



NASEF Y4 Research: Student Outcomes and Attitudes

JeSeok Lee, Garrison Wells, Craig G. Anderson, & Constance Steinkuehler

Youth interest in esports (competitive videogames) is on the rise globally and, with it, organized esports programming for teens. The North American Scholastic Esports Federation (NASEF) is one such program, providing enriched, school-affiliated esports programming for students led by trained teachers and coaches. NASEF includes both a competitive league and supporting clubs at multiple implementation sites that engage students in esports-related scholastic programming to foster academic, social, and emotional positive outcomes.

For the last three years, our research team has investigated the outcomes of the NASEF program. Our early quantitative and qualitative findings suggest positive outcomes for participating youth in areas such as STEM interest, school affiliation, emotional self-regulation, and positive relationships with adults and peers. Yet, more research is needed to better test the veracity of these student gains and to better understand their specific nature, causes, and consequences of the pedagogical design. This quasi-experimental study continues our empirical investigation into these dynamics.

Research Questions

Four research questions drove this investigation:

1. What are the impacts of the NASEF program on key variables representing NASEF's mission for ***student academic, social and emotional growth***, including:
 - a. STEM interest
 - b. Communication
 - c. Constructive Mindset (toward learning)
 - d. School Engagement
 - e. Self-Regulation
 - f. Relationships with Others
 - g. Wellness
 - h. GPA
 - i. School Attendance
2. What are ***students' attitudes*** towards the NASEF program and its features, including:
 - a. Satisfaction with general managers (GM), coach, club, league, communications, award programs, and championship events
 - b. Feelings of connectedness to NASEF peers, team, club, and league
 - c. Any feedback on the program (open response)

- d. Whether they plan to continue in NASEF (and, if not, why)
- 3. If positive outcomes are observed, are they **equitable** for all students across gender, ethnicity, socio-economic status (SES), and disability status?
- 4. What **program features** impact student outcomes?

Study Goals & Design

Toward these ends, we conducted two surveys, one for students and one for staff. For the main instrument, the student survey, an *alternative one sample pretest-posttest design* was used (Figure 1). In it, students completed either a pretest or posttest survey instrument administered during four separate time windows:

- *pre-pretest & pretest* before each league season (Fall and Spring), and
- *posttest & post-posttest* after each league season.

By comparing changes across the NASEF season (from pretest and posttest) to change over comparable periods of time both before (pre-pre and pre) and after (post and post-post) the season, this design has greater internal and external validity, allowing us to eliminate multiple potential threats to validity at once or competing explanation for any effects we find, including testing effects (when exposure to the pretest effects posttest scores), historical effects (when external circumstances such as a global pandemic inadvertently influences the results), most importantly, maturation effects (a significant risk when studying teens since natural cognitive, social and emotional development happen rapidly over time as a natural consequence of time). If students show significantly greater change from pretest to posttest across the NASEF season than they do during the time periods when the league is *not* active, then we have strong evidence to support the claim that the NASEF program is indeed causing student outcomes.

TABLE 1
(Using Campbell and Stanley's Notation)

Pre-Pre	Pre	Post	(Follow up)*
R	0 ₁	X	
R	0 ₂	X	
R		X	0 ₃
(R)		(X)	(0 ₄)
.....			
R			0 ₁
R			0 ₂
R			X
(R)			(X)
			0 ₃
			(0 ₄)
.....			
	Pre-Pre	Pre	Post
R	0 ₁	X	
R		X	
R		X	0 ₃
(R)		(X)	(0 ₄)

The sets of all scores for each 0_i (i = 1, 2, 3, 4) form the three (four) independent groups.
* Parentheses in the table indicate observations made on optional follow-up occasion.

Figure 1. Alternative one sample pretest-posttest design (Johnson, 1986).

The student pretest and posttest survey instruments were similar to that used in prior research on NASEF with slight modifications and improvements (such as reducing its length by consolidating items) and measured: student academic, social and emotional outcomes (RQ1); student attitudes toward specific features of the program (RQ2); and basic demographic information, including age, grade year, gender, ethnicity, socioeconomic status (SES), GPA, and disability status (RQ3). All four student surveys (pre-pretest, pretest, posttest, and post-posttest) were isomorphic except for sections measuring attitudes toward NASEF, which were included only on the posttests after students had experiences with the program to draw from. The pre-pretest and pretest instrument consisted of 65 items including 15 questions regarding demographic and socioeconomic factors and 50 developmental and socio-emotional measures presented on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”. The posttest and post-posttest survey instrument was similar to the pre-survey, containing the same demographic questions and Likert scale items, but with additional questions measuring the nature of their engagement in NASEF and overall satisfaction with key aspects of the program.

The staff survey was constructed to gather details on program implementation at each site that could be used to examine whether specific features of the NASEF program led to greater student outcomes and, if so, which. The staff instrument included items measuring:

1. *Club Size*, including: number of students in the club, number of students on the team.
2. *Extent of Program*, including: club staffing, student roles offered, competitions attended, resources used, number of game titles played, presence of in-school curriculum.
3. *Site Support* in terms of their site facilities and home organization's leadership

The staff survey also included additional items designed to measure staff attitudes toward NASEF program features and communication and their perceived impacts of NASEF on students. Here we focus on those data related to program implementation; in a separate staff survey report, we detail their perceptions of and attitudes toward the program.

Data Collection

The 4-group student study design was conducted twice over the academic year — once over the Spring NASEF league season and once over the Fall NASEF league season. Data from both seasons were pooled together to create a single data corpus. The staff survey was administered at the conclusion of the academic year (after spring league season).

Sample

A minimum of 50 students were sampled during each phase of data collection, with n=102 completing the pre-pretest, n=107 completing the pretest, n=106 completing the posttest, and n=106 completing the post-posttest. The resulting total sample size was n=421 students across the fall and spring seasons combined.

