**Final Report for Period:** 06/2009 - 05/2010 **Submitted on:** 06/01/2010

**Principal Investigator:** Sinn, Robb . **Award ID:** 0633264

**Organization:** North Georgia College

**Submitted By:**

Spence, Dianna - Co-Principal Investigator

**Title:**

Authentic, Career-Based, Discovery Learning Projects in Introductory Statistics

## Project Participants

**Senior Personnel**

**Name:** Sinn, Robb

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Spence, Dianna

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Hendricks, Todd

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Bailey, Brad

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

After serving as one of our pilot instructors, Dr. Bailey took part in our data analysis and dissemination efforts. He received no course release time or other funding for his role, but he did share the budgeted professional development funds to present our findings at conferences.

**Post-doc**

**Graduate Student**

**Undergraduate Student**

**Technician, Programmer**

**Other Participant**

**Research Experience for Undergraduates**

## Organizational Partners

## Other Collaborators or Contacts

We had 3 collaborators in our department (Mathematics) to teach control and treatment sections of the statistics course. They each taught one or more control sections (Spring 2008) and one or more treatment sections (Fall 2008-Fall 2009). These Mathematics faculty are Brad Bailey, Karen Briggs, and John Holliday. All 3 of these collaborators also assisted with the workshop held for statistics instructors. After teaching his treatment sections, Brad Bailey became more involved with the data analysis and dissemination efforts and was therefore considered senior project personnel during the final year of the project.

We also worked with an interdisciplinary team of 8 faculty from other departments who assisted in developing authentic research constructs and project ideas from their fields that were used in the project guides that have been developed. The team members (and related disciplines) are: Stuart Batchelder (Criminal Justice), Nancy Dalman (Biology), Giovanna Follo (Sociology), Steven Lloyd (Psychology), James Martin (Psych./Counseling), Mohammad Nourbakhsh (Physical Therapy), Brent Paterline (Criminal Justice), and Marina Slemmons (Nursing).

## Activities and Findings

**Research and Education Activities: (See PDF version submitted by PI at the end of the report)**

Project activities are described in the attached document.

**Findings: (See PDF version submitted by PI at the end of the report)**

Project findings are described in the attached document.

**Training and Development:**

Instructors participating in the control and treatment groups were trained to use the teaching methods and materials developed during this project. This includes 3 professors at NGCSU, 1 professor at GPC, and 1 high school teacher at FCHS. (See activities report document for full details.) These instructors developed professionally during their first semester of implementing these materials in their classes, and subsequently assisted in conducting training sessions during the workshop we hosted in January 2009 for other statistics instructors.

**Outreach Activities:**

Our primary outreach activity was the 'Make It Real' Statistics Workshop that we hosted in January 2009 for teachers of statistics. This one day professional development workshop was advertised in advance at local and regional conferences, as well as through our local RESAs (regional educational service agencies). Nineteen high school teachers of AP statistics attended. The workshop was one full day onsite, with online followup assignments and discussion; optional professional learning units (PLU's) were available to participants. Full details of the workshop are given in the Project Activities document.

## Journal Publications

Sinn, A. R.; Spence, D. J., "Authentic discovery experiences in elementary statistics", Journal of Statistics Education, p. , vol. , (2010). Submitted,

Spence, D. J.; Sharp, J.; Sinn, A. R., "Investigation of factors mediating the effectiveness of authentic projects in the teaching of elementary statistics", Journal of Mathematical Behavior, p. , vol. , (2010). Submitted,

Spence, D. J.; Sinn, A. R.; Sharp, J., "Measurement of student outcomes in elementary statistics: Development of multiple instruments", Applied Measurement in Education, p. , vol. , (2010). In preparation for submission,

## Books or Other One-time Publications

Spence, D. J.; Sinn, A. R., "Authentic Discovery Projects in Statistics", (2009). Conference Proceedings, Published
Editor(s): M. Garner
Collection: Proceedings of the 3rd Annual Meeting of the Georgia Association of Mathematics Teacher Educators, October 14, 2009
Bibliography: Spence, D. J., & Sinn, A. R. (2009). Authentic Discovery Projects in Statistics. In M. Garner (Ed.), Proceedings of the 3rd Annual Meeting of the Georgia Association of Mathematics

Spence, D. J.; Sinn, A. R.; Briggs, K. S., "Authentic Discovery Projects in Elementary Statistics", (2009). Conference Proceedings, Published
Editor(s): J. Horn

Collection: Engaging Approaches: Proceedings of the 7th Annual Teaching Matters Conference, March 27-28, 2009
Bibliography: Authentic Discovery Projects in Elementary Statistics. In J. Horn (Ed.), Engaging Approaches: Proceedings of the 7th Annual Teaching Matters Conference, March 27-28, 2009 (pp. 107

Sinn, A. R.; Spence, D. J., "Discovering Statistics Project Guide", (2010). Online Text, Publication pending revisions
Bibliography: Freeman Publisher

## Web/Internet Site

**URL(s):**

http://radar.northgeorgia.edu/~djspence/nsf/

**Description:**

This site is a gateway to all materials associated with the project. From the main page, users can link to curriculum materials that have been developed (instructor and student guides), research publications, conference presentation abstracts and slides, and reference materials about the grant itself, including annual activity reports.

The page detailing publications will be updated as additional manuscripts currently submitted are accepted to journals (see Project Activities document for details on these pending submissions).

The link to the curriculum materials will be redirected when the instructor and student guides become part of Freeman Publisher's suite of web-based suite of textbooks, supplements, and resources for statistics instruction.

Throughout the grant, the site also served as a resource for our interdisciplinary team members, and provided information and a registration gateway for the professional development workshop that was offered for statistics instructors. Workshop details have since been removed, but artifacts of interest from the interdisciplinary team remain on the site, including contact information (if still current) and copies of their original contributions to the project, where available in electronic form.

## Other Specific Products

**Product Type:**

**Instruments or equipment developed**

**Product Description:**

Perceived Usefulness of Statistics scale:
This is the instrument we developed to measure perceived usefulness of statistics, as described in the section on project activities.

**Sharing Information:**

The instrument has been shared with the "Statistics Education NSF PIs Community" during the meeting organized by Dennis Pearl in January 2010 at the MAA/AMS Joint Mathematics Meetings. The instrument is also part of the focus of an instrumentation article currently in preparation for submission to Applied Measurement in Education.

**Product Type:**

**Instruments or equipment developed**

**Product Description:**

Statistics Self-Efficacy scale:
This is the instrument we developed to measure self-beliefs about statistical knowledge and skills, as described in the section on project activities.

**Sharing Information:**

The instrument has been shared with the "Statistics Education NSF PIs Community" during the meeting organized by Dennis Pearl in January 2010 at the MAA/AMS Joint Mathematics Meetings. The instrument is also part of the focus of an instrumentation article currently in preparation for submission to Applied Measurement in Education.

**Product Type:**

**Instruments or equipment developed**

**Product Description:**

Statistics Content Knowledge test:
This is the instrument we developed to measure content knowledge specifically in the areas of linear regression and t-tests, as described in the section on project activities.

**Sharing Information:**

The instrument has been shared with the "Statistics Education NSF PIs Community" during the meeting organized by Dennis Pearl in January 2010 at the MAA/AMS Joint Mathematics Meetings. The instrument is also part of the focus of an instrumentation article currently in preparation for submission to Applied Measurement in Education.

**Product Type:**

**Teaching aids**

**Product Description:**

Instructor Project Guide:
This resource provides guidelines and resources to help instructors to implement discovery projects in their statistics classes.

**Sharing Information:**

This guide is shared with instructors on the Internet at
http://radar.northgeorgia.edu/~rsinn/nsf/IG.html

We have also signed jointly with Freeman Publisher a "letter of intent to publish" this guide in conjunction with educational resources accompanying the statistics text "The Basic Practice of Statistics" (by David Moore). Publication is expected pending revisions now in progress.

**Product Type:**

**Teaching aids**

**Product Description:**

Student Project Guide:

This guide provides information and assistance to students implementing discovery projects in statistics.

**Sharing Information:**

This guide is currently available on the Internet at:

http://radar.northgeorgia.edu/~rsinn/nsf/StudentGuide/StudentGuide.pdf

We have also signed jointly with Freeman Publisher a "letter of intent to publish" this guide in conjunction with educational resources accompanying the statistics text "The Basic Practice of Statistics" (by David Moore). Publication is expected pending revisions now in progress.

## Contributions

**Contributions within Discipline:**

Contributions within Discipline

The co-PI's on this project are mathematics educators who teach introductory statistics. The discipline for the project has therefore been a confluence of statistics, a core course in many STEM and even non-STEM plans of study, and mathematics education. The curriculum developed and the dispersal of teaching methods has improved learning outcomes for our statistics courses and has the capacity to offer similar improvements to many other types of statistics courses.

The instrumentation and educational research results are particularly of interest to the mathematics education community, especially those who investigate secondary and post-secondary mathematics settings.

The best practice teaching modalities that inspired this project include discovery learning, apprentice learning, and attention to student motivation and attitudes. These practices have been widely studied by mathematics education researchers and implemented across many levels and types of courses. The co-PI's agreed to take these focal points of best practice mathematics education and apply them to introductory statistics courses. Such an initiative has been supported repeatedly in mathematics and statistics education literature.

Researchers and educators have often suggested improvements to statistics teaching methods, especially those that focus on implementing the scientific method through authentic statistical experiences (Bryce, 2005). The consensus among many researchers is that statistics is taught most effectively with real data (Cobb & Moore, 1997). In particular, there is greater benefit to students' learning when they collect their own data rather than merely working with data already collected by others (Hogg, 1991). This finding parallels the suggestion by many researchers that statistics education should be student-centered (Roseth, Garfield, & Ben-Zvi, 2008).

When best-practice pedagogies have been implemented in statistics courses, the results have been positive for achievement and for improved attitudes toward statistics. For instance, students who have participated in all aspects of statistical research? collecting data, performing analyses, and communicating results? have demonstrated benefits in exam performance and in students' evaluations of the course (Smith, 1998). Further, research suggests that apprentice learning, wherein students complete real-world mathematics in authentic settings, develops better conceptual understanding and better knowledge transfer to non-mathematical and non-school settings (Boaler, 1998). Findings also suggest that statistics courses based on more constructivist models improve student attitudes toward statistics and that personal relevance is important for successful learning in statistics (Mvududu, 2003). One case study revealed that students learned more from a real-world project than from any other instructional component of a statistics course; the project also fostered an increase in student motivation (Yesilcay, 2000).

The American Statistical Association sponsored a project, the Guidelines for Assessment and Instruction in Statistics Education (GAISE), through which recommendations were developed for best practices in statistics teaching (Franklin & Garfield, 2006). These recommendations included the use of real data and the fostering of active learning. Also among the guidelines offered was the stipulation that 'teachers of statistics should rely much less on lecturing, [and] much more on the alternatives such as projects' (Guidelines for Assessment and Instruction in Statistics Education [GAISE], 2005, p. 4). In agreement with this recommendation, Landrum and Smith (2007) suggest as a best practice 'that students receive some 'hands-on' experience with a research project. An ideal situation would be to finish a complete project that included data collection and analysis' (p. 52). Nevertheless, although the use of projects has been increasingly recommended as a sound pedagogical practice in statistics, many instructors still do not incorporate projects into their statistics courses (Landrum & Smith).

This project addressed both the need for these types of curricular advances and the means for measuring student progress in skills, knowledge, attitudes, and self-beliefs. The curriculum development portion of the project addressed the needs for more authentic statistics experiences while the scale development addressed two aspects of student motivation for statistics: students' perceived usefulness of statistics and students' self-efficacy for statistics tasks.

