness; however this effect was not evident until the end of the school year (and it is difficult to attribute this decrease in Alertness solely to the Mindful Schools treatment – there are many student behavior changes that are apparent as the school year winds down that may have impacted the study results). Furthermore, the program did not have any short-term effects on students' orienting or conflict attention scores or students' self-report of Mindfulness. Similarly, there were no significant short-term effects on teachers' compassion satisfaction, feelings of burnout, or self-efficacy.

Finally, the evaluation results indicate that, in general, teachers were satisfied with the program, found it to be beneficial, and plan to utilize Mindfulness practice in the future. Teachers reported that Mindfulness helps both teachers and students improve their awareness, stay grounded in the moment, deal with stress, and regain focus. In addition, the teachers reported that the program helped to improve student behavior and self-control. Finally, the majority of teachers noticed at least a few of their students using Mindfulness strategies on their own.

Limitations of the Evaluation

Although the evaluation was unable to detect many statistically significant effects of the program, this is likely due, at least in part, to limitations of the evaluation. First, the sample size (47 teachers) is fairly small, limiting the statistical power of the analyses, and consequently limiting the ability to detect statistically significant effects. Furthermore, as discussed in the description of the student survey, there were some limitations of the survey instrument which may bias the results. The questions included on the survey were likely too advanced for the students' reading level, which threatens the validity of the data, and anecdotal reports suggest that teachers may not have been consistent in their survey administration practice. Finally, administration of the ANT-C in school settings is vulnerable to many extraneous factors that may impact student attention and compliance with test procedures.

References

B. Hudnall Stamm, 2009. Professional Quality of Life: Compassion Satisfaction and Fatigue
    Version 5 (ProQOL). /www.isu.edu/~bhstamm or www.proqol.org

Greco, L. A., Baer, R. A., & Smith, G. T. (2011, April 11). Assessing Mindfulness in Children
    and Adolescents: Development and Validation of the Child and Adolescent Mindfulness
    Measure (CAMM). *Psychological Assessment*. Advance online publication. doi:
    10.1037/a0022819

Johnson, K.A., Robertson, I. H., Barry, E., Mulligan, A.,Dáibhis, A., Daly, M., Watchorn, A.,
    Gill, M. & Bellgrove, M.A. (2008). Impaired conflict resolution and alerting in children
    with ADHD: evidence from the Attention Network Task (ANT). *Journal of Child
    Psychology and Psychiatry, 49* (12), 1339-1347.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: developmental properties
    and sociodemographic correlates in an epidemiological sample of young, urban children.
    Child Development, 75 (5), 1373-1386.

Tschannen-Moran, M., Woolfolk}Hoy, A., & Hoy, W. K. (1998). Teacher Efficacy: Its meaning
    and measure. Review of Educational Research, 68(2), 202}248.

Appendix A: Examining Non-Response Bias in Teacher Survey Data

As in most studies, there were missing data in each data set. In order to address whether these missing data points might be a source of bias in the results, the pattern of missing data is examined in this appendix.

To test whether there was any non-response bias, T-test were conducted to compare teachers who had missing data for Time 2 and/or time 3 to their Time 1 score. There were no statistically significant differences between the teachers who had missing data and those that did not; this is also true for the remaining measures.

Teacher Survey Missing Data

Of the 8 teachers with missing data on the teacher's survey, 7 had the majority of missing data from Time 2; one teacher had the majority of their missing data from Time 3. Half of the missing responses were from control group teachers, the other half from treatment teachers, with the majority of those among the intervention with classroom follow-up group.   See Table A1 below for group assignment frequencies.

Table A1. Missing Data by Group

Group	Frequency	Percentage
Control	4	50.0
Intervention with PD	1	12.5
Intervention with Follow-up	3	37.5

Most of the teachers with missing data (75%) taught grade 4 or higher (see Table A2 below).