Figure 1 gives details on the demographics of the overall sample. Most of the participants came from California (226) with much smaller numbers coming from Illinois (43), Pennsylvania (32), Missouri (22), New Jersey (18), and Florida (14). The vast majority of participants identified as male (87.2%) as compared those who identified as female (10.5%), nonbinary (0.8%), or preferred not to disclose (0.8%). Students identified themselves as Asian (42.3%), Caucasian (28%), Hispanic/Latinx (13.8%), Black/African (1.7%) and Other (14%). Average GPA of participants was 3.51 out of 4.0, and the distribution was skewed toward higher grades (mode=4.0).

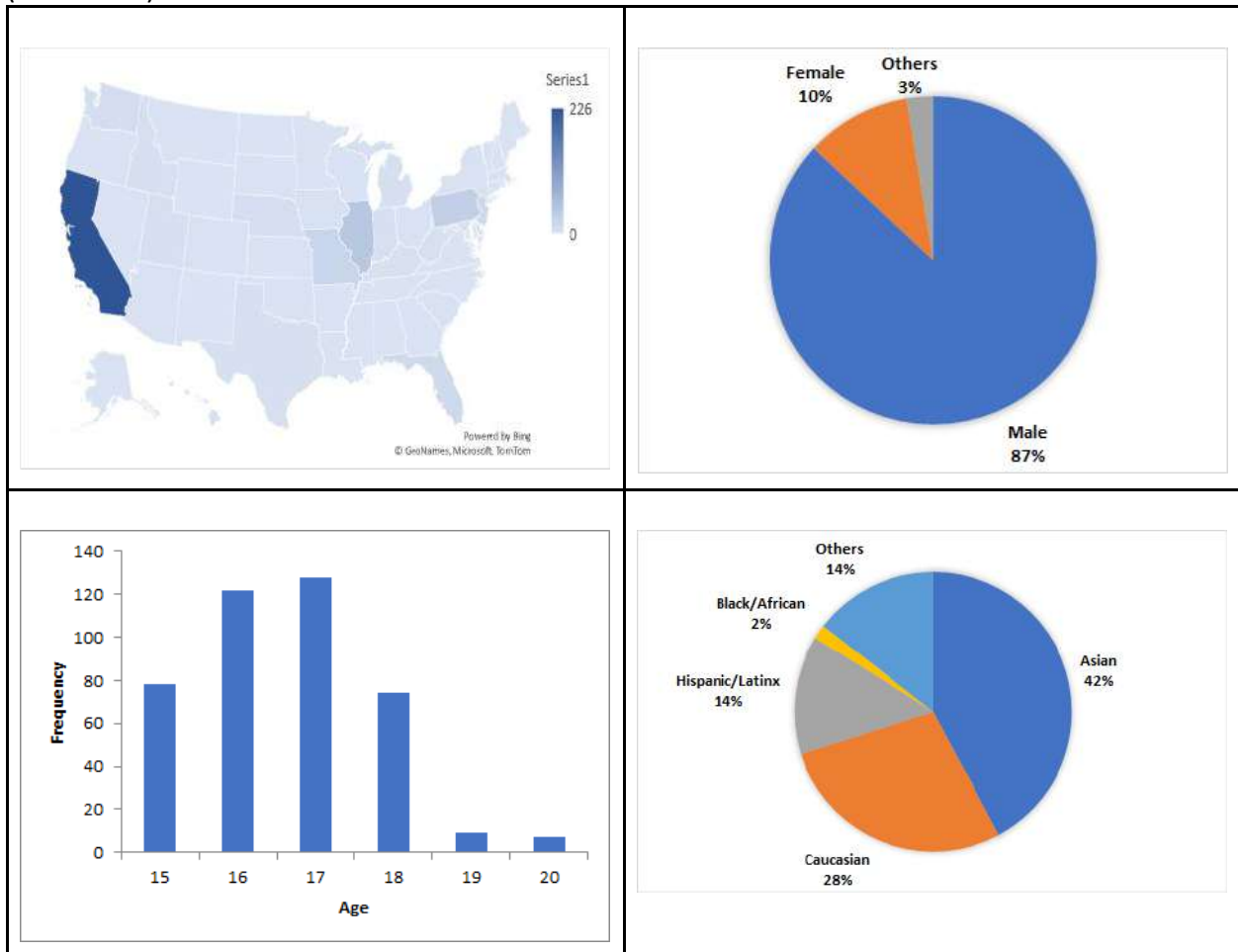


Figure 1. Sample demographics, including geographic location (top left), gender (top right), age (bottom left), and ethnicity (bottom right).

While the gender ratio of our sample is similar to the NASEF population based on registration data (90% male and 9% female), ethnicity distribution was not (table 1). Thus, cases were weighted according to the ethnicity ratio of the entire NASEF population. These weights were then taken into account in all subsequent analyses so that the results reported are not affected.

Table 1. Weights Applied to Sample to Insure Representativeness

	Population Sample Weight		
Asian	16.5%	42.3%	0.40
Black / African	3.9%	1.7%	2.35
Caucasian / White	49.1%	28.0%	1.76
Hispanic / Latinx	16.7%	13.8%	1.22
Others	13.8%	14.2%	0.90

A total of 93 NASEF staff members responded to our survey: 98% of respondents were general managers, 41% were student presidents, and 29% were virtual coaches.

FINDINGS

Student Cognitive, Social and Emotional Gains

Figure 2 below shows the mean change from pretest to posttest only on 7 Likert scale variables representing NASEF’s mission to support student cognitive and socio-emotional development. Variables are grouped by color to indicate the five (5) broader constructs that these variables together represent.