These scales address research which has suggested that instructors in statistics courses would do well to consider variables from the affective domain as an integral (not peripheral) part of the statistical learning process. A meta-analysis of 113 mathematics education studies found a significant influence of attitude toward mathematics upon achievement in mathematics (Ma & Kishor, 1997). In a review of the literature surrounding motivation to learn mathematics, Middleton and Spanias (1999) report that careful design of instruction can strongly influence student motivation for mathematics achievement, which increases the likelihood students will choose to take future mathematics courses. A wealth of evidence also connects students' mathematics achievement with their self-beliefs in mathematics (e.g., Pajares & Schunk, 2001, 2002). Of particular interest in the current study was the construct of students' mathematics self-efficacy, or students' beliefs about their ability to carry out mathematical tasks (Pajares & Miller, 1994). Findings suggest that self-efficacy beliefs play a role in people's career choices, especially in mathematics and science related fields (e.g., Hackett, 1995; Zeldin & Pajares, 2000). Further, research suggests that mathematics self-efficacy may be as reliable as mental ability in predicting mathematics performance (Pajares & Kranzler, 1995). Well-established social-cognitive theory maintains that authentic mastery experiences exert the strongest influence on the development of one's self-efficacy beliefs (Bandura, 1997). Confirming this assertion, the authentic research experiences encountered by students in this study were shown to have a positive impact on students' statistical self-beliefs.

The synergy of studying both achievement and attitudinal variables in this project facilitated two meaningful positive outcomes. First, performance indicators did increase. Second, the experience of authentic statistics practices improved student self-beliefs. According to the literature, such improved attitudinal variables are linked to higher effort levels, better transference of knowledge to other domains, and higher likelihood to take future classes in the discipline.

The contributions of the two attitudes and beliefs scales are new to the research literature and will interest mathematics educators who focus on secondary and post-secondary mathematics instruction and specifically those who study statistics education. Both have high reliability and validity. The Perceived Usefulness scale specifically measures students' beliefs about the usefulness of statistics in their careers and lives. The literature suggests high perceived utility for a subject is connected to higher effort levels and increased likelihood for further coursework in related areas. The Self-Efficacy for Statistical Tasks scale measures students' beliefs about their capabilities for mastering statistics concepts and processes. Self-efficacy is a task-specific construct which correlates highly with performance measures. Self-efficacy measures also correlate highly with many affective domain variables and attitudinal variables. However, self-efficacy has often been shown to mediate the affects of other affective domain and attitudinal variables upon performance. Our examination of the relationship between Statistics

Self-Efficacy, Perceived Usefulness for Statistics and Performance is a valuable contribution in the field of statistics education. The new scales will be of use to other researchers in this field interested in comparing and contrasting teaching methods and instructional approaches.

The content test developed by the project team addresses very specific topics and will not be useful for widely testing content knowledge in statistics. However, the instrument has proved valuable in assessing content knowledge for two essential topics at the heart of the introductory statistics experience: regression and t-tests. These topics are central to basic statistics courses as well as more advanced statistics courses and to statistics courses in other disciplines. The instrumentation will prove valuable to explore the portability of these authentic discovery methods to other types of statistics classes on many types of secondary and collegiate campuses, which has formed the target for a recently submitted Type II project proposal.

To conclude, the curriculum development fills an observed need in statistics education circles, and the instrumentation provides valuable tools for researchers studying secondary and post-secondary mathematics education in general and statistics education in particular. The research findings also contribute to the current scholarship in mathematics and statistics education. Manuscripts detailing various facets of the project activities and findings have been submitted to JSE (Journal of Statistics Education), JRME (Journal for Research in Mathematics Education), and AME (Applied Measurement in Education), and the PI's expect publication of these manuscripts. Full details of these submissions are given in the Project Activities document.

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

Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. School Science and Mathematics, 108, 355-361.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. Journal for Research in Mathematics Education, 29, 41-62.

Bryce, G. R. (2005). Developing tomorrow's statistician. Journal of Statistics Education, 13(1). Retrieved from http://www.amstat.org/publications/jse/v13n1/bryce.html

Cobb, G. W., & Moore, D. (1997). Mathematics, statistics, and teaching. The American Mathematical Monthly, 104, 801-823.

Franklin, C., & Garfield, J. B. (2006). The GAISE Project: Developing statistics education guidelines for pre K-12 and college courses. In G. Burrill (Ed.), Thinking and reasoning with data and chance: 68th NCTM yearbook (pp. 345-375). Reston, VA: National Council of Teachers of Mathematics.

Guidelines for Assessment and Instruction in Statistics Education [GAISE] (2005). Guidelines for assessment and instruction in statistics education (GAISE) college report. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/gaise/GAISECollege.pdf

Hackett, G. (1995). Self-efficacy in career choice and development. In A. Bandura (Ed.), Self-efficacy in changing societies (pp. 231-258). New York: Cambridge University Press.

Hogg, R. V. (1991). Statistical education: Improvements are badly needed. The American Statistician, 45, 342-343.

Kirk, R. E. (2007). Changing topics and trends in introductory statistics. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), Best practices for teaching statistics and research methods in the behavioral sciences (pp. 25-44). Mahwah, NJ: Lawrence Erlbaum.

Landrum, R. E., & Smith, R. A. (2007) Creating syllabi for statistics and research methods courses. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), Best practices for teaching statistics and research methods in the behavioral sciences (pp. 45-57). Mahwah, NJ: Lawrence Erlbaum.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. Journal for Research in Mathematics Education, 28, 26-47.

Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. Journal for Research in Mathematics Education, 30, 65-88.

Mvududu, N. (2003). A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course. Journal of Statistics Education, 11(3). Retrieved from http://www.amstat.org/publications/jse/v11n3/mvududu.html

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. Contemporary Educational Psychology, 20, 426-443.

Pajares, F., & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. Journal of Educational Psychology, 86, 193-203.

Pajares, F., & Schunk, D. (2001). Self-beliefs and school success: Self-efficacy and self-concept in academic settings. In R. Riding & S. Rayner (Eds.), Self-perception (pp. 239-266). London: Ablex.

Pajares, F., & Schunk, D. (2002). Self and self-belief in psychology and education: A historical perspective. In J. Aronson (Ed.), Improving academic achievement: Impact of psychological factors on education (pp. 3-21). New York: Academic Press.

Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. Journal of Statistics Education, 16(1). Retrieved from http://www.amstat.org/publications/jse/v16n1/roseth.html

Smith, G., & Colleger, P. (1998). Learning statistics by doing statistics. Journal of Statistics Education, 6(3). Retrieved from http://www.amstat.org/publications/jse/v6n3/smith.html

Yesilcay, Y. (2000). Research project in statistics: Implications of a case study for the undergraduate statistics curriculum. Journal of Statistics Education, 8(2). Retrieved from http://www.amstat.org/publications/jse/secure/v8n2/yesilcay.cfm

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. American Educational Research Journal, 37, 215-246.

**Contributions to Other Disciplines:**

Contributions to Other Disciplines

The Interdisciplinary Statistics Team (IST) formed an integral part of this project and provided strong evidence of the potential impact far beyond mathematics education or statistics. Statistics instruction forms an important part of the basic education in many disciplines, both STEM and non-STEM. The IST was formed from scholars in various disciplines who conduct quantitative, statistics-based research and/or teach statistics courses within their disciplines. From STEM disciplines, the IST included representatives from biology, nursing and physical therapy along with the co-PI's who all work in mathematics departments. Other (non-STEM) disciplines represented on the IST include business, sociology, psychology, criminal justice, political science and community counseling. It bears noting that the members of the IST expressed enthusiasm for involving their students in statistical research projects of the type that they helped to define and develop. They have noted that such projects will not only deepen their students' understanding of statistics itself, but also bring more meaning to the constructs in their own disciplines. This observation highlights the interdisciplinary nature of this educational approach.

As the methods developed with the help of the IST have proved effective, results have been disseminated to instructors of basic statistics with potential for positive impact on all students who take statistics courses. The materials developed have helped statistics instructors to connect the content they teach more fully with other disciplines in a meaningful way. Thus, the positive impact is not merely one of understanding statistical tests and procedures better, but of understanding the importance, the role, and the relationship of statistics in other fields of study.

Mathematicians, as the co-PI's noted in the grant proposal for this project, are often ill-equipped to teach statistics courses having rarely conducted any survey-based studies or statistics-based research themselves. One key idea behind the Instructor's Guide (Fall 2008) and the workshop (January 2009) was to provide teachers of mathematics (both secondary and post-secondary) with all the support necessary to include these projects in their statistics courses. Inclusion of these projects by any individual instructor, however, will include training, tips and suggestions for engaging students in collaborative learning, utilizing relevant technology, writing across the curriculum, communicating mathematical ideas to others both orally and in written form, alternative assessment practices that enhance learning, and thoroughly testing a hypothesis using the Scientific Method. These best practice teaching methods will provide ideas and insights for mathematics instructors of a wide variety of courses in mathematics and other STEM disciplines.

As the current project draws to a successful close, an obvious question arises. What other statistics courses might benefit from similar instructional methods? Dozens of different types of undergraduate statistics courses exist in the context of other disciplines, and most have regression topics and t-tests as important topics. A major benefit of the instrumentation progress made thus far by the current project is that it will allow further study of the impact of authentic discovery in other statistics courses in a wide variety of undergraduate disciplines, including engineering, environmental studies, health sciences, and social sciences.

**Contributions to Human Resource Development:**

Contributions to Human Resource Development

Professional development for mathematics instructors lies at the heart of this project. Mathematicians are often tasked with teaching statistics despite the relative rarity of mathematicians who have personal experience with survey-based studies or statistics-based research. The concept of this project was not only to develop and test an approach that improved statistics teaching and learning but also to develop and test concrete materials that are comprehensive and of sufficiently high quality that instructors can incorporate these methods easily, with reasonable effort and limited need for additional professional development.

The five pilot instructors attended half a day of training before implementing the materials in their own courses for the first time. This training, in conjunction with individual guidance offered to the instructors before and during their first implementation, was a significant source of professional development for them.

Students have now utilized the Student Project Guide in several statistics classes, under the guidance of several different instructors. The instructors report overall success using the materials to guide their students. Many students have indicated that the Student Project Guide has enough information for them to brainstorm a research topic related to their career interests, develop a survey that includes demographics questions and questions targeting research variables, and develop their sampling design.

The written materials themselves will provide an immense contribution to the professional development opportunities for teachers of statistics. A full-day workshop was conducted with 19 secondary mathematics instructors, and follow-up online sessions were held in the following weeks. The professional development offered in this workshop was informed by the successes and struggles of the five instructors during the pilot phase. Attendees of this workshop were teachers of AP Statistics. However, it was noted that some of the content of the workshop could be adapted easily to benefit teachers of non-AP mathematics courses in the high school curriculum, particularly for the linear regression projects.

Abbreviated workshop-style conference sessions, smaller in scope, were conducted at local and national conferences for mathematics teachers, including the GCTM and NCTM annual conferences. These sessions were designed to target high school teachers of both AP and non-AP courses, leveraging what we learned and accomplished by conducting the first workshop.

We also occasionally hear from both high school and college instructors who have found our instructional materials through a web search. These instructors indicate that they have used the materials for their own professional development and to enhance their teaching. It is reasonable to assume that for everyone who takes the time to contact us when they choose to use these materials, there are several others who are using them without our explicit knowledge.

Thus, the project has provided professional development for the five pilot phase instructors, workshop and conference attendees, and several others who have accessed (and continue to access) the online and/or printed materials resulting from the project.

**Contributions to Resources for Research and Education:**

Contributions to Resources for Research and Education

Because the research and instrumentation in this project are based on mathematics education principles and apply to curriculum development in mathematics and statistics instruction, nearly all outcomes listed under 'Contributions within Discipline' apply to this section as well. Briefly summarizing, that section detailed the value of the educational research conducted; the value of the instrumentation developed for mathematics educators who study secondary and post-secondary instruction; and the value of the curriculum materials developed.

