Prairie Journal of Educational Research

Volume 1 | Issue 1 Article 7

2016

Mind the Gap: Student Researchers Use Secondary Data to Explore Disparities in STEM Education

Nathan Bean
Kansas State University, nhb7817@ksu.edu

Amanda Gnadt
Kansas State University, amanda.gnadt@gmail.com

Nicole Maupin
Kansas State University, nmaupin@ksu.edu

See next page for additional authors

Recommended Citation
Bean, N., Gnadt, A., Maupin, N., White, S. A., & Andersen, L. (2016). Mind the Gap: Student Researchers Use Secondary Data to Explore Disparities in STEM Education. *Prairie Journal of Educational Research, 1*(1). https://doi.org/10.4148/2373-0994.1002

Mind the Gap: Student Researchers Use Secondary Data to Explore Disparities in STEM Education

Abstract

Large data sets offer opportunities for graduate students to become involved in meaningful research, but also comes with a unique set of challenges. This paper seeks to examine that relationship through utilizing the High School Longitudinal Study 2009 – representative of US ninth graders in 2009 (n = 21,444) – to examine a set of research questions about STEM interest and preparation amongst secondary students. Student researchers identified gaps in plans and outcomes with regards to race, gender, exceptionalities, and socioeconomic status. Findings indicated inequities that affect STEM outcomes. A significant interaction was found between students education expectations by gender on science self-efficacy [$F(4,1264) = 2.797$, $p = .025$]. This interaction was not observed for math self-efficacy. Females and underrepresented minorities were less likely to pursue computer science courses and computer science careers [Females: X^2 (2, N = 20,594) = 111.500, $p < .0001$; Minorities: X^2 (2, N = 13,069) = 6.455, $p = .040$]. Students' expectations for post-secondary education differed by IEP status and socioeconomic status [X^2 (3, n =165,684) = 26.886, $p = 0.001$]. Finally, time spent in extracurricular activities impacted academic achievement and students in lower socioeconomic groups were less involved in extracurricular activities [X^2 (4, n = 20,598) = 132.298, $p < .0001$].

Keywords

STEM, equity, under representation, gender, disability, extracurricular activities, K-12, graduate education, large data sets

Authors

Nathan Bean, Amanda Gnadt, Nicole Maupin, Sherry A. White, and Lori Andersen

Mind the Gap: Student Researchers Use Secondary Data to Explore Disparities in STEM Education

Nathan Bean, Amanda Gnadt, Nicole Maupin, Sherry A. White, & Lori Andersen

Large data sets offer opportunities for graduate students to become involved in meaningful research, but also comes with a unique set of challenges. This paper seeks to examine that relationship through utilizing the High School Longitudinal Study 2009 – representative of US ninth graders in 2009 (n = 21,444) – to examine a set of research questions about STEM interest and preparation amongst secondary students. Student researchers identified gaps in plans and outcomes with regards to race, gender, exceptionalities, and socioeconomic status. Findings indicated inequities that affect STEM outcomes. A significant interaction was found between students education expectations by gender on science self-efficacy [$F(4,1264) = 2.797$, $p = .025$]. This interaction was not observed for math self-efficacy. Females and underrepresented minorities were less likely to pursue computer science courses and computer science careers [Females: X^2 (2, N = 20,594) = 111.500, $p < .0001$; Minorities: X^2 (2, N = 13,069) = 6.455, $p = .040$]. Students' expectations for post-secondary education differed by IEP status and socioeconomic status [X^2 (3, n =165,684) = 26.886, $p = 0.001$]. Finally, time spent in extracurricular activities impacted academic achievement and students in lower socioeconomic groups were less involved in extracurricular activities [X^2 (4, n = 20,598) = 132.298, $p < .0001$].

Introduction

It can be difficult for a single researcher to collect and analyze a large data set. Fortunately, organizations like the National Center for Education Statistics (NCES) make secondary data sets available to faculty, staff, and students for conducting research. The use of secondary data helps to advance the field because it allows for replication and confirmation of studies and findings (Tresniewski, Donnellan, & Lucas, 2011). Additionally, secondary data allows researchers to examine samples representative of a larger population in greater detail. Large data sets can be particularly useful for undergraduate and graduate students learning best practices of research and analysis during a semester-long course.

To develop a better understanding of how existing large data sets can be used to answer research questions, the research team will work with a secondary data set to illustrate common issues of working with secondary data, including: missing data, sampling and weighting, and statistical analysis of large data sets. The team will be exploring the High School Longitudinal Study (HSLS:09) collected by the National Center for Education Statistics and representative of the US population of high school students in 2009. This data set will be used to answer questions about gaps in STEM education.

The National Science Foundation's Center for Science and Engineering Statistics shows clear trends of underrepresentation of women in engineering and the computer sciences, and disparities across all sciences, engineering, and mathematics for minorities and people with disabilities (NSF, 2013). At a time when STEM-ready student production is at an all-time low, this is a serious problem for the nation in terms of infrastructure, security, and social equality. Addressing these questions is therefore a pertinent subject for investigation with a nationally representative dataset like HSLS:09.

The researchers identified four focus areas within the umbrella of STEM disparities: math and science self-efficacy, interest in computer science, extracurricular participation, and post-secondary plans for students with disabilities. These areas will be addressed in each section using the preceding sequential order.

Review of Literature

There is a growing concern in the United States about the performance and interest of elementary and secondary school students in the science, technology, engineering and math (STEM) fields. According to the President's Council of Advisors on Science and Technology (2010), the future of the United States as a leader in technological advancements is dependent upon the education and success of students in these areas, within which women and minorities are known to be underrepresented. As the Council argues, "we must *prepare* all students, including girls and minorities who are underrepresented in these fields, to be proficient in STEM subjects and we must *inspire* all students to learn STEM and, in the process motivate them to pursue STEM careers" (President's Council of Advisors on Science and Technology, 2010, p. 44-45, emphasis theirs).

Math and Science Self-Efficacy and Postsecondary Education Plans by Gender

The President's Council identifies the primary concern related to STEM fields, and brings about the bigger question of how can this be accomplished? If the United States is going to remain competitive in the international community we must increase student achievement and interest in the STEM fields. Developing a better understanding of self-efficacy, social support and the educational and career goals of students is one place to start.

Students' perceptions about their abilities in math and science play a major role in their persistence in STEM fields. Based on the existing gender gaps in the STEM field it may be possible that self-efficacy for males and females varies. According to Rice, Barth, Guadagno, Smith, and McCallum (2013), self-efficacy for girls remains stable or decreases and self-efficacy for boys tends to decrease throughout adolescence. Support from parents, teachers and peers can affect a student's self-efficacy in science and math. Students who perceive support and encouragement of their math and science abilities reported higher self-efficacy in math and science (Rice et al., 2013).

