Annual Report for Period: 06/2007 - 05/2008
Submitted on: 04/13/2008
Principal Investigator: Sinn, Robb .
Award ID: 0633264
Organization: North Georgia College
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, Andrea
Worked for more than 160 Hours: No
Contribution to Project:
Andrea Hendricks has not been able to participate in the project as planned. After having a baby, she had her teaching assignment changed to exclusively online classes. We learned this in Fall 2007 and invited another instructor at Georgia Perimeter College to take her place as our co-PI. This instructor is Todd Hendricks. He has accepted our invitation and has become involved in the project. We have submitted a formal request (through FastLane) to change our co-PI from Andrea Hendricks to Todd Hendricks, and the FastLane system indicated that this request has been forwarded to the Sponsored Programs Office.

Name: Hendricks, Todd
Worked for more than 160 Hours: No
Contribution to Project:
As noted in the participant notes for our original co-PI Andrea Hendricks, we have been in the process of replacing Andrea Hendricks with Todd Hendricks as our co-PI.

Post-doc

Graduate Student

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

We have 3 collaborators in our department (Mathematics) to teach control and treatment sections of the statistics course. They are teaching control sections this semester and will subsequently use the materials we are developing to teach treatment sections. These Mathematics faculty are Brad Bailey, Karen Briggs, and John Holliday.

We also have an interdisciplinary team of 10 faculty from other departments who are developing authentic research constructs and project ideas from their fields that can be used in the project workbook being 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), Kim Melton (Business Administration), Mohammad Nourbakhsh (Physical Therapy), Brent Paterline (Criminal Justice), Robert Rogan (Education), 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:

Teaching skills and experience are expected to be significant for project participants when the treatment groups are implemented in fall 2008.

Outreach Activities:

Increasing students' understanding, participation, and appreciation of statistics is the primary goal of this project. The treatment groups scheduled for fall 2008 will be intended to foster this increase.

Journal Publications

Books or Other One-time Publications

Web/Internet Site

URL(s):

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

Description:

This site gives a high level overview of the project and provides links to reference materials and relevant documents. In particular, it has served as a resource for our interdisciplinary team members. Planned additions to the site include one or more pages with resources for instructors as they implement the treatment groups; and links to appropriate conference presentation and publication abstracts as they become available.

This site may also house information about the instructor workshops to be arranged in fall 2008; however, we may decide to place workshop information on a separate site altogether.

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:

To date, the instrument has not been shared with anyone outside this project, nor has it been used outside the context of this project. The instrument will be shared closer to the end of the research phase of the project, both in conference presentations and in relevant articles that are submitted for publication.

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:

To date, the instrument has not been shared with anyone outside this project, nor has it been used outside the context of this project. The instrument will be shared closer to the end of the research phase of the project, both in conference presentations and in relevant articles that are submitted for publication.

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:

To date, the instrument has not been shared with anyone outside this project, nor has it been used outside the context of this project. The instrument will be shared at instructor workshops, in conference presentations, and in relevant articles that are submitted for publication.

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 is therefore 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 will improve learning outcomes for many types of statistics courses. The instrumentation and educational research results will be of more interest to the mathematics education community at large, 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. There would appear to be both need for these types of curricular advances together with means for measuring student progress broadly-defined.

Statistics educators have repeatedly suggested improvements ought to be made, especially focusing on implementation of the scientific method utilizing authentic statistical experiences (as is the case herein), but these calls for improvement have not been widely heeded (Bryce, 2005). Researchers have found statistics courses based on more constructivist models improved student attitudes toward statistics and that personal relevance is important for successful learning in statistics (Mvududu, 2003). A researcher used case-study methodology to evaluate a real-world, project-based approach to learning statistics and found that students learned more from the project than from any other instructional component of the course. The researcher further reported improvements in student motivation (Yesilcay, 2000).

The curriculum development portion of the project addresses the needs for more authentic statistics experiences while the scale development addresses two aspects of student motivation for statistics: students' perceived usefulness of statistics and students' self-efficacy for statistics tasks. The synergy of studying both achievement and attitudinal variables as the pilot phase moves forward is that three meaningful positive outcomes can result. First, performance indicators may increase. Second, performance may be equivalent to traditional methods, but the experience of authentic statistics practices may improve the attitudinal variables. According to the literature, 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. Third, the project may find that both performance and attitudinal variables are improved.