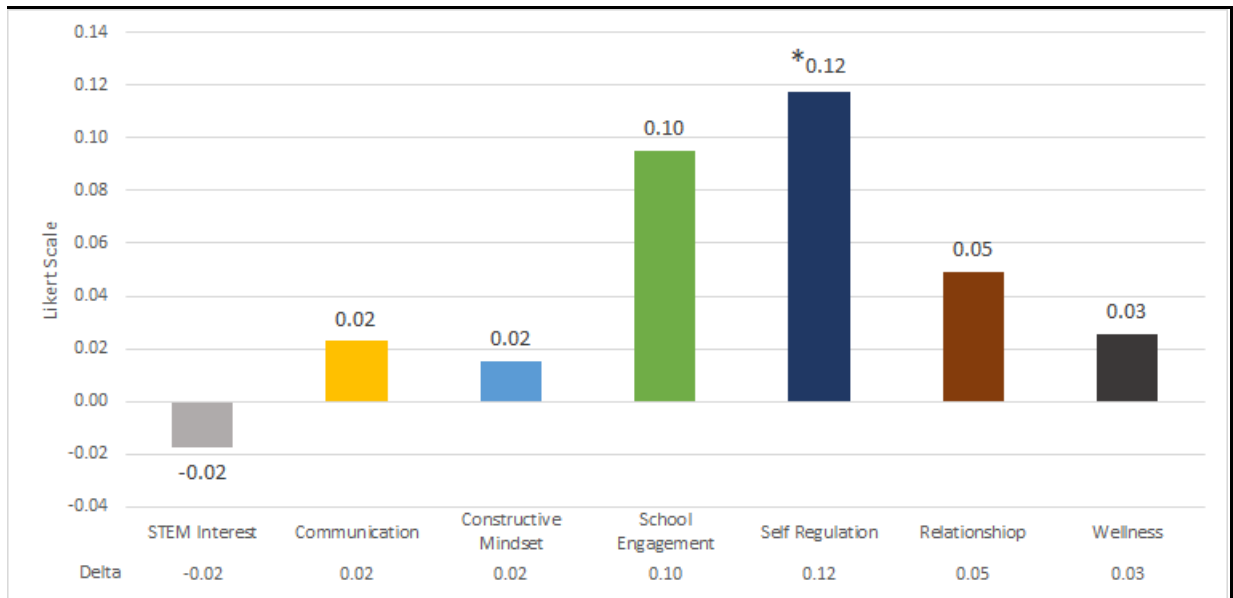


Figure 2. Means change from pretest to posttest only across 7 (of 9 total) variables representing NASEF program goals with statistically significant variables asterisked (*).

Figure 3 below shows the mean change from pretest to posttest only in GPA and school attendance (measured in days missed from school).

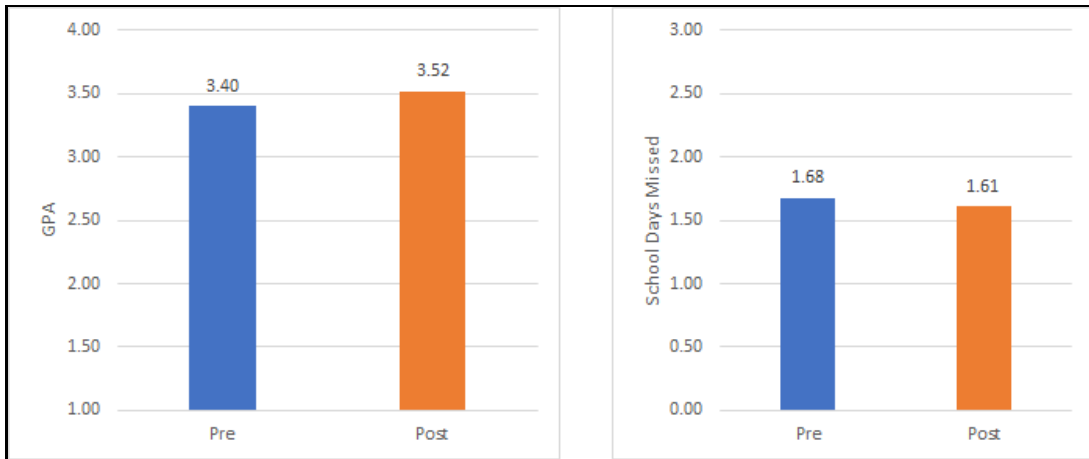


Figure 3. Means pretest and posttest GPA (left panel) and attendance (right panel).

Positive changes were found in communication ($\Delta=0.02$), constructive mindset ($\Delta=0.01$), school engagement ($\Delta=0.09$), self regulation ($\Delta=0.12$), relationships with peers and mentors ($\Delta=0.05$), overall wellness ($\Delta=0.02$), GPA ($\Delta=0.11$), and school attendance ($\Delta=0.07$).

We conduct two-sample t-tests for independent samples on each variable to determine whether the changes observed were significant. There was a significant difference in the scores for *self-regulation* between pretest ($M=3.50$, $SD=0.54$) and posttest ($M=3.62$, $SD=0.55$); $t(258)=-1.743$, $p=0.08$. No other variables were statistically significant (see Table 2).

Table 2. Results of Two-Sample T-Tests of 9 Student Outcomes Variables

	t	df	One-Sided p	Mean Difference
GPA	-1.500	251.138	0.068	+0.11
Days missed school	0.291	237.632	0.386	-0.07
STEM Interest	0.220	260.247	0.413	-0.02
Communication	-0.370	260	0.356	+0.02
Constructive Mindset	-0.268	260	0.394	+0.02
School Engagement	-1.079	260	0.141	+0.10
Self Regulation	-1.743	258	0.041*	+0.12
Relationship	-0.458	259	0.324	+0.05
Wellness	-0.339	260	0.367	+0.03

Next, we conducted a Kruskal-Wallis test, a nonparametric independent samples test, to compare changes across the NASEF season (pretest to posttest) to changes before the season (pre-pretest to pretest) and changes after the season (post-posttest to posttest). Comparing change across these three windows of time allows us to test whether NASEF caused any student gains seen rather than alternative potential causes such as maturation (described in detail above). Figure 4 and 5 below show trends in each variable over time.