Developing an understanding of students' self-efficacy in the STEM fields and how it changes over time will provide a foundation for identifying strategies to increase student interest. According to Rice et al. (2013), students need support from parents, teachers and peers to increase self-efficacy. Encouragement and positive reinforcement from these groups should lead to students' increased perceptions about their abilities in science, technology, math and engineering. HSLS:09 offers the chance to examine the relationship between supporting factors and STEM-centric math and science self-efficacy.

Disparities in Computer Science Participation

Computer science (CS) and its tools, computational thinking and programming, have grown into invaluable resources for STEM practitioners, reshaping the way we theorize, form and test hypotheses, and even how we carry out and disseminate research (Bundy, 2007). Jeanette Wing, former CS professor and current Vice President of Microsoft Research has suggested “computational thinking will be instrumental to new discovery and innovation in all fields of endeavor (Wing, 2008, p. 3717).” Given this importance, the current gaps in undergraduate degrees awarded for CS is disconcerting: 61.2% were White, and 85.8% were male (Zweben & Bizot, 2013). While ethnic minorities have been making slow gains, female participation in computing science has actually been declining (National Science Foundation [NSF], 2013). This raises serious concerns for both the United States’ position in the computing field and for gender and racial equality within it. As Brigid Barron expresses, the question becomes “who will have the knowledge that will position them to design, create, invent, and use the [computational] technologies to enhance their personal lives and social worlds?” (2004, pp. 1-2).

Numerous organizations, including the NSF, ACM, CSTA, Code.org, Kahn Academy, and Google have embarked upon ambitious efforts to interest and involve more diverse audiences in CS. HSLS:09 offers the chance to see what impact these efforts have had on current high-school populations, helping to guide future efforts.

Post-Secondary Involvement of Students with Disabilities

Students receiving special education services throughout high school often are conflicted over post secondary plans and have anxiety about the transition after high school. Students with disabilities often pursue employment opportunities and shy away from obtaining a further degree (VanBergeijk, 2012). Sitlington, Frank, and Carson (1993) investigated post secondary adjustments of young who received special education services throughout high school. Of the sampled participants, 49% of those with behavior disorders, 70% of those with learning disabilities, and 54% of those with mental disabilities reported that they received no postsecondary education or training of any kind. When examined by gender 73% of the males and 68% of the females had no postsecondary training. The most frequently reported educational experience was a community college program. The second most commonly reported option for males was military training. Fewer than 5% of individuals in any disability category had attended a 4-year college (Sitlington et al., 1993).

As participation within STEM disciplines is predicated upon receiving specialized post-secondary training, typically through a 4-year college, it is important to understand current trends amongst this population’s post secondary plans. HSLS:09 offers a window into just that for the current cohort of high-school students.

Extracurricular Participation and Postsecondary Education Plans

Fredricks (2012) found involvement in EAs have a positive impact on academic outcomes despite concerns about time required and impact on family dynamics, up to a certain level of participation (more than 20 hours). Participation in Extracurricular activities (EAs) may also result in the ability to demonstrate grit, various levels of independence, leadership skills, and positive social outcomes (Covay & Carbonaro, 2012). These skills and academic improvement can help prepare students for the demands of post secondary education in STEM fields.

However, the accessibility to school-based EAs for students of lower-socioeconomic status (SES) may be reduced compared to those of higher-SES. Opportunities to participate depend on how involved the family is, transportation factors, and the overall financial requirements. Covay and Carbonaro (2012) expressed the need for further research to gain a deeper understanding of students from lower socio-economic status (SES) and the relationship between EAs and achievement gaps. HSLS:09 offers current data bearing on this relationship.

Research Questions

The research team examined four different research questions related to equity issues in the STEM disciplines, self-efficacy in these fields, and students' post-secondary plans. The following research questions were addressed:

- How do students' math and science self-efficacies relate to students' postsecondary education plans? Are there differences by gender?
- Is gender or race related to students' taking of computer science courses? In the student's choice of a computer science career?
- Do students with individualized education plans (IEPs) differ from general education students in their expectations to obtain a degree post high school? Of the students that have an IEP, are there differences in their expectations for postsecondary plans by socioeconomic status?
- Does participating in extracurricular activities have an effect on a student's plans to attend college? Does SES status affect the relationship between participation and educational plans?

Method

This study used data from the High School Longitudinal Study of 2009 (HSLS:09) conducted by the National Center for Education Statistics (NCES). This is a comprehensive longitudinal quantitative study with follow-up quantitative surveys planned throughout the secondary and postsecondary years (NCES, 2014). Currently, survey data is available for the base year and the follow-up survey in the 11th grade. Data for this study was downloaded using the Educational Data Analysis Tool, an online tool provided by NCES to access HSLS data. This study was found to be exempt from review by the IRB at Kansas State University.

Sampling Plan

HSLS:09 utilizes a two-stage sampling design. Public and private schools were selected using stratified random sampling, resulting in 1,889 schools with 944 (55.5%) choosing to participate. In the second stage, a total of 25,206 students were randomly sampled. Approximately 27 students per school participated resulting in 21,444 respondents (Ingels et al., 2011). Because this sampling plan involved stratified and clustered data, additional steps were taken by the researchers to apply appropriate weights to the data.

Sample populations who share experiences and culture tend to be more similar, displaying a narrower data distributions and smaller standard deviations. As the HSLS:09 sampling plan used participants who attended the same schools and classes, the raw data suffers from clustering effects arising from these shared-experience groups. These can be countered by applying a *design effect weight* to the data, which is computed as part of the sampling effort (Trzesniewski, Donnellan, & Lucas, 2011). Researchers used two different design effect weights from HSLS:09 – one for the base year data and one for the first follow-up, which the researchers applied based on which data set they were examining when needing to compare measures of variance (measures of centrality are not affected by clustering).

https://newprairiepress.org/pjer/vol1/iss1/7
DOI: 10.4148/2373-0994.1002

To be generalizable to the national population, the respondent pool needs to also be representative of that population. In practice response rates vary across subgroups of interest, like ethnicity and SES. Further, some groups may be deliberately oversampled to provide the statistical strength needed for within and between-groups analyses – as is the case with HSLS:09, which oversampled the Asian student population (Ingels et al., 2011). To bring the data back into a generalizable format, *sample weights* must be applied to each response based on the respondent's subgroups to adjust their aggregate contributions to the analysis to be proportional to that subgroup's national representation. These weights are provided with the HSLS:09 and vary by survey instrument as the base-year, first follow-up, teacher, administrator, and parent surveys all had different response rates. Researchers therefore also had to select and apply the sample weight appropriate for the data they were working with. Details on specific sample weights used by the researchers can be found in the Appendix.

Secondary Data Challenges

In addition to the need to apply weights to ensure statistical analyses were generalizable to the national population and that standard errors were adjusted for the complex sampling design, other challenges arose from working with secondary data. These centered around what data was collected, and how much of it was available.

