

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,800

Open access books available

142,000

International authors and editors

180M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Chapter

# The Power in Groups: Using Cluster Analysis to Critically Quantify Women's STEM Enrollment

*Ann M. Gansemer-Topf, Ulrike Genschel, Xuan Hien Nguyen, Jasmine Sourwine and Yuchen Wang*

## Abstract

Despite efforts to close the gender gap in science, technology, engineering, and math (STEM), disparities still exist, especially in math intensive STEM (MISTEM) majors. Females and males receive similar academic preparation and overall, perform similarly, yet females continue to enroll in STEM majors less frequently than men. In examining academic preparation, most research considers performance measures individually, ignoring the possible interrelationships between these measures. We address this problem by using hierarchical agglomerative clustering – a statistical technique which allows for identifying groups (i.e., clusters) of students who are similar in multiple factors. We first apply this technique to readily available institutional data to determine if we could identify distinct groups. Results illustrated that it was possible to identify nine unique groups. We then examined differences in STEM enrollment by group and by gender. We found that the proportion of females differed by group, and the gap between males and females also varied by group. Overall, males enrolled in STEM at a higher proportion than females and did so regardless of the strength of their academic preparation. Our results provide a novel yet feasible approach to examining gender differences in STEM enrollment in post-secondary education.

**Keywords:** STEM, gender, cluster analysis, female, enrollment

## 1. Introduction

Gender disparities in STEM enrollment in college continue to receive a significant amount of attention [1]. Research investigating reasons for this disparity in enrollment by gender highlights three consistent themes: a) females who are less academically prepared in math and sciences than males are less likely to enroll in STEM; b) despite comparable academic preparation, females, on average, enroll in STEM majors in smaller percentages than males, and c) females' participation varies by major, with some majors, such as biology seeing a higher proportion of females than in math intensive fields (MI STEM) such as engineering or computer science [2–5]. Despite the abundance of research on this topic, there lacks a more detailed yet

cohesive look at the interrelationships among high school academic preparation, ACT scores, and STEM enrollment.

This study utilized hierarchical agglomerative clustering to analyze high school and college data from a cohort ( $n = 3104$ ) of students at a large, public research institution in the Midwest. Students entering this institution tend to arrive with a broad academic background and varying levels of readiness to pursue a STEM degree. Students receive guidance from professional academic advising staff and faculty about course enrollment and academic trajectories. Frequently these decisions are made by considering one or two individual data points such as a student's GPA or ACT score, yet research in this area consistently demonstrates that it is a combination of several of these factors that more accurately represents students' preparedness and ability to succeed [6–8]. By clustering, our data analysis will provide more nuanced insights than traditional statistical analyses. Traditional analyses tend to focus on averages and distributions of individual student characteristics across all students. Clustering algorithms seek to identify natural groupings (clusters) of students such that students within a cluster are academically more similar to each other regarding their pre-collegiate training than they are to students from any other cluster. Thus, clustering better accounts for the interrelationships of several factors and offers much more robust information than the more common approach to examining individual variables.

As it relates to our study, we were interested in the potential for this technique to provide a more in-depth understanding of female students' enrollment in STEM based on several factors of academic preparation. This approach has been used to study gender inequality in the STEM workforce (see, for example, [9]); we apply a similar method to examine gender disparities in STEM choice.

We begin by investigating the potential of identifying unique groups of students using hierarchical agglomerative clustering. We then examine enrollment in STEM by gender and clusters. The following questions guide our analyses:

1. What are pre-collegiate grades, academic rank, and academic mathematical and science courses, of males and females in NonSTEM, math-intensive STEM (MISTEM), and other STEM (OSTEM)?
2. Can hierarchical agglomerative clustering meaningfully identify unique groups of students based on precollege student characteristics? If yes, do gender differences in academic background exist?
3. Given cluster membership, are there gender differences in enrollment into STEM, specifically into MISTEM and OSTEM majors?

The results of our study have implications for secondary and postsecondary education. A more robust understanding of the interrelationships among variables that contribute to enrollment in STEM areas can be used to develop strategies that can enhance STEM enrollment. Postsecondary institutions can gain a better and more sophisticated understanding of the academic readiness of a cohort of incoming students. This information can be used strategically by institutions and departments to tailor course offerings to the immediate academic needs of students allowing for possibly critical adjustments in the present time. This approach can also be used to better understand not only enrollment, but also retention and completion of STEM degrees. Further, using the clusters to measure the pursuit of STEM degrees, retention, and

completion of STEM degrees permits a novel use of clusters as a statistical predictive model in STEM education. Our data analysis approach is innovative and differs from traditional statistical analyses by using hierarchical agglomerative clustering that allows students with similar characteristics to be grouped together. Through the use of clustering, this study more closely examines gender differences in academic preparation, STEM interest, and enrollment in STEM.

## 2. Literature review

Females have been narrowing the gaps in math and science achievement and have seen more participation in STEM enrollment and careers [10]. However, the rate of participation is still drastically disproportionate, with only 27% of all STEM careers being occupied by women [11]. These statistics also vary based on major. For example, women represent over half of all bachelor's degree recipients in biology but are significantly underrepresented in physical sciences, engineering, mathematics, and computer sciences (MISTEM) [12–14]. Even at selective institutions with a large pool of interested students willing to enroll, females represent between 15 and 28% for Bachelor's degrees and only 13–20% for Ph.D.s in math departments [15]. We drew upon past literature to better understand differences in STEM and guide our selection of variables.

### 2.1 Pre-collegiate preparation and STEM enrollment

According to Weeden, Gelbgiser, and Morgan [16], between 19% and 32% of the gender gap for STEM degree completion can be attributed to the gender gap in STEM career interest in high school. Additionally, only 13% of female high school graduates expressed an intent to pursue a STEM career compared to 26% of their male counterparts [16]. This gap in interest as early as high school indicates that the STEM career gap is not solely caused by attrition in college or women exiting STEM careers post-graduation alone.

In high school, females consistently outperform males in their core classes, including math and science [17–19]. Despite earning high grades, females do not perform quite as well on high stakes standardized tests in math and science, scoring 0.7, 0.2, and 0.4 fewer points than males on the math, science, and STEM portions of the ACTs, respectively [20]. In spite of earning better grades in math and science, course selection among males and females shows some discrepancies. For example, female graduates took roughly the same number of advanced math courses as their male counterparts, experiencing an overrepresentation only in Algebra II. Science, however, shows more variations, with ten percent more females taking Advanced Biology, about six percent more females taking Chemistry, about 6 percent more males taking Physics, and males enrolling in engineering courses five times higher than females [21]. Furthermore, correlational research supports that participation in Advanced Placement (AP) STEM courses and STEM career interest are associated [22] and that students with high math abilities and exposure to rigorous courses were more likely to enroll in STEM majors [23].