The co-PI's, informed by the experiences of the five pilot instructors, designed professional development experiences, both in the form of professional presentations at conferences and in the form of a teacher workshop. Both the research results and the curriculum materials resulting from this project have been disseminated fully at eight conference sessions, four of which were national conferences. Details of these presentations are given in the Project Activities document.

These presentations, corresponding papers published in conference proceedings, and curriculum materials for instructors are also all available on the project website, as noted in a prior section of this report.

The curriculum materials will be an ongoing educational resource with an even broader base of users when publication details with Freeman publisher are finalized.

**Contributions Beyond Science and Engineering:**

Contributions Beyond Science and Engineering

This project focuses on a course typically taught by mathematics or statistics departments. Yet, many students taking an introductory statistics course are not STEM majors. Therefore, the benefits of the approach and materials developed will often carry over to non-STEM undergraduate programs like teacher education, business-related majors (accounting, marketing, finance, administration), psychology, communication, sociology, clinical counseling, political science, and criminal justice.

Inviting a high school to join the project also created a broader impact. Demonstrating the efficacy of these materials for high school AP Statistics courses has enable innovative and interesting STEM content to be absorbed into secondary schools, introducing high school students to a research-based, apprentice-learning approach appropriate for undergraduates. A key feature of this project has been to infuse authentic statistics with an explicit connection to the Scientific Method. Engaging high school students in the practice of developing and testing hypotheses has deepened their experience of science and mathematics and has hopefully encouraged more of them to pursue intriguing career opportunities in the many high-need STEM disciplines.

The interdisciplinary statistics team who contributed to this project was a group of professors and professional clinicians who teach statistics and/or research classes within their discipline. The design of the statistics projects for this study impressed many of them. For example, beginning with a regression project which requires larger sample sizes early in the semester produces estimates of population parameters that make one-sample t-test projects feasible later in the semester. Several commented that they intended to utilize some of these ideas to improve and add to the research and apprentice-learning in their own more advanced statistics courses. The interdisciplinary team also discussed the fact that these research experiences are more realistic and authentic than any others that many students receive prior to their graduate studies.

The inclusion of apprentice-learning and discovery-learning modalities in these research simulation projects has influenced how our colleagues think about their own courses, content and methods. If successful in securing Type II funding for this project, we plan to explore how well these materials can be exported to other undergraduate statistics courses and to a greater variety of high school settings as part of the evaluation design.

## Conference Proceedings

## Categories for which nothing is reported:

Organizational Partners

Any Conference

## Project Activities

1) ***Instrument development***: In summer and fall 2007, we researched existing instruments in the areas of statistics attitudes, confidence, self-beliefs, and content knowledge. We used these to inform our own development of 3 instruments appropriate for use in our data collection phase. The first instrument targets perceived usefulness of statistics; the second instrument targets student self-beliefs about statistics skills (statistics self-efficacy); and the third instrument targets statistical content knowledge. Because the target statistics topics in this project are linear regression and t-tests, the self-belief and content knowledge instruments focus specifically on these concepts.

2) ***Instrument validation***: Near the end of fall semester 2007, we administered these instruments to 328 students in sections of elementary statistics at NGCSU, for the purpose of validating the instruments. With the data collected, we ran analyses for reliability and conducted KR-20 and exploratory factor analyses (EFA) on the instruments, with assistance from Frank Pajares (our instrument validation expert). Based on the findings from these analyses (see Project Findings section), minimal revision was called for; however, we did revise all three instruments as the analyses seemed to indicate appropriate.

3) ***Development of constructs and projects***: During fall 2007, the interdisciplinary team was formed (see Participants section). We met with the team late in fall 2007 to communicate the nature and goals of the project, to convey their roles, and to define their deliverables. The team members worked first individually to identify authentic research constructs in their respective fields, including guidelines on how the constructs should be defined, operationalized, measured, and quantified. In several instances, team members provided established instruments in their field that have been used to measure the construct in question. Then the team met during spring 2008 with the construct ideas that they had documented and exchanged ideas, both about constructs and about reasonable project ideas that incorporate these constructs. A partial list of constructs submitted is given below, with project ideas sometimes noted where necessary to clarify the reason for measuring the construct. These constructs and research project ideas formed a component of the teaching materials that were developed (see item #4, '***Development of Teaching Materials and Resources***'). The specific instruments and/or instructions for measuring the construct are not shown, but are given and/or referenced (where intellectual property is a potential issue) in the student project guide. Naturally, there is some overlap in constructs among related fields.

**Field: Counseling**

**<u>Suggested constructs</u>**

1. Perceived Stress Scale (PSS)
2. Satisfaction with Life Scale (SWLS)
3. Perfectionism – Almost Perfect Scale, Revised (APS-R)
4. Depression Scale (CES-D)
5. Life Orientation Test (LOT-R)
6. Rosenberg Self-Esteem Inventory (RSI)
7. Screening Test for Alcohol Abuse (CAGE)
8. Internet Addictions Test (IAT)
9. Self-Rating Anxiety Scale (SAS)

**Field: Psychology**

**Suggested constructs**

1. Math Anxiety Scale
2. Short Computer Anxiety Scale
3. Self-esteem Scale
4. Social Phobia Inventory (SPIN and Mini-SPIN)
5. Type A personality Inventory
6. Attitudes Towards Homosexuality
7. Narcissistic Personality Inventory
8. Toxic Relationship Beliefs
9. Positive and Negative Affect Survey (PANAS)
10. Achievement Motivation
11. Satisfaction with Life Scale
12. Self-Monitoring Scale
13. The General Self-Efficacy Scale
14. Locus of Control Scale
15. Emotional Intelligence Scale (Social/Interpersonal Intelligence)
16. Letter Comparison (Speed of Processing) (Intelligence – mental ability)
17. Pattern Comparison (Speed of Processing) (Intelligence – mental ability)
18. EAT-26 (Eating Attitudes Test)
19. Big Five Personality Inventory
    a. Openness to Experience/Intellect
    b. Conscientiousness
    c. Extraversion
    d. Agreeableness
    e. Neuroticism

**Field: Nursing**

**Suggested constructs**

1. Body Mass Index (BMI)
2. Rosenberg Self-Esteem Scale
3. Youth Risk Behaviors Scale
4. Blood Pressure Reading
5. Risk Score for Cardiovascular Disease
6. Pulse (resting and after exercise)
7. Nutrition Risk Scale
8. Alcohol Risk (CAGE)
9. Depression Scale
10. Pain Intensity

**Field: Physical Therapy**

**Suggested constructs**

1. Measurements of muscle strength (compare at different join positions)
2. Measurement of hand strength (e.g., left versus right)
3. Lumbar curvature measurement
4. Physical activity level
5. Hamstring tightness measurement
6. Measures of joint flexibility
7. Frequency of pain in a joint
8. Measures of balance under various conditions (with eyes closed, with dome placed over one's head, standing on a flexible surface, etc.)
9. Heart rate at different levels of physical activity
10. Heart rate as response to specific situations (not limited to physical activity)
11. Frequency of exercise

**Field: Criminal Justice**
**Suggested constructs**

1. Attitude toward Death Penalty
2. Attitude toward Gun Control
3. Racism
4. Attitude toward Spanking
5. Attitude toward Pornography
6. Attitude toward Homosexuality
7. Attitude toward Abortion
8. Attitude toward Environmental Protection
9. Attitude toward Legalization of Marijuana
10. Attitude toward Legal Drinking Age
11. Fear of Crime
12. Attitude towards Women
13. Religious Conservatism
14. Belief in the Bible
15. Belief in God
16. Religiousness
17. Measure of Binge Drinking
18. Problem Drinking Scale
19. Obsessive Compulsive Disorder (OCD) Scale
20. Attention Deficit Hyperactivity Disorder (ADHD) Scale
21. Depression Scale
22. Self-Esteem Scale
23. Stress Scale
24. Family Variables
    - Employment of mother
    - Education level of mother
    - Education level of father
    - Drinking behavior of parents
    - Divorce between parents
    - Closeness to parents
25. Life Style Variables (drinking, smoking, religion, extra-curricular activities, etc.)

**Field: Biology**
**Suggested constructs**

1. Calories consumed per day (e.g., gender or other group comparison)
2. Protein consumed per day
3. Recommended protein consumption per day (calculation)
4. Amount (e.g., weight) of trash generated in one week
5. Amount of specific trash types (plastic, paper, etc.) generated in one week
6. Participation in recycling
7. Lichen growth as indicator of air quality
8. Perceived environmental risk rankings (e.g., of specific factors such as pesticides, burning fossil fuels, etc.) – perceptions can be compared between groups and compared to established beliefs in the field
9. Attitudes of students toward environmental concerns before and after taking an environmental science course
10. Measures of physical fitness (compare between groups; e.g., nursing vs. non-nursing or athletes vs. non-athletes)
11. Heart rate (compare between persons with and without caffeine consumption)

4) ***Development of teaching materials and resources***: An *Instructor Project Guide* and *Student Project Guide* were developed to facilitate the discovery learning projects in statistics classes. These materials include a catalog of constructs and project ideas proposed by the interdisciplinary team. The *Instructor Project Guide* includes sections on project overview, literature review, course design considerations, guidelines for assessment, guidelines for assigning and directing collaborative teams, research techniques for projects, data analysis and reporting considerations, and an extensive technology guide. Several instructor resources are included, such as sample assignment pages, sample hypothesis and proposal documents, project instructions for students, and sample scoring rubrics. The Student Project Guide includes an explanation of both projects (scope and sequence), and detailed guidelines for selecting a research topic, crafting a research question, defining appropriate variables, composing a suitable survey or instrument, using sound sampling techniques, and avoiding pitfalls that introduce bias or bad data. Also included are specific assignments to guide students through the process, a technology guide, and guidelines for presentation of research results.

   As topics in the student and instructor guides are interdependent, both guides are written with extensive hyperlinks and are available in a web-based format.

5) ***Control group implementation***: The control and treatment group design was improved to help control for variability between instructors. Participating instructors taught their sections of statistics the way they would ordinarily teach any mathematics course assigned to them (before later teaching the course with the materials developed in the project.) The design was structured this way so that the control and treatment groups would not inadvertently compare instructors rather than teaching methods. Three statistics instructors (other than co-PIs Sinn and Spence) participated at NGCSU; co-PI Todd Hendricks participated at Georgia Perimeter College (GPC); and Debbie Barrineau participated at Forsyth Central High School (FCHS). Control sections were taught in spring semester 2008 for all college instructors; because the high school course is a 1-year course, the FCHS control sections spanned the full 2007-2008 school year.

6) ***Training of participating instructors***: The 5 participating instructors (3 at NGCSU, 1 at GPC, and 1 at FCHS) attended a training session with Robb Sinn and Dianna Spence in August 2008. Because each of these instructors had already met with us individually several times to learn about the project, only about half a day's formal training was needed. During training, participating instructors became acquainted with the teaching materials, practiced working through various phases of the projects (such as instrument design and data analysis), and covered logistical and administrative details of implementation and data collection.

7) ***Treatment group implementation***: Each of the instructors noted above used the materials we developed to teach one or more treatment sections. Because Robb Sinn and Dianna Spence had previously used these teaching techniques and supplied the content for the instructor and student guides, they assisted all of the participating instructors and were available as resources to them throughout this phase. The college level treatment sections (NGCSU and GPC) were taught during one or more semesters from fall semester 2008 through fall semester 2009. The high school treatment section (at FCHS) spanned the full 2008-2009 academic year, because the corresponding course is a full year course.