Table A2. Missing Data by Grade Taught by Teachers

Grade	Frequency	Percentage
K	1	12.5
1	1	12.5
2	1	12.5
4	2	25.0
5	3	37.5

The number of years a teacher taught was also examined to determine if there were any trends in teachers with missing data. There appeared to be a wide spread of years of experience, ranging from 4 to 16 years. Additionally, teacher experience years were fairly evenly spread throughout the teachers with missing data. See Table A3 below for frequency of number of years taught.

Table A3. Missing Data by Number of Years Taught

Number of Years Teaching	Frequency	Percentage
4	2	25.0
5	2	25.0
8	2	25.0
13	1	12.5
16	1	12.5

Appendix B: Examining Non-Response Bias

Kinder Rubric Missing Data

There were a total of 137 students who had missing scores on the Kinder rubric at Time 1, Time 2, and Time 3, because of the nature of the analysis (regression) only students who had data points for all three time periods were included in the analysis. Therefore 137 students are the ones being described in the following tables. Missing data are again almost evenly split between control and treatment, with the classroom intervention treatment group having more missing data than those with follow-up PD (Table B1).[16]

Table B1. Missing Data by Group

Group	Frequency	Percentage
Control	72	52.6%
Intervention with PD	21	15.3%
Intervention with Follow-up	44	32.1%

Missing students' grade levels were fairly evenly spread out (see Table B2 below).

Table B2. Missing Data by Grade

Grade	Frequency	Percentage
K	26	19.0%
1	28	20.4%
2	9	6.6%
3	29	21.2%
4	30	22.0%
5	15	10.9%

Student gender was also examined to see if there were any trends in which students did not have Kinder data. . There were slightly more males that were missing data (see Table 3 below).

Table B3. Missing Data by Student Gender

Gender	Frequency	Percentage
Female	46	42.2%
Male	63	57.7%

---

[16] To test whether there was any non-response bias, T-test were conducted to compare teachers who had missing data for Time 2 and/or time 3 to their Time 1 score. There were no statistically significant differences between the teachers who had missing data and those that did not

Appendix C

ANT-C Missing Data

There were a total of 85 students who had missing data for the ANT-C measure at Time 1, Time 2, and Time 3, because of the nature of the analysis (regression) only students who had data point for all three time periods were included in the analysis. Therefore 85 students failed to have data for all three time periods, and are the ones being described in the following tables. The largest group of the students who were missing data (42%) came from the Intervention with PD group (Table C1).[17]

Table C1. Missing Data by Group

Group	Frequency	Percentage
Control	28	32.9%
Intervention with PD	36	42.4%
Intervention with Follow-up	21	24.7%

Most (77%) of students missing ANT-C data were in grade 4 or higher (see Table C2 below). Perhaps this was due to the fact that older students are under more time constraints, and trying to fit in the ANT-C was not possible.