GPA is positively correlated with the pursuit of a STEM major [24], and it is a better predictor for college success than the ACTs [25]. The rigor of math and science courses is a better predictor for enrollment in a STEM major than the number of courses alone [26]. Class rank, on the other hand, is more complex. When ranks are

calculated by subject (i.e., math and reading) and communicated to students, ranking can have statistically significant effects on students' career choices. For example, a study conducted in Ireland found that students ranked highly in math had a positive association with STEM career choice and a negative association with careers in the arts and social sciences, while those who were highly ranked in English had a positive association with arts and social sciences and a negative association with STEM careers [27]. A study performed in Florida found that high school class rank and GPA, which are higher for females, were the best predictors of collegiate GPA and the number of credits earned in college [28]. But, as detailed above, males and females experience similar pre-collegiate STEM preparation in many respects with small differences in math, science, and STEM test scores and some discrepancy in enrollment of advanced science courses. Despite these similarities, males are twice as likely to intend to declare a STEM major than females. A closer look at STEM enrollment is required.

## **2.2 STEM enrollment**

In high school, females are less likely to be interested in STEM and more likely to lose interest over time [29]. Controlling for math achievement and aptitude, females are still less likely than males to be interested in STEM [30]. In fact, one of the best predictors for enrollment and persistence through a STEM major is an individual's desire to pursue a STEM career in high school, with those expressing interest in high school completing degrees at three times the rate of those who do not express this interest in high school [31]. Among females who intend to major in a STEM field in college, nearly half of them switch majors to non-STEM fields compared with only a third of males [21].

That is not to say women are not enrolling in STEM majors; in fact, women earned 53% of STEM degrees (short of their 58% share of all degrees that would be proportional to their overall makeup of the workforce) [32]. However, there are significant disparities among the types of STEM degrees women choose to pursue. For example, women are overrepresented in health-related STEM careers with, 85% of Bachelor's degrees being awarded to women, but they are awarded less than 45% in Mathematics and Physical Science and less than 25% in Engineering and Computer Science [32]. Because interest in STEM begins before students enroll in postsecondary education and gender gaps in STEM still persist, we consider new ways to understand reasons for these gaps based on high school preparation.

## **3. Methods**

### **3.1 Data source and sample**

Data for this study were provided by the institution's Office of the Registrar at one large public, research-intensive institution. To ensure our research conformed to standards and guidelines of ethical research practice, we received approval from the Institutional Review Board at the study's institution. Per written agreement with the Office of the Registrar, all students enrolled in an introductory level mathematics or statistics course in the Spring 12, Fall 12, or Spring 13 semester were eligible for the parent study. Students were given the opportunity to opt out of the study, and of 16,401 eligible students, 32 chose to opt out. We focused on these courses as they often serve as gatekeeper courses to a STEM degree.



Because we were interested in the relationship between high school preparation and enrollment in a STEM major, we focused only on first-semester, degree-seeking students who entered the institution directly from high school and were enrolled in an introductory level math course. We excluded students who transferred from another post-secondary institution because we did not have access to pre-college data for many of these students. Students who were classified as non-degree-seeking or international were also removed prior to the analysis as these students are likely to differ in their academic background or degree goal. Of the initial 3219 eligible students, we had complete data on the variables of primary interest for 3104 students.

Using the STEM Designated Degree Program List (2012 revised list) provided by the Department of Homeland Security, we categorized students into STEM and NonSTEM majors. We further split STEM majors into math-intensive STEM (MISTEM) and Other STEM (OSTEM) majors. For the purpose of this study, a STEM major was considered math-intensive if it required at least one semester of science or engineering calculus. This definition is similar to the definitions used by Ceci and Williams [33] and by Bressoud [34]. This differentiation served distinct purposes: (a) the gender gap has historically been more pronounced in MISTEM majors such as engineering, computer science, or physics, whereas fields like biology or chemistry (OSTEM) have increasingly grown the proportion of women to the extent that women are now in the majority of degree earners [35]; (b) definitions of STEM vary greatly and can range from more inclusive by considering fields such as psychology, dietetics majors or kinesiology as STEM fields to less inclusive lists, which consider mainly engineering, mathematics, physics, natural sciences, and computer sciences. Distinguishing MISTEM majors define majors represented in most STEM field definitions.

Our analysis included the following variables: gender, major, high school rank (HS Rank), grade point average (GPA), number of high school credits earned in mathematics courses, including algebra, geometry, trigonometry, and calculus, and credits earned in biology, chemistry, and physics, ACT composite, ACT English, and ACT Math scores. The ACT is a national standardized test commonly used in college admissions decisions.

### **3.2 Statistical methodology**

Using student demographic characteristics and pre-college academic background variables, we conducted an agglomerative hierarchical cluster analysis. "Cluster analysis is a data-mining technique that allows researchers to cluster a set of observations into similar (homogeneous) groupings based on a set of features" [36]. It accounts for the different high school experiences and preparation with which students enter their first year of college and can provide a more complex description of students than a comparison on a variable-by-variable basis. Clustering students reduces the focus on mean comparisons, which captures a population's average behavior, but less on how students compare at the individual level. Although cluster analysis has been used in a variety of academic settings, its use to investigate female enrollment discrepancies in STEM vs. non-STEM fields is novel. Cluster analysis has been used to develop classroom observation tools [36], reveal different learner profiles based on motivation, achievement, needs satisfaction, etc. [37], and the differences between females and males who succeed within higher technical education [38]. Using similar methods as these studies, we will compare clusters of similar students to determine any trends among factors such as gender, preparation, and STEM enrollment.

### **3.3 Data analysis**

All statistical analyses were run in SAS/STAT software, Version 9.4 of the SAS system and RStudio Version 1.3.1073. To address the first research question, we calculated the means and standard deviations of each of the variables used in the cluster analysis by type of major (NonSTEM, MISTEM, OSTEM) and by gender. To address the second research question, we utilized PROC Cluster and Ward's minimum-variance method [39]. Ward's minimum-variance method is based on the total error sum of squares that arises by grouping observations into distinct clusters where the total sum of squares corresponds to the sum of the within-cluster sum of squares [40]. Merging a set of observations into a cluster can be considered a loss of information. Ward's method seeks to minimize the loss of information from merging any two clusters at a given step in the clustering algorithm. That is, the two clusters whose merging will lead to the smallest increase in the total error sum of squares will be combined into a new cluster [40]. Initially, each student represents a single cluster. At each step of the algorithm, two existing clusters are merged until only one cluster remains. The number of clusters is unknown prior to the analysis and an appropriate cluster solution is typically based on a set of clustering criteria such as the cubic clustering criterion (CCC), the Pseudo-F, and Pseudo T2 statistic [41, 42].