8) ***Workshop for statistics instructors***: We hosted a one-day workshop for statistics instructors on January 30, 2009. Nineteen participants attended the workshop, which was entitled the "Make It Real" Statistics Instruction Workshop. Although the workshop was open to both high school and college instructors, all participants were high school teachers of Advanced Placement (AP) Statistics. Sessions conducted during the workshop included:

I – Designing Quality Variables and Constructs
II – Hands-on Survey Design Session
III – Project Organization, Phases, Assessment, and Rubrics
IV – Best Practices and Avoiding Pitfalls (Panel Discussion)
V – Technology Tools and Hands-On Data Analysis
VI – Team Presentations (Participants share their work product)
VII – Instructor Observations from First Implementations

In early February 2009, we conducted a series of follow-up sessions online to allow participants to discuss the material further, including details of their own plans for implementation. Instructors who so desired were eligible to earn 1 PLU (Professional Learning Unit) credit by participating in the workshop.

9) ***Data analysis of treatment and control sections***: Data analysis was conducted on data collected from the treatment and control sections for each instructor. Analyses included examination of setting (high school, 2-year college, 4-year college); instructor; student achievement level; and the three outcome variables identified at the start of the project and measured with the instruments described above (content knowledge, perceived usefulness, and self beliefs.) Among the analyses conducted were t-test comparisons of control and treatment groups, correlations among the research variables, comparisons of outcomes by setting and by instructor, and multivariate models examining possible interactions among treatment, setting, instructor, and student achievement level, and the contribution of these variables and their interactions to the three student outcomes in question. The findings of these analyses are detailed in the project findings section.

10) ***Dissemination***: Dissemination includes three journal articles (two submitted and a third in preparation), several conference presentations, and pending publication of the curriculum materials developed during the project. All of these are detailed below.

**Journal Submissions**

a) Sinn, A. R., & Spence, D. J., Authentic discovery experiences in elementary statistics. Submitted to the *Journal of Statistics Education*, April 2010.

b) Spence, D. J., Sharp, J., & Sinn, A. R., Investigation of factors mediating the effectiveness of authentic projects in the teaching of elementary statistics. Submitted to the *Journal of Mathematic al Behavior*, June 2010.

c) Spence, D. J., Sinn, A. R., & Sharp, J., Measurement of student outcomes in elementary statistics: Development of multiple instruments. In preparation for submission to *Applied Measurement in Education*, anticipated June 2010.

## Conference Presentations

*Venue was a national meeting or conference.
†Work is published in conference proceedings.

a) * Spence, D. J., *Authentic Discovery Learning Projects in Statistics*. MathFest 2008 (a conference of MAA, the Mathematical Association of America). Madison, WI, July 2008.
b) Spence, D. J., Briggs, K. S., & Sinn, A. R., *Discovering How to Discover: Collaborative Development of Discovery Learning Materials for Statistics*. Accepting the STEM Challenge Conference (hosted by PRISM – Partnership for Reform in Science and Mathematics). Atlanta, GA, September 2008.
c) † Spence, D. J., Sinn, A. R., & Briggs, K. S., *Authentic Discovery Projects in Elementary Statistics*. Engaging Approaches: Teaching Matters Annual Interdisciplinary Conference for Teachers of Undergraduates, Barnesville, GA, April 2009.
d) † Spence, D. J., & Sinn, A. R., *Authentic Discovery Projects in Statistics*. Annual Meeting of Georgia Association of Mathematics Teacher Educators (GAMTE), Eatonton, GA, October 2009.
e) Spence, D. J., *Authentic Discovery Projects in Statistics*. Georgia Mathematics Conference (annual conference of GCTM – Georgia Council for Teachers of Mathematics). Eatonton, GA, October 2009.
f) * Spence, D. J., Bailey, B., & Sinn, A. R., *Using Authentic Discovery Projects to Improve Student Outcomes in Statistics*. Joint Mathematics Meetings of the American Mathematical Society (AMS) and Mathematical Association of America (MAA), San Francisco, CA, January 2010.
g) * Sinn, A. R., & Spence, D. J., *Training Teachers to Use Authentic Discovery Learning Projects in Statistics*. Association of Mathematics Teacher Educators (AMTE) Annual Conference, Irvine, CA, January 2010.
h) * Spence, D. J., & Sinn, A. R., *Authentic Discovery Learning Projects in Statistics*. National Council for Teachers of Mathematics (NCTM) National Conference, San Diego, CA, April 2010.

## Pending Publication of Curriculum Materials

W. H. Freeman Publisher has signed jointly with Spence and Sinn a formal letter of intent for publication for the student and instructor project guides that were developed for this grant. The publisher intends to offer the combined works as part of their online library of statistics materials and texts. The proposed title of the combined work is the *Discovering Statistics Project Guide*. Publication is pending additional formatting and content edits currently under negotiation between the authors (Spence and Sinn) and the publisher (W. H. Freeman).

# Project Findings

## I. Instrument Validation

Of the 328 statistics students who were given the instruments during the validation phase, 51 declined to participate. The data analysis was conducted for data collected from the remaining 277 students.

**A. Reliability Analyses for Perceived Usefulness and Self-Efficacy Scales**
The reliability analyses for the perceived usefulness and self-efficacy scales are shown below. The perceived usefulness scale had an overall reliability (Cronbach's alpha) of .93. The self-efficacy scale had an overall reliability of .95. Both of these are extremely high reliability coefficients.

**<u>Perceived Usefulness of Statistics Scale</u>**

- 12 items, all using 6-pt Likert scale responses
- 5 items reverse-scored (items 2, 5, 7, 8, and 10)

**Case Processing Summary: PU**

| | | N | % |
|---|---|---|---|
| Cases | Valid | 275 | 99.3 |
| | Excluded[a] | 2 | .7 |
| | Total | 277 | 100.0 |

a. Listwise deletion based on all variables in the procedure.

**Item Statistics: PU**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| PU01 | 3.56 | 1.356 | 275 |
| PU02R | 4.21 | 1.359 | 275 |
| PU03 | 3.99 | 1.345 | 275 |
| PU04 | 4.67 | 1.172 | 275 |
| PU05R | 4.43 | 1.246 | 275 |
| PU06 | 4.41 | 1.160 | 275 |
| PU07R | 4.57 | 1.237 | 275 |
| PU08R | 5.00 | 1.016 | 275 |
| PU09 | 4.73 | 1.150 | 275 |
| PU10R | 4.65 | 1.212 | 275 |
| PU11 | 3.74 | 1.544 | 275 |
| PU12 | 4.25 | 1.359 | 275 |

**Reliability Statistics: PU**

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|---|---|---|
| .932 | .933 | 12 |

**Summary Item Statistics: PU**

| | Mean | Minimum | Maximum | Range | Maximum / Minimum | Variance | N of Items |
|---|---|---|---|---|---|---|---|
| Item Means | 4.352 | 3.556 | 5.004 | 1.447 | 1.407 | .180 | 12 |
| Item Variances | 1.613 | 1.033 | 2.384 | 1.351 | 2.308 | .125 | 12 |

**Scale Statistics: PU**

| Mean | Variance | Std. Deviation | N of Items |
|---|---|---|---|
| 52.22 | 133.239 | 11.543 | 12 |

**Item-Total Statistics: PU**

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|---|
| PU01 | 48.67 | 112.829 | .645 | .474 | .929 |
| PU02R | 48.01 | 111.022 | .711 | .546 | .927 |
| PU03 | 48.23 | 110.031 | .758 | .648 | .925 |
| PU04 | 47.55 | 113.124 | .752 | .604 | .925 |
| PU05R | 47.80 | 111.236 | .778 | .662 | .924 |
| PU06 | 47.81 | 112.375 | .794 | .665 | .924 |
| PU07R | 47.65 | 111.256 | .784 | .682 | .924 |
| PU08R | 47.22 | 118.916 | .600 | .428 | .931 |
| PU09 | 47.49 | 116.609 | .616 | .448 | .930 |
| PU10R | 47.57 | 116.232 | .595 | .406 | .931 |
| PU11 | 48.48 | 107.783 | .720 | .672 | .927 |
| PU12 | 47.97 | 110.331 | .738 | .663 | .925 |

## Statistics Self-Efficacy Scale

- 15 items, all using 6-pt Likert scale responses
- No reverse-scored items

**Case Processing Summary: SE**

| | | N | % |
|---|---|---|---|
| Cases | Valid | 266 | 96.0 |
| | Excluded[a] | 11 | 4.0 |
| | Total | 277 | 100.0 |

a. Listwise deletion based on all variables in the procedure.

**Reliability Statistics: SE**

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|---|---|---|
| .954 | .954 | 15 |

**Item Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| SE01 | 4.62 | 1.072 | 266 |
| SE02 | 4.60 | 1.028 | 266 |
| SE03 | 5.03 | .929 | 266 |
| SE04 | 5.05 | .899 | 266 |
| SE05 | 4.65 | .972 | 266 |
| SE06 | 5.03 | .870 | 266 |
| SE07 | 4.55 | 1.088 | 266 |
| SE08 | 4.68 | .983 | 266 |
| SE09 | 4.54 | 1.032 | 266 |
| SE10 | 4.55 | 1.106 | 266 |
| SE11 | 4.92 | .993 | 266 |
| SE12 | 4.82 | .938 | 266 |
| SE13 | 4.73 | 1.029 | 266 |
| SE14 | 4.83 | 1.059 | 266 |
| SE15 | 4.62 | 1.073 | 266 |

**Summary Item Statistics: SE**

| | Mean | Minimum | Maximum | Range | Maximum / Minimum | Variance | N of Items |
|---|---|---|---|---|---|---|---|
| Item Means | 4.748 | 4.538 | 5.049 | .511 | 1.113 | .035 | 15 |
| Item Variances | 1.014 | .757 | 1.222 | .465 | 1.614 | .020 | 15 |

**Scale Statistics: SE**

| Mean | Variance | Std. Deviation | N of Items |
|---|---|---|---|
| 71.22 | 138.603 | 11.773 | 15 |

**Item-Total Statistics: SE**

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|---|
| SE01 | 66.60 | 119.720 | .756 | .793 | .950 |
| SE02 | 66.62 | 119.677 | .795 | .826 | .950 |
| SE03 | 66.19 | 122.229 | .755 | .706 | .950 |
| SE04 | 66.17 | 123.004 | .741 | .719 | .951 |
| SE05 | 66.57 | 120.450 | .806 | .683 | .949 |
| SE06 | 66.19 | 123.866 | .722 | .655 | .951 |
| SE07 | 66.67 | 118.660 | .791 | .712 | .950 |
| SE08 | 66.54 | 123.087 | .667 | .622 | .952 |
| SE09 | 66.68 | 120.270 | .763 | .764 | .950 |
| SE10 | 66.67 | 120.228 | .708 | .737 | .952 |
| SE11 | 66.30 | 121.570 | .733 | .703 | .951 |
| SE12 | 66.40 | 121.887 | .764 | .845 | .950 |
| SE13 | 66.49 | 121.164 | .723 | .847 | .951 |
| SE14 | 66.39 | 120.375 | .736 | .697 | .951 |
| SE15 | 66.61 | 120.874 | .703 | .669 | .952 |

**B. Exploratory Factor Analysis for Perceived Usefulness and Statistics Self-Efficacy Scales**

First, an Exploratory Factor Analysis (EFA) was conducted on the perceived usefulness and self-efficacy items together. The Perceived Usefulness items all loaded on one factor, and the first 10 items showed very strong loadings, as shown below.

| | |
|---|---|
| P1 | 58 * |
| P2 | 77 * |
| P3 | 56 * |
| P4 | 70 * |
| P5 | 95 * |
| P6 | 63 * |
| P7 | 90 * |
| P8 | 66 * |
| P9 | 49 * |
| P10 | 63 * |
| P11 | 36 * |
| P12 | 37 * |

Items 11 and 12 loaded in the high 30s, hence they were the weakest items. These items were examined to determine how they differed from the other 10. Each item referred to training in statistics being "required" for a career or making one "more valuable" in a future profession. These were the only items on which words such as “required” and “valuable” were used in the context of one’s future profession. It appears that students viewed those two items in different ways than they viewed the other 10. Supporting this observation is the fact that when these two items are correlated with the content test scores and with the self-efficacy scores, the correlations obtained are lower than the correlations for the other 10 items with these two instruments. In particular the first 10 Perceived Usefulness items correlate .26 with the content test and .56 with the self-efficacy measure. By contrast, items 11 and 12 correlate .21 with the test and .43 with self-efficacy, both of which are markedly lower than the corresponding correlations for the first 10 items.

