ATE
Materials Outcome
Comparison

Technical Report

Submitted by
The Evaluation Center
Western Michigan University
Kalamazoo, MI 49008-5237

September 2005

Prepared by
James J. Appleton
Frances Lawrenz

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

This ATE materials development evaluation report is the fourth in a series. The initial report described the creation of a rubric to rate the development of materials and the results of an external expert review of ATE developed materials using the developed rubric. The second report detailed the development and validation procedures for the Biotechnology Problem Solving Skills Assessment (BPSSA). The materials development processes used by ATE sites that developed materials rated as high quality by external reviewers were the subject of the third report. That report compared applied processes and theoretical recommendations and utilized the comparison to create a template for future development efforts.

This fourth report examines the effectiveness of two curriculum materials developed by ATE sites to improve student learning. Achievement of students taught using two of the four highest rated ATE developed curricular materials (i.e., Environmental Science (ATE-EnvSci) and Engineering Technology (ATE-EngTech)) was compared with the achievement of students taught using other materials. In addition to the effectiveness data, this report describes the evaluation procedures used in order to provide examples for other ATE projects to adapt and implement.

Four achievement measures were employed for the study; three were constructed and one standardized achievement measure was purchased. The reliability and validity information about each of the four measures is provided. All measures meet minimal standards for validity and reliability although suggestions about improving the measures are provided. All constructed measures were pilot tested and revised before being used for the study.

One measure was used to assess the ATE-EnvSci materials and three were used for the ATE-EngTech materials. The Environmental Science Assessment (ESA) was constructed by selecting items matched to the AAAS 2061 and NRC Science Education Standards for environmental science from existing national tests. The three measures to assess student understanding of engineering technology used for the ATE-EngTech comparison included: the commercially available ACT WorkKeys Applied Technology Assessment (AWATA), and two forms of the instrument, Engineering Technology Problem-Solving Assessment (ET PSA), developed expressly for this study. The ET PSA was developed based on the biotechnology assessment instrument described in previous reports. The developed instruments were designed to assess STEM content in an integrated fashion consistent with real world experience.

Although the researchers tried to ensure that the characteristics of the two groups were the same, students were not randomly assigned to treatment and comparison groups. Because of this group characteristics were assessed to determine the need to account and adjust for potential selection effects. Group comparisons revealed that students receiving the ATE Environmental Science curriculum (ATE-EnvSci) differed significantly ($p<.05$ to $p<.001$) from Non-ATE students by having

- a greater number of college science courses,
- higher science-related grades, and
- higher degree aspirations.

To ensure consideration of these and other educational background differences in outcome analyses, researchers created a single number (using propensity score methodology) that reflected these disparities between curricular groups. Even after including pre-existing educational background differences in the analyses, ATE-EnvSci students (N=77) performed significantly better than Non-ATE students (N=68, $p<.001$) on the ESA with what would be considered a medium effect size (r = .328) by Cohen's (1992) criteria.

Results of the engineering technology materials outcome comparison revealed that in comparison to Non-ATE students, ATE Engineering Technology (ATE-EngTech) students

- had taken more of their program-specific coursework,
- had taken a significantly greater number of high school science courses, and
- reported significantly lower scientific career aspirations.

Propensity score methodology was again attempted to adjust for group differences, but the sample size was too small. Instead, researchers used the educational background variables that differed significantly between the groups as covariates. There were no significant differences between the groups using the ATE (students=37) and non ATE (students=21) materials on the AWATA Secondary Scale Score or on either form of the ET PSA.

Findings and Recommendations:

Because this is the last report in the series and is a culmination of the materials development evaluation work, the recommendations presented here incorporate the understandings developed over the three years of the evaluation. The findings and recommendations are provided in two sets. The first set is recommendations to the ATE program. The second set is recommendations for improving this type of evaluation.

Recommendations for the ATE Program

Use of the selected ATE developed materials produced students with equal or higher levels of achievement than use of traditional materials. Therefore, it seems reasonable that the ATE program should continue to promote materials development. Given the recent changes to the ATE program solicitation, this type of materials development would most likely be couched within program improvement.

However even within program improvement, the recommendation to continue materials development has some caveats. It must be remembered that the materials tested here were representative not of ATE materials development as a whole but of the "best" materials. Materials development efforts in future ATE projects should be informed by the processes used to develop and identify the exceptional materials reported on here. The ATE Materials Development Processes Report (Lawrenz & Appleton, 2004) provides insight into the processes used to develop the materials rated as excellent (pgs 26-27) as well as an integrated model for guiding the development of technological education materials. Furthermore the Evaluation of Materials Produced by the ATE Program report (Keiser, Lawrenz and Appleton, 2003) outlines a process and provides a rubric for external assessment of the quality of curricular materials.

Students in classes using ATE Environmental science materials had higher achievement than students in classes using other environmental science materials. Although this was only one limited study of the effectiveness of these materials, it has several implications:

- Because these materials appear to help students learn environmental science content, efforts could be made to promote the availability of these materials across the country. The accessibility of these materials via the National Center for Sustainable Resources (NCSR) website provides one means of dissemination. However, although the materials were tested at different sites, all sites were in the northwestern part of the US. Additionally, although some evidence of the portability of the materials is provided by the similar findings of effectiveness across two instructors, more evidence is necessary. Therefore the efficacy of these materials for different areas of the country and with a variety of instructors should be examined. Future ATE projects developing materials should provide plans for future dissemination and for guaranteeing portability should their curriculum be shown to be effective.

- Because the process used to develop the environmental science materials resulted in an effective curriculum, that process might serve as a model for other materials development. The description of the development process provided in the ATE Materials Development Processes Report (Lawrenz & Appleton, 2004) indicates that the key features to the success of the development of the environmental science materials were the expertise and personal commitment to the materials by the single developer and the extensive effort he exerted to utilize the results of several iterations of pilot testing to refine the materials over a long period of time. Having a single person develop the materials was a unique element compared to the processes used to develop the other highly rated materials (Lawrenz & Appleton, 2004). However, the use of expertise, extensive pilot testing and revision over a long period of time were more consistent across the processes used to develop the highly rated materials. Therefore it seems likely that expert attention to all aspects of the materials through a comprehensive process of pilot testing and revision would be most likely to result in effective materials.

- Making the assumption that the success of the environmental science materials could be replicated by other ATE materials development efforts lends support to the recommendation above that ATE continue supporting materials development.

Recommendations for evaluations of ATE developed materials

There are several different implications about the evaluation of materials developed by ATE projects that arose throughout our three year effort. Our evaluation process provides a model for any materials evaluation effort. The suggestions and the findings that lead to them are included below in three sections: before field testing, during field testing and research related to field testing.

Before Field Testing

It is important to recognize that the process of evaluating materials will include several steps. The two comparison studies presented in this report highlight the potential and feasibility for careful student outcome based evaluations of ATE curricula, but this type of study is costly and most relevant for materials in their final stages. Other types of evaluation should be used at different stages in the development of materials.