Table C2. Missing Data by Grade

Grade	Frequency	Percentage
3	19	22.4%
4	28	32.9%
5	38	44.7%

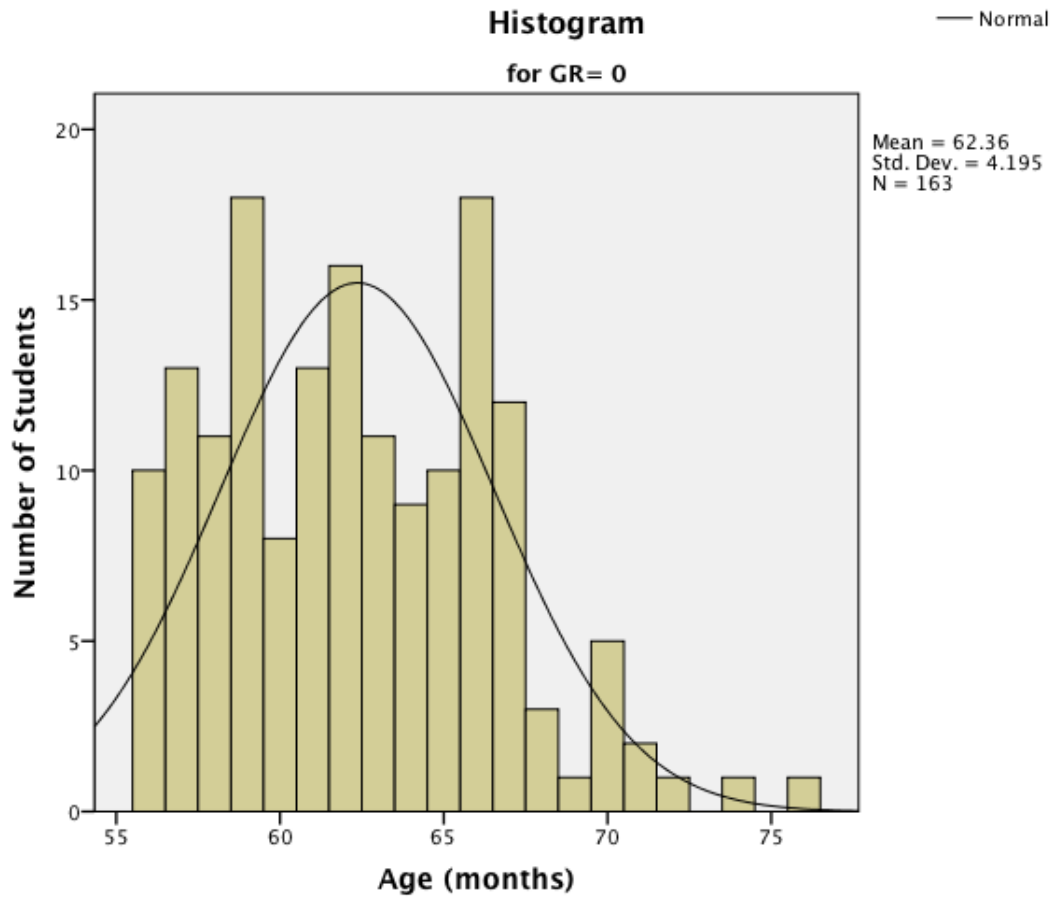Student gender was also examined to see if there were any trends in which students participated (Table C3).

Table C3. Missing Data by Gender

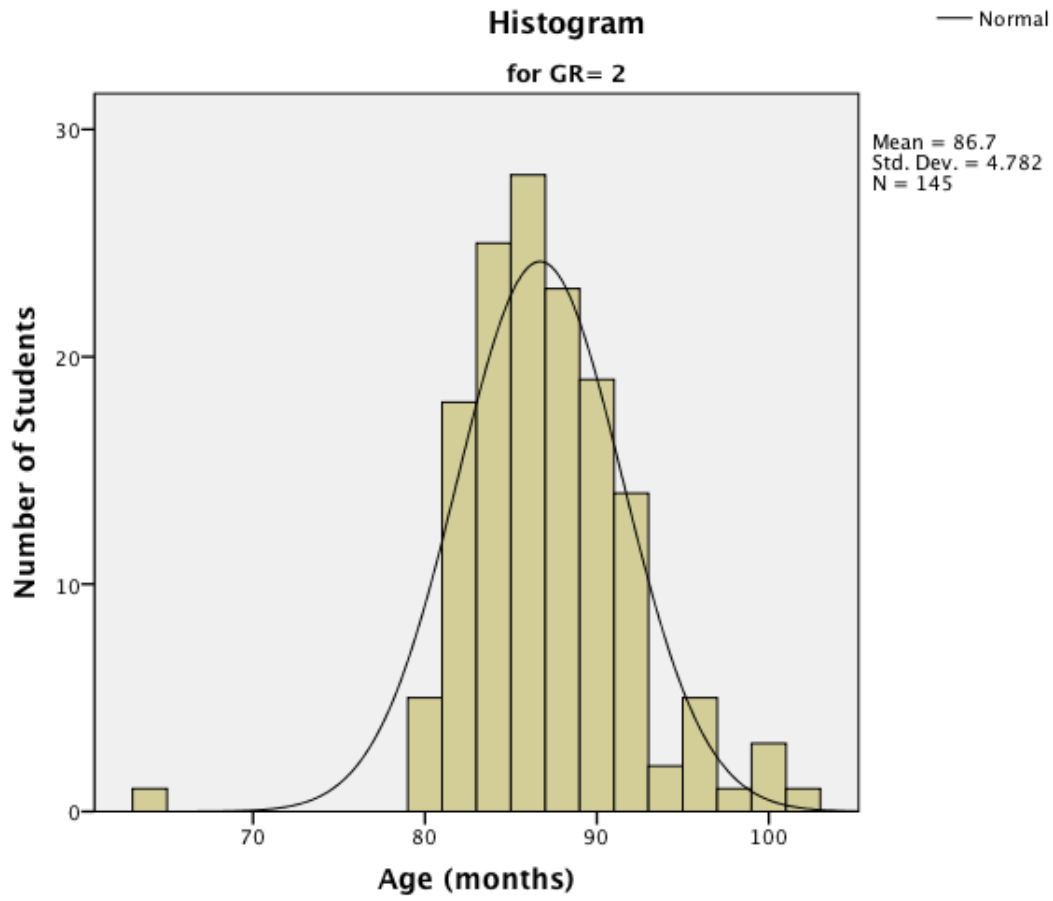| Gender	Frequency	Percentage
Female	34	43.0%
Male	45	57.0%

