Possible Titles:

Identifying Early Interest in Mathematics and Science

Using Latent Class Analysis to Analyze Seventh Grade Science and Mathematics Attitudes

Similarities in Seventh Grade: Using Latent Class Analysis to Analyze Mathematics and Science Attitudes by Gender and Ethnicity

Using Latent Class Analysis to Identify Early Interest in Mathematics and Science

A Method to Identify Early Interest in Mathematics and Science

Supporting and Fostering STEM Career Aspirations

Abstract

Blah blah blah.

*Keywords*: attitudes, STEM, achievement, mathematics, science

Title

There is a projected need for students to fill STEM-related occupations (add references) and a projected shortage of students who have the interest and adequate preparation to fill these occupations (add references). To address this projected need and shortage, much can be done to support the early interest and persistence of STEM careers in K-12 settings. However, …

This is an even more acute issue for females and underrepresented minority students. Research suggests that underrepresented minority students (such as Black and Hispanic students) are more susceptible to dropping out of STEM careers despite expressing initial interest in STEM careers (Bonous-Hammarth, 2000; Daempfle, 2003; Elliott, Strenta, Adair, Matier, & Scott, 1996; Gándara & Maxwell-Jolly, 1999; National Academy of Sciences, 2011; Grandy, 1998; Oseguera, Hurtado, Denson, Cerna, & Saenz, 2006; Russell, & Atwater, 2005; White, J. Altschuld, J. & Lee, 2006). Recent research finds that underrepresented minorities aspire to major in STEM in college at the same rates as their white and Asian American peers but have lower four- and five-year completion rates relative to those of whites and Asian Americans (HERI, 2010, Huang, Taddese, & Walter, 2000). Hurtado, Cabrera, Lin, Arellano, Espinosa (2009) raise the issue that common assumptions regarding racial and ethnic inequality focus on a perceived lack of motivation and preparation within minority populations but that studies have shown many academically well-prepared high school URM students are interested in pursuing scientific and engineering careers (see for example, College Board 2005; Hurtado et al. 2006; NSF 2004).

Add literature about importance of pre-college (k-12) settings (see for example, Sax & Harper, 2007).

**Importance of Attitudes in STEM Career Aspirations (1 page)**

Add recent literature (Riegle-Crumb et al., 2012) about lack of achievement differences between males/females; underrepresented minorities/Whites and Asians in K-12 settings. Differences in achievement fail to account for differences in STEM career aspirations.

Add literature about importance of attitudes and aspirations in STEM persistence.

**Mathematics and Science Attitudes (1 page)**

Add literature from science identity from Aschbacher et al. (2010), Archer et al. (2012), Calabrese-Barton et al. (in press)

Add literature from math attitudes and aspirations

This study builds on existing literature about differences in mathematics and science attitudes by examining differences between seventh graders in terms of their mathematics and science attitudes. This study explores whether such differences in attitudes vary by gender, ethnicity and achievement and relates these differences to whether or not these students were employed in a STEM career by their mid-30’s. This study uses a methodology that helps make sense of underlying differences in student mathematics and science attitudes. In doing so, this study contributes to the existing literature in the following ways: includes longitudinal data of a national sample of seventh graders, includes mathematics and science attitude variables, and includes outcome of actual STEM career (and not intentions to pursue STEM degree). Information from this study can be used to help target programs or interventions to particular groups of students. MI: Maybe give hypothetical example of what this means in terms of specific interventions that could raise attitudes for students who are interested in math but not interested in science.

**Method**

**Sample**

This sample includes participants from the Longitudinal Study of American Youth (LSAY). The LSAY is a national sample of public school students in 1987 who are now in their late-30s. The LSAY was funded by the National Science Foundation in 1986 to examine the development of student achievement in middle and high school and the relationship of those patterns to career choices. The students included in this particular study were from the seventh grade cohort. The cohort consists of students from 52 middle schools across the United States in 1987. Approximately 60 students were randomly selected from each school. The sample is predominantly White (70%) with approximately equal numbers of females (48%) and males (52%). The sample included 9% Hispanic, 11% African American, 4% Asian, and 2% Native American (5% of students did not indicate any race/ethnicity). Thirty-one percent of the students in the sample had at least one parent who completed college, while the other 69% did not. This study included data from students who completed an attitudinal questionnaire and mathematics achievement test in seventh grade (*N* = 3,116). In 2007, more than 95% of the original sample completed a questionnaire about their educational and occupational outcomes (Miller, 2010).

**Measures**

**Mathematics and science attitudes**. Ten items related to mathematics and science attitudes were included (Table X). These items were selected based on prior literature (see for example, XX, ??). MI: I don’t know if these items were used in another LSAY study? Do you? But if not, we can bring in some theory about why attitudes are important and why these particular items were selected to measure attitudes? Would it make sense to include confirmatory factor analysis results here?