Researchers working with secondary data are limited to what items the original researchers chose to include in the study (Trzesniewski, Donnellan, & Lucas, 2011). To address research questions that were not part of the study's original focus can therefore require significant creativity on the part of the researcher. For example, HSLS:09 did not collect data concerning participant's disability status. However, the student data did include a flag to indicate if a student had an individualized education plan (IEP) [x1iepflag], which the researchers used as an indicator of disability status.

As with any survey-based research, missing data also becomes a problem. HSLS:09 preparers imputed missing values in the fields they considered most critical, like ethnicity, gender, and mathematical ability, but other measures did not receive this treatment (Ingels et al., 2011). Returning to the previous example, the parent survey contained an item asking if their student received Special Education Services [p1specialed]; while the expectation would be any student receiving such aid *should have* an IEP, in a significant number of cases the x1iepflag variable did not reflect the parent's report. Accordingly, the researchers created a new variable in which either a positive response to either the x1iepflag or p1specialed variables indicated disability status.

To protect participants, potentially identifying information related to schools and the participants were removed by NCES from the public dataset. This was not limited to participants' identities, but extended to any item that involved a small enough group that responses could be used for identification. In these instances, the response was replaced with a *data redacted* code. For example, NCES collected information on participant's career plans and coded them using both the general 2-digit [X2STU30OCC02] and the more specific 6-digit O*NET [X2STU30OCC06] occupational codes (Ingels et al., 2011). The 6-digit code was redacted from the public data set, so the researchers used the more generic 2-digit code, which reduced the specificity of that particular analysis.

Math and Science Self-Efficacy and Postsecondary Education Plans by Sex

The first research question examined the relationship between math and science self-efficacies on student postsecondary education plans. *Self-efficacy* can be defined as an individual's belief in his ability to complete tasks and reach goals (Bandura, 1997). Self-efficacy measures were scale scores [X1MTHEFF, X1SCIEFF] computed by NCES using four survey items, which asked about students' confidence in their ability to master skills, successfully complete assignments, understand textbooks and do well on tests in either math or science. Student educational expectations were obtained from a survey item [X1STUEDEXPCT] asking students to report how far in school he or she thinks they will go in school. For this study, the researcher collapsed the student educational expectations categories from 11 categories to five categories: (1) high school or less, (2) associates, (3) bachelors, (4) advanced degree, and (5) don't know.

Base-year data were pulled from HSLS:09 for student math self-efficacy, student science self-efficacy and student educational expectations. Mean self-efficacies were compared between each of the student education expectation categories for math and science self-efficacy. A two-way ANOVA was conducted to examine the interaction between student educational expectations and sex on math self-efficacy. A second two-way ANOVA was conducted to examine the interaction between student educational expectations and sex on science self-efficacy.

Disparities in Computer Science Participation

The second research question examined the relationship between race, gender (X2SEX), and student's plans to take a computer science course and anticipation of a computer science career. For this study, the researcher collapsed race (x2race) into three categories: (1) White, (2) Asian, and (3) Underrepresented minorities. Underrepresented minorities consisted of Black and Hispanic students. All other races were not included in subsequent analyses.

In the first follow-up study, students were asked if they were taking a computer-related course in Spring 2012. The options were: a computer applications course (S2COMPAPP12), a computer programming course [S2COMPPROG12], an AP Computer Science course [S2APCOMPSCI12], or other computer or information science course [S2OTHCOMP12]. The researcher collapsed the latter three categories into a single category — taking a computer science course. The computer applications course option was intentionally left out, as these courses were likely to focus on basic computer literacy, not computer science topics.

In the same survey, students were also asked to write in their anticipated career choices in the base year and follow-up survey, these responses were coded by NCES using 2-digit O*NET occupational codes [X2STU30OCC02]. The researcher created a binary variable with a value of 1 when the student indicated a career with an O*NET code of 15 (Computer & Mathematics Occupations) and a value of 0 for all other values. Undecided, missing, and uncodable responses were left out of the subsequent analyses.

Tables were constructed and Chi-Squared values computed to determine if computer science course taking and computer science career plans were related to student's race/ethnicity or gender (Tables 3-6).

Post-Secondary Involvement of Students with Disabilities

To answer the third research question regarding students' post-secondary expectations by disability status, disability status was operationalized as the student having an individualized education plan (IEP). IEP status was determined from x1iepflag and p1specialed. If a conflict of reporting was present (i.e. parents reported no IEP but school enrollment stated that a student did have an IEP), the researcher used the information from the school enrollment forms as detailed in the Secondary Data Challenges section, above.

Students were also asked to self-report on how far they expected to get in school [X1STUEDEXCPT], which the researcher collapsed into four different categories, (1) high school or less, (2) associates/bachelor's degree, (3) advanced degree (Masters or PhD), and (4) students who did not know.

Socioeconomic status (SES) for students was divided by NCES into 5 different levels [X1SESQ1, X1SESQ2, X1SESQ3, X1SESQ4, X1SESQ5], conceptualized by responses using five components: highest education among parents, education level of the other parent (if applicable), highest occupational prestige score of parent, occupation prestige of the other parent (if applicable), and the family income. To better examine differences in the expectations of students who had an IEP based on socioeconomic status, these SES quintiles were collapsed into three groups; low, middle, and high.

A table was created to identify students postsecondary expectations based on IEP status in each SES group. A chi-square was conducted to determine the difference between postsecondary expectations of students with an IEP in the lowest and highest socioeconomic quintiles (Table 8).

Involvement in Extracurricular Activities and Student Plans on Achievement Outcomes

To answer the fourth research question examining how far students planned to go in school in relation to EC involvement, the researcher drew upon the base year data. Involvement in EAs was measured by NCES as the amount of time spent in EAs [S1HRACTIVITY]. For the analysis the eleven educational expectation categories [x1studedexpct] were collapsed by the researcher into five: (1) High school (HS) or less, (2) Associates Degree (AA), (3) Bachelor's Degree (BA), (4) Advanced Degrees, and (5) Don't know.

As this analysis did not indicate a significant difference between the number of hours that a student participated in EAs and his or her plans to complete an AA or a BA (Table 9), the researcher collapsed the data pertaining to the educational expectation levels again. Five categories were collapsed into three: (1) No College, (2) College, and (3) Don't Know. Additionally, this analysis included collapsed levels of SES quintiles (Table 10).

Results

Missing Data and Imputed Values

NCES identified a number of variables as critical to future analyses, and took great steps to ensure that data was available in these categories, including combining items from multiple surveys, telephone follow-ups, scouring school records, and, as a last resort, imputation (Ingels et al., 2011). Among these variables were several used by the researchers – gender [X1SEX, X2SEX], race/ethnicity [X1RACE, X2RACE], SES quintile [X1SESQ1, X1SESQ2, X1SESQ3, X1SESQ4, X1SESQ5], IEP Status [X1IEPFLAG]. Specific details on how missing values were handled by NCES for these variables can be found in the Appendix.