---

[17] Based on the nature of the data non- response bias analyses were not conducted. But based on the results of the non-response analysis for the other measures, there is likely to be no such bias in the ANT-C, because the previous analysis did not yield significant results
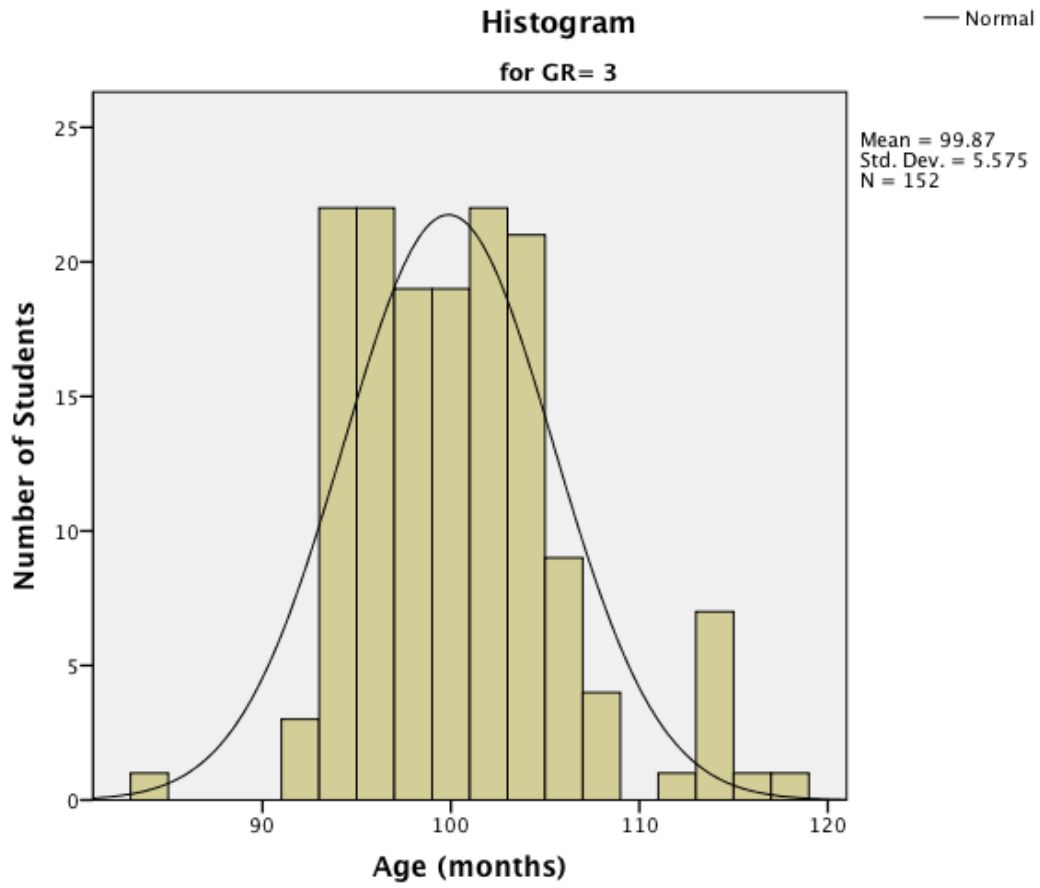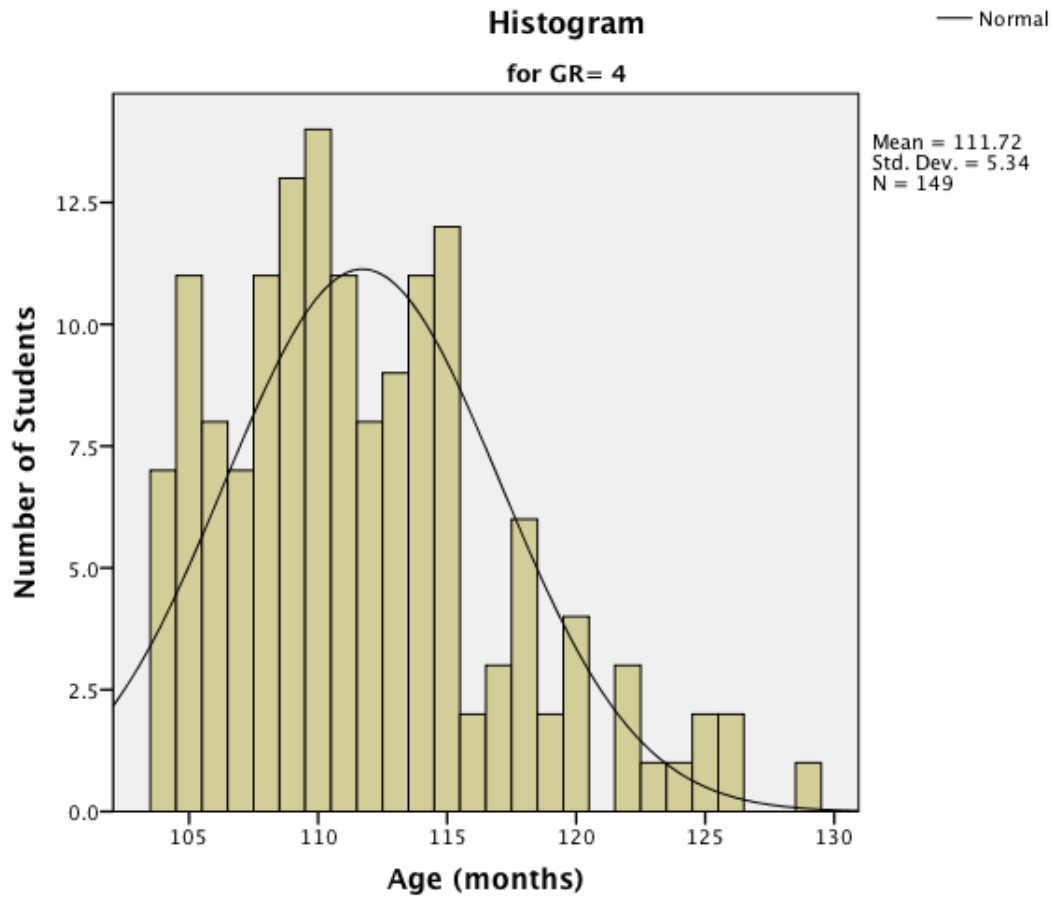
Appendix D: Student Age by Grade