Involve experts. This is already common in ATE materials development in the form of content experts and industry standards but less common in terms of educational or instructional development expertise.

Conduct iterative pilot testing. The materials field tested in this study had undergone extensive pilot testing where the materials were tried out, student and instructor opinions were gathered and modifications were made. Pilot testing should continue until the materials appear to be meeting the prescribed outcomes. The integrated development process provided in the ATE Materials Development Processes Report, (Lawrenz & Appleton, 2004) can serve as a model to help identify all of the issues to consider when developing materials.

Submit the materials to external review. The rubric provided in the Evaluation of Materials Produced by the ATE Program report (Keiser, Lawrenz, & Appleton, 2003) can serve as a model for conducting this type of multiple expert review and to help identify the issues relevant to high quality technological education materials.

During the Field Test

Once materials have evolved past pilot testing and external review, the next expectation could be field tests. Substantial funds need to be available to conduct these. Although as suggested by the comparison studies presented here, it is possible to create reliable (ESA and ET PSA) and valid (ESA and potentially the ET PSA) assessments as well as recruit adequate comparison groups with which to examine the effectiveness of developed materials, it was a very time consuming, intense, process requiring substantial expertise in research design, sampling, measurement, data analysis and reporting. This type of field testing is designed to showcase how the results produced by using newly developed materials compare to the results produced by other materials. Therefore use of field testing should be restricted to instances where this is an important question.

Plan in advance. By planning for comparative studies at the outset and getting agreement from participating courses, it may be possible to both increase participation and reduce costs by building the comparative testing procedures into regular course expectations.

Ensure comparability of the sites. Extensive efforts were used in this evaluation to ensure that the sites selected were as similar to each other as possible. Even with these careful selection procedures, the students at the sites showed some differences which were adjusted for statistically. Anticipating and collecting data on variables related to the outcomes of interest are critical to obtaining meaningful comparisons.

Use valid and reliable assessment instruments. Three instruments were constructed for use in this evaluation and one was purchased. Despite the careful construction and pilot testing of the developed instruments, they may not have worked in exactly the ways envisioned. Furthermore, commercially available instruments may not fit the exact goals of the courses being examined. Appropriate instrumentation is the key to meaningful interpretation.

Consider outcomes in addition to written achievement tests. The sole use of a written response achievement test to measure curriculum effectiveness:

- may overlook materials' effectiveness at attaining other developer goals (e.g., retaining students)
- may not consider other ways of measuring achievement (e.g., different types of tests or the inclusion of other content) that might have resulted in different findings
- may miss the successes apparent when using different definitions of achievement (e.g., narrowing achievement discrepancies across ethnicities)

Therefore, other indices of achievement and other outcomes besides achievement should be considered in determining materials' effectiveness.

Investigate transferability. The present study investigated the portability of materials across instructors. These results are promising, but future studies should more directly investigate the mechanisms for, and the effects of, transporting and using materials in varied locations as well as with different instructors.

Recruit large samples. The samples included in these studies were sufficient to detect substantial to small main effects for the ATE-EnvSci curriculum comparison and sufficient to detect substantial to moderate main effects with the ATE-EngTech curriculum using the AWATA. The samples were not sufficient to detect moderate effects with the ET PSA. The sample sizes were not sufficient to test other important questions such as special contexts or populations (e.g., ethnic groups or sex) which might yield significant differences. Although increasing sample size is difficult, it would enable a broader array of analyses. Recruiting methods should be carefully considered. In designing studies, it is recommended that researchers consider the tradeoff between the examination of longer term programs, the level of similarity of participants, the value of a longer assessment, and the available advertising avenues on the one hand and sample size on the other.

Research Related to Field Testing

Research factors affecting recruitment. The greater prevalence of introductory environmental science courses, the attempt to recruit subjects from similar geographical areas, the length of the assessment (30 vs. 60 minutes), and the method of advertising the opportunity to participate (in or outside of regular class) each seemed to impact the size of the comparison site sample. For at least these reasons, the ATE-EnvSci sample is much larger than the ATE-EngTech (145 vs. 58, respectively). It should also be noted that compensation, while important, may not offset these other factors. This evaluation recruited more students to complete a 30 minute exam for $10 than a 60 minute exam for $20.

Develop more instruments that are relevant to ATE needs. Only one instrument was available commercially that could be used in this study. Development of instruments is costly and time consuming. Developing and providing more instruments aligned with the goals of ATE projects would facilitate comparison studies.

Improve the measurement tools and sources for gathering background data. Researchers encountered challenges in attempting to gather appropriate data as proxies for educational background and future aspiration differences. For instance, non-ATE students may have misinterpreted a question on courses taken and/or to have forgotten the exact names of courses completed. In addition, self-report methods for obtaining student educational background information (e.g., G.P.A., other test performances, and the rigor of high school and college science courses) may be less objective than sources such as transcripts. Therefore more research about how best to measure background variables either through the use of better survey items or the use of different sources of evidence should be conducted.

ATE Materials Outcome Comparison Report

The National Science Foundation's Advanced Technological Education (ATE) program stems from a national interest in developing and using technology to meet the nation's educational and workforce needs. Funded via a Congressional mandate, the ATE program was designed to (1) produce more science and engineering technicians to meet workforce demands and (2) improve the technical skills and the general science, technology, engineering, and mathematics (STEM) preparation of new technicians and the educators who prepare them. The majority of ATE funding is directed at the community college level in order to strengthen and expand the scientific and technical education and training capabilities of individuals at these institutions. More specifically, the objectives of the ATE program are to

- Develop model instructional programs in advanced-technology fields
- Provide professional development for college faculty and secondary school teachers in advanced-technology fields
- Establish career routes from secondary schools to two-year colleges and from two-year colleges to four-year schools with a secondary goal of articulation among two-year to four-year programs that specifically focuses on K-12 prospective teachers in technological education.
- Conduct applied research on technical education.
- Develop and disseminate instructional materials

As part of the ATE program, NSF included funding for evaluation to assess the impact and effectiveness of the ATE program. The evaluation, conducted by The Evaluation Center at Western Michigan University and the University of Minnesota, has sought to answer four basic questions deemed important to ATE and its stakeholders:

1. To what degree is the program achieving its goals?
2. Is the ATE program making an impact and reaching the intended individuals and groups?
3. How effective is the ATE program when it reaches its constituents?
4. Are there ways the program can be improved significantly?