Math and Science Self-Efficacy and Postsecondary Education Plans by Gender

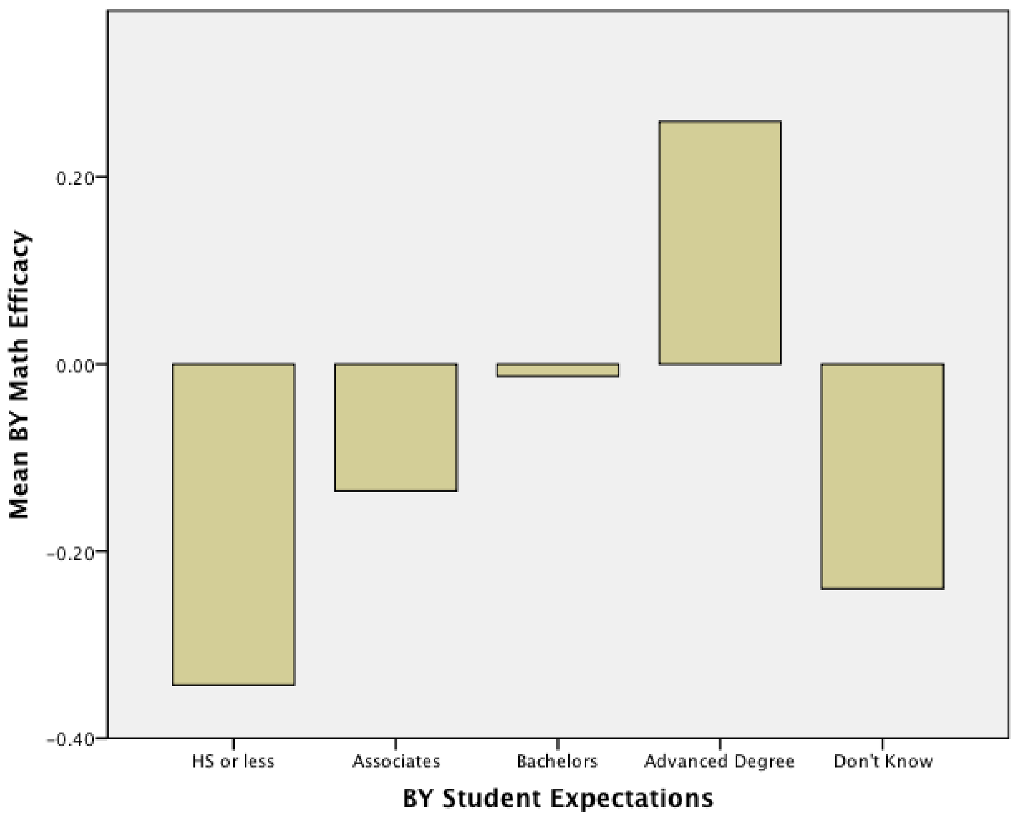
This analysis is dependent upon survey results for math and science self-efficacy. The math self-efficacy score is based on a scale computed by NCES using four survey items. This scale score had 13.1% missing values, most likely because of block scheduling. Students who were not enrolled in math or science did not respond to any of the self-efficacy questions. These cases were omitted from analysis. When comparing the respondent group and the non-respondent group by gender there were more males in the non-respondent group (54%) compared to the respondent group (49.4%).

Table 1
Student Educational Expectations.

Educational Expectations	Frequency	Percent
HS or Less	606,356	14.4%
Associates	279,940	6.8%
Bachelors	682,255	16.6%
Adv. Degree	1,653,843	40.2%
Don't Know	892,566	21.7%
Total	4,114,690	100.0%

* Weighted by W1STUDENT

Figure 1
Student education expectations by math self-efficacy.



Science self-efficacy was also computed by NCES as a scale score based on four survey items. The science self-efficacy scale had 20.5% missing values, most likely due to block scheduling. Students who were not enrolled in a science course during the fall 2009 term would not have completed these survey items. These cases were omitted from analysis. When comparing the respondent group and the non-respondent group there were more males in the non-respondent group (54.6%) compared to the respondent group (49.7%).

The missing data for math and science self-efficacy were examined by gender. The missing data did not indicate notable differences between the two genders. The missing data appears to be at random and those cases were omitted from analysis.

A 2 x 1 ANOVA on student education expectations and sex on math self-efficacy showed significant differences for some of the five education expectation categories. Students with expectations of earning an advanced degree had a higher self-efficacy than all of the other categories (Figure 1). There was a statistically significant effect for student education expectations [F (4,1455) = 23.606, *p* = .001] and a significant main effect for sex [F (1,1455) = 4.572, *p* = .033], but no significant interaction [F(4,1455) = 1.026, *p* = .392] between education expectations and sex. Math self-efficacy was different between categories of student education expectations. Math self-efficacy was different for students based on sex. However, there was no significant interaction between education expectations and sex on math self-efficacy.

Table 2

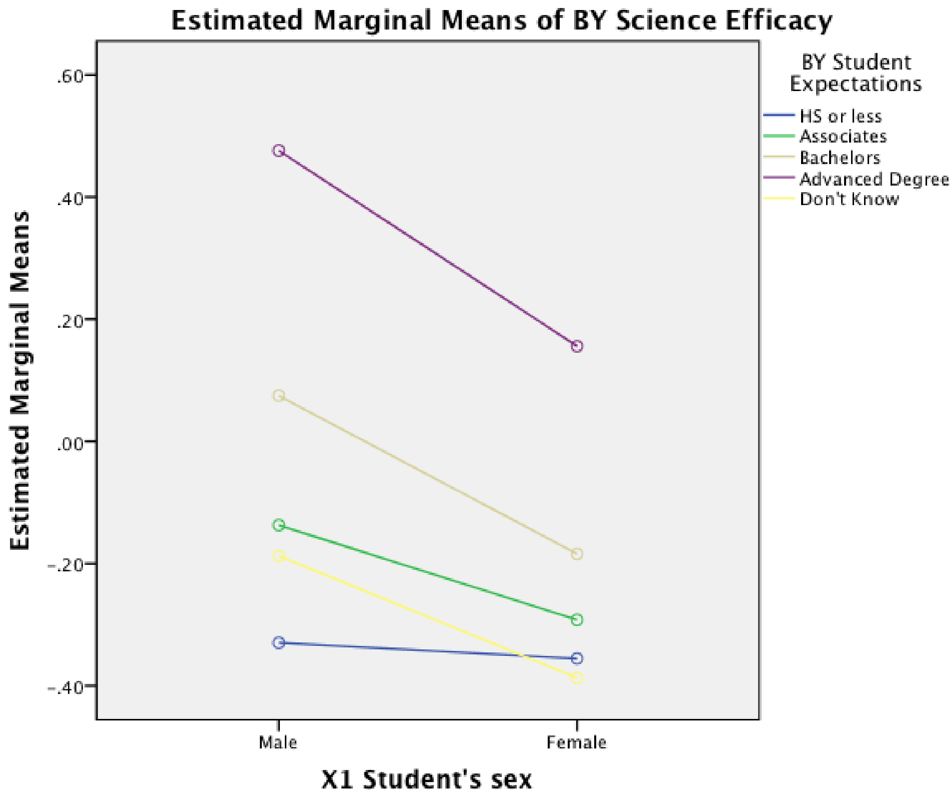
Two-Way Analyses of Variance for Math Self-Efficacy and Science Self-Efficacy as a Function of Student Educational Expectations and Sex.