Because age in months was a variable in several analyses, there was some concern that students of the same age might be in different grades, making interpretation of the results, particularly for the Kinder rubric scores, which are provided by the classroom teacher, somewhat problematic. For example, if a 5th grade child who is "young" compared to their classmates is the same age as a 4th grade child who is "old" compared to their classmates, there is some question about the teachers' standard of reference when scoring the Kinder rubric. We have no easy way of addressing this issue or even identifying if it is an issue. Histograms were run for each separate grade level to determine if there was a large amount of overlap of age in months of students between grade levels. Based on the histograms presented below, there does not appear to be a substantial amount of overlap in children's ages between grade levels, although this does vary somewhat between grades with the standard deviation of age per grade increasing with grade level. Consequently, if there is any bias due to overlap of age between grades, it is most likely to occur in the higher grades – particularly in grades 4 and 5. Because there were so few classrooms per treatment group per grade (see Table 1 in the main report), it is not methodologically sound to enter both grade and age in regression equations. Exploration of the impact of age within grades on outcome measures will require a larger sample of classrooms.

Histogram

for GR= 0

— Normal

Mean = 62.36
Std. Dev. = 4.195
N = 163

Number of Students

Age (months)

Histogram

for GR= 1

—— Normal

Mean = 74.32
Std. Dev. = 4.36
N = 157

Age (months)

Number of Students

Histogram

for GR= 2

— Normal

Mean = 86.7
Std. Dev. = 4.782
N = 145

Number of Students

Age (months)

Histogram

for GR= 3

Mean = 99.87
Std. Dev. = 5.575
N = 152

— Normal

Number of Students

Age (months)

Histogram

for GR= 4

— Normal

Mean = 111.72
Std. Dev. = 5.34
N = 149

Number of Students (y-axis)

Age (months) (x-axis)

