An Evaluation of NJSO CHAMPS

Findings from the 2014-2015 Season

Steven J. Holochwost

WolfBrown

September 30, 2015

Table of Contents

Executive Summary

The New Jersey Symphony Orchestra's Character, Achievement, and Music Project (NJSO CHAMPS) offers high-quality music education to children who would otherwise not receive it. It currently serves 50 students in grades 4 through 7 attending the University Heights Charter School. In addition to teaching students how to play a string instrument, both as an individual musician and as a member of an ensemble, CHAMPS seeks to foster students' personal and academic growth by teaching perseverance, social skills, and positive attitudes about the future.

Our evaluation of the 2014-15 season of NJSO CHAMPS was guided by three questions. We wanted to know whether students enrolled in CHAMPS:

- Made progress in their musical achievement.
- Improved in aspects of their character.
- Exhibited higher levels of academic achievement than their peers.

To address these questions, data were collected from CHAMPS students, and report cards were delivered for these students and their peers at UHCS. Students enrolled in CHAMPS also sat for two juries at the beginning and end of the season, at which time they also completed written tests of musical knowledge, and their primary teachers in the program rated students' engagement in program at the beginning and end of the program year.

Did Students Progress Musically?

Our results indicated that students did progress musically over the course of the CHAMPS season. Nearly two-thirds of CHAMPS students achieved higher scores on their end-of-year juries when compared to their juries at the beginning of the year, and nearly three quarters of students earned higher scores on their written tests of musical knowledge. The average increase in jury scores was approximately 7% over baseline; for the written tests, it was in excess of 8%.

Figure S1. Jury Scores

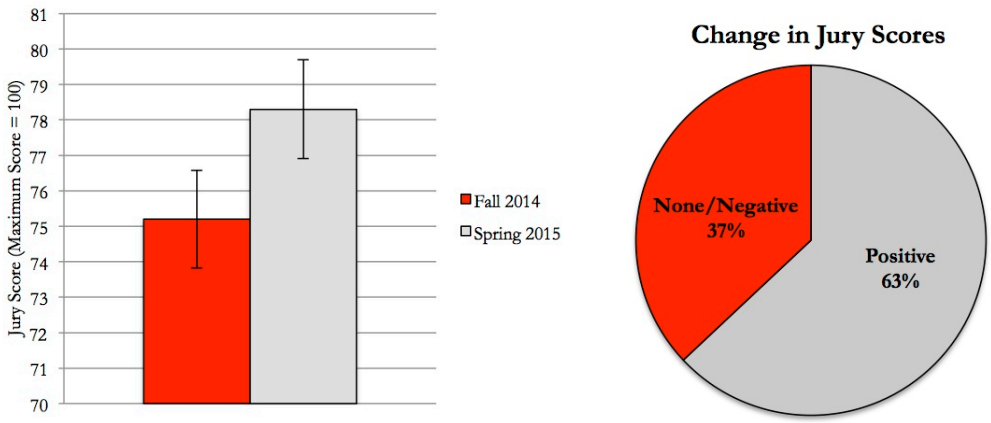
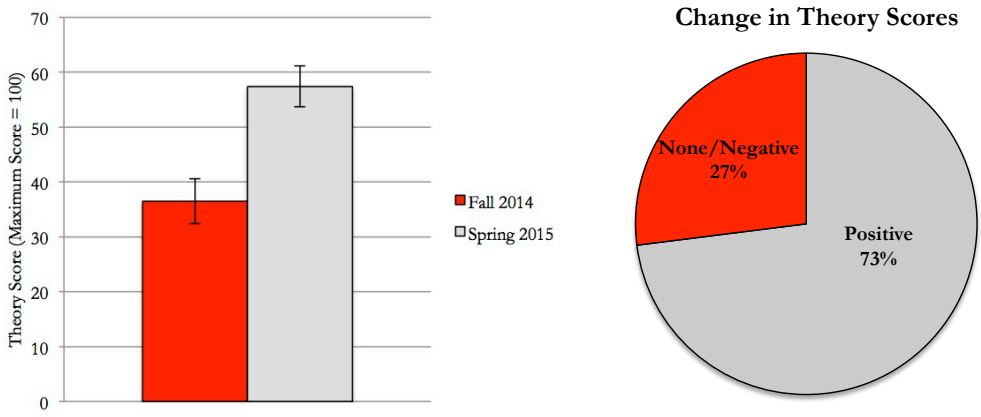