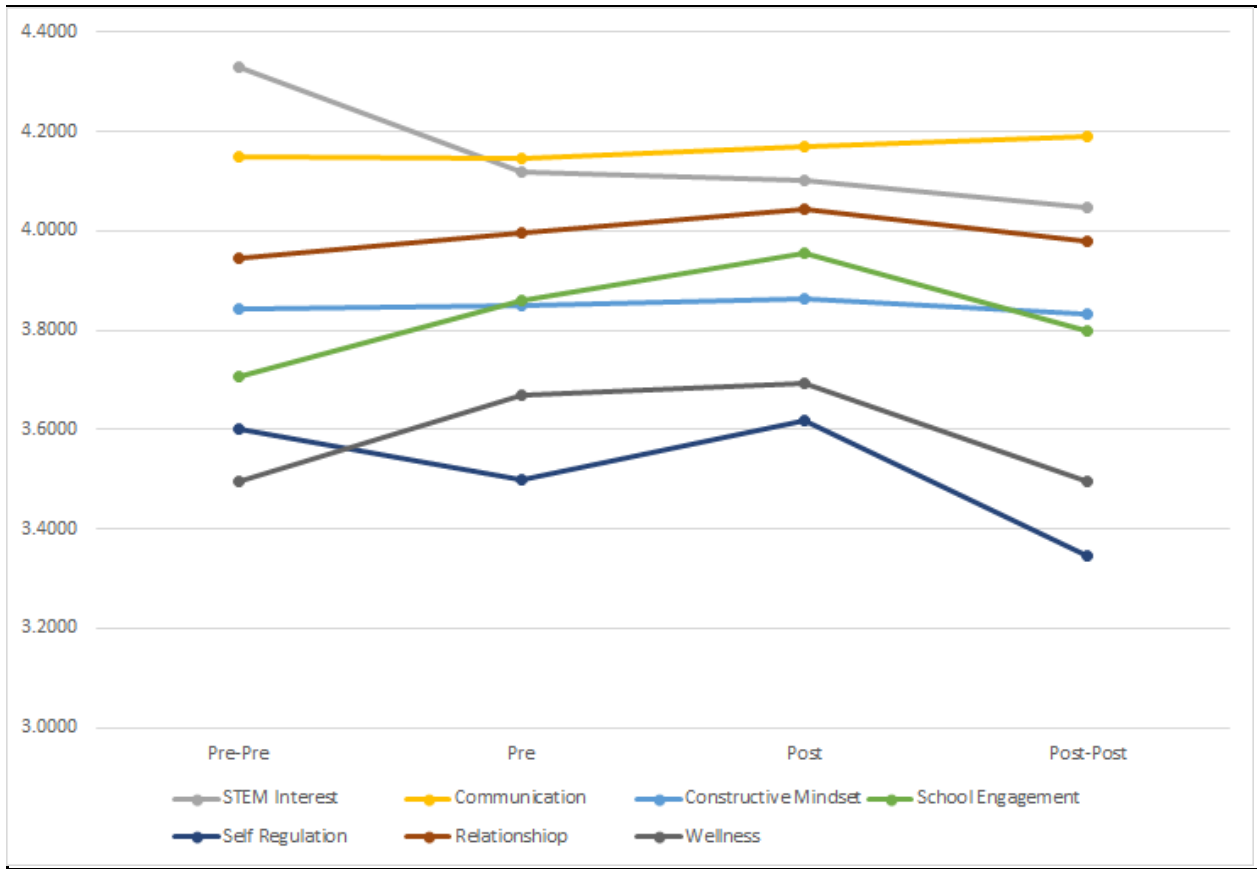


Figure 4. Change in the means of 7 (of 9 total) variables representing NASEF program goals by time of measurement, indicating trends over time.

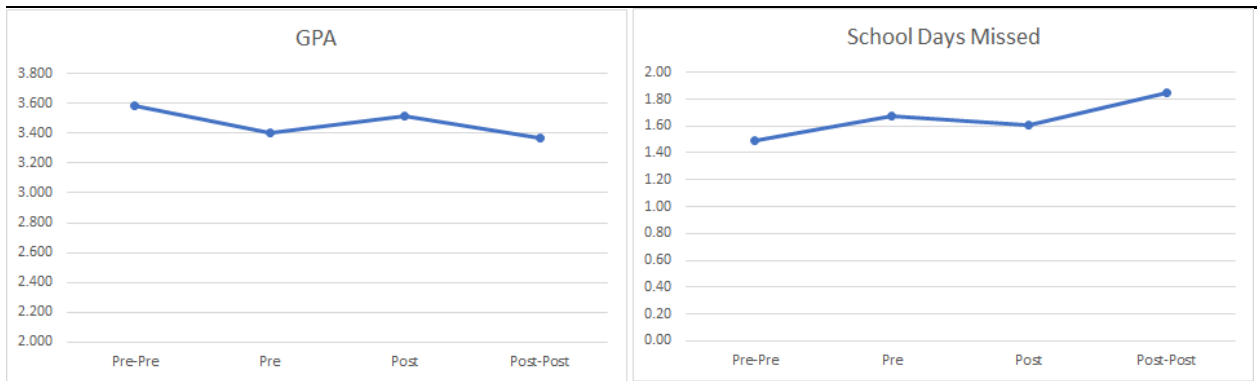


Figure 5. Change in the mean GPA and School Days Missed (2 of 9 total variables) by time of measurement, indicating trends over time.