Source	df	MS	F	P	η^2
On Math Self-Efficacy					
Student Expect	4	21.904	23.606	< .05	.061
Sex	1	4.242	4.572	< .05	.003
Student Expect x Sex	4	.952	1.026	.392	.003
Error	1445	.928			
On Science Self-Efficacy					
Student Expect	4	30.955	34.764	< .05	.099
Sex	1	.091	.102	.749	.000
Student Expect x Sex	4	2.491	2.797	< .05	.009
Error	1264	.890			

* Weighted by DEFF Weight

There was a statistically significant effect for student education expectations [F (4,1264) = 34.764, *p* = .001] and a significant interaction between education expectations and sex on science self-efficacy [F (4,1264) = 2.797, *p* = .025], but no statistically significant effect for sex [F (1,1264) = .102, *p* = .749]. Science self-efficacy was different between categories of student education expectations (Figure 2). There was a significant interaction between student education expectations and sex on science self-efficacy. There was not a significant difference by sex on science self-efficacy. Table 2 shows the effects of student education expectations and sex on math and science self-efficacy.

Figure 2
Interaction between science self-efficacy and education expectations.



Disparities in Computer Science Participation

This analysis focused on two different subgroups – gender and ethnicity – and their relationship with CS course and career plans. Of the career choice responses, 354 were missing. These and an additional 5,550 "don't know", 131 uncodable, and 2,821 non-response responses were left out of the analysis. In determining which students were taking computer science courses in Spring 2012, 583 missing and 2,821 non-respondents were left out of the analysis, while 4,023 legitimate skips were merged into the "No" category. In analyzing relationships with race, 105 American Indian/Alaskan, 1,287 multi-race non-Hispanic, and Native Hawaiian/Pacific Islander, Non-Hispanic respondents were left out of the analysis.

Race and computer science career plans

A chi-square test of independence was performed to examine the relationship between race and anticipated computer science career among students who had developed career plans. The relation between these variables was statistically significant [X^2 (2, N = 13,069) = 6.455, p = .040]. The odds ratio for underrepresented minorities to White students was 1.699 with a 95%

confidence interval of [1.017, 2.840]. This suggests that White students are 69.9% more likely to plan for a computer science career as their underrepresented minority counterparts.

Table 3

Anticipated Computer Science Career vs. Race.

		White	Asian	Underrepresented Minorities
Not Anticipating Computer Science Career	Count % within race	1,469,434 97.1%	83,974 95.3%	1,027,594 98.2%
Anticipating Computer Science Career	Count % within race	44,557 2.9%	4,121 4.7%	18,311 1.8%
Total	Count	2,151,495	147,067	1,498,620

*Weighted by W2STUDENT

Gender and computer science career plans

A chi-square test of independence was performed to examine the relationship between gender and anticipated computer science career among students with career plans. The relation between these variables was statistically significant [X^2 (1, N = 14,529) = 63.252, p < .0001]. The odds ratio for the Males is 10.017 with a 95% confidence interval of [4.997, 20.078]. This suggests that males are 10 times more likely than females to express plans for a career in computer science.

Table 4

Anticipated Computer Science Career vs. Sex.

	Male	Female
Not Anticipating Computer Science Career	1,281,212 95.2%	1,560,822 4.8%
Anticipating Computer Science Career	64,964 2.6%	7,894 0.5%
Total	1,346,176	1,568,716

*Weighted by W2STUDENT

Race and computer science course plans

A chi-square test of independence was performed to examine the relationship between race and taking a computer science course. The relation between these variables was not statistically significant [X^2 (2, N = 18,120) = 1.286, p = .526]. There was no statistically significant relationship found between a student's race and likelihood of taking a computer science course.

Table 5

Plans for Taking a Computer Science Course in Spring 2012 vs. Race.

		White	Asian	Underrepresented Minorities
Not Taking Computer Science Course	Count % within race	2,036,181 96.7%	136,860 95.7%	1,409,203 97.1%
Taking Computer Science Course	Count % within race	68,633 3.3%	6,147 4.3%	41,674 3.1%
Total	Count	2,104,814	143,034	1,450,877

* Weighted by W2STUDENT

Gender and computer science course plans

A chi-square test of independence was performed to examine the relationship between gender and taking a computer science course. The relation between these variables was statistically significant [X^2 (1, N = 20,071) = 17.394, p < .0001]. This suggests that males are more likely than females to take a course in computer science. The odds ratio for the males is 2.066 with a 95% confidence interval of [1.456, 2.932]. This suggests that males are more than twice as likely as females to take a course in computer science.

Table 6

Plans for Taking a Computer Science Course in Spring 2012 vs. Gender.

		Male	Female
Not Taking a Computer Science Course	Count % within sex	1,944,471 95.7%	1,976,737 97.9%
Taking a Computer Science Course	Count % within sex	86,532 4.3%	42,228 2.1%
Total	Count	2,031,003	2,018,965

* Weighted by W2STUDENT

Post-Secondary Involvement of Students with Disabilities

To determine the amount of missing data in students IEP status the researcher combined reports of IEP status from school enrollment and parent responses to whether or not students had an IEP; 5% of students had data missing for IEP status. Missing values were examined for parent responses in each of the SES groups to identify patterns. Each SES group had a similar percentage of missing values, indicating that removal of these cases would not bias the results by SES. The cases without data were removed.

IEP status and educational expectations

The percentage and frequencies of students' reports for how far they expected to go in school by IEP status was examined. The largest difference was observed in the percentage of students who expected to get through high school or less. Of students with an IEP, 25.8% reported that they only expected to get through high school or less compared to the 11.8% of students without an IEP. There was also a large difference in the plans for advanced degrees. Of the students without IEPs, 44.3% reported an expectation to obtain an advanced degree compared to only 22.9% of students with IEPs expectation for an advanced degree. There was also a large percentage of students with IEPs that are still unsure and/or do not know how far they plan to go in school (table 7).

Table 7

Student educational expectations by IEP Status.