The evaluation study results presented here are related to the materials development portion of the ATE program and therefore only one part of the evaluation of the overall ATE program. Furthermore, this study is only one report in a series of reports concerning the evaluation of materials development in the ATE program. The evaluation of the materials development portion of the ATE program has several components:

- Development of a curricular materials evaluation system
- Use of the system by external experts to evaluate selected ATE-developed materials
- Consideration of the processes used by ATE projects to develop materials
- Development of a device to assess student facility in workplace-based problem solving
- Development of other assessment devices
- Implementation of a quasi-experimental study examining the effect of ATE-developed materials on student achievement in comparison with the effect of traditional materials

The first report in the series of ATE materials development evaluation reports described the development of a material's rating rubric and the results from external expert use of the rubric to rate ATE developed materials. Four-fifths of the materials received "adequate" or better ratings. Comparisons of the highest and lowest rated materials suggested that much of the range in ratings was due to the differential ways in which materials addressed industrial and content issues such as quality performance, rigorous content and relevant applications, as well as curricular issues such as assessments and integration of general education skills (Evaluation of Materials Produced By the ATE Program, Keiser, Lawrenz, & Appleton, 2003).
The second report described the development and validation of the Biotechnology Problem Solving Skills Assessment (BPSSA) device. This report detailed the content knowledge base and reliability and validity analyses for parallel forms of a 17-item, work related, problem-solving assessment (Summary Report for the Development and Validation of the Biotechnology Problem-Solving Skills Assessment, Lavoie, 2003). The third report detailed the materials development processes used by ATE sites which developed highly rated materials, compared those applied processes to theoretical recommendations, and used both applied and theoretical processes to create a template to guide future development efforts (ATE Materials Development Processes Report, Lawrenz & Appleton, 2004).

This report continues the investigation of the effectiveness of the materials development portion of the ATE program by examining the effectiveness of the ATE developed materials in promoting student achievement. It describes the assessment of outcomes associated with implementations of two of the highly rated ATE materials (i.e., Environmental Science and Engineering Technology) and matched comparison sites. A secondary purpose of this report is to provide data on the measurement properties of the Engineering Technology Problem Solving Assessment. This new assessment instrument was modeled after the BPSSA and utilized as one of two measures to gauge the outcomes associated with the Engineering Technology materials. A third purpose is to describe the processes used so that they could serve as an example for others conducting evaluations of materials development.

METHODOLOGY

As summarized in a previous ATE materials development evaluation report (Keiser, Lawrenz, & Appleton, 2003), 65 projects and centers had reported being involved in materials development on the yearly ATE survey. Of these, 37 responded to a request to send in a copy of their best material to be reviewed. Preliminary review reduced the number of materials to 27, judged suitable and sufficiently complete for review. Of the 27 materials, 23 were judged adequate or better overall by a team of experts trained to use a comprehensive assessment rubric. Of these 23, 14 received overall ratings of good or better; two were judged excellent, 2 as good to excellent and 10 as good. The four materials rated most highly were considered to be of exceptional quality. Table 1 provides an overview of these four materials.

Table 1: ATE Exemplary Materials Ratings

Overall Team Rating[a]	Subject Area	Material Type	Material Format	Funding Type	Start Date	Award Amount to Date
4.0	*Engineering Technology (ATE-EngTech)*	*Multiple Modules*	*Combination: Texts & Packets*	*Center*	*Sep-99*	*$ 2,000,000*
4.0	Electrical - Mechanical Engineering With Ethical Case Studies	Multiple Modules	Text(s)	Center	Sep-98	$ 2,000,000
3.5	Engineering Technology (Marine)	Course	Packet	Center	Sep-2000	$ 2,000,000
3.5	*Environmental Science (ATE-EnvSci)*	*Course*	*Text*	*Project*	*Oct-2001*	*$ 1,000,000*

[a]*Team ratings could range from 0–4*

To conduct the next portion of the evaluation of the ATE materials development, researchers contacted the developers of the four curriculum materials described above to determine the specific sites where the materials were being used. This process revealed that only two of the four materials were currently being implemented in locations beyond the developer's site. In order to assess outcomes related to the curriculum, we required that it be in operation at more than one site. Additionally, appropriate comparison groups (described in the "comparison processes" section below) had to be available. Appropriate comparison classrooms were available for two of the four materials; the same two which had multiple implementation sites. Therefore, these two materials (those shown above in italics) were used in this investigation: the Environmental Science (ATE-EnvSci) and the Engineering Technology (ATE-EngTech) materials.

The investigations of the effect of the two materials on student outcomes followed similar but not exactly the same procedures because of differences in the type of materials and the sites in which the materials were used. The four major components of the investigation process were "Sample Selection", "Instrument Development", "Collection of Data and Sample Characteristics" and "Data Analyses". The sample selection component provides a description of the curriculum being examined and of the methods used to ensure the ATE and non ATE curricular materials addressed similar goals. The instrument development component provides information on the three instruments that were constructed for these investigations and the one extant instrument that was selected. The collection of data and sample characteristics section provides a description of how the data were collected for each curriculum and information on the participants in the samples. The data analyses component provides an overview of the data analysis plans that are then followed in the results section.

Sample Selection

An appropriate sample is necessary in order to conduct an adequate comparative study. In this case we could not randomly assign teachers or students to different curricular materials. Therefore careful matching of the participants in the two groups was necessary. The experimental or ATE sites were predetermined as those sites using either one of the ATE materials being tested. Therefore sites comparable to those using the ATE materials had to be identified. This was accomplished in slightly different ways with the two different sets of materials but the primary goal in each was to ensure that the context, teachers and students at the comparison sites were as similar as possible to those at the experimental sites. In other words the sample selection process attempted to control for all variables that might affect the measured student outcomes except those related to the curriculum being used.

To locate sites for the Environmental Science curricula comparison, researchers utilized developer recommendations of sites implementing similar curricula, an internet search of environmental science programs and courses, and recommendations by other environmental science experts. These resources provided approximately 40 potential sites for comparison with ATE-EnvSci materials. Examinations of course content, credit hours, laboratory components, and course duration provided criteria enabling the researchers to reduce the pool of potential sites to 11. Researchers initiated contact with each of these sites to involve them in the study. Of these eleven sites, 4 expressed interest in being involved and took part in the study. Of these 4 recruited sites and the developer's site, 3 contributed 68 comparison students and 2 (one recruited site and the developer's site) contributed 77 experimental students (i.e., instructed using the ATE-EnvSci materials).

To locate sites for the Engineering Technology curricula comparison, recommendations from developers were again used, this time to locate and secure sites both implementing ATE-EngTech materials as well as comparison sites. A total of 9 sites was recommended. Of these 9 sites, 7 agreed to participate with 6 eventually administering assessments to 58 students. Three sites were implementing the ATE-EngTech materials (37 students) and three sites were implementing other similar materials (21 students).

A description of each ATE curriculum as well as the process for determining the similarity of comparison groups is described in the following paragraphs.

<u>ATE-EnvSci:</u>

Curriculum: The ATE-EnvSci materials could comprise a single course. The focus of this course is on ecosystem management, goals of maintaining existing biodiversity, evolutionary and ecological processes within ecosystems, and accommodating human uses within these constraints. The ATE EnvSci materials are in the form of a lab manual/enhanced syllabus with a primary audience of 1st year CC/TC students. A textbook, <u>Environmental Science: Earth as a Living Planet</u> (2003, 4th ed.) by D. Botkin and E. Keller is also used in the course.