**Mathematics achievement**. Student mathematics achievement was assessed in the fall of seventh grade (*M* = 50.47, *SD* = 10.20). The scores were calculated using an Item Response Theory model (Lord, 1980), with a scale ranging from 0 to 100. There is a mean score of 50 and a standard deviation of 10 for the seventh grade students. Each test consisted of items from the National Assessment of Educational Progress (NAEP, 1986) and was designed to measure basic skills, algebra, geometry, and quantitative literacy.

**STEM Career**. In the 2007 questionnaire, respondents were asked about the industry of their current occupation. LSAY created a dichotomous variable to indicate whether or not the respondent was currently employed in a STEM occupation (such engineering) or not. The particular definition of a STEM career used as an outcome variable in this study includes the full range of STEM occupations but excludes social science occupations. Fifteen percent (*n* = 275) of the sample was employed in a STEM or STEM support occupation.

**Demographics**. Students self-reported their gender and ethnicity. There were roughly similar number of males (51%) and females (49%) included. A dichotomous variable of ethnicity was created: White and Asian (not underrepresented minority); and African American, Hispanic/Latino, and Native American (underrepresented minorities). This grouping is included because underrepresented minority students have lower representation in STEM careers compared to other students (see for example, Huang, Taddese, & Walter, 2000). In 2004, for example, African-Americans made up 12.8% of the population, but only 3.1% of engineers identified as African American in that year. Additionally, Hispanics made up 14.1% of the population in 2004, but only 4.9% of engineers identified as Hispanic in 2004 (National Science Foundation, 2005). This sample includes 77% White and Asian students and 23% underrepresented minority students.

**Analysis**

Data was analyzed using Latent Class Analysis (LCA), an analytic technique used to classify groups of individuals into latent classes based on their responses to the set of indicators (Bartholomew, 1987; Collins & Lanza, 2010; Goodman, 1974; Heinen, 1996; Lazarsfeld & Henry, 1968; Muthén, 1992, 2001). LCA is an exploratory multivariate analysis—that is, there is no apriori specification of the number or type of classes that emerge. LCA differs from the more commonly used factor analysis in that factor analysis clusters items and LCA clusters individuals.

Traditionally, LCA models are fit in a series of steps. First, a one-class model is fit and then the number of classes is increased. The fit of each new model, that difers by the specification on one more class, is compared to the previous model. With parsimony in mind, a model with the greater number of classes is selected only if increasing the number of classes results in conceptually meaningful groupings and provides good statistical fit. In the present analysis, once the number of classes was decided upon, covariates (gender, race, and previous math achievement) were included in the model (Nylund-Gibson & Masyn, 2011) and the means of the distal outcomes (STEM and WHAT) were estimated for each class using pseudoclass draws. Pseudoclass draws are a preferred method for comparing the mean of distal outcomes compared to say a regression, because this method takes into account the fact that individuals are not assigned with 100% certainly into each of the latent classes (Clark & Muthén, 2009). The *p-*values from a series of pairwise Wald tests (e.g., “auxiliary (e)” in M*plus*) were used to test for significant differences in means across the five readiness classes based on the means from the pseudoclass draws (seeMuthén & Asparouhov, 2010).

In this LCA application, several indicators of model fit were used since there is no single statistical indicator is recommended to assess model fit. We used a combination of statistical indicators and substantive theory to decide on the best fitting model (Nylund, Asparouhov, & Muthén, 2007). The Bayesian Information Criterion (BIC; Schwartz, 1978), the most commonly used and trusted fit indices for model comparison was used, where lower values of the BIC indicated better fit. Along with the BIC, we compared models that differed in the number of classes using the Lo-Mendell-Rubin (LMR) and the bootstrap likelihood ratio test (BLRT) to evaluate if adding an additional class significantly improved model fit (for more on these fit indices see Nylund et al. [2007]). The entropy of the final model selected is reported in the text, but not used for model fit because it describes the overall classification of students into the latent classes. Entropy ranges between 0 and 1, where 1 is perfect classification and values approaching 1 indicate clear delineation of classes (Celeux & Soromenho, 1996).