IEP Status	HS or Less		Assoc./Bach		Adv. Degree		Don't Know	
	%	N	%	N	%	N	%	N
Y	26.40%	102,580	21.70%	84,349	22.20%	86,476	29.70%	115,725
N	12.10%	359,613	24.40%	723,030	43.10%	1,279,325	20.30%	602,919
Total	14.70%	606,355	23.40%	962,195	40.20%	1,653,843	21.70%	892,567

* Weighted by W1PARENT

SES, IEP status, and educational expectations

To address the second portion of the research question, the effect of SES categories on expectations by both IEP status was examined. A chi-square test of independence was performed to examine the relation between expectations of students with an IEP in the highest and the lowest SES groups. The relationship between these variables was significant [X^2 (3, n = 165,684) = 26.886, p = 0.001].

Students with an IEP in the lowest quintile had differences in their postsecondary plans compared to students with an IEP in the highest quintile (See table 8). Students from a high socioeconomic status who had an IEP had a higher expectation to attend some type of postsecondary education. Of students with an IEP from a lower socioeconomic status, a large percentage only expected to obtain a high school diploma or less.

Table 8

Educational Expectations by IEP status grouped in SES groups.

SES	IEP Status	HS or Less	Assoc./Bach.	Adv. Degree	Don't Know
Lowest Quintile	Y	34% 44,779	20.6% 27,073	16.2% 21,372	29.2% 38,433
	N	22% 151.482	25.5% 175,669	30.6% 210,733	22% 151,558
Middle 3 Quintiles	Y	26.1% 61,456	22.5% 52,866	19.8% 46,643	31.5% 74,087
	N	11.7% 222,307	25.5% 484,682	42.8% 813,412	20% 379,750
Highest Quintile	Y	8.5% 5,642	23.6% 15,610	47.1% 31,137	20.8% 13,741
	N	3% 26,645	22.2% 197,038	58.3% 516,764	16.5% 146,023
Total	Y	25.8% 111,877	22.1% 95,549	22.9% 99,152	29.2% 126,261
	N	11.5% 400,434	24.7% 857,389	44.3% 1,540,909	19.5% 677,331

* Weighted by W1PARENT

Extracurricular Participation and Postsecondary Education Plans

Findings suggest that time spent in EAs is associated with higher rates of planning to pursue advanced degrees. Overall, 20% of the ninth graders do not have plans for post secondary education. Of the students that were sampled, 50% participated in over three hours SIEA per day. The relationship between SAPO and seeking an advanced degree was significant [X^2 (4, n = 20,598) = 132.298, p = < .0001] (See table 9).

Table 9

Student Involvement in EA and Educational Expectations.

SIEA (hrs)	HS<	AA	BA	Advanced	Unknown	Total
<1	21%	8%	15%	31%	26%	1,369,077
1-3	12%	7%	18%	44%	20%	1,772,482
3+	9%	5%	18%	50%	18%	790,032
Total						3,931,519

* Weighted by W1STUDENT

SIEA and SAPO reports were combined with the quintile of SES for analysis. Students in the lowest quintile participated in less than one hour of SIEA compared to 5% of students who participate in less than one hour from the highest quintile [X^2 (4, n = 2,550) = 123.988, p = .0001] (See table 10).

Of students who did not plan to go to college 32% of students from the lowest SES quintile participated in less than one hour of EAs in comparison with 5% of students from the highest quintile [X^2 (4, n=2550), p=<.0001].

Table 10:

SIEA and SAPO by Socioeconomic Status

	SES & SIEA (hrs.)	No College	College	Don't Know
Lowest Quintile	<1	32%	43%	26%
	1-3	25%	55%	21%
	3+	20%	63%	17%
	Total			758,253
Middle Quintiles	<1	20%	54%	26%
	1-3	12%	67%	21%
	3+	9%	72%	19%
	Total			2,367,693
Highest Quintile	<1	5%	71%	25%
	1-3	3%	81%	16%
	3+	3%	82%	15%
	Total			815,646

* Weighted by W1PARENT

Discussion

The following section includes discussion about the conclusions and implications for each of the four research questions. This is followed by future research recommendations.

Math and Science Self-Efficacy and Postsecondary Education Plans by Gender

When looking at the frequency distribution for student education expectations a large group (40%) of 9th grade students indicated plans to earn an advanced degree; either a master's or doctoral degree. The second largest group of students did not have educational plans (21%). This group of students is particularly interesting.

The two-way ANOVAs conducted on math and science self-efficacy by sex and education expectations yielded significant results in some areas. In the math self-efficacy ANOVA, females indicated lower self-efficacies than males. Additionally, students with higher self-efficacy in math indicated higher education expectations. As math self-efficacy decreased education expectations decreased as well.

The two-way ANOVA conducted on science self-efficacy yielded a significant interaction between sex and education expectations. Significant differences were observed for males in the "don't know" group and the "high school or less" group. The males in the education expectation group "don't know" had a higher self-efficacy than those who indicated they planned to complete high school or less. This interaction should be re-examined in future studies.

It would be interesting to examine the follow-up survey data to see how the males in the "don't know" and "high school or less" groups reported self-efficacy and education expectations in the eleventh grade. According to Rice et al. (2013), males may experience a decrease in self-efficacy throughout adolescence while females' self-efficacy often remains stable over the adolescent years. Future research could include examination of student perceptions of the support and interactions with their teachers, parents, and peers. Rice et al. (2013), report perceived support from these individuals leads students to report higher self-efficacies in math and science. This indicates self-efficacy may be affected by feedback received from teachers, parents and peers. Encouragement from these groups may make a difference in student self-efficacy.

The finding that males in the "don't know" education expectation category had higher self-efficacies than those reporting education expectations of "high school or less" calls for a closer look at the education system. The President's Council of Advisors on Science and Technology calls for an increased effort to encourage and educate students in the STEM fields (2010). It may be necessary to examine the K-12 curriculum to identify how science and math are treated and taught. Ensuring that schools are afforded the time and resources to teach science and math may be one place to start.

Disparities in Computer Science Participation

No statistically significant relationships were found between a student's race and their likelihood to plan on a CS career. Additionally, the percentages were very similar between White and underrepresented students. These results suggest that national efforts to recruit underrepresented minorities into high school CS courses are succeeding. A good follow-up question would be if this holds true across SES categories, as a more subtle and profound source of disparity is the lack availability of CS courses in poorer schools which generally have a higher proportion of underrepresented minorities (Margolis & Fischer, 2003).

https://newprairiepress.org/pjer/vol1/iss1/7
DOI: 10.4148/2373-0994.1002

Nonetheless, even while efforts at recruiting underrepresented minorities into CS coursework at the high school level has been successful, the significant relationship that was found between race and anticipated career at 30 – with minorities about half as likely to choose a CS career – raises questions about how well these efforts will succeed at changing disparities within the professional field. This particular finding would benefit from further analysis with the restricted HSLS:09 data set, where the more-specific 6-digit O*NET occupational codes could be used to separate students interested in CS careers from those interested in math and other computing careers. Additionally, as further follow-up data becomes available it would be valuable to see if this population of students does pursue higher education and eventual careers in the field, despite their lack of current intent.