While there were some fluctuations in student outcomes over time that suggest more positive outcomes over the pretest-to-posttest time period compared to preseason and postseason, no significant differences were found. *STEM Interest* dropped dramatically during preseason, with measures as pre-pretest significantly greater than measures at pretest, posttest, and post-posttest, but again we found no significant changes during the league season itself. *Self-*

Regulation dropped during preseason and postseason significantly, and rose during the league season, but not significantly. *Wellness* showed a similar pattern, with significant gains between pre-pretest and posttest and between posttest and post-posttest but again, the gains over the league season were not significant compared to preseason or postseason. Overall, despite securing a sizable sample with at least n=100 participants for each time of measurement, it may be that the positive trends we see in the data are simply too small for us to detect as significant. Post-hoc power analysis for this comparison reveals that power for this study hovers around 40% when it should be at least 80%.

Student Attitudes toward NASEF Program Features

Figure 6 below shows the mean student ratings of key features of the NASEF program at posttest (immediately after participation).

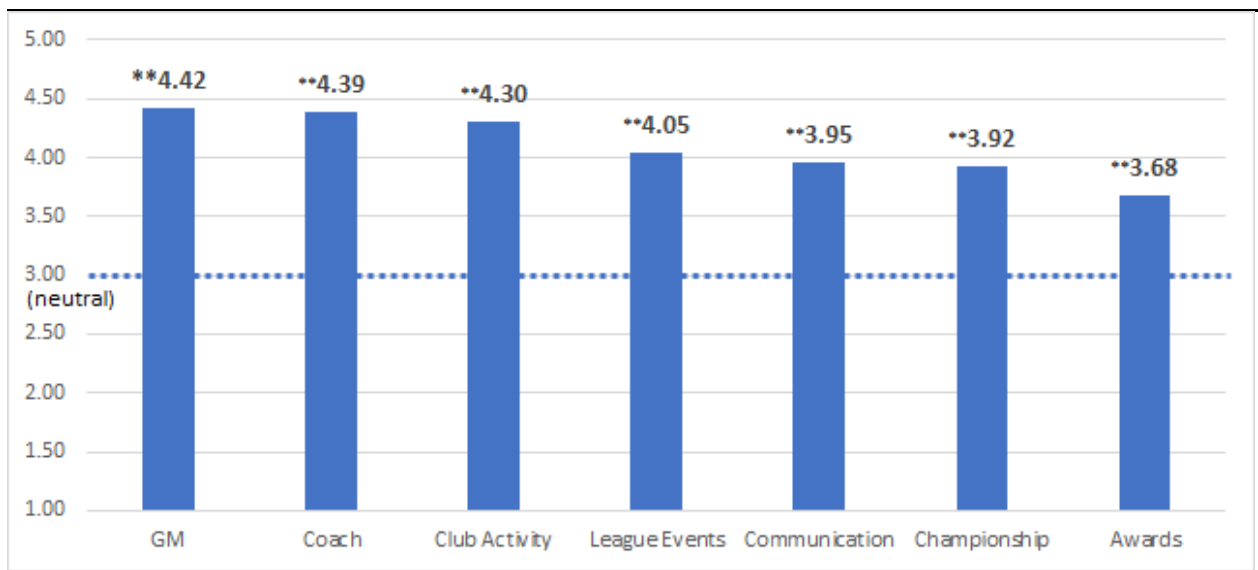


Figure 6. Mean student ratings of NASEF program features after participation with the midline 3.0 marked to represent “neutral” and statistically significant variables asterisked (*).

We conduct one-tailed single sample t-tests on each variable to determine whether the mean score differed significantly from “neutral” (midpoint score of 3 on a 1-5 scale with 1 representing “very dissatisfied” and 5 representing “very satisfied”). As shown in Table 3 below, all 7 variables were statistically significant (p<0.001). Students were very satisfied with all aspects of the program, particularly its mentoring structures. On average, they rated general managers and coaches in NASEF more highly than even the championships and awards.

Table 3. Results of One-Sample T-Tests of Student Attitudes Toward Program Features

	t	df	One-Sided p	Mean difference
GM	19.695	116	<0.001**	1.423
Coach	18.697	125	<0.001**	1.394
Club Activity	19.139	126	<0.001**	1.299
League Events	12.751	129	<0.001**	1.046
Communication	12.958	128	<0.001**	0.952
Awards	7.545	117	<0.001**	0.678
Championship	10.077	117	<0.001**	0.924

Student Feelings of Connectedness

A similar pattern characterizes students' feelings of connectedness within the program. Figure 7 below shows students' mean self-report of feelings of connectedness within the program at posttest (immediately after participation).