Histogram
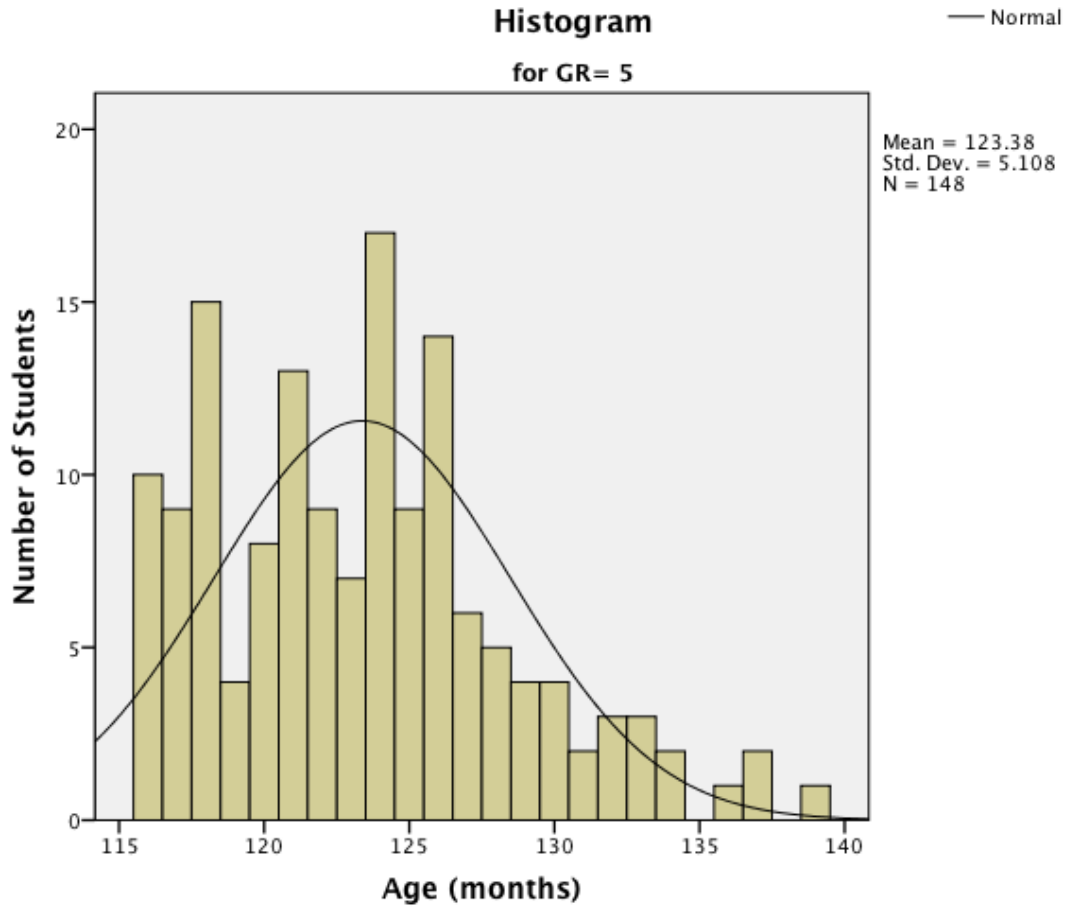
for GR= 5

Mean = 123.38
Std. Dev. = 5.108
N = 148

Appendix E: Kinder Survey Regressions by Gender

Short-Term Effects of the Mindful Schools program participation on Student Behavior (Females)

	β
Constant	.194
Mindful Schools	.091
Pre-Behavior Score	.565***
Age in Months	-.001

*p ≤.05; **p ≤.01; *** p ≤ .001, + p ≤ .1

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Short-Term Effects of the Mindful Schools program participation on Student Behavior (Males)

	β
Constant	-.480*
Mindful Schools	.119
Pre-Behavior Score	.635***
Age in Months	.003

*p ≤.05; **p ≤.01; *** p ≤ .001, + p ≤ .1

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Long-Term Effects of the Mindful Schools program participation on Student Behavior (Females)

	β
Constant	1.03***
Mindful Schools	.096
Pre-Behavior Score	.547***
Age in Months	-.009***

*p ≤.05; **p ≤.01; *** p ≤ .001, ⁺p ≤.1

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Long-Term Effects of the Mindful Schools program participation on Student Behavior (Males)

	β
Constant	.278
Mindful Schools	.139
Pre-Behavior Score	.603***
Age in Months	-.005*

*p ≤.05; **p ≤.01; *** p ≤ .001, ⁺p ≤.1

Note: The reference group includes students in the control group. The Mindful schools coefficient represents the difference in the behavior ratings of the treatment and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Effect of Mindful Schools program plus follow up or professional development on Student Behavior (Females)

	β
Constant	1.03***
Mindful Schools + PD	.115
Mindful Schools + Follow Up	.075
Pre-Behavior Score	.547***
Age in Months	-.009***

*p ≤.05; **p ≤.01; *** p ≤ .001, ⁺p ≤.1