In order to see if the gender disparity in STEM enrollment is associated with gender or merely high school preparation, we clustered students according to their high school science, mathematics, and standardized test score data. Taking calculus in high school is a strong predictor of STEM interest and success [43]. Therefore, we separated students into two groups prior to clustering: students with calculus in high school (Calc group) and students without (NonCalc group). We then ran a separate cluster analysis based on high school rank, ACT English, ACT Math, and the sum of high school science credits in biology, physics, and chemistry. We decided on an initial number of clusters in each group based on the CCC, Pseudo-F, and Pseudo-T2 clustering criteria. To address the final research question, we examined the proportion of males and females in each cluster that chose a major in NonSTEM and STEM. We then limited our sample only to those who chose STEM and examined the proportion of males and females in each cluster who chose MISTEM or OSTEM.

### **3.4 Limitations**

We wish to acknowledge some methodological limitations. To be included in the sample, a student had to take a mathematics or statistics course during their first semester in college. We are therefore missing students who may have transferred credits into college or postponed taking a mathematics or statistics course in their first semester. Our sample also only included students who chose to major in STEM in their first semester. Additional research that examines students who may decide to major in STEM after their first semester would provide additional insights.

A cluster analysis using different variables would likely result in different clusters. For example, if we were to treat the numbers of biology credits, chemistry credits, and physics credits as separable variables rather than use their sum, clusters would likely form around differences between students with respect to the individual variables such as students with many biology credits versus students with few biology credits. We chose the sum of all science credits for two reasons. From a methodology point of view, the sum of science credits is more preferable as a variable because it has

a greater range of values. Secondly, the choice of variables depends on the characteristics deemed meaningful to identify differences between students.

Additionally, we wish to acknowledge the limitations and ethical considerations of using quantitatively techniques to group students and subsequent interpretations of these efforts. Quantitatively analysis affords an opportunity to see patterns that may otherwise be unclear, yet this approach can also minimize nuances within clusters and overlook significant implications of variables that were not included. For example, our study focused on gender but did not consider variables such as socioeconomic status, nationality or race, or secondary school quality. Our results also should not be used to imply causality or judgment [44, 45]— we seek to understand possible associations between variables but cannot conclude that one set of patterns causes a specific outcome or that one is qualitatively better than others.

## 4. Results

### 4.1 Research question 1. what are advanced placement scores, pre-collegiate grades, academic rank, and academic mathematics and science courses, of males and females in NonSTEM, math-intensive stem (MISTEM), and other stem (OSTEM)

Across all fields, NonSTEM, MISTEM, and OSTEM, female students are equally prepared as men in the mathematics and sciences courses (see **Table 1**). Females have consistently higher high school ranks and GPA scores compared to their male peers, which is consistent with the results of the American Association of University Women Educational Foundation [17], Degol et al., [18], and Voyer and Voyer [19]. Females who enroll in MISTEM also score on average as well as their male peers and slightly outperform them on the English ACT placement test. Men enrolling in NonSTEM and OSTEM majors show a slight advantage on the Math ACT placement test, which is consistent with prior research [20].

### 4.2 Research question 2: can hierarchical agglomerative clustering meaningfully identify unique groups of students based on pre-college student characteristics? If yes, do gender differences in academic background exist?

Based on the clustering criteria, four or five clusters were reasonable choices for students with and without calculus. To arrive at the most meaningful number of clusters for each group, we plotted each clustering variable using side-by-side boxplots (see **Figure 1**). Each boxplot shows the distribution of the variables for all students in the respective cluster, while the horizontal line inside the box represents the median value observed for these students. We based our decision on the final number of clusters for each group of students on what we considered to be meaningful differences in the distribution and median value for each cluster in the context of our research questions [46]. Due to the agglomerative nature of the clustering procedure, a solution consisting of four clusters arises from the merging of the two closest clusters in the solution consisting of five clusters while the remaining clusters remain unchanged. Thus, we will decide, for example, on four clusters if merging the two closest clusters in the five-cluster solution does not result in a sufficiently large loss of information but changing from four to three clusters would. In essence, we



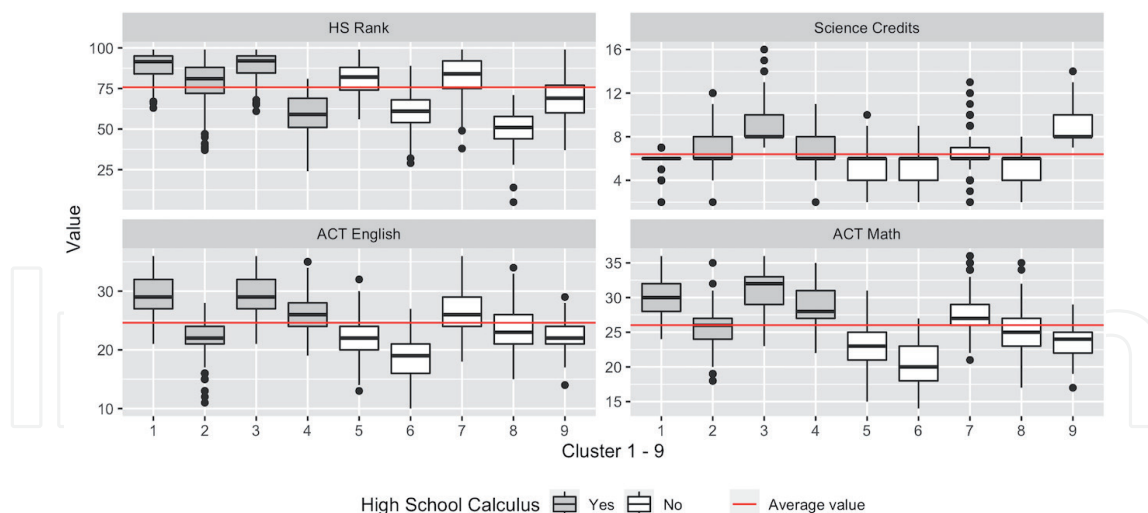
Variable	NonSTEM NF=537, N <sub>M</sub> =542		MISTEM NF=273, N <sub>M</sub> =1385		OSTEM NF=216, N <sub>M</sub> =151	
	Female	Male	Female	Male	Female	Male
High School						
GPA	3.62 (0.36)	3.38 (0.42)	3.81 (0.38)	3.63 (0.4)	3.65 (0.34)	3.5 (0.4)
Rank	77 (14.89)	68.1 (16.23)	83.9 (14.36)	76.6 (16.17)	77.9 (14.46)	72.3 (17.17)
Calculus Cr	0.4 (0.8)	0.5 (0.91)	1.3 (1.17)	1.2 (1.18)	0.6 (0.98)	0.5 (0.89)
Algebra Cr	4 (0.26)	4 (0.28)	4 (0.35)	4 (0.21)	4 (0.26)	4 (0.27)
Geometry Cr	2.6 (0.51)	2.6 (0.5)	2.7 (0.46)	2.8 (0.45)	2.6 (0.49)	2.6 (0.51)
Trigonometry Cr	1 (0.68)	1 (0.66)	1.2 (0.57)	1.2 (0.6)	1 (0.63)	1 (0.67)
Statistics Cr	0.3 (0.64)	0.4 (0.78)	0.4 (0.77)	0.3 (0.69)	0.3 (0.67)	0.2 (0.57)
Adv. Math Cr	1.3 (1.05)	1.2 (0.99)	1.8 (0.95)	1.7 (0.90)	1.3 (1.31)	1.3 (1.12)
Physics Cr	1 (1.03)	1.3 (1.05)	1.9 (1.09)	2 (1.05)	1.1 (0.99)	1.5 (1.01)
Biology Cr	2.7 (1.07)	2.5 (0.95)	2.6 (1.19)	2.4 (0.91)	3.1 (1.36)	2.8 (1.32)
Chemistry Cr	2.1 (0.63)	2 (0.71)	2.4 (1.03)	2.3 (0.95)	2.2 (0.69)	2.1 (0.94)
Science Cr	5.8 (1.6)	5.9 (1.47)	6.9 (2.01)	6.7 (1.66)	6.4 (1.83)	6.4 (1.92)
ACT						
Math	23.3 (3.71)	24.3 (3.68)	27.3 (4.27)	28.1 (3.89)	23.3 (4.1)	24.8 (4.12)
English	24 (4.56)	22.9 (4.28)	26.7 (5.06)	25.4 (4.6)	23.6 (4.57)	23.4 (4.8)
Composite	23.8 (3.48)	24.1 (3.47)	26.9 (4.01)	26.9 (3.63)	23.9 (3.77)	24.6 (3.62)