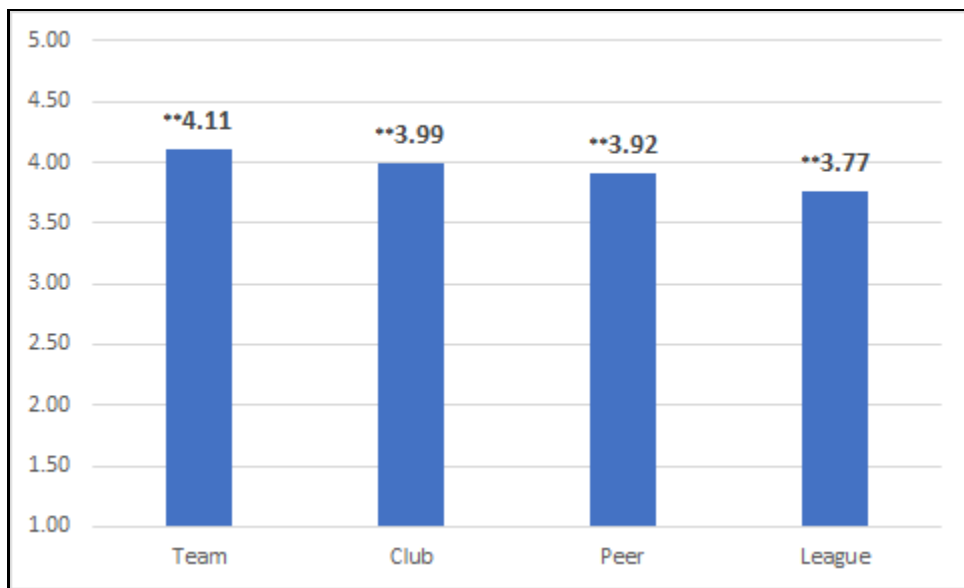


Figure 7. Students' mean self-report of feeling of connectedness in NASEF with the midline 3.0 marked to represent "neutral" and statistically significant variables asterisked (*).

One-tailed single sample t-tests were again used to determine whether the mean score on each variable differed significantly from "neutral" (midpoint score of 3). All 4 variables were statistically significant ($p < 0.001$) as shown in Table 4 below. Students felt connected at all levels of the program, in order from the competing team ($M = 4.11$, $SD = 0.89$) to the broader club ($M = 3.99$, $SD = 0.92$) to the peers in the program generally ($M = 3.92$, $SD = 0.82$) to the NASEF league as a whole ($M = 3.77$, $SD = 1.03$). Findings here appear to reflect expanding levels of the NASEF community in which individual students are nested.

Table 4. Results of One-Sample T-Tests of Student Feelings of Connectedness

	t	df	One-Sided p	Mean Difference from Neutral (3)
Team	14.147	130	<0.001**	1.111
Club	12.060	129	<0.001**	0.986
Peer	12.737	127	<0.001**	0.915
League	8.453	129	<0.001**	0.766

The Import of Length of Time in Program

Next, we used Spearman’s Rho to measure associations between students’ length of time in the program and both student ratings of NASEF program features (Figure 6 above) and student self-reported feelings of connectedness (Figure 7 above). Results show that the amount of time students spent in NASEF was significantly and positively correlated with evaluations of the coach and all three of the four measures of connectedness (Table 5). As student time investment in NASEF increases, their ratings of the coach increase as well as their feelings of connectedness to their NASEF team, club and peers.

Table 5. Measures of Relationship Between NASEF Participation (in Hours) and Both Student Ratings of Program Features and Student Self-Reported Connectedness

Spearman's Rho	How do you rate...	GM	Coach	Club Activity	League Events	Communication	Championship	Awards
Weekly hours spent on NASEF Activities	Correlation Coefficient	0.148	.227	0.200	-0.045	0.122	0.049	0.017
	Sig. (2-tailed)	0.170	0.027	0.050	0.659	0.232	0.648	0.874
	N	87	95	96	98	97	88	88
	How connected do you feel to....		Team	Club	Peer	League		
Weekly hours spent on NASEF Activities	Correlation Coefficient		.212	.272	.286	.190		
	Sig. (2-tailed)		0.037	0.007	0.004	0.061		
	N		98	96	98	98		

The Import of Student’s Roles in the Club

The overwhelming majority of NASEF students (97%) served as members on the competitive team; however, nearly a quarter (22%) served in other roles as well (e.g. manager, analyst, artist, streamer). In order to test for differences in student outcomes by the kinds of roles engaged in throughout the season, we conducted a two-sample t-test to compare outcomes for students who were only a competitive team member versus those who also served in at least one additional role. Results show that students with multiple roles show significantly higher

communication (M=4.34, SD=0.56). The results for constructive mindset ((M=3.95, SD=0.62) also indicate a potential relationship, but are not significant in this sample. (see Table 6). It appears that broadening student engagement from competition alone may increase (and possibly diversify) student outcomes as a result.

Table 6. Results of Two-Sample T-Test of Student Outcomes by Club Roles Served

	Posttest Mean (Competitive player only)	Posttest Mean (Multi-role)	t	df	One-sided p
Communication	4.10	4.34	-2.513	62	0.007**
Constructive Mindset	3.82	3.95	-1.564	65	0.061

Qualitative Feedback from Students

When students were asked for feedback and suggestions for the program, several recurring topics of discussion arose. One of these topics was game selection; several respondents asked for new games to be added and for more cross-platform options to be implemented so that a broader range of students can participate. Some students asked for the rulesets and tournament structures to be adjusted to more closely match those being used at the professional level. A few comments noted communication issues between their teams and NASEF regarding program logistics and tournament rewards, asking for greater clarity and transparency. The majority of students who responded, though, expressed satisfaction with how the program is currently run, with some simply requesting longer seasons and more opportunities to play.

Participant Churn

NASEF participant turnover or “churn” remains remarkably low with no substantive problems interfering with student retention. Of the 212 students who completed the posttest and post-posttest surveys, only 34 stated that they planned to leave the NASEF program. Eighty-two percent (28 of 34) gave graduation as their main reason for leaving. Of the six students remaining, two expected to have too much work during the upcoming seasons to participate, two hinted at problems with their club making the experience unenjoyable, one expressed disinterest in the games available (but a willingness to return if more games are added), and one was transferring to a new high school without an esports program.

Student Attitudes & Outcomes by Demographic Group

One key question is whether student attitudes and outcomes are equitable across demographic groups. To answer this question, we conducted two sample t-tests to compare student outcomes (changes in target variables from pretest to posttest) and student attitudes (toward NASEF program features, feelings of connectedness, and plan to continue) by gender, ethnicity, SES, and disability status.