Two quasi-Bayesian information-heuristic model fit comparisons, both functions of the individual model BIC values, are also included that have shown promise in latent class growth model selection (Nagin, 1999) and have been proposed for use with latent class analysis (Masyn, 2012). The Bayes Factor (*BF*) is a pairwise comparison of relative fit between two models. In our study, the computed *BF* approximates the ratio of the probability two models under the assumption that one of the two models being compared is the “true” model. The value of the ratio is then compared to the Jeffery’s Scale of Evident (Wasserman, 2000), for which 1 < *BF* < 3 is considered weak evidence for Model *K* over Model *K +1*, 3 < *BF* < 10 is moderate evidence for Model *K*, and *BF*  > 10 is strong evidence for Model *K*. The other comparison is the approximate correct model probability (*cmP*), which estimates the probability that each model out of a given set of latent class models being considered is correct, under the assumption that the “true” model is contained within that same set of models; thus, the *cmP* values across the given set of models sum to 1.00. The model with the largest *cmP* value is then the model that has the highest probably of being the correct model among the set of models under consideration. See Maysn (2012) for more on these two fit comparisons and their calculations.

**Results (2 pages)**

The results are divided into three sections: (a) identifying the attitudinal profile groups (latent classes), (b) assessing attitudinal group differences with respect to the covariates, and (c) assessing the attitudinal profiles with respect to long term job choice.

**Understanding Math and Science Attitude Profiles**

As described before, a series of LCA models were fit specifying 1-6 latent classes. For each model, fit statistics were collected and used to help inform the decision about how many classes were sufficient to describe the heterogeneity in studnets attitudes towards math and science. Table 1 presents model fit information for the LCA models with the latent classes that were considered. The lowest value of the BIC (13223.58), currently the most trusted fit statistic used for LCA models, indicated a 5-class model (Masyn, 2012; Nylund et al., 2007). The significant *p*-value of the LMR indicated that a 2-class model provided superior fit to a 3-class model. There was never a nonsignificant *p*-value for the BLRT, so this did not inform our decision. The Bayes Factor value went from being very small to very large at the 5-class model, and was in the strong fit range for the 5- and 6-class models. Given parsimony, the 5-class model is preferred. Also, the *cmP* value was highest for the 5-class model indicating that this model had the highest probability of being the correct model given the 1-7-class models considered. Given the statistical support of the 5-class model, and the substantive plausibility of the solution, this model was considered the final model. The entropy for this model was .81, an acceptable value.

* Relationship to gender, ethnicity, math achievement (Table 3)
* Relationship to distal outcome(s). Students with more positive attitudes are more likely to work in STEM. Students with less positive attitudes are less likely to work in STEM.
* Percent of female and underrepresented minority students in Class 1

**Discussion (2 pages)**

In this sample, seventh grade females and underrepresented minority students were not only similar in their attitudes towards mathematics and science compared to males and White and Asian students but, were actually more likely to be in Class 1 compared to other students. However, we are all familiar with how this story ends. Despite this initial early interest in mathematics and science, females and underrepresented students were less likely to persist in a STEM career. There might be several reasons for this early interest or initial similarities. Flush this out: Students express interest but don’t have interest in an actual career that uses math and science; maybe they don’t know what careers are that use math and science; or maybe they just don’t have the support for this initial interest to sustain them through college. We know that females opt out at higher rates than males and that these differences are most pronounced in college and (as we say with this data) with who actually attains STEM employment. However we don’t know where the breakdown occurs or whether this breakdown is systematic enough to detect and do something about. Future research on this longitudinal data will help identify at what time point this interest starts to dip and gaps between females/males, underrepresented minorities/white and Asians become more pronounced. Identifying this could help target interventions and programs (give concrete examples).

References

Bonous-Hammarth, M. (2000). Pathways to success: Affirming opportunities for science, mathematics, and engineering majors. *Journal of Negro Education*, *69*(1/2), 92-111.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. Journal of Research in Science Teaching, 44(8), 1187-1218.

College Board. (2005). *2005 College-bound seniors: Total group profile report*. Retrieved February 25, 2008 from <http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2005/2005-college-bound-seniors.pdf>

Daempfle, P. A. (2003). An analysis of the high attrition rates among first year college science, math, and engineering majors. *Journal of College Student Retention*, *5*(1), 37-52.

DeWitt, J., Archer, L., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2011). High aspirations but low progression: The science aspirations-careers paradox among minority ethnic students. *International Journal of Science and Mathematics Education*, *9*(2), 243–271

Elliott, R., Strenta, A. C., Adair, R., Matier, M., & Scott, J. (1996). The role of ethnicity in choosing and leaving science in highly selective institutions. *Research in Higher Education, 37*(6), 681-709.

Gándara, P., & Maxwell-Jolly, J. (1999). *Priming the pump: Strategies for increasing the achievement of underrepresented minority undergraduates*. New York: The College Board.

Grandy, J. (1998). Persistence in science of high-ability minority students. *The Journal of Higher Education*, 69(6), 589-620.

Higher Education Research Institute. (2010). *Degrees of success: Bachelor’s degree completion rates among initial STEM majors*. Los Angeles: Higher Education Research Institute.

Huang, G., Taddese, N., & Walter, E. (2000). *Entry and persistence of women and minorities in college science and engineering education* (No. NCES 2000601). Washington, D.C.: National Center for Education Statistics.