**Table 1.** High School mathematical and science background of incoming students by enrollment into MISTEM, OSTEM and NonSTEM. All values are rounded to the nearest decimal place.

are looking for a solution that is both inclusive and parsimonious. For this reason, we included the three- and six-cluster solutions in the decision-making process. For simplicity, we discuss the different cluster solutions in terms of one cluster being split as opposed to two clusters being merged.

#### 4.2.1 Clustering for students with calculus

We begin with the interpretation of the three-cluster solution and describe which of the three clusters is divided in the transition from three clusters to four.



**Figure 1.**  
*Finalized cluster solution.*

For students with calculus ( $N = 1280$ ), a three-cluster solution identifies three types of students. A first cluster consists of 627 students with noticeably lower HS ranks, ACT English and ACT Math scores compared to the students in the remaining two clusters. Among the latter, the more academically prepared students, two groups emerge; a cluster with students who have substantially more science credits and a slightly better but noticeable ACT Math score ( $N = 335$ ) compared to the students in the second group ( $N = 318$ ). Both groups have comparable ACT English scores and HS Ranks.

The four-cluster solution arises from splitting the lower performing group of 627 students into two distinct groups based on differences in ACT scores and HS Ranks. Students in both groups have the lowest HS ranks out of all calculus students but the first group ( $N = 213$ ) performs substantially better on the ACT English and ACT Math exams than the second group ( $N = 414$ ). We deemed this split meaningful. A plausible interpretation could be that the second group ( $N = 414$ ) consists of students who tend to underperform on standardized tests for a variety of reasons. Alternatively, the smaller group of students ( $N = 213$ ) excels on standardized exams relative to their overall high school performance as reflected by high school rank. Due to the hierarchical nature of the clustering algorithm, the other two clusters remained unchanged.

In a five-cluster solution, the group of 414 students is broken up into students with slightly higher ACT scores, noticeably more science credits, and better HS ranks ( $N = 99$ ) compared to the second group ( $N = 315$ ). Although this distinction might be relevant, it was not relevant for our research questions, and we decided against this additional split; both groups of students maintained their relative, below-average performance on the ACT exams. Consequently, we determined four clusters to be the most appropriate number of clusters for students with calculus. For this reason, we did not consider the six-cluster solution.

#### 4.2.2 Clustering students without calculus

For students without calculus ( $N = 1284$ ), a three-cluster solution distinguishes three groups. Students in group 1 ( $N = 663$ ) have overall the lowest high school ranks but perform otherwise similar to students in a second group ( $N = 423$ ) with respect to the number of science credits taken in high school and scores on the ACT exams.

The third group (N = 738) outperforms the first and second group in the number of science credits and on the ACT components but shares HS ranks similar to those in the group of 423 students.

In a four-cluster solution, the third group of 738 students is divided into two separate groups. The first group (N = 603) includes the best students with respect to HS rank and ACT English and Math scores, but students in this cluster tend to have taken fewer science credits. The remaining students (N = 135) have the most science credits among all students without calculus. Their ACT scores and HS ranks are, however, much lower compared to the top students, and they also tend to have lower HS ranks compared to the second group (N = 423) in the three-cluster solution.

To move from four clusters to five the group of students with below average ACT scores and lower HS ranks (N = 663) are divided. This split reflects the same pattern we saw in the group of students with calculus. The 663 students are separated into a group of students (N = 258) who have overall the low high school ranks but who score better on the ACT English and Math placement tests and a second group, whose students have higher high school ranks but do not perform below average on the placement tests. Because we already saw a similar distinction among the calculus students, we consider the five-cluster solution meaningful and retain it over the four-cluster solution.

The six-cluster solution focused on the cluster previously consisting of the overall most prepared students (N = 603). These students are broken up into two clusters. The smaller of the two clusters retains the best students overall (N = 228), while students in the second cluster (N = 375) perform slightly worse than the top students, they still do better than the cluster of N = 423 students. Because both groups still outperform any of the remaining clusters overall, we decided against this additional

Cluster	Name	N	Cluster description
1	Calc. Strong, Less Science	318	Highest ACT English and HS Rank, Second highest ACT Math scores, below average (less than 6) science credits
2	Calc, Average	414	Typically, about average, showing slightly above average HS Rank and science credits but tend to fall short of average ACT English and Math scores
3	Calc, Strong Overall	335	Best students overall with all students having taken above average number of science credits
4	Calc, Low HS Rank	213	Students with far below average HS Ranks that have above average ACT English and Math scores
5	No Calc, Average	423	Above average HS Rank, very few science credits, below average ACT English and Math scores
6	No Calc, Low ACTs	405	About 50% of students have HS rank 1 standard deviation below average, lowest ACT English and Math scores
7	No Calc, Strong Overall	603	Strongest performers out of all non-calculus groups but students do not score as high as two top calculus clusters
8	No Calc, Low HS Rank	258	Lowest HS Rank, below average science credits but almost all students have ACT English and Math scores within 1 standard deviation below average
9	No Calc, Average, More Science	135	Worst performers on ACT English and Math, low HS Rank but take many science credits, credits comparable to the top cluster (Cluster 3)

**Table 2.**  
*Description of clusters and respective sample sizes.*

split and retain the 5-cluster solution as the final number of clusters. A description of the finalized cluster solution is given in **Table 2**.