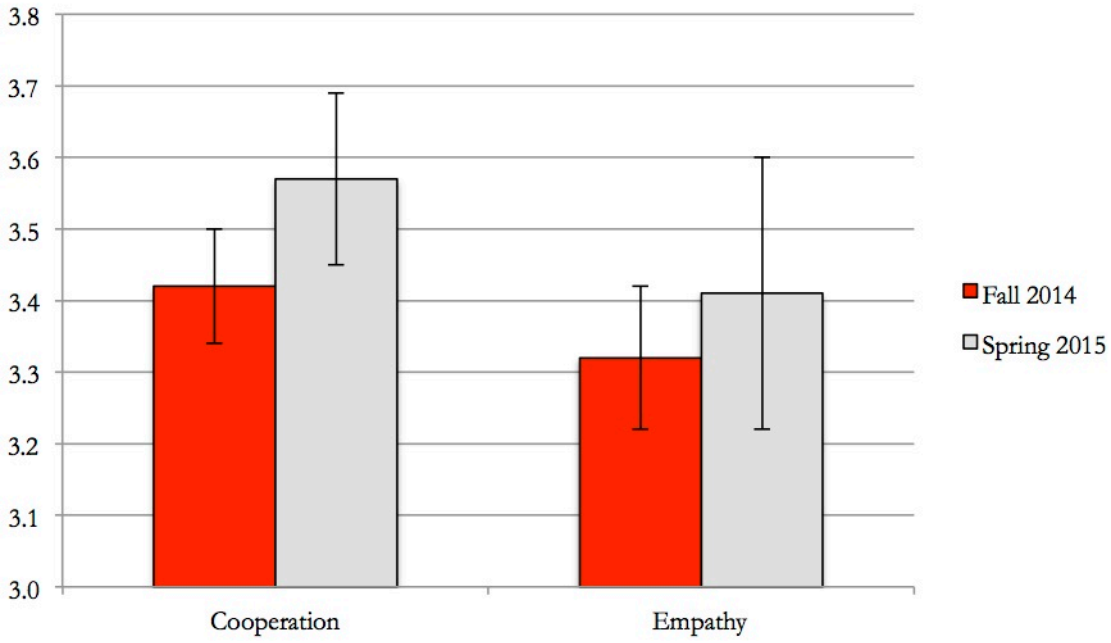
Figure S2. Theory Test Scores

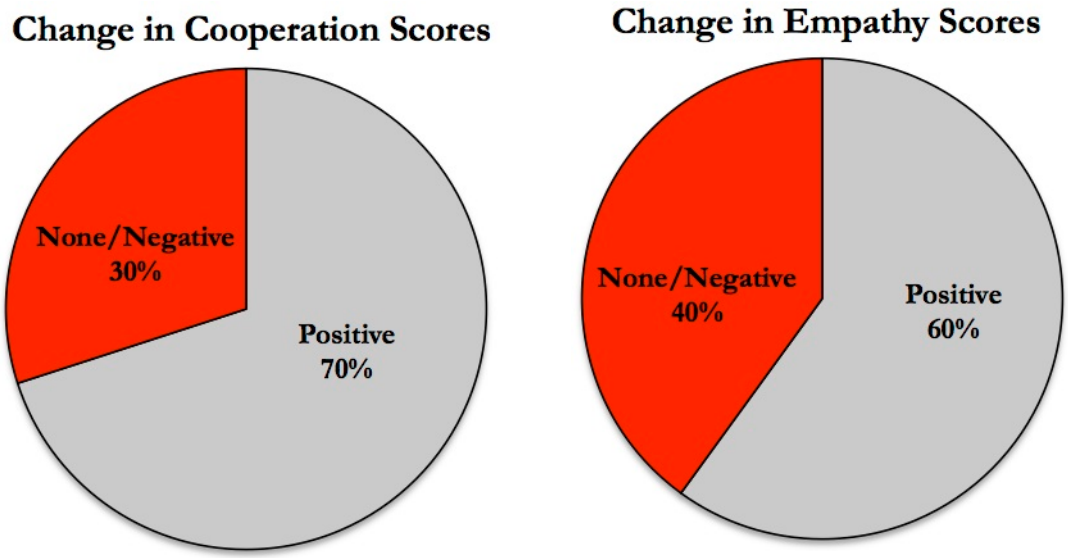


Did Students' Character Improve?

One aspect of character is prosocial behavior – the ability to work productively with peers – and two components of prosocial behavior are the capacities to cooperate and empathize with others. When offered in a supportive, ensemble setting, music education may be particularly effective in fostering these domains of character. Our results support this argument: 7 of every ten children enrolled in CHAMPS exhibited positive change in their cooperation, while 6 in ten exhibited positive change in their empathy. For both cooperation and empathy, the average change over baseline was approximately 11%.

Figure S3. Cooperation and Empathy Scores



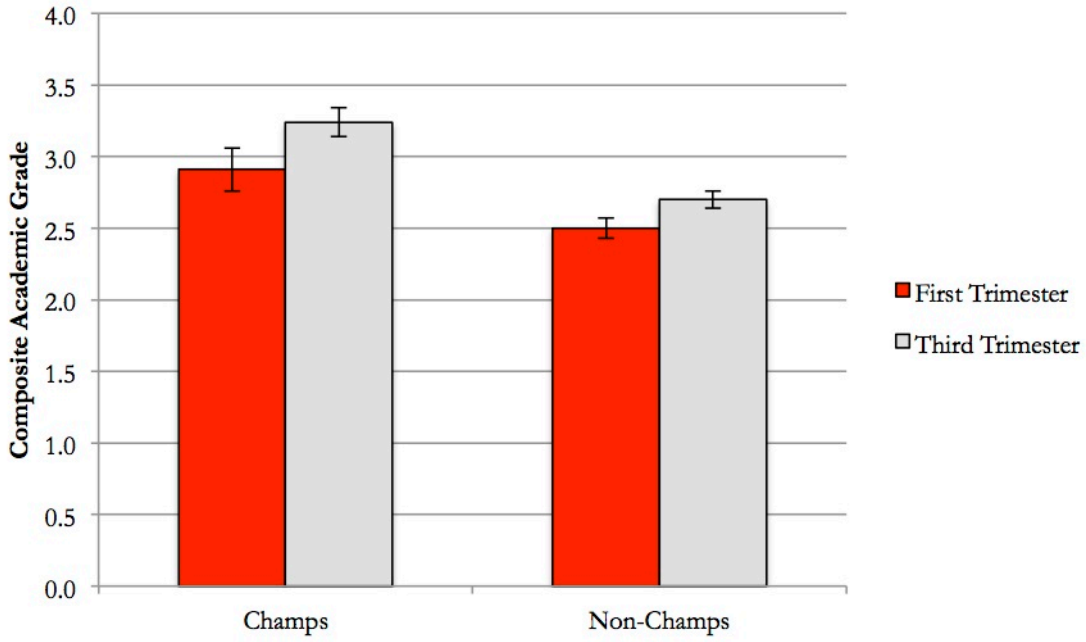


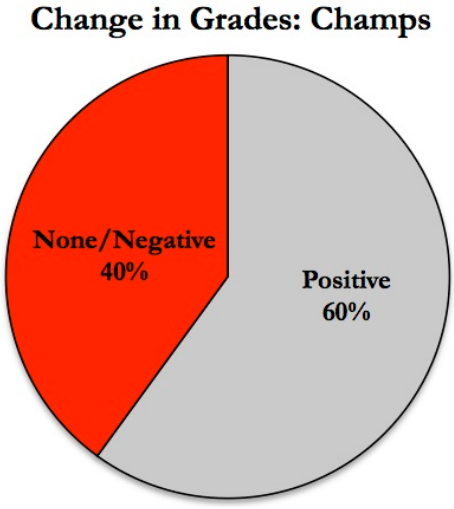
In addition, students enrolled in CHAMPS missed significantly fewer days of school over the course of the school year than their peers. While one could argue that CHAMPS students were simply drawn from a larger pool of students who were more likely to attend school in any event, CHAMPS students also missed significantly fewer days of school from the end of the first trimester to the end of the school year. On average, students who were not in CHAMPS missed six days of school, while students who were in CHAMPS missed four.

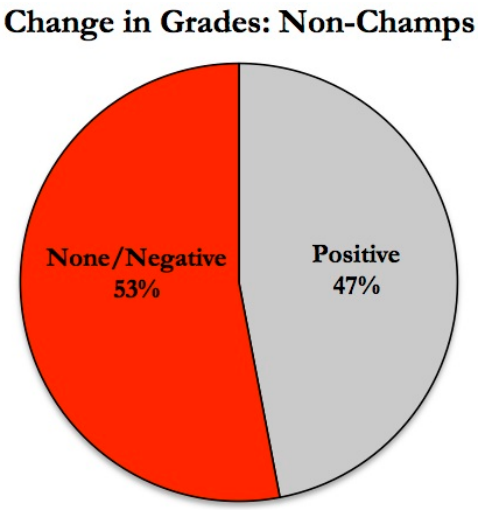
Did Students Enrolled in CHAMPS Outperform their Peers Academically?

Students in CHAMPS earned significantly higher grades in both the first and third trimesters than their peers. More importantly, CHAMPS students' grades increased more sharply over this period than did their peers. That is, CHAMPS students not only earned higher grades over the first third of the school year, but they earned even higher grades still over the remainder of the year. Over half of CHAMPS students exhibited positive change in their grades, while fewer than half of students who were not enrolled in CHAMPS did so. On average, CHAMPS students grades increased nearly 30% over baseline, or approximately three times as much as students who were not enrolled in CHAMPS, whose grades increased 12% over baseline.