Gender. Comparing student outcomes by gender reveals that female students showed significant gains in *school engagement* ($\Delta=0.52$) while male students showed significant gains in

self-regulation ($\Delta=0.14$) across the league season (from pretest to posttest) (Table 7 below). No other variables were statistically significant.

Table 7. Results of Two-Sample T-Tests of Student Outcomes by Gender

		Mean (pre)	Mean (post)	t	df	One-sided p
Female	School engagement	3.75	4.27	-3.046	21	0.003**
Male	Self-regulation	3.42	3.56	-1.933	220	0.027*

Ethnicity. Similar analysis by ethnicity reveals that Caucasian students in the program made significant positive gains in *GPA* ($\Delta=0.23$) and in *self-regulation* ($\Delta=0.23$) while Hispanic/Latinx students made significant positive gains in *school engagement* ($\Delta=0.41$) (Table 8 below). No other differences by ethnicity were found.

Table 8. Results of Two-Sample T-Tests of Student Outcomes by Ethnicity

		Mean (pre)	Mean (post)	t	df	One-sided p
Caucasian	GPA	3.39	3.62	-2.626	153	0.005**
	Self-regulation	3.44	3.68	-2.629	157	0.005**
Hispanic/Latinx	School engagement	3.83	4.24	-2.239	29	0.016*

Socio-Economic Status (SES). Tests of differences in student outcomes by socioeconomic status show that students in the lower SES group showed significant positive gains in *GPA* ($\Delta=0.36$) and *school engagement* ($\Delta=0.23$) across the league season (Table 9 below). In contrast, students in the upper SES group showed significant losses in *GPA* ($\Delta=-0.20$) across the league season.

Table 9. Results of Two-Sample T-Tests of Student Outcomes by SES

		Mean (pre)	Mean (post)	t	df	One-sided p
Lower SES	School Engagement	3.70	3.93	-1.97	141	0.05*
	GPA	3.26	3.62	-3.73	136	<0.001***
Upper SES	GPA	3.60	3.40	1.797	112	0.037*

Disability Status. Tests did not reveal significant differences in student outcomes by disability status, although it is possible that the lack of findings is due to low sample size (only 6.6% of the full sample reported receiving accommodation for a documented learning disability).

NASEF Program Features that Impact Student Outcomes & Attitudes

To test how varying program implementation characteristics mediate student outcomes and attitudes (at posttest), we nested students within high school/organization and then tested for associations (for interval variables) and group differences (for categorical variables) of the three key variables of interest: club size, extent of program, and site support. Roughly 19% ($n=20$) of post-test students were successfully mapped to programs (12 sites total) for which we were able to gather implementation data from the staff survey.

Club size. Measures of association between student outcomes and club size reveals that the larger the club, the greater the student outcomes in terms of *relationships* (Table 10). One caveat here, however: It is reasonable to expect that this relationship may well break down for extremely large club sizes. No other associations between student club size and student outcomes were significant.

Table 10. Results of Spearman’s Rho test of Student Outcomes (at Posttest) by Club Size

	Correlation Coefficient	Sig. (1-tailed)	N
# of students in club Relationship	.378	0.05*	20

Extent of Program. We defined “extent of program” quantitatively in terms of the richness of resources and amount of activity, including: the number of club staff members, student roles occupied, competitions attended, and game titles played. Tests examining the relationship between these variables and student outcomes reveal multiple significant, positive patterns of association. Among the patterns found (Table 11), club staff size correlates positively with *STEM interest, school engagement, and relationships*, suggesting that the more adults involved, the more students engage in academic content. Higher numbers of differentiated student club roles correlate with higher *GPA*, implying that students who are able to participate in more roles within their club perform better academically. Students in clubs that participate in a larger number of competitions showed higher *school engagement*, and students in clubs that play a larger variety of game titles show higher levels of *communication and social relationships*. Finally, clubs that take advantage of more of NASEF’s resources show greater student *wellness* overall.

Table 11. Results of Spearman’s Rho test of Student Outcomes (at Posttest) by Extent of Program

		Correlation Coefficient	Sig. (1-tailed)	N
# of club staffing	STEM Interest	.406*	0.038*	20
	School Engagement	.623	0.002**	20
	Relationship	.582	0.004**	20
# of current student roles	GPA	.397*	0.046*	19
# of competitions attended	School Engagement	.402*	0.039*	20
# of games played	Communication	.387	0.046*	20
	Relationship	.391*	0.044*	20
# resources used	Wellness	.470*	0.018*	20

To test for differences in student outcomes between schools and organizations with the classroom esports-based English Language Arts or Career Technical Education curricula implemented versus without it, we conducted a two-sample t-test (Table 12). Students in school sites where an esports classroom curriculum is available show higher gains in

communication and *relationships* than students in schools without it. These findings suggest that schools with greater commitment to esports may foster greater academic and social gains.

Table 12. Results of Two-Samples T-Tests of Student Outcomes (at Posttest) by In-School Curricula

	t	df	One-Sided p	Mean Difference
Communication	2.379	17	0.015*	0.749
Relationship	3.048	17	0.004*	0.853

Site Support. Measures of association also reveal that students attending schools with greater support from site leadership showed significant gains in both *STEM interest* and *communication*. Site facilities, however, show no significant associations with student outcomes overall.

Table 13. Results of Spearman’s Rho test of Student Outcomes by Site Support

	Correlation Coefficient	Sig. (2-tailed)	N
School Leadership STEM interest	.484*	0.030*	20
Communication	.485*	0.030*	20

Implications

Based on these findings, we *cannot conclude that any positive changes over time that arose among our sampled students are due to the program* and not other alternative explanations like maturation, history, or testing effects. While positive changes were found for 7 of 9 target outcome variables, only *self-regulation* was significant from pretest to posttest or across the league season itself. While there appear to be changes during the league season on each of the 9 variables in the direction sought (most notably, Figure 4), none are statistically significant. There are, however, several findings that do point to noteworthy patterns in terms of student outcomes, attitudes and experiences and the variables that shape them.