These materials are intended to be used in a 4-credit course with weekly requirements of three hours of lecture, a three-hour lab and a duration of one 11-week quarter

Final enhanced syllabi are either posted on a web site by a web consultant or produced in hard copy or both.

Comparability: To ensure comparability, each of the 5 sites included in study (i.e., 3 comparison, 2 experimental) were verified as offering 11 week courses on a quarter system calendar. Additionally, researchers emailed the consolidated standards to each site to verify that the content taught in the courses would be satisfactorily measured with the assessment that the researchers had constructed. All replied and agreed. Also, course syllabi were requested from each site and a comparison conducted to examine the similarity between the 7 consolidated standards and the content of the courses at each site. For one site, a course webpage was examined in addition to the syllabus as the website provided more detailed information regarding course content. A description of the syllabi comparison was emailed to the sites with the request that instructors return email comments regarding any missing information or misunderstandings of the course content represented in the syllabi comparison. Table 2 provides the information obtained from this syllabi comparison.

Table 2: Environmental Science Syllabi Comparison

INSTRUCTOR:																							
SITE A		**Introduction to Environmental Science**				**The Nature of Science**		**Env. Attitudes and Values**	**History of Natural Resource Use/Cons./E.P. in the U.S.**		**Basic Ecological Concepts**												
		Sustainability	*Ecosystem Management*		*Env. Unity*	*Sci. Method*	*Sci. vs. Other disc.*		*Major Events*	*Impacts on Pub opin and policy*	*Biosphere*		*Ecosystems*					*Communities*			*Populations*		
			Elements and goals	Ex. Of Implementation							"Gaia Hypothesis"	Role/nature of solar radia.	Char. Of systems	Biotic and Abiotic Components	Biogeochem. Cycles	Primary Productivity and energy flow	X btwn climate and biomes	Biological Diversity	Species Interactions	Ecological Succession	Bio props. of pops.	Pop. Growth	Hum. Pop. Growth as env issue
Standard		2	2	2	2	7	7	5	5	5	2	3	2	2	4	3	2	1	1	2	6	6	5, 6
SITE B	**Similar**	Content determined comparable via examination of course website (currently not available)												Content determined comparable via examination of course website (currently not available)									
	Different: Assigned projects, but no formal lab component																						
Standard																							
SITE C	**Similar**	Ecology and life on land and in water (9/29/04)	Ecology and life on land and in water (9/29/04)		Ecology and life on land and in water (9/29/04)	Research / sampling statistics (10/11, 10/13, 11/15, 11/17, 11/22, 11/29, 11/30, 12/1/04)								Ecology and life on land and in water (9/29/04)	Nutrient Cycling (10/25, 10/27/04)	Intro to energy and nutrient relations (10/11, 10/13/04); Primary production and flow of energy (11/18, 11/20/04)	Effects of temperature, microclimates to global patterns (10/4, 10/6/04)	Species abundance and diversity (11/1, 11/3/04); Natural selection (12/8/04)	Population density and dynamics (11/22/04)	Ecology and life on land and in water (9/29/04)	Population genetics, distributions, density, abundance, and dynamics (11/15, 11/17, 11/22, 12/6, 12/8/04)	Population genetics, distributions, density, abundance, and dynamics (11/15, 11/17, 11/22, 12/6, 12/8/04)	Population genetics, distributions, density, abundance, and dynamics (11/15, 11/17, 11/22, 12/6, 12/8/04)
	Different: Introduction to water relations																						
Standard	2	2	2		2	7	7					3		2	4	3	2	1	6	2	5, 6	5, 6	5, 6
SITE D	**Similar**	Sustainability (9/27/04)	Ecosystems and Ecosystem Management (10/4/04)		Env. Unity (9/27/04)	Scientific Method and decision-making (9/27/04)	Scientific Method and decision-making (9/27/04)	Values and Science (9/27/04)	Basic Issues in Env. Sci. (9/27/04)	Basic Issues in Env. Sci. (9/27/04)	"Gaia Hypothesis" (9/27/04)		Char of systems (9/27/04)		Biogeo-chemical cycles (10/04/04)	Biological Productivity and Energy Flow (10/4/04)	Climate and biomes (10/11/04)	Biological diversity /evolution (10/11/04)	Interactions between species (10/11/04)	Ecological Restoration (10/11/04)	Human Population Growth (9/27/04)	Human Population Growth (9/27/04)	Human Population problem (9/27, 10/18, 10/25, 11/1/04)
	Different	2	2		2	7	7	5	5	5	2		2		4	3	2	1	1	2	6	6	5, 6
Standard																							
SITE E	**Similar**	Ecosystems (11/17/04)				Experimentation, observation, data gathering, computation, conclusions	Experimentation, observation, data gathering, computation, conclusions	Social Interactions / Behavior (11/29/04)			The Biosphere (11/22/04)			Ecology (12/1/04)	Biogeochem. Cycles (11/19/04)	Ecology (12/1/04)	Ecology (12/1/04)	Evolutionary Theory (9/27-9/29, 10/6-10/8-10/11-10/13/04)	Community Interactions (11/15/04)	Ecology (12/1/04)	Populations (11/19/04)		Human Impact on the Biosphere (11/24/04)
	Different: Speciation, Viruses and Bacteria, Protists, Fungi, Lower Vascular Plants, Gymnosperms and Angiosperms, Sponges to roundworms, molluscs, annelids, arthropods, echinoderms, chordates																						
Standard	7	2	2	2	2	7	7	5			2	2		2	4	2	2	1	1	2	6	6	5

Standards
1) Biodiversity/Natural selection
2) Ecosystems
3) Energy and the environment
4) Geochemical cycles
5) Human effects on the environment
6) Population size and rate of growth (influences and results)
7) Scientific method/scientific inquiry

ATE-EngTech:

Curriculum: The ATE-EngTech materials (in the form of modules, an instructor guide, and student handouts) were designed for engineering technology students as a first-year, problem based general education curriculum to fulfill foundational requirements prior to specialization within an engineering field. These materials integrate mathematics, physics, communication and engineering technology. These modules are typically taught over the course of three semesters. The primary audience is first year community college or technical college and high school students.

Comparability: To ensure comparability, sites that were currently using materials purporting to cover content similar to that in the ATE-EngTech curriculum were selected based upon developer recommendations. Suggested sites included some that were considering implementing ATE-EngTech materials in the future. Subject area experts also provided the names of other potentially comparable sites. In each case recommenders attested to the comparability of the sites. The fact that some sites that were deemed similar intended to implement ATE-EngTech materials in the future further supported claims of comparability. Researchers also examined comparison site program information via the internet to confirm the number and type of credits earned for the courses typically taken by students at comparison sites. Separate demographic forms were created for each comparison site and emailed to site contacts to verify that the appropriate courses were listed. The number of courses completed over both a two and three semester time period or equivalent to a two or three semester time period were examined in analyzing these data.