Figure S4. Academic Achievement







Conclusion

The results of this evaluation, together with those from the evaluation of the 2013-14 season, provide considerable evidence for the efficacy of CHAMPS, particularly with respect to students' musical achievement and character. As detailed below, more work is necessary to fully understand the how the program benefits participating students. But clearly CHAMPS is a program with great promise to improve the lives of the children it serves, both now and in the future.

Introduction

Over the course of the 2013 – 14 academic year, WolfBrown conducted its first full evaluation of the New Jersey Symphony Orchestra's Character, Achievement, and Music Project (NJSO CHAMPS). Our work was guided by a set of guiding questions:

- First, we wanted to know if CHAMPS improved students' academic achievement, and, if so, whether those effects were mediated or "carried by" the effects of CHAMPS on students' character: their capacity to persevere in the face of difficulty, the way in which they interacted with the peers (i.e., prosocial behaviors), their ability to control their feelings and actions (intrapersonal skills), and their attitudes about the future.

- Second we wanted to know if CHAMPS might buffer the effects of risk – poverty and its correlates – on academic achievement or these aspects of character, or if CHAMPS might amplify the effects of positive factors present in a child's environment, such as a particularly close relationship between a child and their parent.

- Finally, we wanted to know if students' musical achievement changed over the course of the season, and, if so, the extent to which that change was predicted by each student's level of engagement in the program.

To address these questions, data were collected from students enrolled in CHAMPS and a comparison group of students enrolled in the same grades and classrooms at University Heights Charter School (UHCS). Students, their parents, and their UHCS teachers were asked to complete a questionnaire at the beginning and end of the CHAMPS season, and students' CHAMPS teachers were asked to rate students' engagement in the program on the same schedule. Students' musical progress was assessed during two juries, the first occurring two months into the CHAMPS season and the second coinciding with its end.

The findings from this initial evaluation were intriguing. At the end of the season, parents rated the prosocial behaviors of students enrolled in CHAMPS more highly, while girls enrolled in CHAMPS reported higher levels of intrapersonal skills and more positive attitudes about the future. These results were obtained using analyses that controlled for (i.e., held constant) students' initial scores in these domains, and thus allowed us to attribute these effects to CHAMPS were more confidence than had we simply looked at year-end scores. Additional analyses revealed that students who were rated as more highly engaged by their CHAMPS teachers were rated as more perseverant and more positive in their peer interactions by their UHCS teachers. These same highly-engaged students also ended the year with more Scholar Dollars then their less engaged peers.

These findings, together with positive results from the assessments of students' musical progress, led us to make two recommendations for the evaluation of CHAMPS in 2014-15. First, we recommended that the same basic evaluation design be employed, but that the second set of guiding questions regarding risk be dropped, given that there were no effects of risk or poverty observed among participants in the 2013-14 sample. Second, we strongly encouraged CHAMPS to do whatever possible to increase levels of participation in the study, leading to a larger sample size and increased statistical power.

Unfortunately circumstances beyond our control, including a series of staff transitions at both NJSO and UHCS, led to disruptions in the recruitment of participants, the administration of measures and the collection of data. As reported in further detail below, a comparison group was not recruited, very few parent measures were returned, and UHCS teachers did not complete measures at either the pre- or post-program collection.

These facts led us to formulate a different evaluation strategy, guided by a different set of questions, than that which was employed in our evaluation of the 2013-14 season.

- We again wanted to know whether students musical achievement increased over the course of the season, and the extent to which this progress was predicted by their engagement in the program. We addressed these questions in two stages: in the first stage, we included all students for whom data were available, regardless of whether informed permission from parents had been obtained. Where results were found, we then re-ran our analyses using only the sub-sample of students for whom permission had been granted. The first stage of analyses yielded data that we could examine internally and share with potential funders, while any results yielded by the second stage of analyses could be shared publicly.

- We again examined whether enrollment in CHAMPS was associated with increases in aspects of character, but we expanded the definition of prosocial behaviors to include measures of cooperation and empathy. However, unlike in our previous evaluation, we were limited to examining how students' perspectives on their own character changed, as data were not available from parents or UHCS teachers. Therefore we were limited to examining whether character for students enrolled in CHAMPS changed over time.

- Finally, we examined whether enrollment in CHAMPS was associated with increases in academic achievement. Because these data were taken from students' report cards, they were available for all students in CHAMPS, as well as all students who were not in CHAMPS but who were enrolled in the same grades and classrooms, and therefore the latter group of students could be considered a comparison group. However, because none of the comparison-groups students parents provided permission for their children to participate in the study, any findings regarding academic achievement could only be used internally.

The remainder of this report outlines the methods used to address these questions, presents the results obtained, and discusses those results, including their implications for the 2015-16 evaluation.

Methods

Participants. As was the case for the 2013-14 season, the 47 students enrolled in CHAMPS for at least of portion of the 2014-15 seasons and their peers attending the same grades (i.e., 4 through 8) at UHCS were invited to participate in the evaluation. The families of 26 CHAMPS students (55% of those enrolled) provided permission for their children to participate in the study, but only a handful of families of non-enrolled students did so. As such, no comparison group could be assembled, obviating concerns about differences between groups in the distribution of students by gender, grade, or classroom.

Procedures. Surveys were distributed to students and their CHAMPS and UHCS teachers within two weeks of the beginning and end of the 2014-15 season. Surveys were returned by 40 (85%) students in the fall but only 14 (30%) students in the spring. CHAMPS teachers returned surveys for all students in the fall but only 25 (53%) of students in the spring. Surveys were not returned by UCHS teachers for any student.

Students' musical progress was assessed twice over the course of the season, in mid-December and late April, through a combination of jury assessments and written theory tests. Forty-two students (89%) completed the December juries, while 33 (70%) students participated in the spring juries; theory test scores were available for 35 (75%) and 31 (66%) students in December and April, respectively.

Measures & Data Reduction.

Musical Progress. Students' musical progress was assessed by a jury evaluation of playing and a written test of music theory. During the jury evaluation, students were asked to perform a piece they were working on in the program for a panel of three musicians: the program's Artistic Director (a violinist by training), as well as a violist and a cellist, both of whom were members of the Orchestra. The panel was comprised of these same individuals for both the December and April assessments, and no panelist was the primary program teacher for any student. Each panel member independently assigned students scores ranging from 1 to 4 on 16 evaluative dimensions, including bow placement, sound, pitch, and musicality. A total score was then calculated as the sum of scores across dimensions, which ranged from 1 to 64.

Inter-rater reliability among the three panelists (i.e., the degree to which panelists agreed in their assessments of each student) was calculated using the intra-class correlation coefficient (ICC), with random effects for raters included in the model. For both assessments, the ICC was very high: for the fall, it was .95, while for the spring it was .94. Therefore a mean of the three panelists scores was calculated for each student and used in all subsequent analyses.