It was determined that these two items should be reworded for greater consistency with the focus of the instrument, namely perceived usefulness. Thus, item 11 (originally “Training in statistics should be required for my career”) was modified to “My training in statistics will prove useful for my career.” Similarly, item 12 (originally “Statistical skills will be valuable in my profession”) was modified to “Having statistical skills will be useful in my profession.” It is believed that removing the terms “required” and “valuable” and focusing instead on the usefulness of statistics would render these items more consistent with the rest of the instrument.

This modification leaves the instrument with three items that explicitly use the word “useful”. This is not inappropriate, as the scale is intended to measure students’ perceptions of the usefulness of statistics. The items on the instrument were re-ordered to avoid proximity of items that sounded too similar to one another; thus, none of the items explicitly using the term “useful” are consecutive items on the instrument.

The EFA on the two scales combined revealed that within the Self-Efficacy Scale, items 11-15 loaded on one Factor, whereas items 5-10 loaded on a separate factor (though #5 loaded weakly). This is not unexpected, since different groups of items target students' self-efficacy about different statistical concepts. In particular, items 11-15 all referred to t-tests, so it makes sense that these would load together. Likewise, items 6-10 were related to correlation/regression, so it also makes sense that these would load nicely. Items 1 and 2 loaded on a third factor, and items 3 and 4 on a fourth factor. Again, this makes sense, given the content focus of these items. Items 1 and 2 rated confidence in understanding very basic concepts; items 3 and 4 measured confidence in understanding descriptive information. Item #5 referred to interpreting the meaning of a z-score; as this concept really is different conceptually than those other four factors, it is not surprising that this item appeared a bit of a misfit (loading only weakly on the same factor with items 6-10).

However, when item #5 was removed from the Self-Efficacy Scale, the factor analysis revealed only two factors: Factor 1 was composed of the 14 (remaining) Self-Efficacy (E) items, and Factor 2 was composed of the 12 Perceived Usefulness (P) items. Moreover, all factor loadings were very strong, as shown below. The interfactor correlation was .52.

| | | |
|---|---|---|
| E1 | 67 * | |
| E2 | 72 * | |
| E3 | 77 * | |
| E4 | 79 * | |
| E6 | 79 * | |
| E7 | 80 * | |
| E8 | 71 * | |
| E9 | 76 * | |
| E10 | 73 * | |
| E11 | 81 * | |
| E12 | 79 * | |
| E13 | 74 * | |
| E14 | 74 * | |
| E15 | 70 * | |
| P1 | | 67 * |
| P2 | | 74 * |
| P3 | | 78 * |
| P4 | | 72 * |
| P5 | | 87 * |
| P6 | | 74 * |
| P7 | | 88 * |
| P8 | | 74 * |
| P9 | | 51 * |
| P10 | | 60 * |
| P11 | | 79 * |
| P12 | | 75 * |

Based on this observation, the decision was made to remove item #5 from the Self-Efficacy scale, leaving the scale with 14 items. Cronbach's Alpha is still .95 with the 14-item scale, so reliability does not suffer at all with the removal of this item. Finally, with the revised self-efficacy instrument, the correlation between the self-efficacy scores and the content test scores increased from .24 to .25.

### C. Item Analysis of Content Test

A preliminary item analysis of the content test was conducted to determine not only the percentage of correct answers (CA) on each item, but also the percentage of primary and secondary distracters selected (PD and SD, respectively). This analysis is shown below.

<u>Content: Topic & Type</u>

- 21-item instrument purports to measure conceptual understanding of linear (bivariate) regression and use of t-tests (1 sample, independent samples and matched pairs varieties)
- Topic: either regression or t-test
- Type: "*Usage*" refers to basic concepts about a statistical procedure (when to use, assumptions, appropriateness, etc.) "*Inference*" refers to correct analysis based upon output of the statistical procedure.

| Item | Topic | Type | Correct Response | Primary Distracter | Secondary Distracter | CA% | PD% | SD% |
|---|---|---|---|---|---|---|---|---|
| 1 | t-Test | Inference | D | B | C | 29.1 | 23.2 | 9.2 |
| 2 | t-Test | Usage | B | D | A | 69.1 | 4.0 | 7.6 |
| 3 | t-Test | Usage | A | C | B | 80.0 | 10.9 | 4.4 |
| 4 | t-Test | Usage | D | C | B | 74.9 | 4.0 | 13.1 |
| 5 | t-Test | Usage | A | E | D | 29.9 | 10.2 | 19.7 |
| 6 | t-Test | Inference | C | B | N/A | 34.7 | 30.7 | N/A |
| 7 | t-Test | Usage | D | A | B | 47.6 | 33.1 | 8.7 |
| 8 | Regression | Usage | D | A | C | 56.4 | 13.1 | 29.6 |
| 9 | Regression | Inference | C | B | D | 14.5 | 42.9 | 11.6 |
| 10 | Regression | Inference | C | B | D | 33.5 | 10.2 | 22.9 |
| 11 | t-Test | Usage | B | C | D | 69.5 | 9.8 | 4.4 |
| 12 | t-Test | Inference | A | C | B | 39.6 | 22.2 | 18.9 |
| 13 | Regression | Inference | B | A | N/A | 50.9 | 17.8 | N/A |
| 14 | Regression | Inference | C | A | B | 40.7 | 27.3 | 24.0 |
| 15 | Regression | Inference | E | C | A | 48.7 | 28.7 | 9.5 |
| 16 | Regression | Usage | C | B | A | 39.9 | 31.9 | 11.4 |
| 17 | t-Test | Usage | D | B | A | 33.5 | 22.5 | 30.5 |
| 18 | t-Test | Inference | C | B | A | 53.3 | 13.1 | 4.0 |
| 19 | Regression | Inference | D | A | B | 55.3 | 11.6 | 16.7 |
| 20 | t-Test | Usage | E | B | C | 41.1 | 20.4 | 14.5 |
| 21 | Regression | Usage | C | D | A | 27.2 | 27.6 | 18.0 |

The varying percentage of correct answer by item was expected, as the items varied in difficulty. The percentage of correct answers ranged from 14.5% (item 9) to 80% (item 3).

## D. Reliability of Content Test

A KR-20 analysis was conducted on the content instrument, with a score of .63. It was expected and acknowledged that the KR-20 analysis would reflect the lack of homogeneity in the test, both in terms of item difficulty and in terms of the variety of concepts covered.

Of the original 21 items, three items had unacceptably low item-total correlations; these were items #16 (.07 correlation with total score), #8 (.12 correlation), and #10 (.13 correlation). It was noted that each of these three items required students to draw a quantitative conclusion based solely on visual interpretation of a scatterplot (e.g., estimating the value of "r" by examining a scatterplot). Clearly, these items draw on a substantially different skill set than does the test as a whole. Removing these items made the instrument psychometrically stronger, with a KR-20 measure of .66. The nature of the concepts covered in these items was such that their removal did not compromise the main content objectives of the test. Therefore, the revised content test consists of 18 items. These items were re-numbered, and in some cases re-ordered, from the original list shown in section III.

## II. Exploratory Data Analysis

Although the primary purpose of administering the instruments in Fall 2007 was to validate and refine the instruments, some exploratory analyses were also conducted with the data collected. The data reflected 10 sections of elementary statistics; 4 of these sections were taught by the co-PI's, who were already employing the authentic discovery project approach which was the focus of this grant. The instructors of the other 6 sections did not use these methods in their instruction. Therefore, statistical comparisons were made between the sections taught by the co-PIs (as a "preliminary treatment group") and the remaining 6 sections (as the control group). The descriptive statistics and comparisons between the groups for perceived usefulness (PU), self-efficacy (SE), and content knowledge (CK) are shown below.

***Descriptive Statistics***

| | Group | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| **PU** | **Control** | 164 | 4.2398 | 1.01126 | .07897 |
| | **Treatment** | 113 | 4.5154 | .85701 | .08062 |
| **SE** | **Control** | 158 | 4.69973 | .875527 | .069653 |
| | **Treatment** | 110 | 4.81636 | .641112 | .061128 |
| **CK** | **Control** | 163 | 8.87 | 3.240 | .254 |
| | **Treatment** | 113 | 10.82 | 3.365 | .317 |

***Independent Samples Tests***

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | |
|---|---|---|---|---|---|---|
| | F | Sig. | t | df | Sig. (*1-tailed*) | Mean Difference |
| PU Equal variances assumed | 4.277 | .040 | –2.369 | 275 | .0095 | –.27552 |
| Equal variances not assumed | | | –2.441 | 263.403 | .0075 | –.27552 |
| SE Equal variances assumed | 6.403 | .012 | –1.192 | 266 | .117 | –.11663 |
| Equal variances not assumed | | | –1.259 | 265.297 | .105 | –.11663 |
| CK Equal variances assumed | .095 | .758 | –4.859 | 274 | .000 | –1.958 |
| Equal variances not assumed | | | –4.825 | 235.095 | .000 | –1.958 |

The t-tests were conducted and reported as 1-tailed because it was hypothesized that the preliminary treatment group would specifically have *higher* scores in perceived usefulness, self-efficacy, and content knowledge. Results show that the preliminary treatment group had significantly higher scores than did the control group on the perceived usefulness scale ($p < .01$) and on the content knowledge test ($p < .001$). The treatment group also had higher self-efficacy, but this difference did not achieve statistical significance. Also noteworthy is that the treatment group had significantly lower variability on both the perceived usefulness and the self-efficacy scales ($p < .05$), whereas the variability on the content knowledge instrument was not significantly different between the two groups.

## III. Full Pilot Study of Curriculum Materials

### Methodology and Data Collection

Five pilot instructors participated, three at a four-year college, one at a two-year college, and one high school teacher of AP Statistics. Each pilot instructor taught one or more sections of their statistics course in their normal fashion and without use of any of the Authentic Discovery materials. These made up the control sections of the study. Each pilot instructor then taught one or more sections of a statistics course in the subsequent semester (or year, in the case of the high school) using the discovery project materials, after a series of meetings and a training session. These sections were the treatment sections. As noted above, the three student outcomes measured were statistics content knowledge, perceived usefulness of statistics, and statistics self-efficacy. These were measured for participating pilot instructors' control and treatment sections of elementary statistics.

By intent, the research design was not fully experimental. The researchers anticipated that the effect of instructor upon instruction would mediate the treatment effect. Moreover, the use of volunteer instructors, individual sections of courses with a common syllabus, and the high school setting required an instructor-focused pilot test design. The quasi-experimental design utilized provided reasonably consistent results across sections and instructors, and there is no evidence from the pilot data that significant class differences were driving the variability. For example, a $\chi^2$ test of independence did not produce evidence of significant differences in grade distribution between control and treatment groups ($p = .261$). The main differences in the grade distributions appeared to be related to the different numbers of sections in the control group compared to the treatment group, not to grade inflation due to project scores, which did comprise a portion of each student's final grade in the treatment groups.