Instrument Development

In terms of assessment devices, one suitable, commercially available instrument was found for the ATE-EngTech assessment process and the three others were constructed (i.e., one for the ATE-EnvSci assessment process and two for the ATE-EngTech assessment process).

ATE-EnvSci Assessment: The Environmental Science Assessment (ESA) instrument was constructed using Bloom's taxonomy of educational objectives (Bloom, Mesia, & Krathwohl, 1964), the National Research Council's National Science Education Standards, and the American Association for the Advancement of Science (AAAS) Project 2061 Benchmarks for Science Literacy. Benchmarks and Standards describing environmental science concepts were compiled and then consolidated into seven categories. The categories of consolidation in combination with the Benchmarks and Standards contained therein are located in Appendix E. Bloom's taxonomy was collapsed from 6 categories to 3 to streamline categorization while still differentiating based upon level of cognitive demand. These three categories were defined as follows: knowledge, comprehension/application, and analysis/synthesis/evaluation. The seven categories of Benchmarks and Standards were used to differentiate potential items along one dimension while the three categories adapted from Bloom's Taxonomy were used to orient potential items along another dimension.

Items were gathered from several sources (e.g., NAEP, TIMSS, NY Regents exams, a standard environmental science text). Items were retained or discarded depending on the clarity of the item and whether it fit within one of the seven consolidated categories. Items with difficulty information (i.e., from NAEP and TIMSS) were utilized when possible. Items (both with and without difficulty information) were gathered until each position along the two dimensions was represented with items. From these gathered items, two forms of the environmental science assessment were constructed with each containing 24 multiple choice items and two constructed response (essay) items. These two forms were administered to 91 post-secondary students (females = 31, males = 59 with one student not replying) with 49 completing Form I and 42 completing Form II. Indices of internal consistency reliability were α =.70 and α =.56 for Form I and Form II, respectively. Item analysis revealed that several multiple choice items either failed to discriminate properly between the highest and lowest scoring students or functioned systematically different across gender. These items were removed from the test. Additionally, one constructed response item had minimal variability, few correct responders, and upon analyzing qualitative data from students was deemed to function poorly and removed.

Since the remaining items (across the levels of Bloom's Taxonomy) represented each consolidated standard relatively well and the total number of items had been reduced substantially, one form of the ESA was created rather than two. Presented in Appendix A, the final version of the ESA contained 26 multiple choice items and 3 constructed response items. This version of the ESA was administered to 42 students in an introductory science course for non-majors. As depicted in Appendix H, the item analyses of this form of the test indicated that the items differentiated well between higher and lower scoring students as well as functioning adequately across gender. Table 3 indicates the items composing this final form, the consolidated standard to which they correspond, and the level of Bloom's Taxonomy that characterizes them. As depicted in Appendix H, internal consistency reliability for this measure was α =.77.

Table 3: Cognitive Levels of ESA Items

		Level of Bloom's Taxonomy		
		(1) Knowledge	**(2) Comprehension, Application**	**(3) Analysis, Synthesis, Evaluation**
Standard	**Total # by Standard**	Total at This Level	Total at This Level	Total at This Level
1)Biodiversity/Natural Selection	*4*	*0*	*3*	*1*[1]
2)Ecosystems	*5*	*3*	*1*	*0*
3)Energy and the Environment	*8*	*5*	*2*	*1*[1]
4)Geochemical Cycles	*2*	*1*	*1*	*0*
5)Human Effects on the Environment	*4*	*2*	*1*	*1*[1]
6)Population Size and Rate of Growth	*3*	*1*	*2*	*0*
7)Scientific Method / Scientific Inquiry	*3*	*0*	*3*	*1*

[1] Constructed Response Item (all items without this designation are multiple choice)

ATE-EngTech Assessment: Three instruments were used in this outcomes assessment. Two instruments were created (Forms I and II of The Engineering Technology Problem-Solving Assessment (ETPSA)) and one was already in existence and commercially available (ACT WorkKeys Applied Technology Assessment (AWATA)).

The constructed assessments (Forms I and II of The Engineering Technology Problem-Solving Assessment (ETPSA)) were modeled after the Biotechnology Problem-Solving Skills Assessment (BPSSA) discussed in a previous report to ATE (Lavoie, 2003). These new engineering problem solving assessments used the same structure (e.g., problem types, rubric design) as the BPPSA, but were adapted by experts who substituted engineering technology scenarios for the biotechnology scenarios. The experts who created the ETPSA forms were in contact with the BPSSA developer to assure adherence to the theoretical underpinnings which provided the foundation for that instrument. To examine the understandability of these items, the ETPSA developers used think aloud procedures to administer the instruments in their entirety to 6 students. Subsequent changes were made based on results. Researchers at the University of Minnesota initially grouped instrument items into three forms and administered these to a total of 16 students to obtain item information. Researchers obtained information on the functioning of each item. While the diversity of subjects (by race and gender) was too limited to examine items across these groups, the data supported conclusions that the assessments discriminated well between high and low performing students. For each item, students in the 75th percentile or higher performed better than those in the 25th percentile or lower. Results are depicted in Appendix G in Table 25. Results also suggested varying levels of internal consistency reliability ranging from .14 to .76. Given the low internal consistency of two of these three scales ($\alpha = .14$ and $\alpha = .24$), researchers consolidated items into two forms of the assessment. The result was two instrument forms deemed to be of similar difficulty. The improved internal consistency reliability ($\alpha = .76$ and $\alpha = .77$) of these two forms is described below in the ATE-EngTech results section. The two forms of the ET PSA are presented in Appendices B and C.

The extant instrument (i.e., AWATA) (http://www.act.org/workkeys/assess/tech/index.html) is designed to measure the skill people use when they solve problems with machines and equipment found in the workplace. This skill includes four areas of technology: electricity, mechanics, fluid dynamics, and thermodynamics. In order to score highly on this assessment, individuals need to know the basic principles of each area. As indicated by ACT, students who use applied technology skills should be able to:

- Analyze a problem by identifying the problem and its parts.
- Decide which parts of a problem are important.
- Decide on the order to follow when dealing with the parts of the problem.
- Apply existing tools, materials, or methods to new situations.

This instrument is a criterion-referenced classification test and indicates student placement in levels ranging from >3 to $\leq$ 7 with 7 indicating the highest level of competence. Scores can also be converted to secondary scale scores to facilitate finer distinctions between test takers. The scale score "is a function of a number-correct (raw) score. The number-correct scores put the examinees in order with respect to ability, although the differences between the numbers do not in any obvious sense measure equal 'distances' (intervals) between the abilities. Therefore, the differences in scale values do not necessarily represent equal differences in ability. The total number of Scale Score points was chosen to be less than the total number-correct score points in the tests, in order to enhance measurement precision. The non-integer Scale Scores were rounded to integer Scale Scores. The scale has a range of 25 points (from 65 to 90) (ACT, 2001, p. 4). More information on the AWATA can be found in Appendix D.

