AC 2009-921: EXPLORATION OF DIFFERENCES IN MALE AND FEMALE STUDENTS OVER A FOUR YEAR PERIOD: DOES THE DATA INDICATE SUPPORT FOR THE GENDER SIMILARITY HYPOTHESIS?

Kenneth Reid, Ohio Northern University

P.K. Imbrie, Purdue University

Exploration of Differences in Male and Female Students over a Four Year Period: Does the Data Indicate Support for the Gender Similarity Hypothesis?

Keywords:

First-year, Freshman, Retention, Gender

Introduction:

Engineering programs continue to see exceptionally low female enrollment. Enrollment in engineering programs is around 20% nationwide, with enrollment in electrical and mechanical engineering around 15%. Other traditionally difficult majors such as medicine and law have seen an increase in female enrollment to around 50% over the past decades, but this trend has not been seen in engineering[1]. A number of reasons have been suggested, including a lack of female faculty in engineering programs, lack of popular media attention towards engineering (as opposed to law and medicine) and a general lack of understanding among students and teachers of engineering during the high school years.

As female enrollment in advanced mathematics in high school has increased, the mathematical performance gap seen in previous decades has disappeared, disproving theories of gender disparity in cognitive ability prevalent through the 1970s. Hyde[2] describes stark differences in math performance prior to 1973 which have decreased since 1974; further results showed that the differences had largely disappeared by 2008[3]. Hyde's Gender Similarity Hypothesis[4], theorizes most psychological gender differences are small or trivial, and in many cases, remaining differences are decreasing. Some areas where differences were found include self perception of some leadership characteristics and self perception of mathematical ability in context: i.e., in situations where a difference in performance in mathematics was "expected". When specific expectations were *not* mentioned, no differences in performance were found.

This study will analyze four years of student self report data collected prior to the beginning of the first year of study to answer the following research questions:

- To what extent do the data collected for this study support the Gender Similarity Hypothesis?

- For characteristics which show a difference, is there evidence that these differences are decreasing over this four year period?

Background:

Prior to the 1980's, theories that male students were superior to female students in mathematical ability were widely accepted. For example, in 1974, Maccoby and Jacklin[5] wrote "Boys excel in mathematical ability" under the heading "Sex Differences That Are Fairly Well Established." They state that, after the age of 12-13 years, boys' mathematical skills increase faster than girls, and not entirely as a function of the number of math courses taken.

Jackson et al.[6] reported results of a self-assessment of affective characteristics in a 1993 survey. The survey factors were designated as past (those perceptions existing prior to college), present (during college) and future (after graduation). Females self perception of their ability in mathematics, science and problem solving and overall academic ability were significantly lower than their male counterparts. Although some constructs relate to the students perception of ability *before* entering engineering, data was collected *during* their college experience; this is typical in self-assessments of affective characteristics

Felder, et al.[7] reported in 1995 that male and female students entered Chemical Engineering with no significant differences in overall SAT score or advanced placement credit. Female students reported a higher level of anxiety entering the second year of study; further, their confidence in their preparation was lower than male students; again, assessed during the college experience. Seymour and Hewitt[8] and Hawkes and Spade[9] found that females in STEM fields had lower confidence in their technical abilities, although there was little to no difference in measures of their high school academic performance. Lack of ability was significantly more likely to be perceived as a barrier to success in female students than male students.

In 2001, Ting[10] found different predictor variables of retention for male and female students. The most significant noncognitive predictors for male students were long term goal preference and positive self-concept, while female predictors were successful leadership experience and positive self-concept. Also in 2001, Besterfield-Sacre, et al.[11] found statistically significant differences in five of 13 noncognitive attitudes between male and female students. Those different attitudes related to how engineers help society, basic engineering knowledge, communication skills, problem solving ability and engineering ability.

Light et al.[12] reported significant differences between male and female students early in the first year in intellectual self-confidence, computer and programming skills and business ability but no significant difference in confidence levels specifically in math or science abilities in a 2007 study. Interestingly, results from the *end* of the first year showed female students rating themselves significantly below their male counterparts in math and science abilities.

These studies explored differences between sexes using data collected during the first year of study. The instrument used in the following analysis was used to collect data *prior to* the beginning of the first year to investigate differences as measured by statistical significance and effect size before they enter college.

Participants and Educational Setting:

The data in this study were collected from students entering a large Midwestern university over a four year period, 2004-2007. Cohorts were as follows:

2004: N = 1615; 312 female (19.3%) and 1303 male (80.6%)
2005: N = 1781; 276 female (15.5%), 1505 male (84.5%)
2006: N = 1779; 297 female (16.7%), 1482 male (83.3%)

2007: N = 1711; 348 female (20.3%), 1363 male (79.7%).