The written test of music theory required students to identify elements of notated music (e.g., barlines, time and key signatures), assign the correct pitch names to notes on different clefs, compose rhythms, and define musical terms.

Program Engagement. Students' engagement in CHAMPS was assessed using their program teachers' responses to a thirteen-item questionnaire (see Appendix A). This measure showed good internal consistency (i.e., internal reliability) for the fall (α = .94) and spring (α = .94) administration. An exploratory factor analysis strongly suggested a one-factor structure for the items, with a single factor accounting for 61% and 62% of the variance in scores for the fall and spring, respectively. Therefore fall and spring composite engagement scores were calculated as the mean of the items.

Perseverance. As was the case for the 2013-14 season, perseverance was assessed using the 12-item grit scale.[1] Internal consistency for the fall administration of the measure was acceptable (α = .78), and a factor analysis suggested a one-factor structure, with a single factor accounting for 37% of the observed variance. Despite the fact that only a small number of students completed the measure in the spring, internal consistency and factor structure were similar (α = .80, and a single factor accounted for 40% of the variance). A perseverance score was calculated as the mean for all students answering at least nine items.

Intrapersonal Skills. This domain was again assessed using the subscale of the same name from the second edition of the Behavioral and Emotional Rating Scale (BERS-2).[2] For both the fall and spring administrations, internal consistency was acceptable (α = .80 for fall, α = .72 for spring), and a single factor accounted for 29% and 30% of the observed variance in fall and spring, respectively. Fall and spring intrapersonal skill scores were therefore calculated as the mean of items for all students answering at least 10 items.

Future Orientation. Students' attitudes about the future were again assessed using seven items adapted from the Longscan Study's Future Events Questionnaire.[3] Internal consistency was acceptable for the fall (α = .71) and spring (α = .74) administrations of the measure. A factor analysis strongly suggested a one-factor structure, with a single factor accounting for 54% of the observed variance in the fall and 57% of the variance in the spring. Future orientation scores were calculated for the fall and spring as the mean for all students answering at least five items.

Cooperation. The cooperation subscale from the Social Skills Improvement System (SSIS)[4] was one of two new sets of questions students answered during the 2014-15 season. Like the empathy subscale taken from the SSIS, this measure displayed good properties during piloting in the spring of 2014. The properties were again evident here: internal consistency was good for the fall (α = .80) and spring administrations of the measure (α = .822) and factor analyses suggested a one-factor structure for both administrations, with a single factor accounting for 46% of the variance in the fall and 56% of the variance in the spring. Cooperation scores were therefore calculated as the mean of items for all students answering at least 5 items.

---

[1] Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92,* 1087-1101.

[2] Epstein, M. H. (2004). *BERS-2: Behavioral and Emotional Rating Scale.* Austin, TX: Pro-Ed.

[3] Longscan. (1998). *Future Events Questionnaire.* Chapel Hill, NC: The Longitudinal Studies of Child Abuse and Neglect.

[4] Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS).* San Antonio, TX: Pearson Assessments.

Empathy. The empathy subscale from the SSIS also displayed good properties. Internal consistency was good for the fall administration (α = .83) and excellent for the spring (α = .92), while factor analyses indicated that a one-factor structure accounted for 56% of the observed variance in the fall and 74% in the spring. Empathy scores were therefore calculated as the mean of items for all students answering at least 4 items.

Other Measures of Prosocial Behavior. Other measures of prosocial behavior were available from students' report cards. Behavior grades were available for all students in CHAMPS and their UHCS peers. These grades were assigned for each of three trimesters in four subjects (English Language Arts (ELA), math, social studies, and science) on a three-point scale, from not-satisfactory to outstanding. Given that the CHAMPS began shortly before the end of the first trimester and ended prior to the close of the third trimester, data reported for these two trimesters were considered pre-program and post-program assessments, respectively.

For the first trimester, behavior grades were internally consistent (α = .85) and a single factor accounted for 69% of the variance in grades; for the third trimester, internal consistency was similar (α = .81) and a single factor accounted for 64% of the variance in grades. We therefore calculated mean behavior grades for the first and third trimesters for all students with grades in at least two subjects. The numbers of days absent or tardy were also reported for each trimester.

Academic Achievement. Academic grades were also available for the first and third trimester in ELA, math, social studies, and science. These grades were assigned on a five-point scale, from F to A. For both the first and third trimester, academic grades were internally consistent (α = .85 for the first trimester, α = .82 for the third) and a single factor accounted for 70% and 67% of variance in grades, respectively. Mean first- and third-trimester academic grades were therefore calculated for students with grades for at least two subjects.

Results

Musical Progress. Table 1 reports descriptive statistics for the jury assessments of playing and the written test of music theory. For both measures, change scores are reported in terms of points and percentile change. Positive change scores indicate the spring score was higher than the fall score.

Table 1

Jury Scores				**Range of Scores**	
	N	*M*	*SD*	Minimum	Maximum
Fall Score	42	75.2	8.95	59.0	94.0
Spring Score	33	78.3	8.00	63.0	95.3
Change in score, points	30	3.70	12.4	-19.3	27.3
Change over fall score, percentile	30	6.70	17.1	-22.0	40.0
Theory Test Scores				**Range of Scores**	
	N	*M*	*SD*	Minimum	Maximum
Fall Score	35	36.5	24.3	0	84.0
Spring Score	31	57.5	20.8	1.0	97.0
Change in score, points	22	23.6	34.5	-56.0	79.0
Change over fall score, percentile	19	8.39	22.4	-1.0	79.0

Note: *N* = number of students, *M* = mean (average) score, *SD* = standard deviation

As is evident from Table 1, the average change in scores on the jury assessments was 3.70 points (*SD* = 12.4), corresponding to an average percent change over baseline of 6.7% (*SD* = 17.1%). Although this change was not statistically significant ($p = .112$), this may be attributable to the very limited statistical power afforded by such a small sample. Among the 30 students for whom fall and spring jury scores were available, 19 students (63%) exhibited a positive change in scores. Both fall ($r = .384$, $p = .012$) and spring ($r = .583$, $p = .004$) jury scores were positively and significantly correlated with teachers' ratings of students engagement in the program.

It is also clear from an inspection of Table 1 that theory scores increased from late fall 2014 to spring 2015. Of the 22 students for whom fall and spring jury scores were available, 16 (73%) exhibited a positive change in scores. On average, the change in scores was 23.6 points (*SD* = 34.5), corresponding to an average percent change over baseline of 8.4% (*SD* = 22.4%). This change *was* statistically significant ($t(21) = 3.21$, $p = .004$), and the effect size (calculated as the difference score divided by the standard deviation for the fall assessment) was large ($d = 0.85$). This pattern of results held among consented students: though the change in scores did not achieve statistical significance ($t(11) = 2.20$, $p = .051$), though the effect size was again large ($d = 0.86$).