Collection of Data and Sample Characteristics

To facilitate comprehensive comparisons of the equivalency of students (e.g., progress within one's program, performance within courses taken, perceived relevance of current course to future endeavors) across sites, demographic information was obtained directly from participating sites and from the students completing the assessment instruments. Both the Environmental Science

Assessment and the Engineering Technology Assessments contained general demographic items to elicit information on gender, home language, race, typical grades in all post-secondary classes, high school science courses taken, level of degree pursued, and aspirations to work in a scientific field upon graduation. In addition, the environmental science demographic questionnaires requested information on post-secondary science courses taken and typical grades received in science-related post-secondary classes. The engineering technology demographic questionnaires requested information on the program-specific courses taken within the semester or quarter sequence equivalent to the ATE-EngTech curriculum sequence and duration, typical grades received in engineering-related courses, and the student's current college grade point average (GPA).

ATE-EnvSci:

To examine curricular outcomes, instructors were sent assessments and compensation funds (at a rate of $10 per student) as well as detailed instructions in order to standardize both the recruitment of subjects and administration of the assessments across sites. Completed assessments were return mailed and included in the present analysis.

One hundred and forty-five students were involved in the outcomes comparison with 143 of these students providing all or most of the demographic information requested on the forms attached to the assessment. Students who reported demographic information in the non-ATE group (66 students with 73% female) contributed the following characteristics: 24% usually speaking a home language other than English, 70.6% White, 20.6% Hispanic/Latino, 4.4% Other, and 1.5% American Indian/Alaskan Native. Students reporting demographic information in the ATE-EngTech group (77 students with 48% female) contributed the following characteristics, 13% usually speaking a home language other than English, 77.9% White, 10.4% Hispanic/Latino, 6.5% Other, 2.6% African American, 2.6% American Indian/Alaskan Native.

ATE-EngTech:

To examine curricular outcomes, instructors were sent the aforementioned demographic forms and the following assessments: constructed response engineering technology problem-solving assessments (ETPSA Forms I and II) and the ACT WorkKeys Applied Technology Assessment (AWATA). In addition, compensation funds (at a rate of $20 per student) as well as detailed instructions in order to standardize both the recruitment of subjects and administration of the assessments across sites were sent. The order of administration of the ETPSA and the AWATA and the order of receiving Form I or Form II of the ETPSA were randomly assigned. All supplied assessments were return mailed and those completed were included in the present analysis.

Fifty-eight students were involved in the Engineering Technology outcomes comparison with 57 of these students providing all or most of the demographic information requested on the forms attached to the assessment. Students who reported demographic information in the non-ATE group (20 students with 10% female) contributed the following characteristics: 17% usually speaking a home language other than English, 61.9% White, 14.3% Asian/Pacific Islander, 9.5%African American, and 4.8% Hispanic/Latino. Students reporting demographic information in the ATE group (37 students with 11% female) contributed the following characteristics: 5% usually speaking a home language other than English, 67.6% White, 32.4% African American, and 2.7% American Indian/Alaskan Native.[1]

Data Analysis

The data analyses were undertaken to determine if there were any differences in the outcomes attained by students being taught using the ATE developed materials and the outcomes attained by students being taught using other materials. This type of investigation uses the comparison of mean scores to determine if any differences between the groups are statistically significant. In other words the question of interest is whether obtained differences between groups is a result of the curriculum difference or due to chance differences in the samples engaged in the study. In the studies presented here, however, because random assignment was not possible statistical techniques were also employed to help account for any prior differences between the groups being compared. To employ these controls the groups were first examined to determine if any prior differences existed. If no differences existed, then the groups could just be compared. If differences did exist, however, a statistical control was included in the comparison. A more detailed look at the analyses conducted to determine what type of statistical control to employ is provided in the results section along with the ultimate results of the comparisons.

RESULTS

In this section the specific results for the ATE-EnvSci outcomes comparison are presented first followed by the results from the ATE-EngTech comparison. Each set of results is preceded by a table which outlines the analyses undertaken, the justification for each analysis and the outcome of each analysis. The processes followed in both sections were (a) first checks for pre existing differences, (b) implementation of statistical procedures to help adjust for those differences and (c) significance testing-- an analysis to determine whether observed student differences resulted from the curriculum materials.

ATE-EnvSci Analyses

As a first step, the data were examined to determine any pre existing differences between the two groups. The first examination determined the quality of the assessment instrument and its scoring. This was measured through internal consistency and intra-rater reliability. Then differences between the groups in their background characteristics such as course taking, ethnicity, gender, etc. were examined. Any differences were related to differences in scores on the outcome assessment. Because it was determined that prior differences existed between the groups, a statistical technique called propensity score analysis was used to help equate the groups. This technique allows a researcher to consider all of the potential variables that are available and assign a predictive score to each person. These scores were then used to modify the between group comparisons to adjust them for the prior differences. Finally the similarity of effects across the different instructors was examined to begin the determination of the effects of

[1] Percentages do not sum to 100% as one student identified as both African American and American Indian/Alaskan Native.

transferring the materials to different sites. The types of analyses, the justification for using them, and the outcomes associated with them are detailed in Table 4.

Table 4: Guide to ATE-EnvSci Analyses

Guide to ATE-EnvSci Analyses		
Analyses	*Justification*	*Outcome*
• Intra-rater and internal consistency reliability tests	• Check scoring consistency and test coherence	• High intra-rater (93%) and adequate internal (α = .77)
• T-tests of differences between ATE and non-ATE students on educational background and aspirations	• Since not true experiment, examine matched sample for pre-existing differences	• Science-related grades, number of college science courses, and degree aspirations sig. different
• ANOVAs with gender, other home language, and Hispanic/Latino ethnicity	• Examine interactions between these differences and curriculum type	• No significant interactions
• Proportions of each gender and Hispanic/Latino ethnicity in each curriculum group	• Examine whether subgroups are more represented in either ATE or Non-ATE groups	• Females were significantly more represented in the Non-ATE group; since males and females had nearly identical test scores, this shouldn't matter
• Propensity Score Methodology	• To stratify participants according to probability of receiving ATE materials	• Levels of observed background variables believed to be similar within strata
• Non-parametric tests of differences on background variables	• To examine the effectiveness of the propensity score stratification process	• No significant differences within strata, suggests control and treatment participants are "balanced"
• Stepwise regression of ESA first on quintile of propensity score and then on curriculum type (N = 136)	• Examine effects of ATE curriculum while controlling for background variables	• The change model is significant with an R^2 change value of .108.
• T-test with instructors	• To check for instructor effects on curriculum type	• No significant differences; the materials were similarly effective across instructors

Test Reliability and Item Functioning
All items had very specific scoring rubrics developed as part of the national assessments in which the items had been used so inter-rater reliability measures were not necessary. All assessments were scored consistently and by one person. An examination of the scorer's consistency with herself resulted in an intra-rater reliability of 93%. For the measure as a whole, the internal consistency reliability was (α =.77).