Given the low numbers of students overall interested in Computing Science (131 underrepresented students taking a CS course and 83 planning on a CS career of 21,095 students included in the analysis) the statistical power for this comparison was very weak. Further work would greatly benefit from oversampling students interested in CS to increase this statistical power.

In examining female interest and participation in CS, strongly significant results were found despite the small numbers. This is due to the sheer size of the disparities, with females half as likely as males to take CS courses and ten times less likely to plan on a CS career. The greater number of females taking high school CS courses again suggests that recruitment efforts at that level have been fruitful, but less so at enticing females to consider a computer science career. This finding does align with the literature, which suggests that females are more likely to have an interest in using the tools of computer science in another career (Barron, 2004; Margolis, 2003). Nonetheless, the computer science industry suffers without gender balance, and research suggests that females are more likely to enter computer science degrees and careers when they are part of a strong cohort of women (Margolis, 2003). Thus, gender disparities in CS remain an open problem amongst the 2009 national high school freshmen cohort.

Post-Secondary Involvement of Students with Disabilities

The findings show a clear disparity in educational expectations for student with disabilities. Students with an IEP do not have as high of expectation for themselves in post secondary education. Further, students with an IEP in lower SES groups had even lower expectations than those in the higher SES groups.

However, when comparing this study to the 1993 study conducted by Sitlington et al. results suggest that current students with IEPs have higher expectations to obtain post secondary education. The 1993 study found that on average 57.7% of the 737 students they surveyed reported no post secondary education or training. The current HSLS:09 study indicated that only 26.4% of students expected to obtain no type of post secondary education or training. This implies that current students receiving special education services have higher expectations from themselves than their counterparts in the 1993 study. They may be more confident in their abilities and/or supports that may be available in their endeavors for post-secondary education.

Though this study found that 43.4% of students with an IEP expect to obtain some type of degree after graduating high school, in reality these students may not have the resources or supports to actually do so. The follow-up surveys planned for these students post graduation will be an important resource for researchers interested in examining these implications further.

Barriers are still very much present for students with an IEP wanting to pursue post-secondary education. It is important individuals are aware of current supports and programs available for special education services post high school. With the boom in distance education, students who require accommodations and modifications to educational programs have more opportunities to receive assistance without the stigma that comes with receiving supports in higher education. Students have the ability to pace themselves and obtain necessary supports to complete courses from within their own home.

Students receiving special education services need to continue to be encouraged and supported in order to see a continued upward trend in expectations for post secondary education. Though it is still important to enter the work field, it is equally important that students with an IEP feel that they can be successful in postsecondary education programs. Anxiety surrounding the transition to college can be reduced if students are more aware of the supports and unique programs available for those who require special services.

Extracurricular Participation and Postsecondary Education Plans

This study examined the relationship between involvement in extracurricular activities and student academic plans. Results support the hypothesis that a positive relationship exists. Further, the relationship between students' involvement in extracurricular activities and their socioeconomic status was examined; low socioeconomic status correlated to less time in EAs.

These findings imply that students from the lowest quintile have unequal opportunities to participate in EAs. Future research and school administrators should consider providing supports to increase equity. Increasing participation opportunities for students in this quintile may result in higher levels of academic achievement goals and outcomes for these students, which in turn will better prepare them for undergraduate STEM programs and eventual careers.

Conclusion

The research conducted in this study provides an example of how secondary data sets can be used by faculty and students to address pressing national problems. The HSLS:09 data obtained from NCES allowed four distinctly different and meaningful research questions to be addressed with enough depth to provide guidance to policy makers and support ongoing research efforts. Further, these results (as captured in this paper) were not confined to the classroom, but shared back into the corpus of educational research.

Large data sets like HSLS:09 provide a veritable wealth of data that can answer a broad range of questions – were the combined analytical powers of the nations' education graduate student population applied to these data sets, the pace and significance of research, both original and conformational, could be vastly increased. Given the slow pace of educational research (limited by resources, researchers, and the growth and learning rates of participants), this kind of crowd-sourced research effort offers the opportunity to accelerate research efforts through parallel inquiries.

Furthermore, this study also prepared four graduate student researchers to conduct future quantitative research efforts, gave them hands-on experience in statistical analysis, and helped them to see the challenges and limitations of such studies. The benefits from this extra depth are two-fold: For those students going into future research, this experience was clearly valuably preparatory and helped establish them as published researchers. For those who intend to return to educational practice, it helped them to understand the role of research in their field, as well as how to evaluate and understand research findings so as to better apply them within their area of practice.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: Freeman.

Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. *Journal of Educational Computing Research, 31*(1), 1-36. doi: 10.2190/1N20-VV12-4RB5-33VA.

Bundy, A. (2007). Computational thinking is pervasive. *Journal of Scientific and Practical Computing, 1*(2), 67–69. doi:10.1155/2014/428080

Covay, E., & Carbonaro, W. (2010). After the bell: Participation in extracurricular activities, classroom behavior, and academic achievement. *Sociology of Education, 83(1)* 20-45. doi: 10.1177/0038040709356565

Fredericks, J. (2012). Extracurricular participation and academic outcomes: Testing the over-scheduling hypothesis. *Journal of Youth and Adolescence, 41*, 295-306 doi:101007/s10964-011-9704-0

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., Rogers, J. E., Jin, Y., & Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSLS:09). Base-year data file documentation (NCES 2011-328). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/surveys/hsls09/pdf/2011328_1.pdf.

Margolis, J., & Fisher, A. (2003). *Unlocking the clubhouse: Women in computing*. Cambridge, MA: The MIT Press.

National Center for Education Statistics (2014). High School Longitudinal Study of 2009 (HSLS:09). Retrieved from http://nces.ed.gov/surveys/hsls09/

National Science Foundation [NSF]. (2013). Women, minorities, and persons with disabilities in science and engineering: 2013. Retrieved from http://www.nsf.gov/statistics/wmpd/2013/pdf/nsf13304_digest.pdf

President's Council of Advisors on Science and Technology. (September 15, 2010). *Prepare and inspire: K-12 science, technology, engineering, and math (STEM) education for America's future*. Retrieved from http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports

Rice, L., Barth, J. M., Guadagno, R. E., Smith, G., & McCallum, D. M. (2013). The role of social support in students' perceived abilities and attitudes toward math and science. *Journal of Youth Adolescence, 42*, 1028-1040. doi: 10.1007/s10964-012-9801-8

Sitlington, P. L., Frank, A. R., & Carson, R. (1993). Adult adjustment among high school graduates with mild disabilities. *Exceptional Children, 59*, 222-223.

Trzesniewski, K., Donnellan, M.B., & Lucas, R.E. (Eds.). (2011). *Secondary data analysis: An introduction for psychologists.* Washington, DC: American Psychological Association.

VanBergeijk, E. (2012, September). NYIT: Providing access to higher education & employment: New York Institute of technology introduction to independence (I to I) and vocational independence programs (VIP). *The Exceptional Parent*, *42*(8), 27-30.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal Society. Series A, Mathematical, Physical, and Engineering Sciences*, *366*(1881), 3717–25. doi:10.1098/rsta.2008.0118.