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. Determining the relationship between Statistics Self-Efficacy, Perceived Usefulness for Statistics and Performance will be 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, one possible project target for a Phase II project proposal.

In sum, the curriculum development fills an observed need in statistical education circles, and the instrumentation provides valuable tools for researchers studying secondary and post-secondary mathematics education in general and statistics education in particular.

References

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

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).
URL: www.amstat.org/publications/jse/v11n3/mvududu.html.

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

Contributions to Other Disciplines:

Contributions to Other Disciplines

The Interdisciplinary Statistics Team (IST) formed an integral part of this project and provides strong evidence of the potential impact far beyond mathematics education or statistics. Statistics instruction forms an important part of the basic education in many STEM disciplines, especially the health sciences and engineering. One goal of future investigations if the current project bears significant fruit will be to ask about the portability of these instructional methods to other types of undergraduate statistics courses both in STEM disciplines and beyond.

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 have already expressed enthusiasm for involving their students in statistical research projects of the type they are helping 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 being developed with the help of the IST prove effective, the results will be disseminated to instructors of basic statistics with potential for positive impact on all students who take statistics courses. The materials being developed will also help 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 and the workshops (Fall 2008) is 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.

If the current project proves to be successful, an obvious question arises. What other statistics courses might benefit from similar instructional methods? Dozens of different types of undergraduate statistics courses exist, and most have regression topics and t-tests as important topics. Research utilizing ANOVA is extremely similar to that for t-tests. 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 STEM settings, for example, engineering, environmental studies and many health 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 so comprehensive and of such high quality that instructors would be able to incorporate these methods easily, with reasonable effort and limited need for additional professional development.

Students in the co-PI's classes have already been utilizing and critiquing portions of the Student Project Guide and report that they could have undertaken their projects without much in-class demonstration or discussion, except for the data entry and data analysis process. They 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, develop their sampling design. Additional high quality materials are also in development.

The written materials themselves will provide an immense contribution to the professional development opportunities for teachers of statistics. Further, full-day workshops are planned with 40 û 50 high school and college mathematics instructors. The professional development within these workshops will be informed by the successes and struggles of the five instructors during the pilot phase.

The project provides professional development for the five pilot phase instructors, the forty or more workshop attendees, and for anyone who accesses the online and/or print 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 instruction generally and statistics instruction specifically, nearly all outcomes listed under 'Contributions within Disclipline' apply to this section as well. Briefly summarizing, that section detailed the value of the instrumentation work to mathematics educators who study secondary and post-secondary instruction and the value of the curriculum materials developed.

The project has evolved slightly since the proposal and will thus provide another benefit to the mathematics educators. The pilot phase now includes three mathematicians from NGCSU who have between one and seven semesters of experience teaching introductory statistics courses. As originally proposed, a high school AP statistics teacher and mathematics professor from a two-year college will also participate in the pilot phase.

The pilot phase utilizes a treatment vs. control design. Each instructor provides his or her own control group by testing their Spring 2008 students on all instruments, teaching without the use of the authentic discovery materials. In Fall 2008, each instructor will be supported by the co-PI's with training and professional development opportunities, and will have access to all project-developed course materials. Then, each instructor's Fall 2008 classes will be tested to determine the impact of the authentic discovery projects in their courses. This quasi-experimental design will provide the best measure of the improvement gains that can be expected from utilizing these authentic discovery projects in statistics courses. The control group phase of the pilot is nearing completion.

The addition of three more instructors to the pilot phase will reduce the individual-specific professional development outcomes. Having five instructors instead of two will allow the co-PI's to compare and contrast the professional development opportunities that afford the best chances of success in implementing these authentic discovery projects into a statistics course.

The co-PI's, informed by the experiences of these five instructors during the pilot phase, will design professional development experiences both in workshop form (as described in the section on Contributions to Human Resources Development) and for professional presentation at conferences (as described below). Thus, in addition to the instrumentation and curriculum contributions mentioned above, this project will add professional development opportunities to inform and enhance the skill sets of mathematics instructors.