Figure 1. Jury Scores[5]

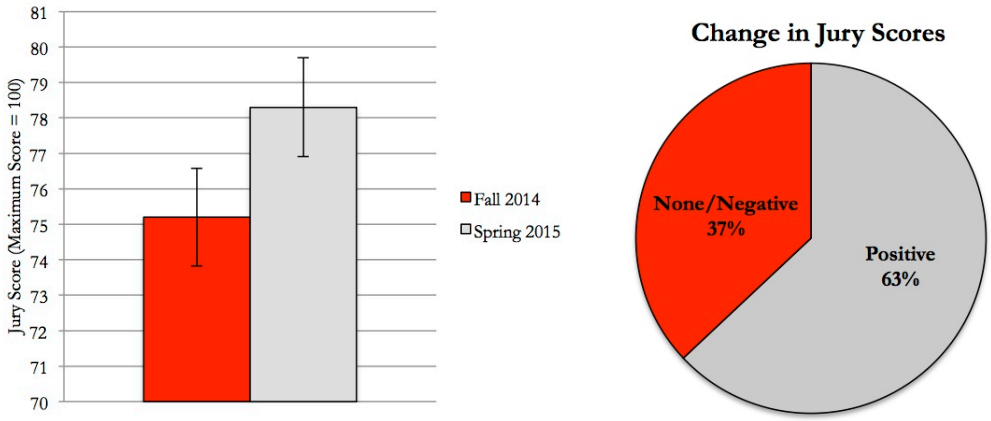
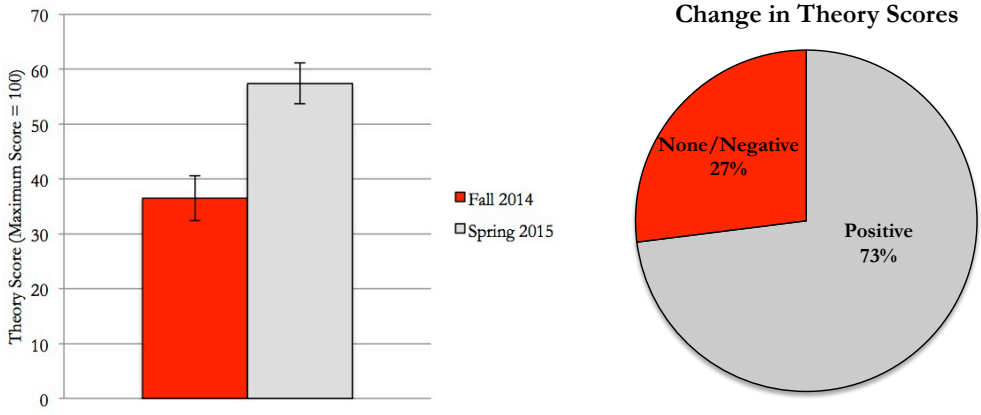


Figure 2. Theory Test Scores



One important point must be made about these results: as in any repeated-measures design, missing data is both a common problem and a cause for concern. The fact that data were missing for the spring assessments of playing and theory are particularly important to acknowledge, given the possibility that the students for whom data are missing are precisely those who would have achieved the lowest scores on the spring assessments. Fortunately, the probability that a child was missing spring scores on these measures was unrelated to the scores on those measures in the fall.

[5] Here and throughout, error bars depict two times the standard error of the estimate.

Character. Data regarding four broad aspects of character were available from measures completed by students: perseverance, intrapersonal skills, future orientation, and prosocial behaviors (and specifically, cooperation and empathy). Table 2 presents descriptive statistics for each of these measures for both the fall and spring, as well as change scores in terms of both points and percentile change.