***Test of Independence***

| Phase / Grade | A | B | C | DFW | $\chi^2$ | *P* |
|---|---|---|---|---|---|---|
| **Control** | 58 | 51 | 14 | 14 | 4.003 | .261 |
| **Treatment** | 77 | 52 | 29 | 14 | | |

For each section in the study, data were collected in the week prior to exams. The only exception was the high school sections where, due to AP testing schedule and nine-month course calendar, students were tested approximately two weeks prior to their final statistics exam. The data collection occurred just after the AP Statistics Exam was administered, but not all students in the AP classes chose to attempt the AP Exam. The numeric variables recorded for the pilot test data set were the students' scores on the three quantitative instruments plus their final grade in the statistics course. One of the 4-year college instructors was unable to complete the scheduled treatment sections due to health complications, leaving data collected from four pilot instructors to be included in the analysis.

## Data Analysis

Data for control and treatment groups were analyzed using t-tests, first for all instructors combined, then by individual instructor. The results of all t-tests are shown below.

***Main Scales – Combined Groups***

| Instrument | Control | | | Treatment | | | *t* | *p* |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | | |
| **Content Knowledge** | 138 | 6.78 | 2.44 | 172 | 7.21 | 3.08 | 1.36 | .088 |
| **Perceived Usefulness** | 137 | 50.42 | 10.05 | 172 | 51.40 | 11.24 | 0.81 | .208 |
| **Self-Efficacy** | 129 | 59.64 | 14.24 | 172 | 62.57 | 11.61 | 1.86 | .032 |

***Sub-Scales – Combined Groups***

| Sub-Scale | Control | | Treatment | | *t* | *p* |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| **Content Knowledge: Regression** | 1.99 | 1.17 | 2.20 | 1.47 | 1.37 | .086 |
| **Content Knowledge: t-Test Usage** | 3.21 | 1.52 | 3.46 | 1.84 | 1.30 | .097 |
| **Content Knowledge: t-Test Inference** | 1.58 | .988 | 1.55 | .951 | -0.25 | .425 |
| **Self-Efficacy: General** | 16.99 | 4.30 | 18.01 | 3.93 | 2.10 | .018 |
| **Self-Efficacy: Regression** | 20.55 | 5.25 | 22.14 | 4.59 | 2.74 | .035 |
| **Self-Efficacy: t-Tests** | 22.10 | 5.73 | 22.42 | 5.22 | 0.50 | .308 |

The self-efficacy for statistics scores showed an overall increase that was significant ($p = 0.032$). The perceived usefulness scores were the most volatile of the three instruments. While the slight gains were in no way statistically significant, the students who experienced the intervention did not on average find statistics less useful than their control group counterparts. The gains on the Content Knowledge test were weakly significant ($p = .088$). One may argue a liberal alpha is appropriate here, despite the reasonable sample size, because the expected effect size was small. The footprint of implementing these projects included a minimum of 90 minutes of class time devoted to presentations and additional class minutes lost to the logistics of assigning and discussing projects. Hence, the intervention was conducted with fewer total minutes of instruction overall, and many fewer minutes of direct instruction. The field experiences associated with the discovery projects appear to have compensated for that removal of direct instruction and do provide evidence of small but significant treatment effect.

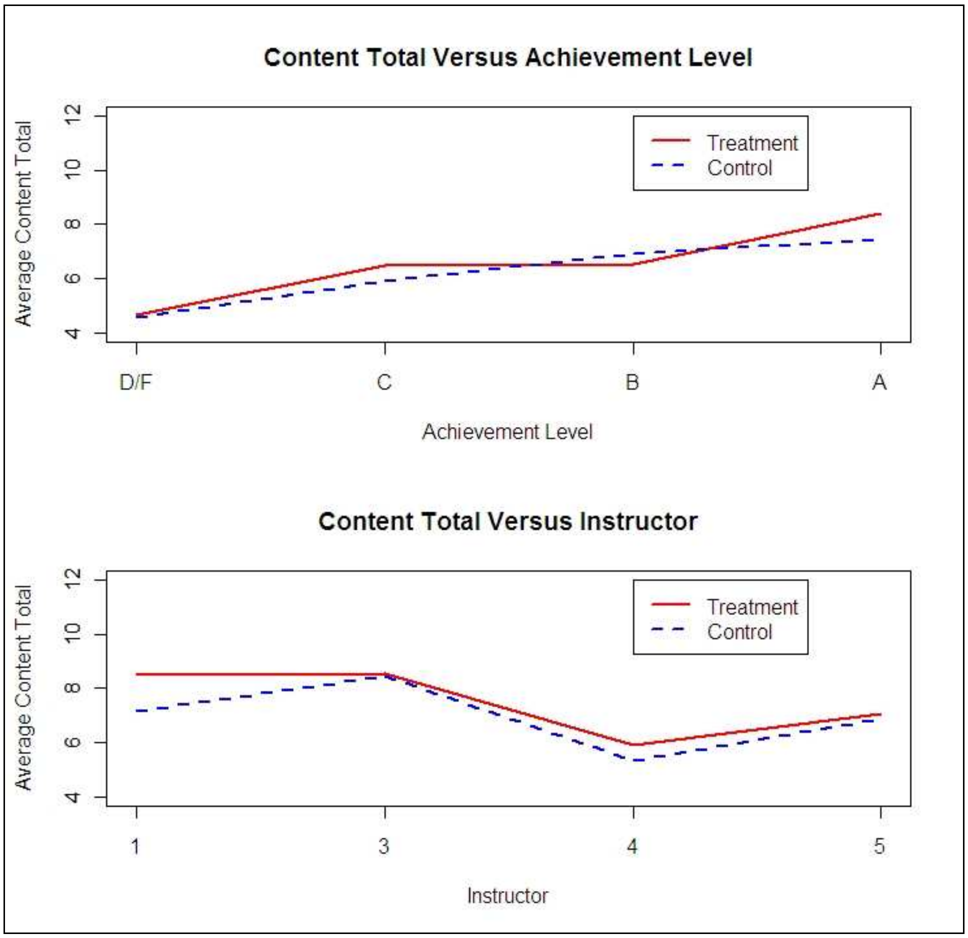
***Instruments and Sub-Scales by Instructor***

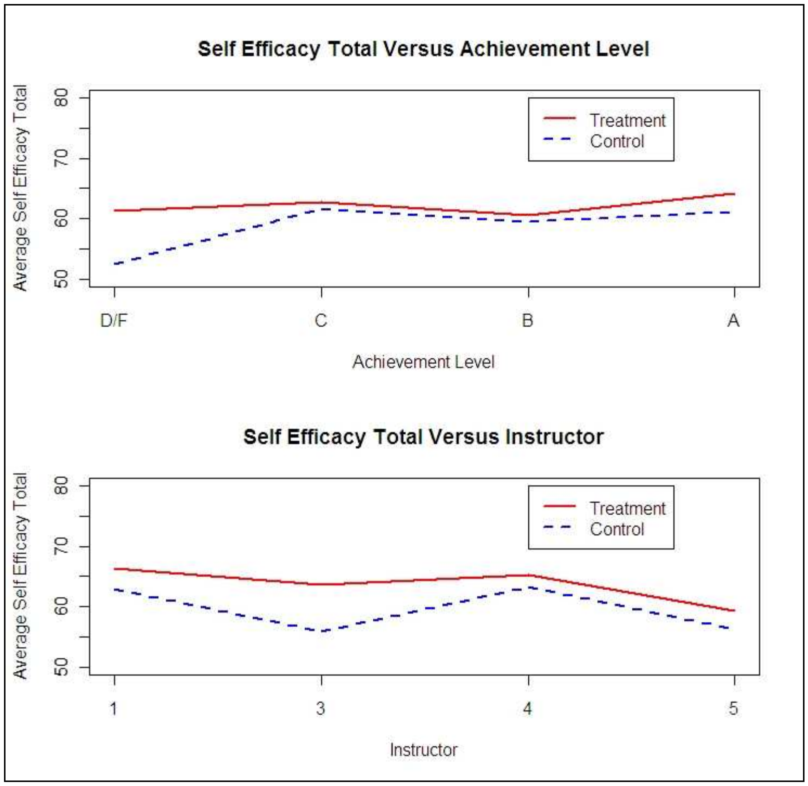
| Scale | Instructor | Control | | | Treatment | | | *t* |
|---|---|---|---|---|---|---|---|---|
| | | N | Mean | SD | N | Mean | SD | |
| **Content Knowledge – Entire Instrument** | **1** | 41 | 7.17 | 2.365 | 24 | 8.50 | 3.093 | 1.817** |
| | **3** | 20 | 8.45 | 2.762 | 28 | 8.54 | 2.502 | 0.110 |
| | **4** | 33 | 5.33 | 1.708 | 44 | 5.93 | 2.546 | 1.233 |
| | **5** | 43 | 6.86 | 2.178 | 76 | 7.05 | 3.253 | 0.385 |
| **Content Knowledge – Linear Regression** | **1** | 41 | 2.12 | 1.029 | 24 | 2.58 | 1.472 | 1.354* |
| | **3** | 20 | 2.10 | 1.119 | 28 | 2.29 | 1.384 | 0.513 |
| | **4** | 33 | 1.24 | 1.032 | 44 | 1.50 | 0.928 | 1.132 |
| | **5** | 43 | 2.42 | 1.180 | 76 | 2.45 | 1.628 | 0.111 |
| **Content Knowledge – Design/Usage of t-Tests** | **1** | 41 | 3.54 | 1.398 | 24 | 4.33 | 1.949 | 1.756** |
| | **3** | 20 | 4.50 | 1.638 | 28 | 4.93 | 1.464 | 0.934 |
| | **4** | 33 | 2.91 | 1.284 | 44 | 3.05 | 1.656 | 0.407 |
| | **5** | 43 | 2.58 | 1.314 | 76 | 2.88 | 1.657 | 1.087 |
| **Content Knowledge – Inference on t-Tests** | **1** | 41 | 1.51 | 0.925 | 24 | 1.58 | 1.018 | 0.281 |
| | **3** | 20 | 1.85 | 1.137 | 28 | 1.32 | 0.819 | -1.776 |
| | **4** | 33 | 1.18 | 0.917 | 44 | 1.39 | 0.920 | 0.967 |
| | **5** | 43 | 1.86 | 0.915 | 76 | 1.72 | 0.974 | -0.765 |
| **Perceived Usefulness – Entire Instrument** | **1** | 41 | 50.59 | 10.361 | 24 | 54.88 | 9.014 | 1.751** |
| | **3** | 20 | 49.35 | 9.016 | 28 | 46.75 | 12.183 | -0.850 |
| | **4** | 32 | 51.41 | 9.339 | 44 | 53.30 | 10.064 | 0.843 |
| | **5** | 43 | 50.23 | 10.963 | 76 | 50.92 | 11.722 | 0.321 |
| **Self-Beliefs – Entire Instrument** | **1** | 32 | 62.97 | 15.073 | 24 | 66.38 | 9.221 | 1.044 |
| | **3** | 20 | 55.95 | 20.289 | 28 | 63.71 | 12.304 | 1.523* |
| | **4** | 33 | 63.21 | 9.512 | 44 | 65.25 | 9.022 | 0.951 |
| | **5** | 43 | 56.23 | 12.569 | 76 | 59.39 | 14.398 | 1.250 |
| **Self-Beliefs – General Statistics Concepts** | **1** | 32 | 17.94 | 4.272 | 24 | 19.33 | 3.116 | 1.414* |
| | **3** | 20 | 16.45 | 6.057 | 28 | 18.64 | 3.540 | 1.452* |
| | **4** | 33 | 17.39 | 3.316 | 44 | 18.95 | 2.853 | 2.168** |
| | **5** | 43 | 16.26 | 4.042 | 76 | 16.80 | 4.505 | 0.680 |
| **Self-Beliefs – Linear Regression** | **1** | 32 | 21.50 | 5.292 | 24 | 22.71 | 3.520 | 1.024 |
| | **3** | 20 | 18.75 | 6.995 | 28 | 23.04 | 4.393 | 2.420** |
| | **4** | 33 | 22.30 | 3.917 | 44 | 22.09 | 3.771 | -0.239 |
| | **5** | 43 | 19.35 | 4.825 | 76 | 21.66 | 5.346 | 2.411** |
| **Self-Beliefs – t-Test Concepts** | **1** | 32 | 23.53 | 6.091 | 24 | 24.33 | 3.964 | 0.596 |
| | **3** | 20 | 20.75 | 7.684 | 28 | 22.04 | 6.009 | 0.624 |
| | **4** | 33 | 23.52 | 3.598 | 44 | 24.20 | 3.638 | 0.828 |
| | **5** | 43 | 20.63 | 5.451 | 76 | 20.93 | 5.594 | 0.292 |