As the pilot phase progresses enough to allow useful information to be shared, the co-PI's will present details of this project at conferences related to mathematics and STEM education. The conference presentations will be assisted by at least two of the pilot phase instructors. The co-PI's and one instructor have already been accepted to present at 'Accepting the STEM Challenge: Preparing K-16 Students for Global Competitiveness in the 21st Century' in Atlanta, GA (September 2008). The conference is sponsored by PRISM (Partnership for Reform in Science and Mathematics). The presentation has been assigned to the topic strand 'Roles & Effectiveness of K-16 STEM Educators' and will focus on the development and implementation of the teaching methods cultivated in this project. The session will offer insights from the co-PI's about the methods and materials that have been developed, as well as a first-hand account from an instructor learning to implement these methods for the first time. Because the treatment groups will be in progress at the time this conference takes place, most of the data will not yet have been collected for the treatment groups (content knowledge, perceived usefulness of statistics, etc.) However, the comparisons already conducted between control groups and 'preliminary treatment groups' (see Project Findings section) will be shared.

The co-PI's and another pilot phase instructor have drafted a proposal for submission to AMTE (Association of Mathematics Teacher Educators). The AMTE annual conference is scheduled for February 5-7, 2009. The advantage to presenting at this time will be that the data generated during the pilot phase (both control and treatment groups) will be available to report. Thus, this session will include both implementation details and specific data regarding the impact of the approach, as analyzed through control and treatment groups for all participating instructors.

The co-PI's are also preparing presentation proposals to a number of other relevant conference organizations, including NCTM (National Council of Teachers of Mathematics), NCSM (National Council of Supervisors of Mathematics), SSMA (School Science and Mathematics Association), GAMTE (Georgia Association of Mathematics Teacher Educators), and GCTM (Georgia Council of Teachers of Mathematics). Similarly, as the data are fully collected and analyzed, the co-PI's will be in a position to submit manuscripts for publication in relevant journals. These journals include JRME (Journal for Research in Mathematics Education), SSM (School Science and Mathematics), and the Journal of Mathematics Education Leadership. As other opportunities arise to present project implementation details and findings, the co-PI's will leverage these opportunities as much as possible.

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. The materials, approach and ideas being developed target the attitudes, beliefs and proficiency of students in STEM disciplines. Specifically, the courses of study for many engineering and health sciences curricula include statistics courses of this type. The benefits will cascade to other 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 will have even wider impact. Demonstrating the efficacy of these materials for high school AP Statistics courses will 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 advanced undergraduates. A key feature of this project is infusing authentic statistics with an overt connection to the Scientific Method. Engaging high school students in the practice of developing and testing hypotheses will deepen their experience of science and mathematics and hopefully encourage more of them to pursue intriguing career opportunities in the many high-need STEM disciplines.

The co-PI's have met with the interdisciplinary statistics team, a group of professors and professional clinicians who teach statistics and/or research classes within their discipline. The design of these projects has impressed 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 have commented that they will 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 has 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 survey-based research simulation projects has influenced how our colleagues think about their own courses, content and methods. If successful in securing Phase II funding for this project, the co-PI's plan to explore how well these materials can be exported to other undergraduate statistics courses as part of the evaluation design.

Special Requirements

Special reporting requirements: None

Change in Objectives or Scope: None

Animal, Human Subjects, Biohazards: None

Categories for which nothing is reported:

Organizational Partners

Any Journal

Any Book

Project Findings

I. Instrument Validation

Of the 328 statistics students who were given the instruments, 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.

Perceived Usefulness of Statistics Scale

- 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.

Reliability Statistics: PU

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

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

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 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 have been re-numbered, and in some cases re-ordered, from the original list shown in section III.

II. Preliminary Data Analysis

Although the primary purpose of administering the instruments in Fall 2007 was to validate and refine the instruments, some preliminary 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 some form of the discovery learning approach being more fully developed as part of this project. The instructors of the other 6 sections did not use these methods in their instruction. Therefore, statistical comparisons were made between the co-PI's sections (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. Interdisciplinary Team Contributions

The interdisciplinary team members suggested research constructs relevant to their 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. Participants also suggested many specific project ideas that might use 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. The specific instruments and/or instructions for measuring the construct are not shown, but will be used 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: 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)
    a. Self Awareness – Aware & in touch with feelings and emotions
    b. Managing Emotions and Self-Regulation – managing w/o suppressing
    c. Self-Motivation – Remaining positive & optimistic
    d. Empathy for Others
    e. Social Skills – having, building, maintaining positive relationships
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: 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: Nursing
<u>Suggested constructs</u>

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: Criminal Justice
<u>Suggested constructs</u>

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
    - Importance of parties
    - Importance of religion
    - Importance of athletics
    - Involvement in fraternity or sorority
    - Participation in smoking/Frequency of smoking
    - Time spent studying
    - Involvement in extra-curricular activities
    - Agreement with political philosophies