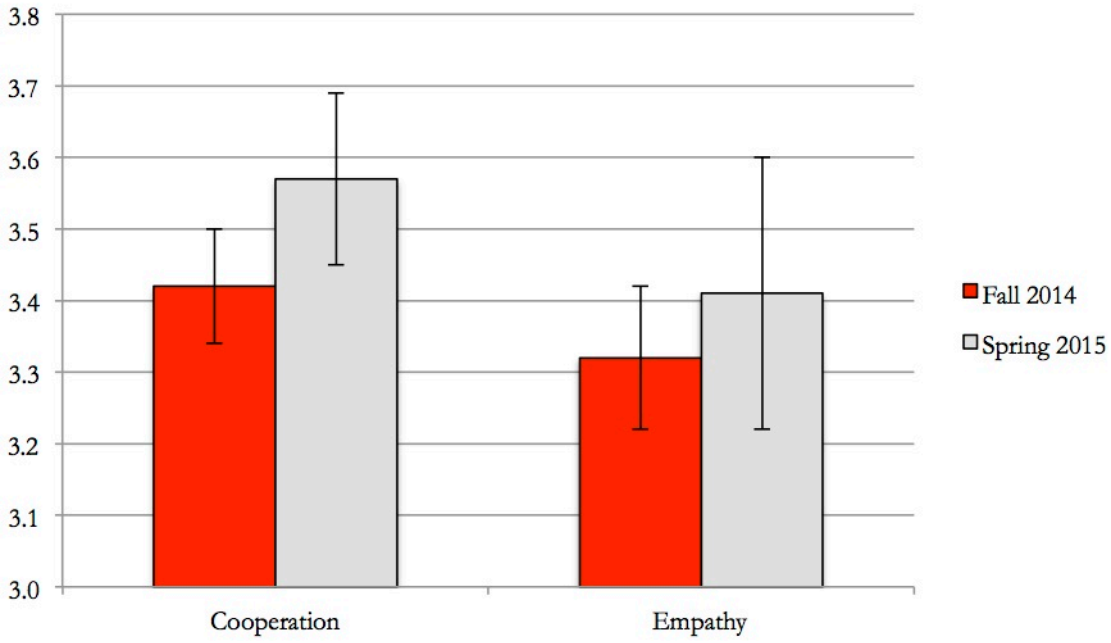
Table 2

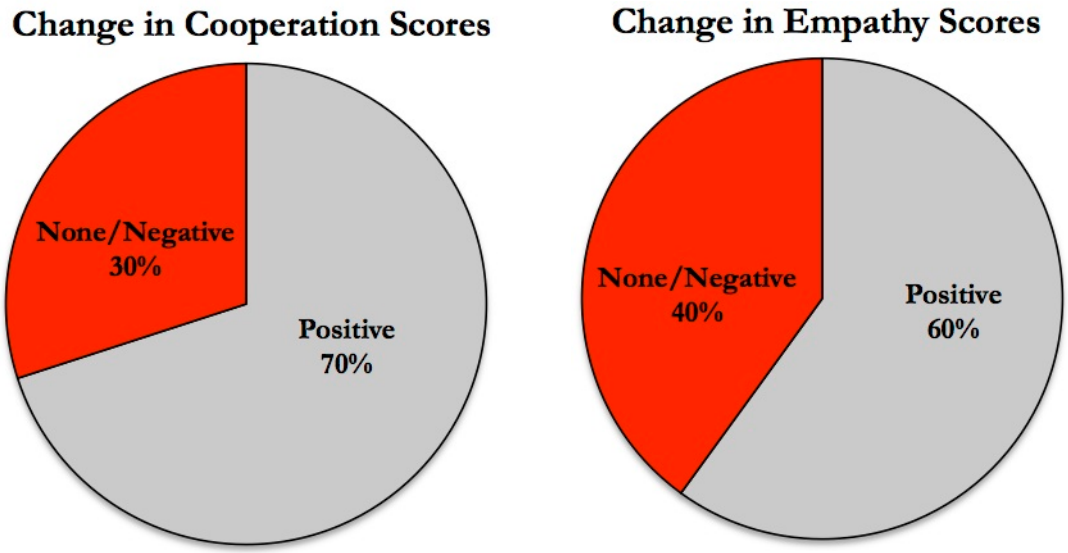
Perseverance (range: 1 to 5)	*N*	*M*	*SD*	**Range of Scores** Minimum	Maximum
Fall Score	40	3.68	.656	2.00	4.67
Spring Score	14	2.83	.600	2.83	4.75
Change in score, points	12	.303	1.01	-1.05	1.92
Change over fall score, percentile	12	.143	.358	-0.26	0.85
Intrapersonal Skills (range: 1 to 4)	*N*	*M*	*SD*	**Range of Scores** Minimum	Maximum
Fall Score	40	3.18	.528	2.07	4.00
Spring Score	14	3.14	.415	2.53	3.73
Change in score, points	12	.272	.562	-0.87	1.13
Change over fall score, percentile	12	.122	.207	-0.24	0.46
Future Orientation (range: 1 to 5)	*N*	*M*	*SD*	**Range of Scores** Minimum	Maximum
Fall Score	40	4.38	.559	2.86	5.00
Spring Score	14	4.50	.522	3.14	5.00
Change in score, points	12	-.022	.828	-1.69	1.71
Change over fall score, percentile	12	.012	.209	-0.35	0.52
Cooperation (range: 1 to 4)	*N*	*M*	*SD*	**Range of Scores** Minimum	Maximum
Fall Score	38	3.42	.522	2.29	4.00
Spring Score	13	3.57	.417	2.86	4.00
Change in score, points	10	.286	.632	-0.71	1.57
Change over fall score, percentile	10	.112	.235	-0.19	0.65
Empathy (range: 1 to 4)	*N*	*M*	*SD*	**Range of Scores** Minimum	Maximum
Fall Score	38	3.32	.590	2.17	4.00
Spring Score	13	3.41	.683	1.67	4.00
Change in score, points	10	.243	.858	-1.33	1.67
Change over fall score, percentile	10	.110	.330	-0.44	0.77

Note: *N* = number of students, *M* = mean (average) score, *SD* = standard deviation

Of the 12 students for whom fall and spring perseverance scores were available, 6 (50%) exhibited a positive change in scores. Across these 12 students, the average change in scores was .303 points (*SD* = 1.01), corresponding to an average percent change over baseline of 14.3% (*SD* = 35.8%), which was not statistically significant (p = .320). A large portion of students (75%) exhibited positive change in intrapersonal skills scores, with an average change in scores of .272 points (*SD* = .562) or 12.2% (*SD* = 20.7%) over baseline that approached statistical significance (p = .122). Although little change was observed in students' future orientation scores (the mean change from fall to spring was nearly zero), majorities of children providing fall and spring scores exhibited positive change in cooperation (70% of children exhibited positive change) and empathy (60%) scores. For cooperation, the average change in scores was .286 points (*SD* = .632), or 11.2% (*SD* = 23.5%) over baseline. For empathy, the average change was .243 points (*SD* = .858), or 11.0% (*SD* = 33.0%) over baseline. The degree of change was not statistically-significant for either cooperation (p = .186) or empathy (p = .393). Fall and spring scores for each measure of character were not associated with contemporaneous ratings of program engagement.

Figure 3. Cooperation and Empathy Scores





The likelihood that spring scores were missing was not related to fall scores for measures of perseverance, future orientation, cooperation, or empathy, but the probability that spring scores of intrapersonal skills were missing was related to fall scores, such that students with lower ratings of intrapersonal skills in the fall were more likely to be missing scores in the spring. Therefore findings regarding intrapersonal skills must be interpreted with caution.

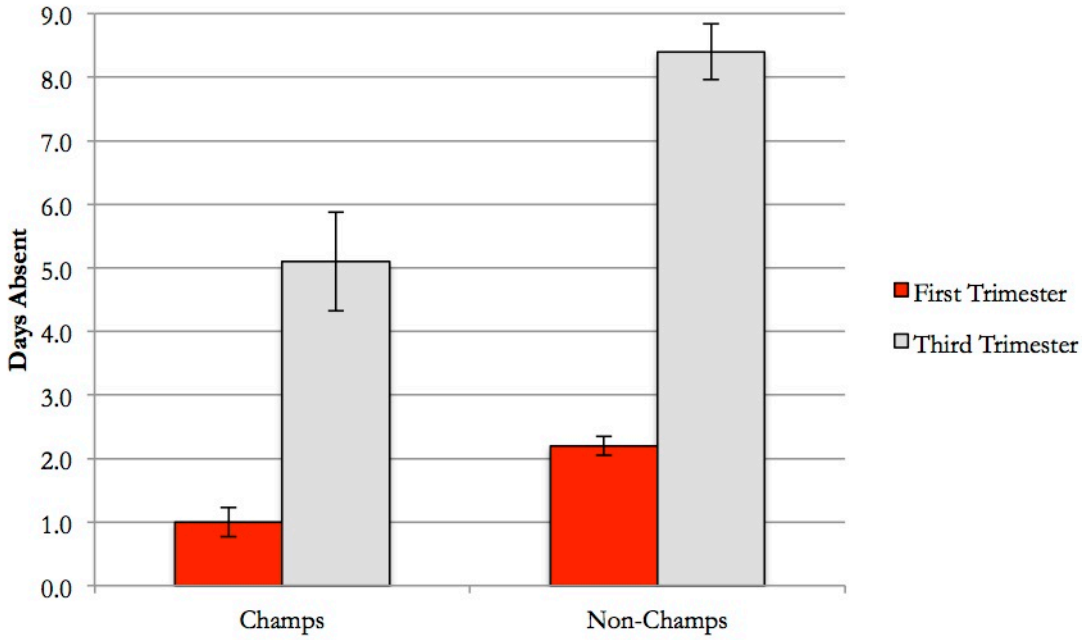
Data regarding another aspect of character – students' prosocial behavior – was available from students' report cards. Given this, the 37 students who completed the CHAMPS program in the 2014-15 academic year could be compared against their peers in the same classrooms at UHCS. Table 3 reports descriptive statistics for composite behavior grades and days absent.