$^{*}p < .10$  $^{**}p < .05$

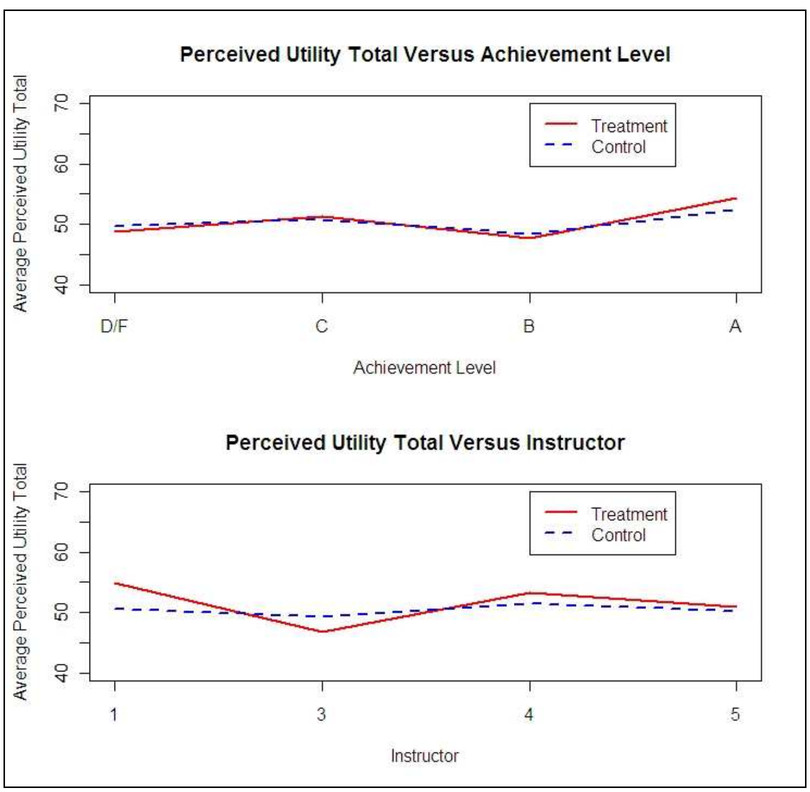
It became clear during the preliminary data analysis that the impact of these curriculum materials and discovery projects varied a great deal by instructor. It also appeared that student outcomes varied by the student's overall achievement level in the class, as represented by the student's final grade in the course; i.e., the impact of the projects may have been different for 'A' students than for 'B' or 'C' students, and so on.

The mean plots below show the control and treatment outcomes by achievement level and by instructor. The ‘D’ and ‘F’ categories in achievement level were combined because relatively few of either grade were assigned. It is worth noting that content knowledge and self-efficacy were higher in the treatment groups for all 4 instructors, though the magnitude of the effect varied by instructor.





Self Efficacy Total Versus Achievement Level
Average Self Efficacy Total
50
60
70
80
Treatment
Control
D/F
C
B
A
Achievement Level
Self Efficacy Total Versus Instructor
Average Self Efficacy Total
50
60
70
80
Treatment
Control
1
3
4
5
Instructor

Perceived Utility Total Versus Achievement Level
Average Perceived Utility Total
40
50
60
70
Treatment
Control
D/F
C
B
A
Achievement Level
Perceived Utility Total Versus Instructor
Average Perceived Utility Total
40
50
60
70
Treatment
Control
1
3
4
5
Instructor

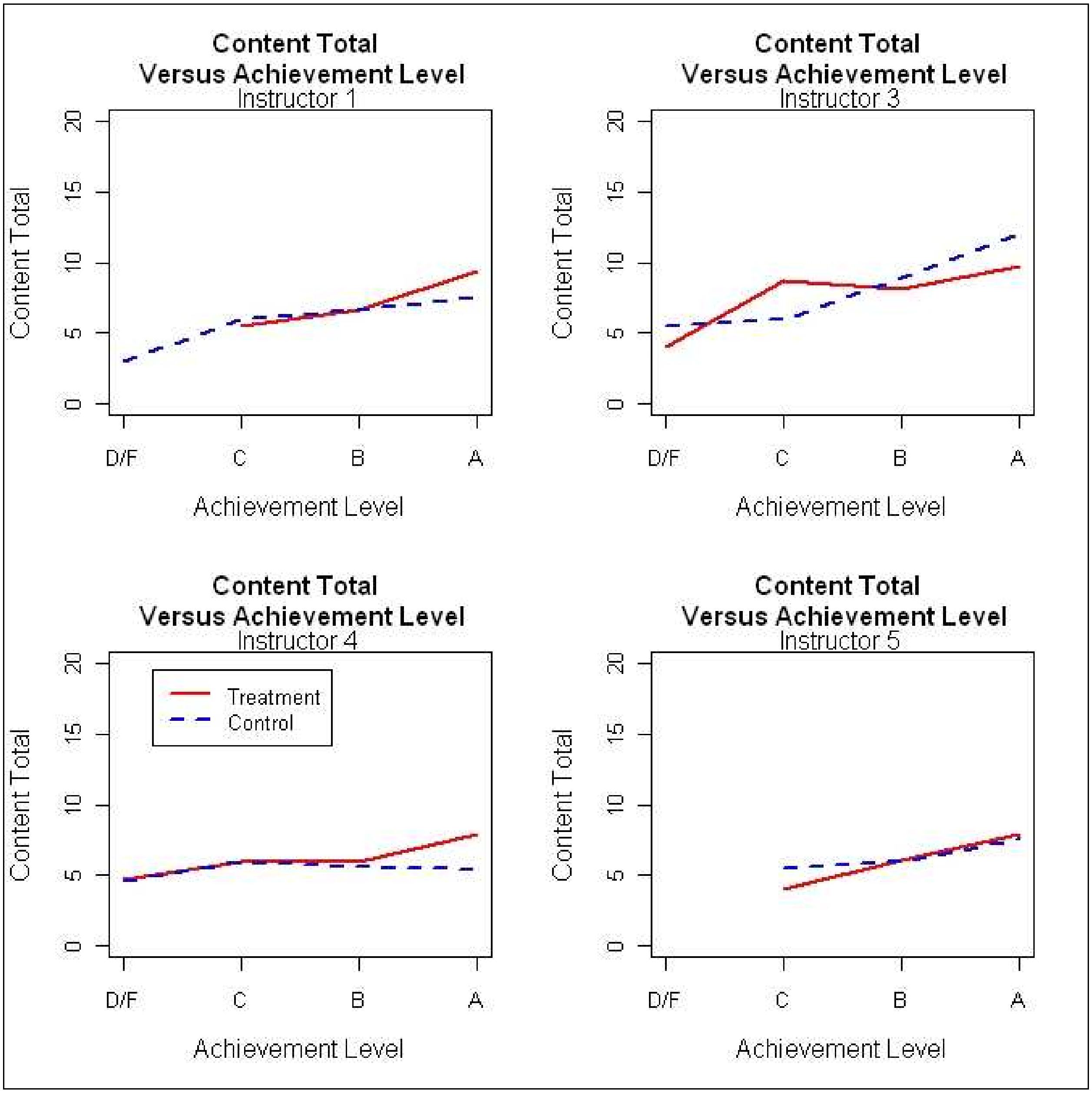
Because there were potential instructor and achievement level effects, data were then analyzed using a multivariate nested design with crossed factors. The crossed factors were instructor and treatment group (control or treatment). Then student achievement level was a nested factor within instructor × treatment. The reason for nesting achievement level within instructor and treatment was that instructors often display different standards for grading. After examining the distribution of grades for each instructor, it appeared that this was the case. Some instructors routinely gave a grade of A to only 2 or 3 students, accounting for less than 10% of the students in the section. Yet others often gave a grade of A to over one third of the students in their sections. Considering this observation combined with the fact that these grades were assigned in different academic settings (high school, 2-year college, 4-year college), it was reasonable to interpret that a given grade for one instructor did not necessarily have the same meaning as that same grade had for another instructor.

The results for the multivariate nested linear model are shown below. For the three main instruments, these include statistics for the model, intercept, main effects (treatment and instructor), interaction effect (treatment × instructor), and achievement nested within the interaction. The p-values given for main effects are bi-directional unless the main effect is consistently unidirectional (i.e., treatment exceeds control for every instructor and at every grade level). Similar analyses are in progress for the instrument sub-scales.

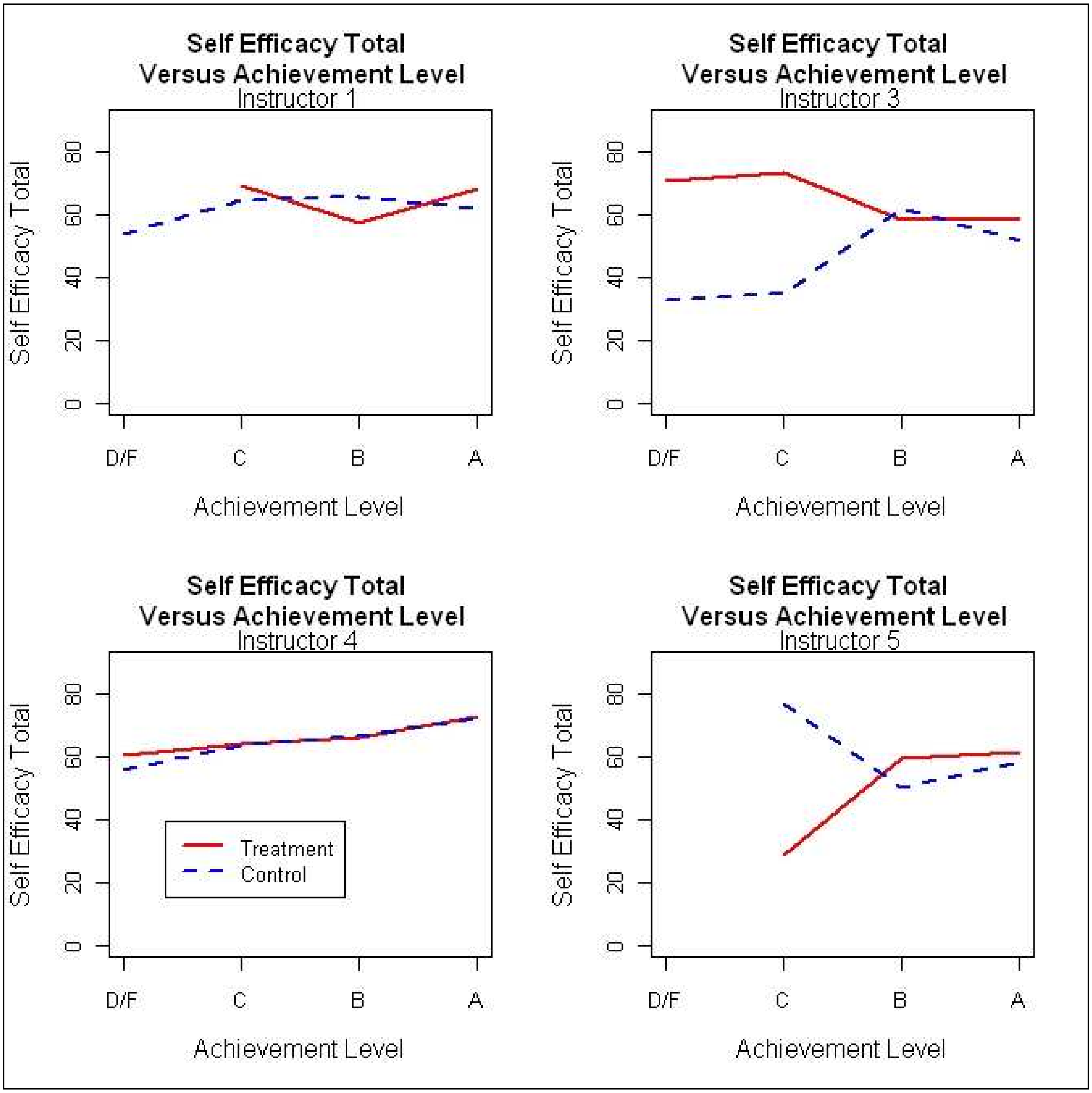
| **Source** | **Dependent Variable** | ***F*** | ***p*** |
|---|---|---|---|
| Model | Content Knowledge | 3.473 | .000 |
| | Self-Efficacy | 4.022 | .000 |
| | Perceived Usefulness | 2.427 | .000 |
| Intercept | Content Knowledge | 707.808 | .000 |
| | Self-Efficacy | 2800.994 | .000 |
| | Perceived Usefulness | 2789.110 | .000 |
| Treatment | Content Knowledge | 0.396 | .530 |
| | Self-Efficacy | 2.393 | .062 |
| | Perceived Usefulness | 2.579 | .109 |
| Instructor | Content Knowledge | 3.392 | .019 |
| | Self-Efficacy | 7.155 | .000 |
| | Perceived Usefulness | 1.006 | .390 |
| Treatment × Instructor | Content Knowledge | 0.749 | .524 |
| | Self-Efficacy | 7.895 | .000 |
| | Perceived Usefulness | 3.619 | .014 |
| Achievement Level ***within*** Trtmt × Instr | Content Knowledge | 2.420 | .001 |
| | Self-Efficacy | 4.122 | .000 |
| | Perceived Usefulness | 2.676 | .000 |