Field: Sociology
<u>Suggested constructs</u>

1. How many hours do you watch reality TV?
2. How many hours of news do you watch?
3. How many hours are you online?
4. How many hours are you on facebook?
5. How many songs do y download?
6. How many movies do you watch a week?
7. How many times do you study a week?
8. How many hours do you watch sports?
9. How many times have you gone to see sports in person?
10. How many times do you watch women's sports in a month?
11. How many times a week are you physically active?
12. How many times do you think about the way you look in a day?
13. How many books have you read in the last month?
14. How many times a day do you use a cell phone?
15. How many do you text in a day?
16. How many people do you consider acquaintances?
17. How many people do you consider close friends?
18. How many times do you see your dad do housework?
19. How many times does your dad come home late from work?
20. How many times does your mom come home late from work?
21. How many times do you clothes shop a month
22. How many times do you commute to school a week?

23. How much money do you spend on gas a week?
24. How many times do you wash your car in a month?
25. How many cups of coffee do you drink a day?
26. How many cups of soda do you drink a day?
27. How many cups of water do you drink a day?
28. How many times do you eat junk food a week?
29. How many times a week do you gossip?
30. How much money do you spend on groceries a month?

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)

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

Project Activities

The scope of the project has not changed, but the schedule of implementation has changed as follows, due to the factors noted. The original plan was to implement control groups in fall 2007 and treatment groups in spring 2008. These would include statistics classes at North Georgia College & State University (NGCSU), Georgia Perimeter College (GPC), and Forsyth Central High School (FCHS). The treatment groups are to use the materials developed with input from our interdisciplinary team. The timeline has been revised so that the control groups are being conducted in spring 2008 and treatment groups will be conducted in fall 2008. There are several factors that made this timeline more appropriate, and they are enumerated below:

1) The plan to collect control group data in fall 2007 would have necessitated using new instruments without any prior means of validation. Although we were reasonably confident in the instruments we developed, we found it better to validate the instruments in the fall, make any necessary modifications, and then use the improved instruments for subsequent control group and treatment group data collection.
2) At the time the proposal was developed, Forsyth Central High School operated on a block schedule, so that courses there were completed in one semester. Such a schedule would enable us to collect control data in fall 2007 and treatment group data in spring 2008, each from a completed section of statistics. After the grant was awarded and we began planning the project implementation stages with FCHS, we found that they had stopped using the block schedule. Their classes now take one full academic year to complete; thus, introducing both a control and a treatment group during the 2007-2008 academic year was not feasible at FCHS.
3) As noted in the participant information section of the report, our co-PI at Georgia Perimeter College was unable to participate in the project as planned. During fall 2007, we secured a different instructor at GPC to participate, but we were not able to work out details of his participation in time to easily facilitate a fall treatment group and a spring control group at GPC.
4) We had hoped to have the interdisciplinary team composed by the end of summer, but fewer faculty initially expressed interest than we had hoped. With the support of various department chairs, appropriate faculty participants for the interdisciplinary team were identified several weeks into fall semester 2007. However, this did not give the team adequate time to develop their ideas, which will be a significant component of the materials used with the treatment groups. The team had most of their productive participation during spring 2008, which gives us time to incorporate their research ideas into the materials in time to use them in the fall 2008 treatment groups.

With the above changes to the project timeline established, our progress on the project is detailed below.

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.

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. We are currently preparing to administer the revised instruments at NGCSU, at Georgia Perimeter College, and at Forsyth Central High School. All three sites are control groups in spring 2008 and will be treatment groups in fall 2008.

The control and treatment group design at NGCSU has been improved to help control for variability between instructors. Three statistics instructors (other than the co-PIs) are participating first as instructors of a control group, for which they will teach the statistics course the way they would ordinarily teach any mathematics course assigned to them. Then each of these instructors will, with our assistance, use the materials we have developed to teach a treatment group of students the following semester. These instructors are noted in the participant information section of the report. Because we have structured the design this way, the control and treatment groups will not inadvertently be comparing instructors rather than teaching methods. The additional participation of these instructors was facilitated with the support of our department head. We have met with each instructor to educate them about the purpose of this project, how it will be conducted, and what their participation entails.

Also 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. 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 projects. These constructs and research project ideas are the basis for the formal teaching materials (Instructor Guidebook and Student Workbook) that are now being developed.