Table 3

Composite Behavior Grade	**First Trimester**			**Third Trimester**		
	N	*M*	*SD*	*N*	*M*	*SD*
CHAMPS	37	1.21	.468	14	1.23	.422
Non-CHAMPS	206	0.91	.464	205	0.95	.464
Days Absent	**First Trimester**			**Third Trimester**		
	N	*M*	*SD*	*N*	*M*	*SD*
CHAMPS	37	1.00	1.41	37	5.08	4.77
Non-CHAMPS	208	2.15	2.12	208	8.43	6.37

Comparable portions of CHAMPS (43%) and non-CHAMPS (32%) students exhibited positive change in their composite behavior grades from the first to the third trimester, and the average change in scores for CHAMPS students – .020 points, or 7.4% over baseline – was similar to that for the non-CHAMPS students (.047 points (*SD* = .434), or 10.2% over baseline (*SD* = 74.1%)). However, CHAMPS students missed significantly fewer days of school in the third trimester than their peers (t (244) = 3.04, p = .003). While it was the case that CHAMPS students also missed significantly fewer days than their peers in the *first* trimester (suggesting that students who choose to be in CHAMPS are also students who choose to attend school), CHAMPS students missed significantly fewer days of school from the end of the first trimester to the end of the third trimester (t (244) = 2.38, p = .018). This may suggest – albeit in a very preliminary fashion – that CHAMPS encourages children to attend school regularly.

Figure 4. Absences for the First and Third Trimesters



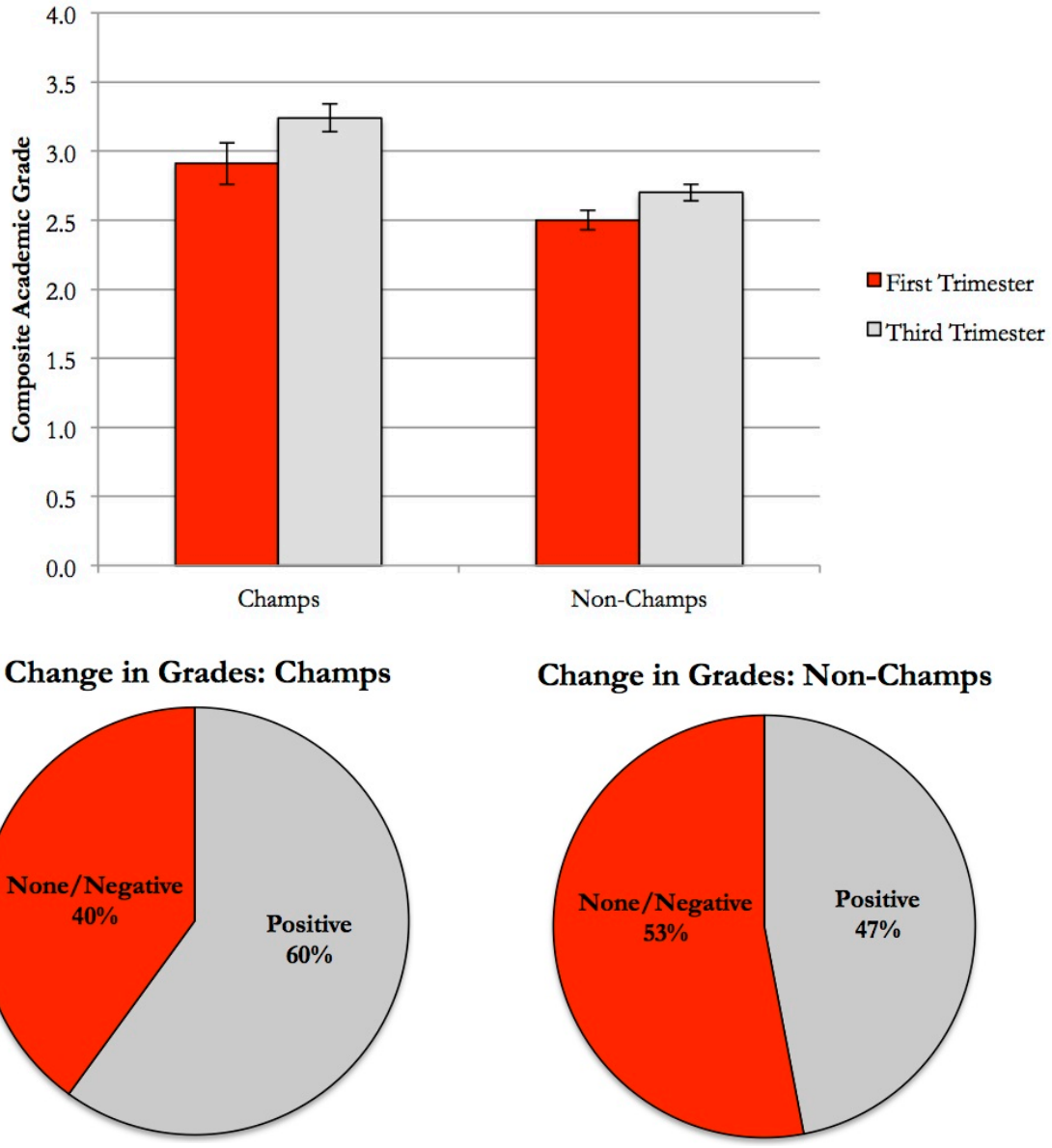
Academic Achievement. Table 4 reports academic grades for ELA, math, social studies and science for CHAMPS students and their peers. CHAMPS students had significantly higher composite grades than their peers in the third semester, but, like behavior grades (t (240) = 3.97, $p < .001$), they also had significantly higher grades in the first trimester (234) = 2.30, $p = .022$), so it is difficult to discern CHAMPS' influence on third trimester grades from this analysis. However, enrollment in CHAMPS was associated with higher third trimester grades after controlling for first trimester grades ($B = .328$, $t = 3.36$, $p = .001$), which indicates that CHAMPS students' grades increased more sharply from the end of the first to the end of third trimester than their peers.

Table 4

CHAMPS (*N* = 37)	**First Trimester**		**Third Trimester**	
	M	*SD*	*M*	*SD*
ELA	2.86	1.06	3.30	.702
Math	2.95	.743	3.08	.829
Social Studies	3.05	1.13	3.46	.836
Science	2.78	1.21	3.14	.673
Composite	**2.91**	**.900**	**3.24**	**.614**
Non-CHAMPS (*N* = 208)	**First Trimester**		**Third Trimester**	
	M	*SD*	*M*	*SD*
ELA	2.34	1.17	2.76	.928
Math	2.49	1.07	2.40	1.13
Social Studies	2.63	1.24	2.67	1.28
Science	2.62	1.16	2.91	.639
Composite	**2.52**	**.963**	**2.68**	**.815**

Both CHAMPS students (t (36) = 3.68, p < .001) and their peers (t (204) = 3.18, p = .002) exhibited significant increases in composite grades over the course of the 2014-15 academic year.[6] However, a larger portion of CHAMPS students (60%) exhibited positive change in these composite grades than their peers (47%), and the magnitude of the change was greater among CHAMPS students. On average, the change in composite academic grades for CHAMPS students was .331 (*SD* = .547), or an increase of 29.5% over baseline (*SD* = 71.8%), whereas the average change for non-CHAMPS students was .148 points (*SD* = .666), or an increase of 11.7% over baseline (*SD* = 54.4%).