The overall model was significant for all three measures, as was the contribution of achievement level within the interaction of treatment × instructor. Further, the interaction of treatment × instructor was significant for self-efficacy and perceived usefulness.

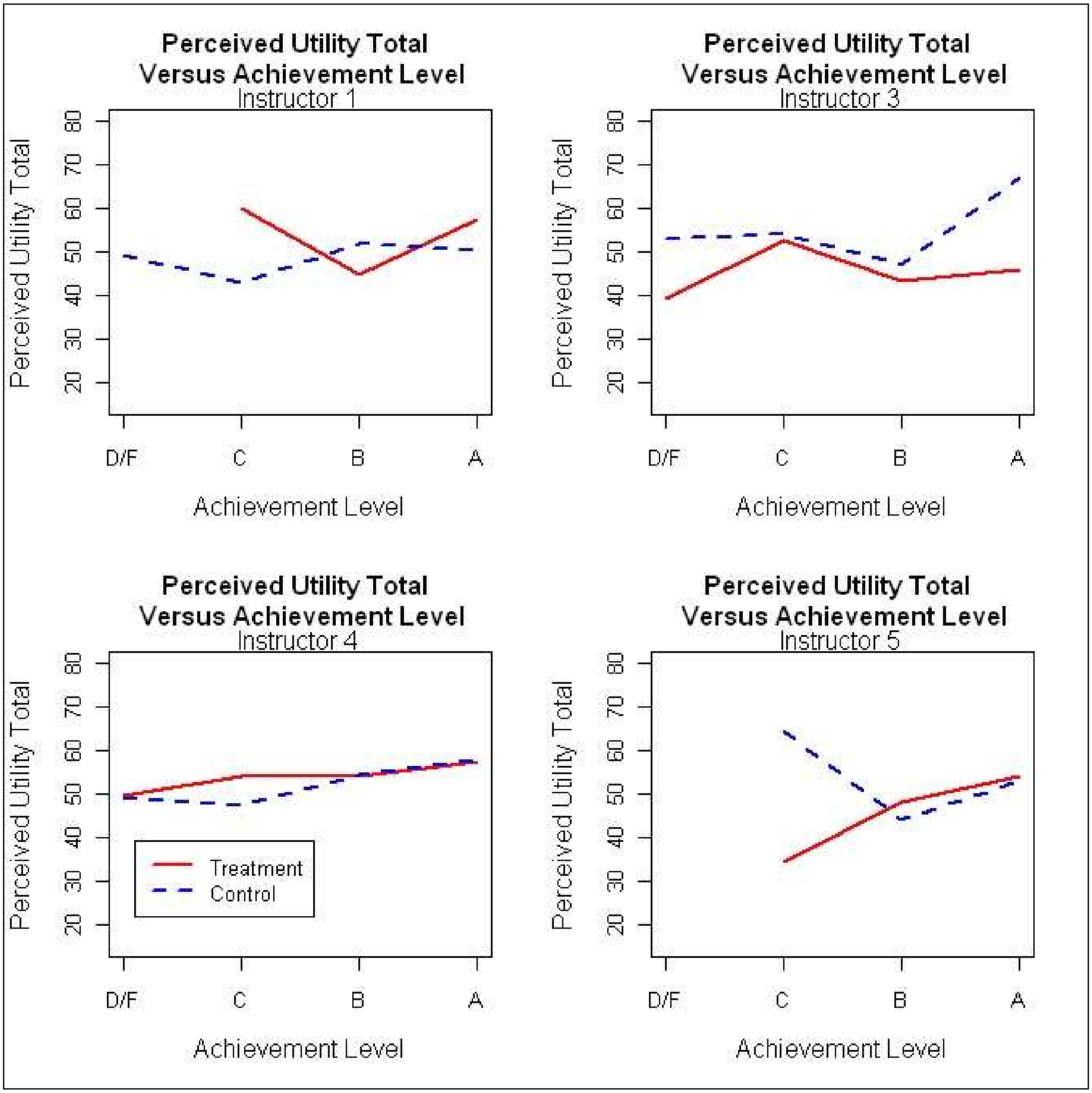
For each instrument, the mean plots for control and treatment groups by achievement level are shown below for each instructor individually.



Content Total
Versus Achievement Level
Instructor 1
Content Total
0
5
10
15
20
D/F
C
B
A
Achievement Level
Content Total
Versus Achievement Level
Instructor 3
Content Total
0
5
10
15
20
D/F
C
B
A
Achievement Level
Content Total
Versus Achievement Level
Instructor 4
Treatment
Control
Content Total
0
5
10
15
20
D/F
C
B
A
Achievement Level
Content Total
Versus Achievement Level
Instructor 5
Content Total
0
5
10
15
20
D/F
C
B
A
Achievement Level



Self Efficacy Total
Versus Achievement Level
Instructor 1
Self Efficacy Total
Achievement Level
D/F
C
B
A
0
20
40
60
80
Self Efficacy Total
Versus Achievement Level
Instructor 3
Self Efficacy Total
Achievement Level
Self Efficacy Total
Versus Achievement Level
Instructor 4
Treatment
Control
Self Efficacy Total
Achievement Level
Self Efficacy Total
Versus Achievement Level
Instructor 5
Self Efficacy Total
Achievement Level




The plots help to visualize some of the interplay between instructor, treatment, and achievement level. For example, instructors 1 and 4 had greater content knowledge gains among A students than among students of other achievement levels. However, instructor 3 only saw content knowledge gains among C students.

## Discussion

The modest gains in content knowledge appeared within two of the three subscales– linear regression and usage of t-tests. No significant gains were obtained for content knowledge in the interpretation of t-test results– i.e., statistical reasoning and inference. The PI's believe this is due in part to the order in which the topics were covered, in conjunction with the compressed class schedule required to accommodate the addition of projects. All participating instructors covered t-tests at the very end of their course and were rushed for time. In most treatment sections, instructors were not able to spend as much instructional time on t-tests as they had in their respective control sections. Thus, students' grasp of the most challenging content– statistical reasoning associated with significance tests– showed no gain in the treatment groups. Fortunately, student understanding in this area showed no significant decline, either.

Given that self-efficacy strongly influences career choices, especially for careers related to mathematics and science, the impact on self-efficacy for statistics is an important finding. It is also worth noting that while the content test focused only on regression and t-test topics, the self-efficacy instrument measured global statistical tasks that would be utilized during the projects. Examples include understanding interpretation and properties of means, medians and standard deviations. The content test did not measure students' more basic knowledge of general statistical concepts, such as interpreting measures of central tendency and spread. This may help to explain why the analysis revealed a clearly significant impact on statistics self-efficacy, but only a marginal impact on content knowledge with borderline significance. For the purposes of measuring the impact of these discovery projects, the content knowledge instrument(s) should be revised to reflect student knowledge of these ideas.

Similarly, anecdotal evidence of improved student learning was provided via pilot instructor reports, journal responses, and follow-up conversations. In particular, the instructors found that significant student attention was focused on topics not necessarily emphasized in an introductory course. Those topics included research design considerations, sampling issues, survey design principles, the importance of data entry accuracy and the difficulties created by "junk" responses on surveys. As noted above, the content knowledge instrument should be updated to measure such serendipitous gains in understanding.

Certain student groups exposed to the authentic discovery process reported lower self-efficacy than did their control group counterparts for statistical tasks relating to real-world statistics-based research. The PI's have conjectured that exposure to the complexities of some aspects of real-world statistics-based research increased the accuracy of these students' self-efficacy judgments which, prior to these authentic experiences, were possibly quite naive. This explanation is particularly likely among high school students; the comparison of self-efficacy scores between treatment and control for instructor #5 (the high school instructor) show a marked drop in self-efficacy among 'C' students in the treatment group. It seems plausible that 'C' students in the high school setting are among those most likely to have unrealistic self-beliefs until they gain more first-hand experience with the tasks in question.

The PI's anticipated that using the discovery projects would make a significant difference on perceived usefulness of statistics. These expectations were based on pre-pilot investigations at their home institution where the Introductory Statistics course is largely populated by juniors and seniors, who often have a better idea of their eventual career trajectory than did many students in the pilot study, where more than half of the participants were at the two-year college or high school test sites. Many of the items in the perceived usefulness instrument make specific reference to applying statistics in the context of the student's future career; if the students cannot clearly imagine their future careers, it is conceivable that they may not be able to imagine such details as using statistics in those careers. Thus, these contrasting populations may have contributed to the vast difference in results between the pilot and the initial exploratory study. In the exploratory study, conducted largely among juniors and seniors in a 4-year college, the impact on perceived usefulness was clearly significant ($p < .01$). Yet in the pilot, conducted in a group heavily populated by high school and two-year college students, no significant impact was detected. To target high school and junior college perceptions of usefulness with more relevance and accuracy, the instrument should be revised to target real world applications not necessarily tied to one's own career plans, which may be defined poorly or not at all.

In summary, the statistical evidence suggests that using discovery projects lends significant benefit to student self-beliefs and potential benefit to student content knowledge in statistics. The PI's have further identified a number of revisions to the instrumentation which may help to reveal more fully the benefit to student content knowledge, and possibly even to student perceptions of the usefulness of statistics. The findings also point to significant interactions between use of the discovery projects, instructor style, and student achievement level; the nature of these interactions and their impact on student learning and attitudes merit further investigation.