All entering students are required to take the Student Attitudinal Success Instrument (SASI), an online self-assessment, including 158 items on multiple Web pages. Students may take all assessments at once, but may log out and back in to avoid completing the assessments during one session. Possible responses to each item are 5-level Likert-scale choice (Strongly Agree to Strongly Disagree). Students who did not complete all sections of the online instrument were not considered in the analysis (1 student in 2004, 7 in 2005, 60 in 2006 and 16 in 2007). Percentages based on sex were consistent within each cohort; the disproportionate number of males in the sample is typical of students entering engineering in this and similar institutions within the United States. Some students who did not subsequently enroll may complete the assessment; therefore, the number of students comprising each cohort is larger than the incoming engineering class.

Constructs and Subfactors:

The analysis consists of student responses to questions measuring nine affective constructs[13]. Each construct is comprised of subfactors identified through factor analysis[14]; each subfactor is in turn comprised of individual items. The constructs include:

- Motivation, consisting of 25 items in four subfactors: *Control*, *Challenge*, *Curiosity* and *Career*.
- Metacognition: consisting of 20 items in four subfactors: *Planning*, *Self-monitoring/Self-Checking*, *Cognitive Strategy* and *Awareness*.
- Deep Learning, consisting of 10 items in two subfactors, *Motive* and *Strategy*.
- Surface Learning, consisting of 10 items in two subfactors, *Memorization* and *Studying*.
- Academic Self-Efficacy, consisting of ten individual items that do not form specific subfactors.
- Leadership, consisting of 20 items with four subfactors, *Motivation*, *Planning*, *Self-Assessment* and *Teammates*
- Team vs. Individual Orientation, consisting of 10 items in two subfactors, *Individual* and *Team Dynamic*.
- Expectancy-Value: consisting of 32 items in five subfactors: *Academic Resources*, *Community Involvement*, *Employment Opportunities*, *Persistence* and *Social Engagement*.
- Major Indecision, consisting of 21 items in four subfactors: *Certainty of Decision*, *Difficulty in Decision*, *Personal Issues*, and *Urgency*. One question was shown not to load to any of the subfactors and is assessed on its own (*Independence*).

The multilevel structure, where each item loads to a superordinate construct (ie: *Major Indecision*) and one subfactor within the domain of the construct (ie: the *Certainty of Decision* subfactor under *Major Indecision*) supports analysis the construct, subfactor and/or item level.

Method:

Significant differences are found using a Mann-Whitney U test for non-parametric data. Mann-Whitney results were found using SAS proc npar1way with the *wilcoxon* and Monte Carlo (*MC*) options. The *MC* option produces Monte Carlo estimates of exact p values and is used specifically for large data sets.

Effect size (Cohen's *d*) is used to assess the magnitude of the effect of a difference. While a measure of statistical significance is important, a statistically significant difference does not *necessarily* imply a meaningful or important difference – only that a true difference most likely exists. Even very small differences tend to become significant as size of the population increases. Cohen's *d*, the effect size, measures the magnitude of the effect or the importance of the difference[15, 16]. Cohen's *d* is found by:

$$d = \frac{(M_1 - M_2)}{\sigma_{pooled}} \tag{1}$$

where M_1 and M_2 are the means of the male and female population. The pooled standard deviation, σ_{pooled}, is the root-mean-square of the standard deviations of the two populations. That is:

$$\sigma_{\text{pooled}} = \sqrt{\frac{\sigma_1^{\,2} + \sigma_2^{\,2}}{2}} \tag{2}$$

Cohen originally, somewhat arbitrarily defined ranges for effect sizes as small: $d = 0.2$, medium, $d = 0.5$; and large, $d = 0.8$, with the caveat that "there is a certain risk in inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science".[17] Ranges for *d* are redefined as part of the Gender Similarity Hypothesis: trivial ($d < 0.10$), small ($0.11 < d < 0.35$), moderate ($0.36 < d < 0.65$) and large ($d > 0.66$).[18]

Results and Discussion:

Table 1 shows each construct: constructs which show statistically significant differences between male and female students are indicated. Five of the nine constructs were found to have a significant difference when examining the aggregate population (2004-2007, N = 6899): these include

- Expectancy-Value
- Motivation
- Surface Learning
- Deep Learning and
- Leadership.