Note: The reference group includes students in the control group. The Mindful Schools + PD coefficient represents the difference in the behavior ratings of the PD and control groups, after controlling for the other variables in the model. The Mindful Schools + Follow Up coefficient represents the difference in the behavior ratings of the follow up and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Effect of Mindful Schools program plus follow up or professional development on Student Behavior (Males)

	β
Constant	.275
Mindful Schools + PD	.154
Mindful Schools + Follow Up	.122
Pre-Behavior Score	.602***
Age in Months	-.005*

*p ≤.05; **p ≤.01; *** p ≤ .001, ⁺p ≤.1

Note: The reference group includes students in the control group. The Mindful Schools + PD coefficient represents the difference in the behavior ratings of the PD and control groups, after controlling for the other variables in the model. The Mindful Schools + Follow Up coefficient represents the difference in the behavior ratings of the follow up and control groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Comparison of the Effects of the Professional Development and Follow Up Sustainability Models (Females)

	β
Constant	1.11***
PD	.122
Pre-Behavior Score	.539***
Age in Months	-.010***

*p ≤.05; **p ≤.01; *** p ≤ .001, + p ≤ .1

Note: The reference group includes the students in the follow up group. The PD coefficient represents the difference in the behavior ratings of the PD and follow up groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Comparison of the Effects of the Professional Development and Follow Up Sustainability Models (Males)

	β
Constant	.177
PD	.147
Pre-Behavior Score	.636***
Age in Months	-.004

*p ≤.05; **p ≤.01; *** p ≤ .001, + p ≤ .1

Note: The reference group includes the students in the follow up group. The PD coefficient represents the difference in the behavior ratings of the PD and follow up groups, after controlling for the other variables in the model. The outcome variable, Time 3 overall Kinder Rubric score, was standardized; therefore the coefficients represent effect sizes in standard deviation units.