First, as one might easily predict after watching this program grow and take shape, *community is key*. Student ratings of all features of the program were consistently significantly positive, but students were especially favorable toward its *social aspects* such as GMs and coaches. This preference toward the social shows up again in terms of students’ significant feelings of connectedness across all levels of the program, from team to club to peers to the league as a whole. In fact, the longer students are in the program, the more highly they appreciate the coaches and the more connected they feel to their team, club and NASEF peers – but not to the league as whole. This last finding should not surprise us, though: the league is a competitive structure in which teams and clubs compete, so it might be expected that the strongest feelings of connectedness are to their home institution’s team and club, not the league as a single entity.

That said, given NASEF’s mission to improve the academic, social and emotional lives of participating youth, it is important to ensure that the program as implemented does not devolve into just another structure for competition despite its efforts to do more for kids. An overwhelming majority (97%) of NASEF students engage directly in competition on the team. While students do occasionally take on additional roles in the club (22%), competition appears to dominate the student experience across schools and sites. Given the educational importance of *club roles* over competition, there may be reason here to be concerned. Staff professional development, increasing the usability and accessibility of educational materials on the website, and continued efforts to overtly incentivize and highlight the career-rich ecosystem over the team alone may be needed to get participating schools and organizations to more fully engage in what NASEF offers. We can bend this recent trend, but not without targeted effort.

In terms of equity of student outcomes and experiences, there are differences across key demographic groups that warrant cautious celebration. Young women in the program show significant gains in school engagement while young men show significant gains in self-regulation, a provocative pattern given the sexist culture for which esports out in the wild is often known. While white students show significant gains in GPA, Hispanic/Latinx students show significant gains in school engagement – an uneasy tradeoff, to be sure, but one that at least suggests that benefits accrue to both subgroups and not only those in the dominant group. Lower SES students show increases in overall GPA across the league season while upper SES students show decreases, a pattern that, while not necessarily ideal, at least suggests that, when student gains between privileged versus minoritized students are unequal, greater benefits tend to accrue to the minoritized. Recruitment and retention of diverse students may be one vital area for subsequent effort in the league; we highlight some effective strategies in our separate “Diversity” report.

After-school, opt-in programs are notoriously hard to evaluate rigorously given variations in how they are implemented across different sites – variations that are, ironically, quite crucial to adapting the program to specific contexts and audiences in ways that maximize impact. Here, we examined the effects of variations in program implementation to see what effect they had on outcomes and found some provocative trends. Not surprisingly, larger clubs bestow greater benefits in terms of social relationships, although there are likely limits to how large a single club can get and still have such positive social effects – for example, “Dunbar’s number” or estimate of the upper limit to the number of individuals one person can maintain relationships with, which is 150 (Dunbar, 1992).

Across the myriad implementations at schools and community organizations, the *more extensive the program version that is implemented, the greater the gains*. The more adults involved in the program, the greater the number of student roles available, and the greater the number of competitions engaged in, the better the academic outcomes. The larger the variety of game titles played, the higher student outcomes in terms of *communication* and *social relationships*. Schools with a greater commitment to esports-related curricula also show greater social gains.

One startling counterpoint to this overall trend relates to NASEF's curricular resources. In contrast to expectations, *the number of educational resources used has no bearing on student academic outcomes but does correlate with increased student overall wellness*. NASEF digital tool kits, online workshops and academies, informational webinars, and other learning materials appear to have no significant bearing on academic achievement, counter to their purpose and program expectations. They do, however, seem to play a role in student wellness. Here, future qualitative research including field observations at representative and diverse program sites will help us better understand *which* specific resources are being used and *how*. Again, staff professional development, increasing the usability and accessibility of educational materials on the website, and continued efforts to overtly incentivize and highlight their use are strongly recommended. NASEF has put much careful thought and effort into the development of such materials; ensuring their access and effective use is well worth the investment.

Limitations

As with any research investigation, our findings here have their limitations. First and most obviously, the 2020-2021 season was conducted during a global pandemic that kept many students at home and during a period of organizational transition as NASEF launched as an independent 501c3 (and subsequently partnered with another large youth esports league). We can be confident that such changes surely impacted the program and its impact on participants, although we cannot easily determine how or to what extent. A second year investigation similar to this year's is well warranted (and our lab is pursuing outside support toward these ends).

Second, power may be a formidable and ongoing issue in program assessment. In this investigation, while our sample size was greater than $n=100$ at each time of measurement, the effects of the program (if any empirically exist) are too small to be detected by standard statistical tests without much larger sample sizes (which may well prove cost prohibitive). We will continue to work on our analyses to see if there are any additional ways to increase power in this analysis.

A third and final limitation is the inconsistent length of intervention between Fall and Spring leagues. The time between pretest and posttest for the Spring season was one month shorter than the time between pretest and posttest for the Fall season. This abbreviated time frame in Spring may underlie notable differences in student outcomes between the two evaluation periods. A longer league season or measurement window may increase effect sizes for key variables above the minimal threshold for detection (power, discussed above).

References

- Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6): 469–493. doi:10.1016/0047-2484(92)90081-J.
- Johnson, C. W. (1986). A more rigorous quasi-experimental alternative to the one-group pretest-posttest design. *Educational and Psychological Measurement*, 46(3), 585-591.

Steinkuehler, C. & Oh, Y. (2012). Apprenticeship in massively multiplayer online games. In Steinkuehler, C., Squire, K., & Barab, S. (Eds.). *Games, learning, and society: Learning and meaning in the digital age* (pp. 154-184). Cambridge: Cambridge University Press.