Multiple-choice items functioned adequately across test performance and gender groups (details are available in Appendix H). Scores on the test were significantly related to the number of college science courses taken (r = .25**) as well as the level of post-secondary degree sought (r = .11*), providing some evidence for the validity of the test.

Demographic Analyses

Using T-tests, researchers examined the educational background and future aspirations information provided by participants. Table 5 shows three significant findings all of which have substantial associated effect sizes (ES). ATE-EnvSci students were stronger in terms of higher science-related grades (ES = .36), more college science courses completed (ES = .68), and higher degree aspirations (ES = .56).

Table 5: Group Means and Standard Deviations for Background Variables

	ATE-EnvSci Curriculum	N	Mean	Std. Deviation
HS Science Courses	**no**	26	2.69	.884
	yes	8	3.00	.756
Science Related Grades	**no**	65	2.86	.682
	yes	74	3.11*	.713
All Grades	**no**	26	3.35	.562
	yes	8	3.25	.707
College Science Courses	**no**	67	1.36	.595
	yes	76	1.87***	.869
Science Field	**no**	67	1.58	.819
	yes	76	1.83	.870
Type of Degree	**no**	67	2.06	.600
	yes	76	2.39**	.568

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

In order to determine whether demographically similar subgroups of ATE-EnvSci and comparison students scored systematically differently on the ESA, univariate examinations of interactions between curriculum type and gender, languages spoken at home, and Hispanic/Latino racial group were conducted. These interactions were not significant. The means and standard deviations used in the analyses are presented in Table 6.

Table 6: Examination of Demographic Variables by Curriculum Type

Descriptive Statistics
Dependent Variable: **Total ESA Score**

ATE-EnvSci Curriculum	**Gender**	Mean	SD	**Other Language in the Home**	Mean	SD	**Hispanic / Latino**	Mean	SD
No	*Male*	17.50	6.327	*Yes*	15.63	5.965	*Yes*	15.21	5.767
	Female	17.54	4.942	*No*	18.14	4.986	*No*	18.20	5.074
	Total	17.53	5.304	*Total*	17.53	5.304	*Total*	17.59	5.320
Yes	*Male*	22.89	4.729	*Yes*	21.70	5.438	*Yes*	21.63	5.290
	Female	22.81	5.358	*No*	23.26	4.970	*No*	23.12	5.002
	Total	22.85	5.015	*Total*	23.04	5.027	*Total*	22.96	5.017

A subsequent chi-square analysis of the proportion of participants in the two curricular groups by gender, other home language, and Hispanic/Latino ethnicity resulted in significance only for the overrepresentation of females in Non-ATE groups. As male and female ESA performance was nearly identical within groups, this instance of disproportional representation was not considered problematic. For details of these analyses including the proportion of subgroups experiencing the ATE-EnvSci curriculum and the gender, home language, and Hispanic/Latino race by curriculum type plots see Appendix F.

Propensity Score Analyses

To control for the possibility that the demographic variables were influencing the observed effect currently attributed to curriculum type, researchers employed propensity score methodology. This methodology enables one to "balance" groups based upon their probability of receiving treatment in this case, the ATE curriculum, given their observed covariates (see Appendix F and Rosenbaum & Rubin, 1983 as well as D'Agostino, 1998, for a detailed explanation of the creation and use of these scores). Somewhat counter to other methods of statistical control, the calculation of propensity scores takes place outside of the data analytic model and can be improved iteratively without inflating Type I error rates. Therefore, it is recommended that as many background variables as possible be included in the initial propensity score calculation (Rubin, 1997) in order to rule out competing explanations for group differences (Leow et al., 2004). Figure 1 depicts the results of propensity score calculation and stratification of subjects according to their predicted probabilities of receiving the ATE curriculum, and Table 7 provides the exact number and type of subjects in each stratum. Stratifying participants among five levels has been shown to reduce approximately 90% of the between group bias (Cochran, 1968). The key is that subjects within a given stratum are expected to be comparable with observed background differences "balanced."

Figure 1: Predicted Probability of Receiving ATE Curriculum Given Propensity Score

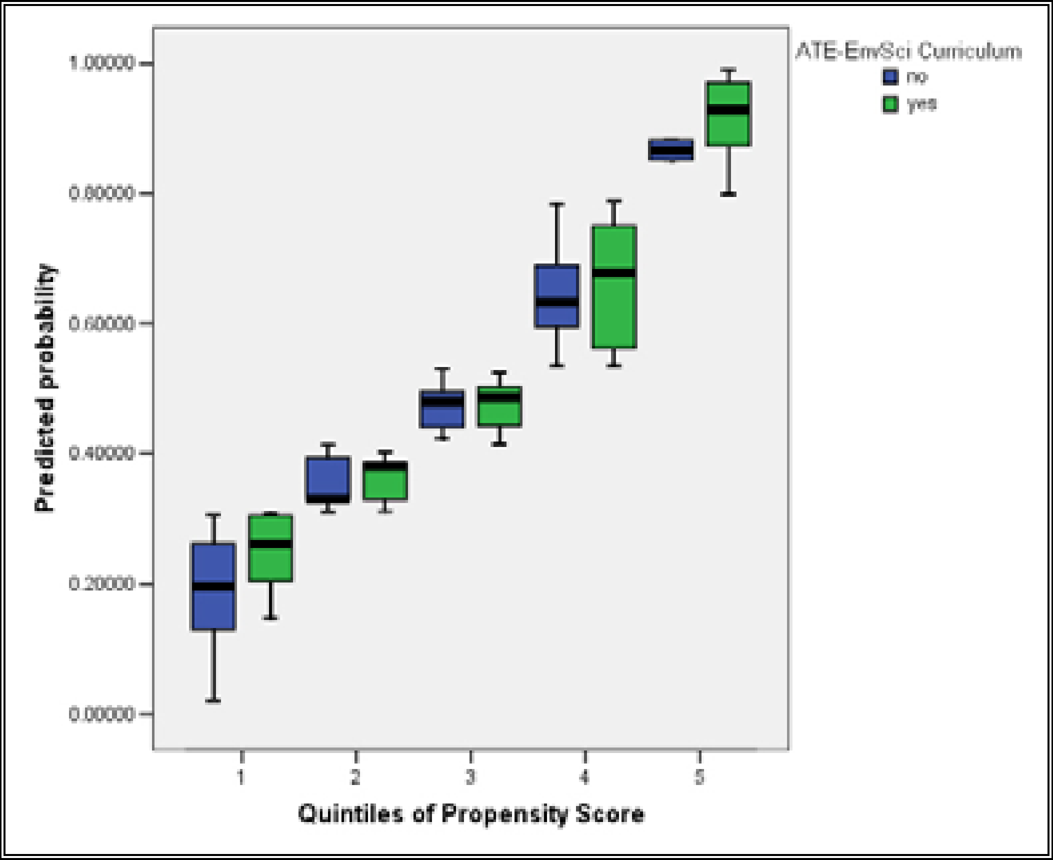


Table 7: Group Membership by Quintile

Quintiles of Propensity Score	ATE-EnvSci Curriculum	N
1	**no**	21
	yes	6
2	**no**	18
	yes	9
3	**no**	13
	yes	14
4	**no**	11
	yes	17
5	**no**	2
	yes	25