Appendix F: Kinder Survey Subscale Mean Change by Gender

The four individual behavior ratings for females on the Kinder Rubric were analyzed separately to examine whether the Mindful Schools program had a short-term impact on female students' emotional, physical, social, and mental behavior. As shown in Table 1, there was no difference in mental (Z= -1.80, p = 0.07, r = -0.09[18]), emotional (Z = -0.89, p = .37, r = -0.04), physical (Z= -1.66, p = 0.10, r = -0.09), or social (Z = -0.53, p = .60, r = -0.03) behavioral changes of the students in the two groups.

Table 1. Female, Mean Change from Time 1 to Time 2 on Kinder Rubric Scales

	Treatment (n =275)	Control (n =106)
Mental	0.31	0.15
Emotional	0.23	0.25
Physical	0.29	0.16
Social	0.20	0.09

* p ≤ .05; ** p ≤ .01; ** p ≤ .001
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

The long-term impact on female students' emotional, physical, social, and mental behavior was also examined. As shown in Table 2, there was a significant difference in mental behavior (Z = -1.98, p = 0.05, r = -0.10) between the two groups. However, there was no difference in emotional (Z = -1.24, p = 0.22, r = -0.06), physical (Z= -0.61, p = 0.55, r = -0.03), or social (Z = -0.05, p = 0.96, r = 0.00) behavioral changes of the students in the two groups.

Table 2. Female, Mean Change from Time 1 to Time 3 on Kinder Rubric Scales

	Treatment (n =275)	Control (n =106)
Mental	0.45	0.26*
Emotional	0.25	0.33
Physical	0.31	0.22
Social	0.17	0.16

* p ≤ .05; ** p ≤ .01; ** p ≤ .001
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

---

[18] Reported effect sizes were calculated using the following formula: $r = \frac{Z}{\sqrt{N}}$

r values are considered to indicate small effect sizes at .1, medium at .3, and large at .5 or above.

The four individual behavior ratings for males on the Kinder Rubric were analyzed separately to examine whether the Mindful Schools program had a short-term impact on male students' emotional, physical, social, and mental behavior. As shown in Table 3, there was a significant difference in mental (Z= -2.23, p = 0.03, r = -0.11) behavior between the two groups. However, there was no difference in emotional (Z = -1.09, p = 0.28, r = -0.05), physical (Z= -1.51, p = 0.13, r = -0.07), or social (Z = -1.37, p = 0.17, r = -0.07) behavioral changes of the students in the two groups.

Table 3. Male, Mean Change from Time 1 to Time 2 on Kinder Rubric Scales

	Treatment (n = 298)	Control (n = 121)
Mental	0.26	0.04*
Emotional	0.31	0.21
Physical	0.28	0.13
Social	0.27	0.11

* $p \leq .05$; ** $p \leq .01$; ** $p \leq .001$
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.

The long-term impact on male students' emotional, physical, social, and mental behavior was also examined. As shown in Table 4, there was a significant difference in social behavior (Z = -3.30, p = .001, r = -0.16) between the two groups. However, there was no difference in mental (Z = -0.84, p = 0.40, r = -0.04), emotional (Z = -0.74, p = 0.46, r = -0.04), or physical (Z= -0.39, p = 0.70, r = -0.02) behavioral changes of the students in the two groups.

Table 4. Male, Mean Change from Time 1 to Time 3 on Kinder Rubric Scales

	Treatment (n = 298)	Control (n = 121)
Mental	0.35	0.26
Emotional	0.32	0.25
Physical	0.30	0.26
Social	0.35	0.04**

* $p \leq .05$; ** $p \leq .01$; ** $p \leq .001$
NOTE: Mann-Whitney Non-parametric tests were used because data are ordinal and not normally distributed.