#### 4.2.3 Gender distribution across clusters

Our analysis revealed that we could use clustering to find meaningful differences in groups. We then examined the proportion of female students in each cluster. Out of 3104 students, 1026 are female representing 33% of the students in the sample. Assuming that there are no systematic differences in the academic background between females and males, we expect to see about 33% of each cluster to be female students. **Figure 2** shows a visualization of the proportion of females in each cluster. In Cluster 7 (No Calculus, Low ACTs), the proportion is close to the target value of 33% with 34.3%; females are slightly below 33% in Clusters 1–3 (Calculus, Strong Less Science – Strong Overall). On the other hand, females are overrepresented in three out of the five non-calculus clusters and underrepresented in Clusters 4 (Calculus, low HS rank) and 8 (No Calculus, low HS rank). Although these students had lower HS ranks they still scored well on the ACT English and Math tests relative to students in clusters that proportionally contain more female students (Clusters 5, 6, and 9). Students in Cluster 8 (No Calculus, Strong Overall) are very similar to students in Clusters 1–3 when it comes to high school rank and performance on the ACT, except they did not have calculus in high school. The proportion of female students in a calculus cluster being average or below average shows that proportionally, fewer female students take calculus in high school (33%) than their male peers (45%).

#### 4.3 Research question 3. Given cluster membership, are there gender differences in enrollment into STEM, specifically into MISTEM, and OSTEM majors?

In the overall sample, 48% of females chose to major in STEM and 74% of males chose to major in STEM. Of those who majored in STEM, 90% of males and 56% of females enrolled in MISTEM.

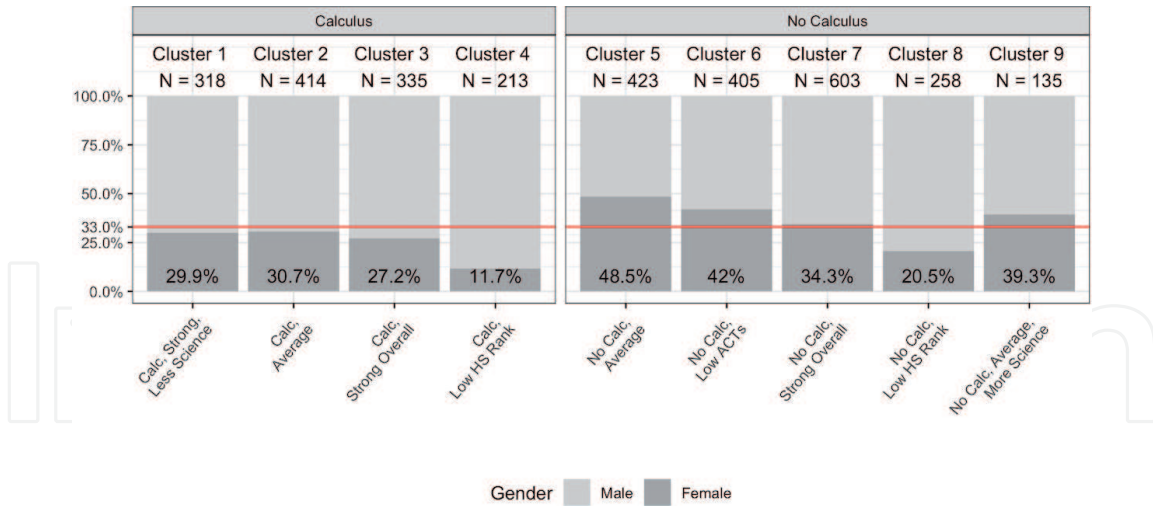
Using the cluster solution identified in **Table 2**, we examine the proportion of female students by cluster. As mentioned previously, if differences in enrollment by gender are within natural variation, we can expect about one-third of the students to be female in each cluster.

Even though the proportion of females in Clusters 1–3 and 7 are similar and close to average (see **Figure 2**), the enrollment in STEM majors is strikingly different. For example, females with a calculus background are consistently more likely than females without a calculus background to choose a STEM major. This is evident when comparing Cluster 7 to Clusters 1–4: female students with no calculus background are less likely than male students to enter STEM.

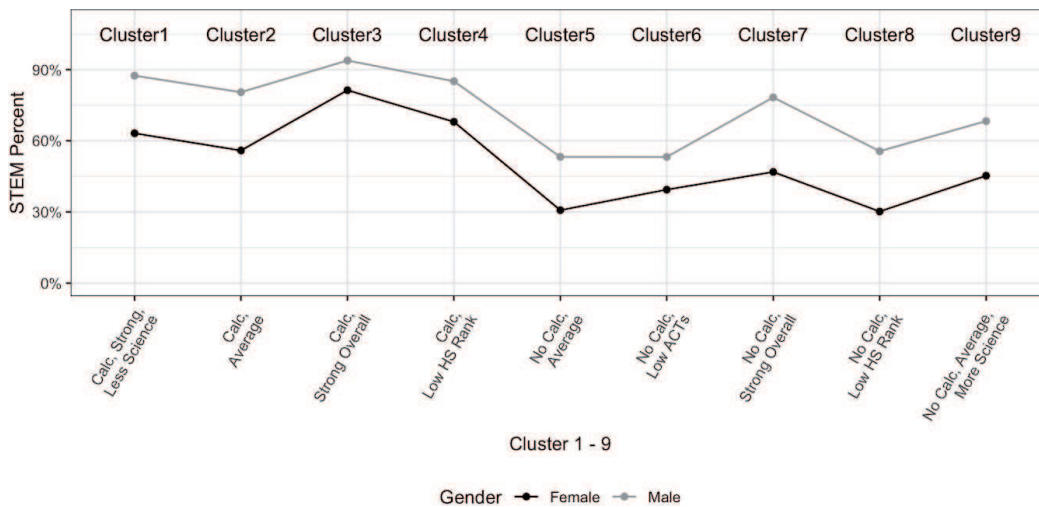
Within the same cluster, thus with similar academic background, a smaller proportion of female students enroll in STEM than male students (see **Figure 3**). When we compare across gender and clusters, we see differences in this gap. For example, for Cluster 3, there is less than a 13% difference between males and females (94% vs. 81%); however, in Cluster 7, this gap increases to 31%.

The lower enrollment rate for females is especially evident for students in the NoCalc, Strong Overall cluster. We mentioned before that this NoCalc cluster is similar in background to the calculus clusters, the only difference being that the students did not have calculus in high school. Males in this cluster, however, enroll in STEM at a similar proportion to their female peers who did have calculus.