Zweben, S. & Bizot, B. (2013). 2013 Taulbee Survey. *Computing Research News, 26*(5), 10-55.

Appendix

Sample Weights

HSLS:09 involved surveys of students, as well as surveys of school administrators, math and science teachers, counselors, and parents. As the base unit of analysis for HSLS:09 was the student, all additional surveys were tied back to individual students. In many cases not all students had a corresponding secondary survey. In order to ensure as much data would be available to the researcher as possible, different sample weights were published with HSLS:09 to correspond to the available survey data. These sample weights appear as extra variables in the data, and have different values for each student. When performing an analysis, the variable in question is multiplied by the weight, which adjusts its influence on the outcome.

For example, if the researcher wanted to ask a question drawing from the base year parent survey, they would apply the weight W1PARENT to their analysis. This weight would zero out the influence of any students whose parents had not filled out a survey, and adjust the influence of the remaining responses to be nationally representative. As many parents did not complete the survey, this necessarily reduces the sample size and statistical power of the analysis. This is why multiple weights are published – by choosing the right sample weight the researcher maximizes the statistical power of their analysis by including all relevant responses.

Similarly, each follow-up survey has a different set of weights. For HSLS:09 the base year survey weight was indicated by a 1 as the second character in the variable name, while the first follow-up survey uses a 2. Additionally, there are weights for longitudinal studies (those comparing students between base year and follow-up surveys), though these were not used by the researchers. The following list describes the sample weights used by the researchers:

Weight	**Description**
W1STUDENT	Base-year sample weight for the items on the student instrument
W1PARENT	Base-year sample weight for items on the parent or parent and student instrument
W2STUDENT	First follow-up sample weight for items on the student instrument

Variables

The HSLS:09 variables used by the researchers were:

Variable	**Description**
P1SPECIALED	Indicates 9th grader is receiving special education services, Based on parents' reports of their students special education services. Item wording is: Does [your 9th grader] currently receive Special Education Services? Students receiving these services often have an Individualized Education Plan (IEP). Yes No
S1HRACTIVITY	Hours student spends on extracurricular activities on the typical schoolday. Item wording is: During a typical weekday during the school year, how many hours do you spend… participating in extracurricular activities

	such as sports teams, clubs, band, student government? Less than 1 hour 1 to 2 hours 2 to 3 hours 3 to 4 hours 4 to 5 hours More than 5 hours
S2COMPAPP12	Indicates student is taking a computer applications course in Spring 2012. Item wording is: What science, computer science, or engineering course or courses are you currently taking? What science, computer Science courses were you taking during the Spring term of 2012? Computer Applications 0 = No 1 = Yes Question wording was customized depending on if the student indicated they were currently attending school. A legitimate skip was generated if the student indicated they weren't taking a science, computer science, or engineering course on a previous item.
S2COMPPROG12	Indicates student is taking a computer programming course in Spring 2012. Item wording is: What science, computer science, or engineering course or courses are you currently taking? What science, computer Science courses were you taking during the Spring term of 2012? Computer Programming 0 = No 1 = Yes Question wording was customized depending on if the student indicated they were currently attending school. A legitimate skip was generated if the student indicated they weren't taking a science, computer science, or engineering course on a previous item.
S2APCOMPSCI12	Indicates student is taking AP computer science in Spring 2012. Item wording is: What science, computer science, or engineering course or courses are you currently taking? What science, computer Science courses were you taking during the Spring term of 2012? Advanced Placement (AP) Computer Science 0 = No 1 = Yes Question wording was customized depending on if the student indicated they were currently attending school. A legitimate skip was

	generated if the student indicated they weren't taking a science, computer science, or engineering course on a previous item.
S2OTHCOMP12	Indicates student is taking other computer or information science course in Spring 2012. Item wording is: What science, computer science, or engineering course or courses are you currently taking? What science, computer Science courses were you taking during the Spring term of 2012? Other computer or information science course 0 = No 1 = Yes Question wording was customized depending on if the student indicated they were currently attending school. A legitimate skip was generated if the student indicated they weren't taking a science, computer science, or engineering course on a previous item.
X1IEPFLAG	Indicates if the student has an individualized education plan (IEP). Information is provided by the 9th grade enrollment lists or subsequent sampled student roster by school personnel.
X1MTHEFF	Scale of student's math self-efficacy; higher X1MTHEFF values represent higher math self-efficacy. Variable was created through principal components factor analysis (weighted by W1STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were Math tests, math textbooks, math skills and math assessment [S1MTESTS, S1MTEXTBOOK, S1MSKILLS, S1MASSEXCL]. Only respondents that provided a full set of responses were assigned a scale value, and legitimate skips were assigned if the student indicated they were not taking a fall math class. Coefficient of reliability for the scale (alpha) was .65.
X1SCIEFF	Scale of student's science self-efficacy; higher X1SCIEFF values represent higher science self-efficacy. Variable was created through principal components factor analysis (weighted by W1STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were Math tests, math textbooks, math skills and math assessment [S1MTESTS, S1MTEXTBOOK, S1MSKILLS, S1MASSEXCL]. Only respondents that provided a full set of responses were assigned a scale value, and legitimate skips were assigned if the student indicated they were not taking a fall science class. Coefficient of reliability for the scale (alpha) was .65.
X1SESQ1 X1SESQ2 X1SESQ3 X1SESQ4	Socioeconomic status coded by quintiles, Based on the following 5 component variables; parents highest level of education, education level of other parent, highest occupation prestige of parent, occupation prestige of other parent, and family income. For cases with

X1SESQ5	nonresponding parent/guardians, imputed values are generated.
X1SEX	Student's sex taken from the base-year student questionnaire, parent questionnaire, and/or school provided sampling roster. If any of these sources was inconsistent, X1SEX was coded based on manual review of student's first name.
X1STUEDEXPCT	The highest level of education the student expects to achieve at 9th grade. If missing from the student questionnaire, it is statistically imputed.
X1RACE	Student race/ethnicity composite determined by NCES based on 6 dichotomous variables (X1HISPANIC, X1WHITE, X1BLACK, X1ASIAN, X1PACISLE, X1AMINDIAN) collected through the student survey. If not present there, the value was based on (in order of preference) data from the school-provided sample roster or data drawn from the parent questionnaire.
X2SEX	Composite based on X1SEX, and when missing updated with data from the first follow-up student questionnaire.
X1STU30OCC02	2-digit Occupational Information Network (O*NET) code of the job the student expects to have at age 30. Students were asked to indicate what job they expected to have at age 30, and the textual responses were coded by NCES into O*NET codes.
X2RACE	Pulled from X1RACE in the base year survey, and if missing drawn from the first follow-up student questionnaire. If still missing, they are based on data from the first follow-up parent questionnaire