Hurtado, S., Cabrera, N. L., Lin, M. H., Arellano, L., & Espinosa, L. L. (2009). Diversifying science: Underrepresented student experiences in structured research programs. *Research in Higher Education*, *50*(2), 189-214.

Hurtado, S., Cerna, O. S., Chang, J. C., Sa´enz, V. B., Lopez, L. R., Mosqueda, C., Oseguera, L., Chang, M. J., & Korn, W. S. (2006). *Aspiring scientists: Characteristics of college freshmen interested in the biomedical and behavioral sciences*. Los Angeles, CA: Higher Education Research Institute, UCLA.

Masyn, K. (2012). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 375-393). Oxford, UK: Oxford University Press.

National Academy of Sciences. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: Author.

National Science Foundation, Division of Science Resources Statistics. (2004). *Women, minorities, and persons with disabilities in science and engineering* (NSF 04–317). Washington, DC: National Science Foundation.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 535–569. doi:10.1111/j.1467-8624.2007.01097.x

Nylund-Gibson, K. L., & Masyn, K. E. (2011, April). Including auxiliary variables in latent class analysis*.* In K. Nylund-Gibson (Chair), *An overview of latent class analysis: Application and issues*. Symposium conducted at the 91st Annual Meeting of the Western Psychological Association, Los Angeles, CA.

Oseguera, L., Hurtado, S., Denson, N., Cerna, O., & Saenz, V. (2006). The characteristics and experiences of minority freshmen committed to biomedical and behavioral science research careers. *Journal of Women and Minorities in Science and Engineering*, *12*(2-3), 155-177.

Russell, M. L., & Atwater, M. M. (2005). Traveling the road to success: A discourse on persistence throughout the science pipeline with African American students at a predominantly White institution. *Journal of Research in Science Teaching*, *42*(6), 691-715.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44*, 92–107.

White, J. L., Altschuld, J. W., & Lee, Y. (2006). Persistence of interest in science, technology, engineering, and mathematics: A minority retention study. *Journal of Women and Minorities in Science and Engineering*, *12*(1), 47-64.

Table 1

*Descriptives of Mathematics and Science Attitude Items*

|  |  |  |  |
| --- | --- | --- | --- |
|  | *N* | *M* | *SD* |
| I enjoy math | 2875 | 3.71 | 1.18 |
| Math is useful I everyday problems | 2819 | 3.87 | 1.01 |
| Math helps a person think logically | 2160 | 3.39 | 0.78 |
| It is important to know math to get a good job | 2824 | 4.05 | 0.96 |
| I will use math in many ways as an adult | 2829 | 4.04 | 0.96 |
| I enjoy science | 2857 | 3.55 | 1.25 |
| Science is useful in everyday problems | 2804 | 3.26 | 1.09 |
| Science helps a person think logically | 2809 | 3.47 | 1.04 |
| It is important to know science to get a good job | 2828 | 3.28 | 1.09 |
| I will use science in many ways as an adult | 2859 | 3.42 | 1.13 |

Table 2

*Summary of Latent Class Analysis Fit Indices with 1-6 Latent Classes*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Number of classes* | *Loglikelihood* | *Number of parameters* | *BIC* | *ABIC* | *VLMR* | *BLRT* | *BF* | *cmP* |
| 1 |  |  |  |  |  |  |  | - - - |
| 2 |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |

Table 3

*Summary of Multinomial Logistic Regression*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Logit* | *SE* | *est/SE* | *Odds Ratio* |
| Class 2: NAME |  |  |  |  |
| Underrepresented minority | -0.48\*\*\* | 0.17 | -2.87 | 0.62 |
| Female | 0.83\*\*\* | 0.13 | 6.43 | 2.28 |
| Mathematics achievement | -0.02\*\* | 0.01 | -2.72 | 0.98 |
| Class 3: NAME |  |  |  |  |
| Underrepresented minority | 0.02 | 0.23 | 0.10 | 1.02 |
| Female | 0.28 | 0.15 | 1.83 | 1.32 |
| Mathematics achievement | -0.06\*\*\* | 0.01 | -5.17 | 0.94 |
| Class 4: NAME |  |  |  |  |
| Underrepresented minority | -0.40\* | 0.18 | -2.22 | 0.67 |
| Female | 0.56\*\*\* | 0.13 | 4.19 | 1.75 |
| Mathematics achievement | -0.06\*\*\* | 0.01 | -5.16 | 0.95 |

*Note*. Comparison group is “Class 1: NAME”.

\**p* < .05. \*\**p* < .01. \*\*\**p* < .001.



*Figure 1*. Item probability profiles. MI: I will change names of classes and make this black and white for publication purposes.