**Figure 2.**  
Distribution of gender across clusters.

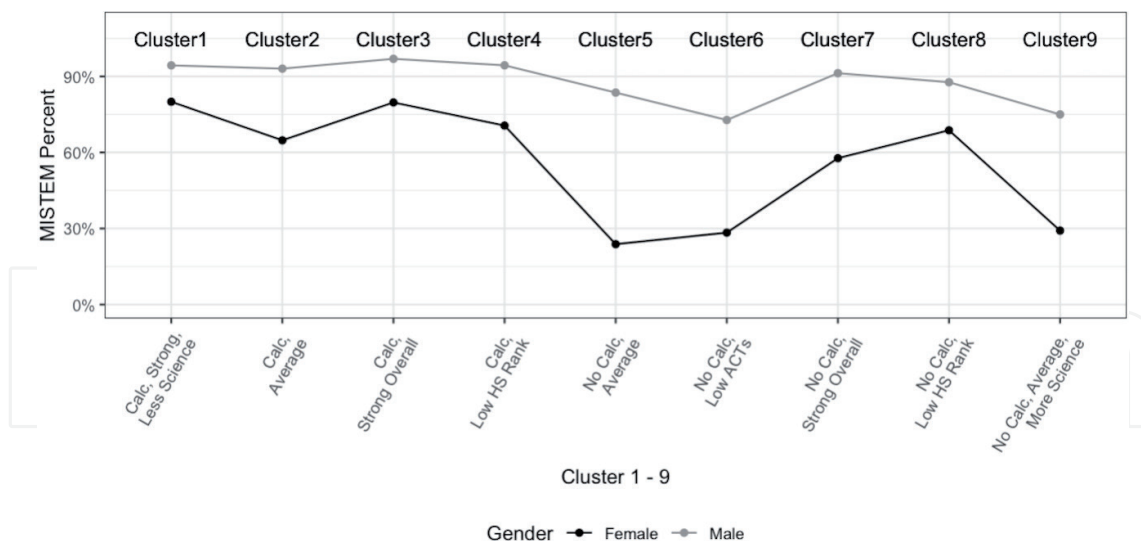


**Figure 3.**  
Enrollment into STEM by gender.

### 4.3.1 MISTEM and OSTEM

We then restricted ourselves to students enrolled in STEM in each cluster. Among those, **Figure 4** shows that a higher proportion of male STEM students choose MISTEM than female STEM students. The trend is very apparent in the NoCalc clusters but is also present in the calculus clusters. Another way of saying this is that more female STEM students choose OSTEM than male students, especially the NoCalc Students. There are two interesting clusters to contrast: the NoCalc, Strong Overall and NoCalc, Average More Science. They have the same proportion of females enrolling into STEM; however, a much higher proportion of females in NoCalc, Strong Overall choose MISTEM than the NoCalc, Average More Science cluster.

Most male students in STEM, independently of their cluster membership, chose to go into MISTEM majors. Female students with a calculus background are more likely to go into MISTEM than female students without calculus in high school. The percentage of females in MISTEM is high for the NoCalc, Low ACT cluster, but because we



**Figure 4.**  
*Proportion of STEM students enrolled in MISTEM by gender.*

limited ourselves to females who chose STEM within the cluster and the females are underrepresented in this cluster, we find an even smaller proportion of females chose STEM.

## 5. Discussion and implications

Despite significant efforts to minimize the gender gap in STEM, differences in interest and enrollment between men and women still exist. Ensuring a globally skilled workforce that meets the needs of the 21st century requires a post-secondary education in STEM fields, yet the interest to pursue STEM begins prior to enrolling in college. Therefore, efforts examining pre-collegiate preparation and STEM enrollment are critical.

We first examined the individual pre-collegiate variables of males and females. Similar to other research in this area [17–19], we found that overall, females and males have similar preparation although males are more likely to take calculus than females [16]. Examining these individual variables may lead to an assumption that these similarities in preparation are consistent across individuals and groups. However, by employing a more advanced statistical technique such as cluster analysis, we notice that when accounting for several pre-collegiate factors simultaneously uncovers (or reveals) marked differences in enrollment patterns within and across gender. For example, if we were to investigate just one variable, such as ACT scores, we get a different picture than when we combine variables such as standardized test scores with rank and/or GPA. For example, in our findings the combination of other factors such as enrollment in science courses, standardized test scores and high school rank results in variations in enrollment in STEM. Relatedly, the grouping of variables reveals differences in STEM participation more broadly and in the types of STEM, i.e., MISTEM and OSTEM.

Females consisted of 33% of our overall sample, yet, their representation in each cluster varies from 12% in Cluster 4 to 49% in Cluster 5 (**Figure 2**). Females never reached 33% in any of the calculus clusters. Enrollment in STEM and type of STEM (MISTEM, OSTEM) also varies by cluster. The results in **Figure 3** add/continue the

pattern of female underrepresentation in STEM. Although males and females follow overall a similar trend across clusters, female enrollment in STEM is consistently below male enrollment, confirming previous results that females chose STEM at lower rates than their male peers even when they are equally prepared [3]. The consistency across clusters also confirms that this holds true for all levels of preparedness in terms of academic high school background. Patterns for males and females, however, were not entirely parallel and equal. While Cluster 3 shows the highest, and Clusters 5 and 8 the lowest enrollment in STEM for both genders differences can be found in Clusters 6 and 7. For men there is little difference between Clusters 5 and 6 and larger differences between Clusters 6 and 7. For females, these percentages increase steadily from Cluster 5 to Cluster 7. Cluster 7 tells us that, despite being strong students, females who do not have Calculus in high school are much less likely to choose STEM than male students with the same background.

Further differences exist in enrollment by type of STEM. **Figure 4** shows percentages for enrollment into MISTEM above 60% for both genders in all calculus clusters. But females in non-calculus clusters show much more variation. Although females in Clusters 5 and 8 are equally unlikely to enroll into STEM having the lowest enrollment rate overall, there is a clear difference in the rate at which females in Cluster 8 enroll into MISTEM compared to Cluster 5. Interestingly, females in Cluster 8 enroll into MISTEM in rates comparable to females in any of the calculus clusters. Female students in Clusters 5, 6 and 9 are more likely to enroll into OSTEM than MISTEM while 75% or more of male students choosing STEM still enroll into MISTEM in those clusters. Overall, male students in STEM overwhelmingly choose MISTEM ranging between 73% in Cluster 6 to 94% in Cluster 8.

Our analysis limits our ability to understand why these differences occur. However, past research may lend some insights. For example, females are more likely to possess both high verbal and high math skills, whereas males are more likely to possess solely high math skills [47]. Due to the discrepancy between math and verbal skills, males tend to choose STEM careers, whereas females, who have a choice between verbal and math-centric careers, tend to choose non-STEM-related fields, opting instead for challenging fields that are more applied and practical rather than theoretical [46]. Of course, there are also work and lifestyle factors to consider; women are looking for work-family balance and value it more highly than men [48]. In addition to lifestyle values, there are also differences between social and moral career preferences with women tending toward occupations with a social, community, or altruistic component and men tending toward careers that require working with objects [49].