Figure 5. Academic Achievement



[6] Among the subset of students for whom parental permission was received, the average change in scores was .300 (*SD* = .441), which was statistically significant (t (19) = 3.04, p = .007).

Discussion

Summary of Findings. We began our analyses by examining the outcome most proximal to CHAMPS: students' musical progress. As noted above, nearly two-thirds of students exhibited positive change in jury scores, and nearly three-quarters exhibited positive change on the written tests of musical knowledge. The magnitude of change over baseline was approximately seven percent for jury scores and eight percent for the theory test, and the change within CHAMPS students approached statistical significance for the jury scores and achieved significance for the theory test. This, together with the results of the evaluation for the previous season, constitutes strong evidence that students enrolled in CHAMPS are making progress in their ability to play and knowledge of music.[7]

In both the fall and spring, students were asked to evaluate themselves on five dimensions of character: perseverance, intrapersonal skills, future orientation, cooperation, and empathy. CHAMPS students exhibited positive increases on four of these five dimensions, with nearly two-thirds to three-quarters of students exhibiting positive change for intrapersonal skills, cooperation, and empathy, the size of which ranged from 10 to 12 percent over baseline. And on one observational measure of prosocial behavior, CHAMPS students outperformed their peers: over the course of the school year, CHAMPS students missed significantly fewer days of school.

The fact that CHAMPS students exhibited positive increases in multiple domains of character is consistent with the results for the 2013-14 season. However, unlike in 2013-14, we have no benchmark against which to measure this character growth, because we have no comparison group. It is possible – though unlikely, given our findings from 2013-14 – that students not enrolled in CHAMPS would have exhibited parallel growth on our measures of character.

Our analyses of academic achievement data yielded interesting results: CHAMPS students earned higher grades at both the end of the first and third trimesters. One could argue that more academically-adept student choose to be in CHAMPS, and that is why we observe higher grades for both trimesters. However, a larger proportion of CHAMPS students exhibited positive change in their grades from the end of the first to the end of the third trimester, and the size of the gain for CHAMPS students was larger than that demonstrated by their peers.

These findings are quite different from those reported for the 2013-14 season, in which both CHAMPS students and their peers exhibited modest increases in overall academic grades. This discrepancy may be attributable to the fact that in this report we are comparing academic grades for CHAMPS students against academic grades for *all* their peers in the same grades and classrooms at UHCS. It is possible that the subset of students in the comparison group for the 2013-14 season were not representative of the broader population of UHCS students. On the other hand, it is also possible that the students in the comparison group *were* more comparable to CHAMPS students; we cannot know, because while we have

[7] The average gain in scores was larger for the 2013-14 season, but this is to be expected: all students were new to the program in 2013-14, but over half (57%) of the students who completed the program in 2014-15 were enrolled during 2013-14 as well.

grades and absences for students who were not enrolled in CHAMPS, we do not have demographic data for them.

Limitations and Future Directions. The questions that arise when we try to compare the results of this year's evaluation with the results obtained for the previous season underscore the limitations of our design, and strongly suggest a course of action for the future:

- First, better rates of return for measures administered to CHAMPS students must be achieved. The rate of return for measures administered in the fall of 2014 was robust, but in the spring of 2015, measures were returned for only 14 students whose families' provided permission for them to be in the study. The fact that for some measures (e.g., empathy) we observed positive change that approached statistical significance is remarkable, given the small size of our sample.

- Second, a more robust sample size is required, and that sample must include a sizable portion of comparison-group students. Given that parents are only being asked to provide permission for their children's participation (and that they are not being asked to complete any measures), obtaining permissions for children enrolled in CHAMPS should be easier than it has been in previous years. Implementing the procedures we have discussed (e.g., having permission forms distributed through students' classroom folders) should increase participation among comparison-group families, but this may not be enough. It may also be necessary to incentivize the participation of comparison-group families, and this would be a worthy use of our research budget.

- Third, we need UHCS teachers to complete measures for students whose parents provide permission for them to participate in the study. If necessary, their participation should be incentivized as well. Having both students' perceptions of how their character changes over the course of a season together with corroborating reports from students' teachers would further strengthen the argument for CHAMPS' effects.

Conclusion. The 2014-15 season was a challenging one, both for CHAMPS and for the work of its evaluation. However, thanks to the dedication of program staff (Ms. Borowski and Kossakowski deserve particular recognition), the continued excellence of the artistic staff and the seriousness with which they approach assessing students' musical progress, and the generosity of UHCS administration, we have a set of promising results that complement those obtained in 2013-14. The coming season is when we should translate this promise into solid results, now that we have the right staff, procedures, and measures in place to capture the effect that CHAMPS has on students.

Works Cited

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92,* 1087-1101.

Epstein, M. H. (2004). *BERS-2: Behavioral and Emotional Rating Scale.* Austin, TX: Pro-Ed.

Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS).* San Antonio, TX: Pearson Assessments.

Longscan. (1998). *Future Events Questionnaire.* Chapel Hill, NC: The Longitudinal Studies of Child Abuse and Neglect.

Appendix A. Teacher Engagement Measure

The following survey asks questions about students in NJSO CHAMPS. Please note that although we ask you to indicate both your name and the name of your student, we will not share your responses with anyone.

Your Name:

Students Name (first name, last initial):

How long have you worked with this student (in months)?

Using the scale below, please indicate how well do you feel you know this student?

☐ Very well	☐ Well	☐ Not well	☐ Hardly at all

In this section you will be asked how strongly you agree or disagree with each statement about this student. Please answer using your impressions of this student throughout **the past program year**.

SA = strongly agree	A = agree	NS = not sure	D = disagree	SD = strongly disagree

1. This student often forgets his or her instrument or music.	SA	A	NS	D	SD
2. This student has made real progress towards mastery of his or her instrument this year.	SA	A	NS	D	SD
3. This student is dedicated to the program.	SA	A	NS	D	SD
4. This student clearly enjoys class.	SA	A	NS	D	SD
5. This student volunteers to answer questions asked in class.	SA	A	NS	D	SD
6. This student takes care of his or her instrument.	SA	A	NS	D	SD
7. This student pays attention during class.	SA	A	NS	D	SD
8. This student is disrespectful and/disruptive during class.	SA	A	NS	D	SD
9. This student is excited about the opportunity to perform in public.	SA	A	NS	D	SD
10. This student acts like they would rather be somewhere else during class.	SA	A	NS	D	SD
11. This student accepts criticism and works to improve his or her technique based upon it.	SA	A	NS	D	SD
12. This student will help others in a constructive way during class.	SA	A	NS	D	SD
13. This student plays like a member of an ensemble during class by listening to others.	SA	A	NS	D	SD