However, the effect sizes were all small ($d < 0.35$) to trivial ($d < 0.11$). Therefore, while there was often a significant difference found in the responses between males and females, the effect – or the importance – was found to be minimal. The statistical significance was an effect of the large sample size rather than the actual magnitude of the effect. These results support the Gender Similarity Hypothesis since most (or in this case, all) of the effect sizes are small.

Table 1: Mean and Standard Deviation, Male and Female Students: Significance and Effect Size

Contruct	Mean, M (N=5665)	σ, M	Mean, F (N=1234)	σ, F	M - F	*p* (MC)	Cohen's *d*	
Expectancy-Value *	3.943	0.360	3.848	0.381	-0.094	<0.0001	-0.254	
Motivation *	4.186	0.391	4.087	0.420	-0.098	<0.0001	-0.243	
Surface Learning *	2.393	0.476	2.486	0.523	0.092	<0.0001	0.185	*small*
Deep Learning *	3.735	0.460	3.652	0.501	-0.082	<0.0001	-0.171	
Leadership *	3.959	0.368	3.910	0.377	-0.048	0.000	-0.129	
Self Efficacy	4.242	0.459	4.214	0.475	-0.029	0.142	-0.061	
Team vs. Individual	3.931	0.381	3.947	0.399	0.016	0.092	0.041	*trivial*
Major Indecision	3.581	0.483	3.580	0.479	-0.001	0.784	-0.003	
Metacognition	3.931	0.406	3.932	0.421	0.001	0.467	0.001	

* = Statistically significant difference, small to trivial effect size

Figure 1 shows a plot of the magnitude of the effect size for each construct in individual years, 2004-2007. Each construct showed a low to trivial effect size when the data is taken in aggregate as well as a low to trivial effect size for each construct in each year.

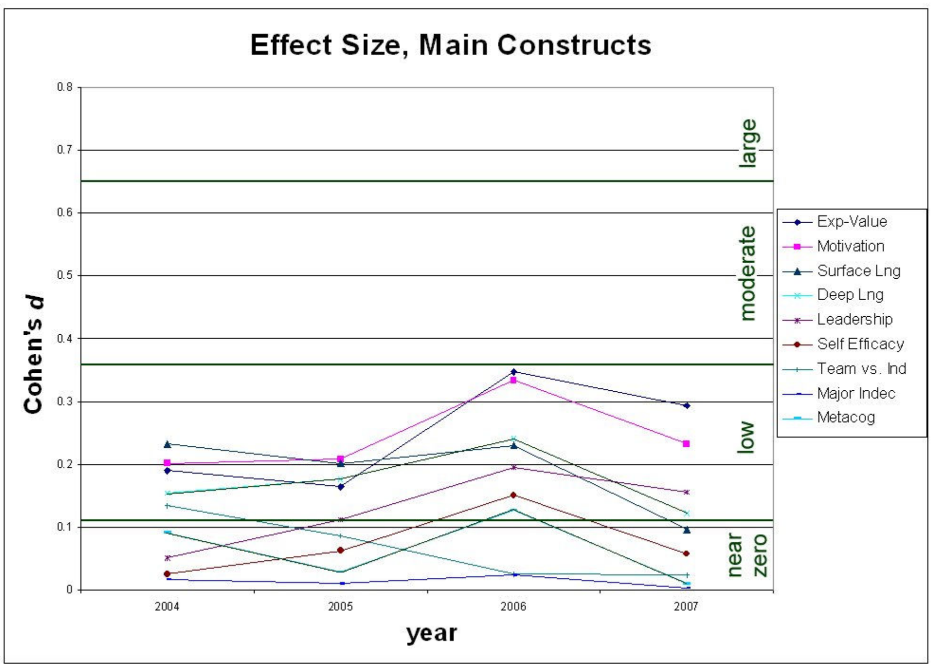

Figure 1: Magnitude of effect sizes (Cohen’s d) for each construct, male / female differences, 2004-2007

Examining the assessment at the item level, only two of the 158 individual items had effect sizes in the moderate range. Both of these items were within the Expectancy-Value construct. The specific questions were:

- As a freshman in college I expect to pursue some sort of community service work outside of my regular course work.
- I expect to make friends with students from my major and form a study group in order to improve my college learning experiences.

Of the remaining items, 76 (48%) had small effect sizes and 80 (51%) had trivial effect sizes.

Small to trivial effect sizes were found from 97% of the subfactors and 99% of the individual items when the data was taken in aggregate. An examination of the data from individual years (2004 – 2007) showed small to trivial effect sizes from 92.1% - 100% of subfactors and 94.3%–99.4% of individual items per each year.