Although our study cannot account for the reasons differences exist, the results have implications for research and practice. Our study illustrates a method that can be adopted by institutional leaders for use on their own campuses. Although this study was limited to one institution, we utilized commonly collected pre-collegiate data. Because of the availability of this data on most campuses, this study can be replicated in a variety of campus contexts. Institutions vary in their enrollment criteria and student populations; cluster analysis techniques afford the ability to select relevant variables and determine if unique groups emerge. Researchers have noted the importance of variables such as students' race, ethnicity, and nationality and non-cognitive variables such as self-efficacy [50]. Future research could incorporate these additional variables.

Identifying these clusters is an important first step in a more comprehensive understanding of STEM interest and success. Once established, future efforts could examine the persistence and graduation rates of students in these clusters. The

research on the relationship between individual measures of academic preparation and persistence and graduation in STEM has produced mixed results. Examining the combination of these measures through cluster analysis would lead to a more nuanced understanding of the role of academic preparation in STEM enrollment. For example, if there was a consistent relationship found between completing calculus and graduation in STEM, regardless of other factors, the availability and enrollment in calculus courses in high school should be encouraged.

Qualitative research methodologies could help address the questions of “why?” Individual interviews or focus groups with students in each cluster could be conducted to understand student choices in academic preparation or what aspects of their academic preparation contributed to their enrollment and success in STEM.

## **6. Conclusion**

Minimizing the gender gap in STEM fields continues to be necessary to meet the needs of the global workforce. Academic preparation prior to enrolling in a post-secondary institution influences students' intent to pursue STEM; yet research efforts that investigate this relationship often are limited by focusing on individual variables. Our study uses an advanced statistical technique - hierarchical agglomerative clustering - that considers multiple factors simultaneously. This technique groups students into distinct categories based on a combination of academic preparation measures and by doing so, paints a different picture of the relationship between academic preparation and STEM enrollment than simply examining individual variables. Subsequently, these inconsistencies reaffirm that narrowing the gap requires a multi-faceted approach that consider academic preparation and non-cognitive factors. In addition to its research significance, there are valuable practical implications from this work. We demonstrate how this technique can be applied to institutional data; thus, we provide a valuable tool that can be utilized in postsecondary institutions for postsecondary leaders to utilize in understanding enrollment patterns within their institutions. Summarily, our study contributes to both research and practice through its use of a robust yet accessible technique that can be widely applied to quantitative data to uncover unique patterns largely overlooked by other approaches.

## **Acknowledgements**

This material is based upon work supported by the National Science Foundation under Grant No. HRD 1036791.

## **Conflict of interest**

The authors declare no conflict of interest.



IntechOpen

### **Author details**

Ann M. Gansemer-Topf<sup>1\*</sup>, Ulrike Genschel<sup>2</sup>, Xuan Hien Nguyen<sup>3</sup>, Jasmine Sourwine<sup>1</sup>  
and Yuchen Wang<sup>2</sup>

1 School of Education, Iowa State University, USA


2 Department of Statistics, Iowa State University, USA

3 Department of Mathematics, Iowa State University, USA

\*Address all correspondence to: [anngt@iastate.edu](mailto:anngt@iastate.edu)

### **IntechOpen**

---

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] National Science Board. Science and Engineering Indicators 2020 [Internet]. 2020. Available from: <https://nces.nsf.gov/pubs/nsb20201>
- [2] Crisp G, Nora A, Taggart A. Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a STEM degree: An analysis of students attending a Hispanic serving institution. *American Educational Research Journal*. 2009;**46**(4):924-942. DOI: 10.3102%2F0002831209349460
- [3] National Center for Education Statistics. Digest of Education Statistics, Table 326.30 [Internet]. 2018. Available from: [https://nces.ed.gov/programs/digest/d18/tables/dt18\\_326.30.asp](https://nces.ed.gov/programs/digest/d18/tables/dt18_326.30.asp)
- [4] Nix S, Perez-Felkner L, Thomas K. Perceived mathematical ability under challenge: A longitudinal perspective on sex segregation among STEM degree fields. *Frontiers in Psychology*. 2015;**6**:530. DOI: 10.3389/fpsyg.2015.00530
- [5] Perez-Felkner L, Nix S, Thomas K. Gendered pathways: How mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in Psychology*. 2017;**8**:386. DOI: 10.3389/fpsyg.2017.00386
- [6] Saunders-Scott D, Braley MB, Stennes-Spidahl N. Traditional and psychological factors associated with academic success: Investigating best predictors of college retention. *Motivation and Emotion*. 2018;**42**:459-465. DOI: 10.1007/s11031-017-9660-4
- [7] Hepworth D, Littlepage B, Hancock K. Factors influencing university student academic success. *Educational Research Quarterly*. 2018;**42**(1):45-61
- [8] Redmond-Sanogo A, Angle J, Davis E. Kinks in the STEM pipeline: Tracking STEM graduation rates using science and mathematics performance. *School Science and Mathematics*. 2016;**116**(7):378-388. DOI: 10.1111/ssm.12195
- [9] Olave BMT. Underestimating the gender gap? An exploratory two-step cluster analysis of STEM labor segmentation and its impact on women. *Journal of Women and Minorities in Science and Engineering*. 2019;**25**(1):53-74. DOI: 10.1615/JWomenMinorScienEng.2019021133
- [10] Para E. Is the Gender Gap Narrowing in Science and Engineering? *Soroptimist International*. 2020 [Internet]. 2020. Available from: <https://www.soroptimistinternational.org/is-the-gender-gap-narrowing-in-science-and-engineering/> [Accessed: January 4, 2022]
- [11] Martinez A, Christnacht C. Women Are Nearly Half of U.S. Workforce but Only 27% of STEM Workers [Internet]. 2021. Available from: <https://www.census.gov/library/stories/2021/01/women-making-gains-in-stem-occupations-but-still-underrepresented.html> [Accessed: December 13, 2021]
- [12] Dika, S. L., D'Amico, M. M. Early experiences and integration in the persistence of first-generation college students in STEM and non-STEM majors. *Journal of Research in Science Teaching* 2016; 53(3): 368-383. DOI: <https://doi.org/10.1002/tea.21301>
- [13] National Science Foundation. Women, Minorities, and Persons with

- Disabilities in Science and Engineering. [Data Set]. [Internet]. 2018. Available from: <https://nces.nsf.gov/pubs/nsf19304/data> [Accessed: December 21, 2021]
- [14] Schneider B, Milesi C, Brown K, Gutin I, Perez-Felkner L. Does the Gender Gap in STEM Majors Vary by Field and Institutional Selectivity? [Internet]. Teachers College Record. 2015. Available from: <https://www.tcrecord.org/books/Content.asp?ContentID=18026>. [Accessed: January 4, 2022]
- [15] Glazer A. National Mathematics Survey [Internet]. 2019. Available from: <https://math.mit.edu/wim/2019/03/10/national-mathematics-survey/> [Accessed: January 4, 2022]
- [16] Weeden KA, Gelbgiser D, Morgan SL. Pipeline dreams: Occupational plans and gender differences in STEM major persistence and completion. *Sociology of Education*. 2020;**93**(4):297-314. DOI: 10.1177/0038040720928484
- [17] Corbett C, Hill C, St. Rose A. Where the Girls Are: The Facts about Gender Equity in Education [Internet]. American Association of University Women. Available from: <https://ww3.aauw.org/files/2013/02/Where-the-Girls-Are-The-Facts-About-Gender-Equity-in-Education.pdf> [Accessed: March 9, 2022]
- [18] Degol JL, Wang M. Te, Zhang Y, Allerton J. Do growth mindsets in math benefit females? Identifying pathways between gender, mindset, and motivation. *Journal of Youth and Adolescence* 2018; 47(5): 976-990. DOI: 10.1007/s10964-017-0739-8
- [19] Voyer D, Voyer SD. Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*. 2014;**140**:1174-1204. DOI: 10.1037/a0036620
- [20] ACT Inc. The ACT Profile Report -National [Data Set] [Internet]. 2020. Available from: <https://www.act.org/content/dam/act/unsecured/documents/2020/2020-National-ACT-Profile-Report.pdf> [Accessed: December 20, 2021]
- [21] U.S. Department of Education, National Center for Education Statistics (NCES). Digest of Education Statistics. [Internet]. 2014. Available from: [https://nces.ed.gov/programs/digest/2014menu\\_tables.asp](https://nces.ed.gov/programs/digest/2014menu_tables.asp)
- [22] Wai J, Lubinski D, Benbow CP, Steiger JH. Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Education & Psychology*. 2010;**102**:860-871. DOI: 10.1037/a0019454
- [23] Bleske-Rechek A, Lubinski D, Benbow CP. Meeting the educational needs of special populations: Advanced placement's role in developing exceptional human capital. *Psychological Science*. 2004;**15**(4):217-224. DOI: 10.1111/j.0956-7976.2004.00655.x
- [24] Ware NC, Lee VE. Sex differences in choice of college science majors. *American Educational Research Journal*. 1988;**25**(4):593-614. DOI: 10.2307/1163131
- [25] Westrick PA, Le H, Robbins SB, Radunzel JM, Schmidt FL. College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*. 2015;**20**(1):23-45. DOI: 10.1080/10627197.2015.997614
- [26] Schneider B, Swanson CB, Rieggle-Crumb C. Opportunities for learning: Course sequences and positional advantages. *Social Psychology*

of Education. 1998;**2**:25-53. DOI:  
10.1023/A:1009601517753

[27] Delaney J, Devereux PJ. The effect of high school rank in English and math on college major choice. ESRI Working Paper No. 650.2020 [Internet]. Available from: <http://aei.pitt.edu/102346/>. [Accessed: January 4, 2021]

[28] Conger D, Long MC. Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science*. 2010;**627**(1):184-214. DOI: 10.1177/00002716209348751

[29] Sadler PM, Sonnert G, Hazari Z, Tai R. Stability and volatility of STEM career interest in high school: A gender study. *Science Education*. 2012;**96**:411-427. DOI: 10.1002/sce.21007

[30] Lubinski D, Benbow CP. Study of mathematically precocious youth after 35 years: Uncovering antecedents for math science expertise. *Perspectives on Psychological Science* 2006; 1: 316-345. DOI:10.1111/1745-6916.2006.00019.x

[31] Maltese AV, Tai RH. Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*. 2011;**95**:877-907. DOI: 10.1002/sce.20441

[32] Fry R, Kennedy B, Funk C. Stem Jobs see Uneven Progress in Increasing Gender, Racial and Ethnic Diversity [Internet]. 2021. Available from: <https://www.pewresearch.org/science/2021/04/01/stem-jobs-see-uneven-progress-in-increasing-gender-racial-and-ethnic-diversity/> [Accessed: December 20, 2021]

[33] Ceci SJ, Williams WM. Sex differences in math-intensive fields.

*Current Directions in Psychological Science*. 2010;**19**(5):275-279. DOI: 10.1177/0963721410383241

[34] Bressoud DM. Historical reflections on teaching the fundamental theorem of integral calculus. *The American Mathematical Monthly*. 2011;**118**(2):99-115. DOI: 10.4169/amer.math.monthly.118.02.099

[35] Cheryan S, Ziegler SA, Montoya AK, Jiang L. Why are some STEM fields more gender balanced than others? *Psychological Bulletin*. 2017;**43**(1):1. DOI: 10.1037/bul0000052

[36] Denaro K, Sato B, Harlow A, Aebersold A, Verma M. Comparison of cluster analysis methodologies for characterization of classroom observation protocol for undergraduate STEM (COPUS) data. *CBE Life Sciences Education*. 2021;**20**(1). DOI: 10.1187/cbe.20-04-0077

[37] Ng BL, Liu WC, Wang JC. Student motivation and learning in mathematics and science: A cluster analysis. *International Journal of Science and Mathematics Education*. 2016;**14**(7):L1359-L1376. DOI: 10.1007/s10763-015-9654-1

[38] Engström S. Differences and similarities between female students and male students that succeed within higher technical education: Profiles emerge through the use of cluster analysis. *International Journal of Technology and Design Education*. 2018;**28**:239-261. DOI: 10.1007/s10798-016-9374-z

[39] SAS Institute Inc. SAS/STAT® 14.2 User's Guide. SAS Institute Inc., 2016

[40] Ward JH Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963;**58**(301):236-244. DOI: 10.1080/01621459.1963.10500845



- [41] SAS/STAT® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc, 2009.
- [42] Sarle WS. Cubic Clustering Criterion. Technical Report A-108, SAS Institute Inc, 1983
- [43] Hill C, Corbett A, Stroese A. Why so Few? Women in Science, Technology, Engineering, and Mathematics [Internet]. 2010. Available from: <https://files.eric.ed.gov/fulltext/ED509653.pdf>
- [44] Petousi V, Sifaki E. Contextualising harm in the framework of research misconduct. Findings from discourse analysis of scientific publications. *International Journal of Sustainable Development*. 2020;**23**(3-4):149-174
- [45] Murnane RJ, Willett JB. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. New York, NY: Oxford University Press; 2010
- [46] Hennig C. What are the true clusters? *Pattern Recognition Letters*. 2015;**64**:53-62. DOI: 10.1016/j.patrec.2015.04.009
- [47] Wang MT, Eccles JS, Kenny S. Not lack of ability but more choice: Individual and gender differences in STEM career choice. *Psychological Science*. 2013;**24**:770-775. DOI: 10.1177/0956797612458937
- [48] Wang MT, Degol JL. Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*. 2017;**229**:119-140. DOI: 10.1007/s10648-015-9355-x
- [49] Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. *Proceedings. National Academy of Sciences. United States of America*. 2011;**108**:3157-3162. DOI: 10.1073/pnas.1014871108
- [50] Zeldin AL, Britner SL, Pajares F. A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*. 2008;**45**(9):1036-1058. DOI: 10.1002/tea.20195