Results certainly support the Gender Similarity Hypothesis as a strong majority of constructs, subfactors and individual items show a small to trivial effect size. However, no evidence of a decreasing trend in effect sizes was seen over the four year period (2004 – 2007). Figure 1 illustrates this lack of trend at the construct level. Examination at the subfactor level and individual item supported the conclusion of a lack of evidence of trends over this period.

Although literature supports a decreasing trend in differences between the sexes in some characteristics, changes may occur over a period of decades; therefore, lack of evidence over a four year period was not necessarily surprising. For example, Hyde shows that differences in math performance between high school male and female students disappeared between two studies from the mid 1990's to the mid 2000's, as the percentage of females enrolled in advanced mathematics classes increased. Changes such as these require time to propagate through a population for trends to become evident.

Conclusion:

Student responses to the Student Attitudinal Success Instrument to assess nine noncognitive characteristics (and their associated subfactors) prior to the beginning of their engineering program were used to study differences between male and female students. Overall, results strongly support the Gender Similarity Hypothesis, which theorizes that most differences between the sexes are small to near zero. Analysis of the data at the construct, subfactor and item level supports this hypothesis. In general, male and female students enter their first year of engineering with very similar self-reported affective characteristics.

There does seem to be evidence to support a widely accepted belief that female students would be well served to emphasize activities associated with community: both learning communities and community service. There also seems to be some evidence to support the importance of investigating characteristics of first year programs on the environment of competition and individual achievement. Female students seem to enter engineering with (slightly) more focus on achieving their goals, but there is evidence that female students have lower self-efficacy by the end of their first year. Identifying characteristics contributing to this change could lead to increased retention in engineering for students who enter college with a strong intent to succeed.

Further work is planned to investigate the normative taxonomy of male vs. female students, investigating whether this cluster analysis also supports the Gender Similarity Hypothesis, and whether trends in affective characteristics are indicated. Further work is also planned to investigate changes in the affective characteristics studied here over the course of the first year.

References:

[1] Sanoff, A (2005). Competing forces. *ASEE Prism*, 15 (2): 24-29.

[2] Hyde, J. S., E. Fennema & S. J. Lamon (1990). Gender differences in math performance: A meta-analysis. *Psychological Bulletin*, 107 (2): 139-155.

[3] Hyde, J. S., S. Lindberg, M. Linn, A. Ellis & C. Williams (2008). Gender similarities characterize math performance. *Science*, 25 July 2008 321: 494-495.

[4] Hyde, J. S. (2005). The gender similarity hypothesis. *American Psychologist*, 60 (6): 581-592.

[5] Maccoby, E.E. & C.N. Jacklin (1974). *The psychology of sex differences.* Stanford, CA, Stanford University Press.

[6] Jackson, L.A., P.D. Gardner & L.A. Sullivan (1993). Engineering persistence: Past, present, and future factors and gender differences. *Higher Education*, 26 (2): 227-246.

[7] Felder, R. M., G. Felder, M. Mauney, C.E. Hamrin, Jr. & E. J. Dietz (1995). A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes. *Journal of Engineering Education*, 84 (2): 151-163.

[8] Seymour, E. & H. Hewitt (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO., Westview Press.

[9] Hawks, B.K. & J.Z. Spade (1998). Women and men engineering students: Anticipation of family and work roles. *Journal of Engineering Education*, 87 (3): 249-256.

[10] Ting, S. R. (2001). Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education*, 17 (1): 75-80.

[11] Besterfield-Sacre, M, M. Moreno, L.J. Shuman & C.J. Atman (2001). Gender and ethnicity differences in freshman engineering student attitudes: A cross-institutional study. *Journal of Engineering Education*, 90 (4): 477-490.

[12] Light, J., R. Korte, K. Yasuhara & D. Kilgore (2007). Exploring relationships among performance on engineering tasks, confidence, gender and first year persistence. *Proceeding of the American Society for Engineering Education Annual Conference*, Honolulu, HI.

[13] Reid, K. J. , P. K. Imbrie (2008). Noncognitive Characteristics of Incoming Engineering Students Compared to Incoming Engineering Technology Students: A Preliminary Examination. *Proceedings of the 2008 American Society for Engineering Education Annual Conference,* Pittsburgh, PA.

[14] Imbrie, P.K., K.J. Reid, J. Immekus & J.J. Lin (2008). *Psychometric Properties and Stability of an Instrument Assessing Noncognitive Characteristics of Engineering Students Prior to their First Year of Study*, in progress.

[15] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

[16] Rosnow, R. & R. Rosenthal (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1 (4): 331-340.

[17] Ref 7.

[18] Ref. 4.