Researchers then evaluated the effectiveness of the stratification by propensity scores for reducing biases between treatment and control groups. Substantial size variation of the two groups (ATE-EnvSci vs. comparison) across quintiles (e.g., 6 vs. 21 in quintile 1 and 25 vs. 2 in quintile 5) and therefore uncertainty whether the assumption of normality would hold, prompted researchers to conduct non parametric tests of background variable differences. The results of these tests were not significant (at $p<.05$ nor the more appropriate alpha-adjusted $.05/30 = p<.0017$) and suggested that within each quintile the ATE-EnvSci subjects no longer differed from the comparison subjects on any of the 6 background variables previously examined (see Table 8 for results).

Table 8: Post-Stratification Nonparametric Test Analyses of Background Variables

Test Statistics(a)

Quintiles	Test	Science Related Grades	All Grades	HS Science Courses	College Science Courses	Science Field	Type of Degree
1	Mann-Whitney U	46.000	62.500	47.500	50.500	53.500	59.000
	Wilcoxon W	67.000	83.500	278.500	281.500	74.500	80.000
2	Mann-Whitney U	53.000	75.000	78.000	81.000	67.000	63.000
	Wilcoxon W	224.000	246.000	249.000	126.000	238.000	234.000
3	Mann-Whitney U	85.000	63.000	72.000	82.500	81.500	90.000
	Wilcoxon W	190.000	168.000	177.000	187.500	186.500	195.000
4	Mann-Whitney U	91.000	76.000	93.000	86.000	85.500	74.000
	Wilcoxon W	244.000	142.000	246.000	152.000	151.500	140.000
5	Mann-Whitney U	24.500	24.000	20.000	22.000	22.500	23.000
	Wilcoxon W	349.500	27.000	23.000	347.000	347.500	26.000

a Grouping Variable: ATE-EnvSci Curriculum

Outcome Comparison Analysis

The propensity score quintile variable was then utilized to control background differences while comparing the ESA scores of students receiving ATE-EnvSci and non-ATE curricula. To this end, researchers conducted a stepwise hierarchical regression analysis using quintile of propensity score in the first model and then including both the propensity score quintile variable and the ATE-EnvSci variable in the second model with total ESA score as the dependent variable. The R^2 change value of the second model suggested that 10.8% of the variance beyond that explained by the propensity score information was explained with the addition of the ATE-EnvSci curriculum variable to the model. The R^2 value of .108 corresponds to an r value that Cohen (1992) would consider a moderate effect ($r = .328$) (see Table 9 for results and Appendix F for information on the fit of the model).

Table 9: Analysis of Propensity Score, ATE-EnvSci Curriculum, and ESA Score.

Model Summary[c]

Model	**R**	R^2	**Change Statistics** R^2 **Change**
1[a]	.374	.140	.140***
2[b]	.498	.248	.108***

p<.001***

a Predictors: (Constant), Quintiles of Propensity Score

b Predictors: (Constant), Quintile of Propensity Score, ATE-EnvSci Curriculum

c Dependent Variable: ESA Total Score

These results suggest that receiving the ATE curriculum had an effect on total ESA score beyond what is explainable by the educational background and future aspiration variables. To enable examination of differences in ESA Total Score by propensity quintile, graphic depictions of mean ESA scores are displayed in Figures 2-3. These graphics show that within each quintile the mean ESA score of those receiving the ATE-EnvSci Curriculum was higher than for those who did not. Given the widely varying numbers of subjects in those ten groups (curriculum received (2) X (5) quintiles) a statistical test of these means would not be appropriate. Rather, the depiction of means is better used as a supplement to the statistical results of the regression analysis described earlier.

Figure 2: Mean Quintile Differences of ESA Score by Curriculum Type

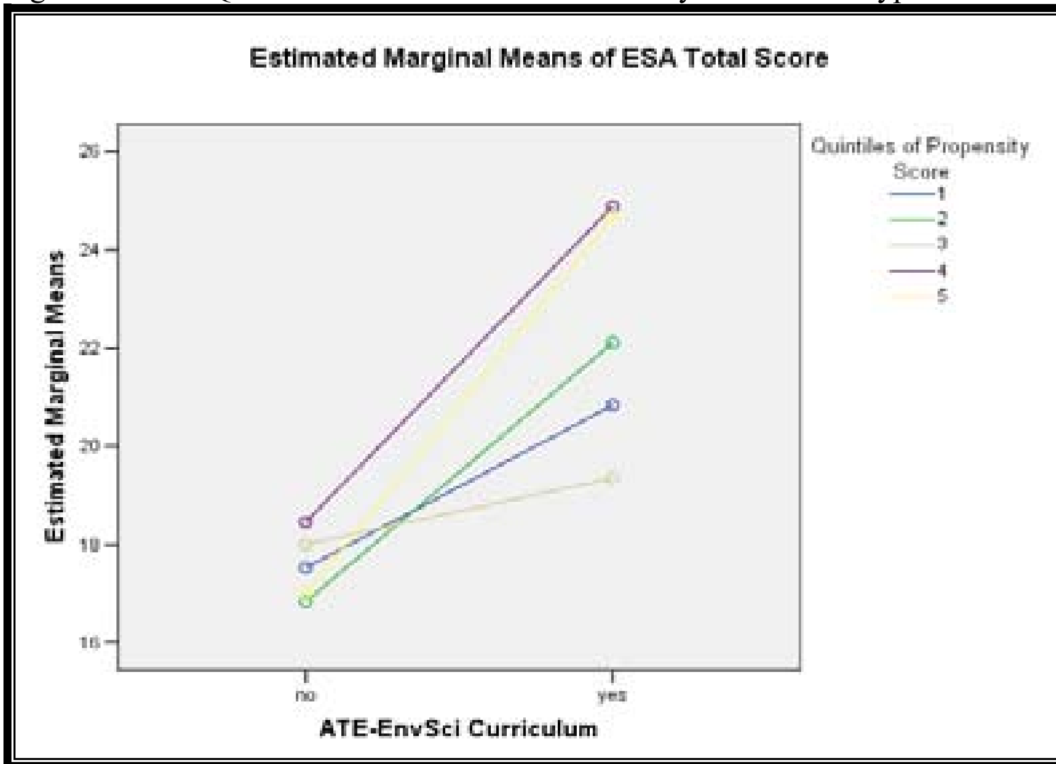
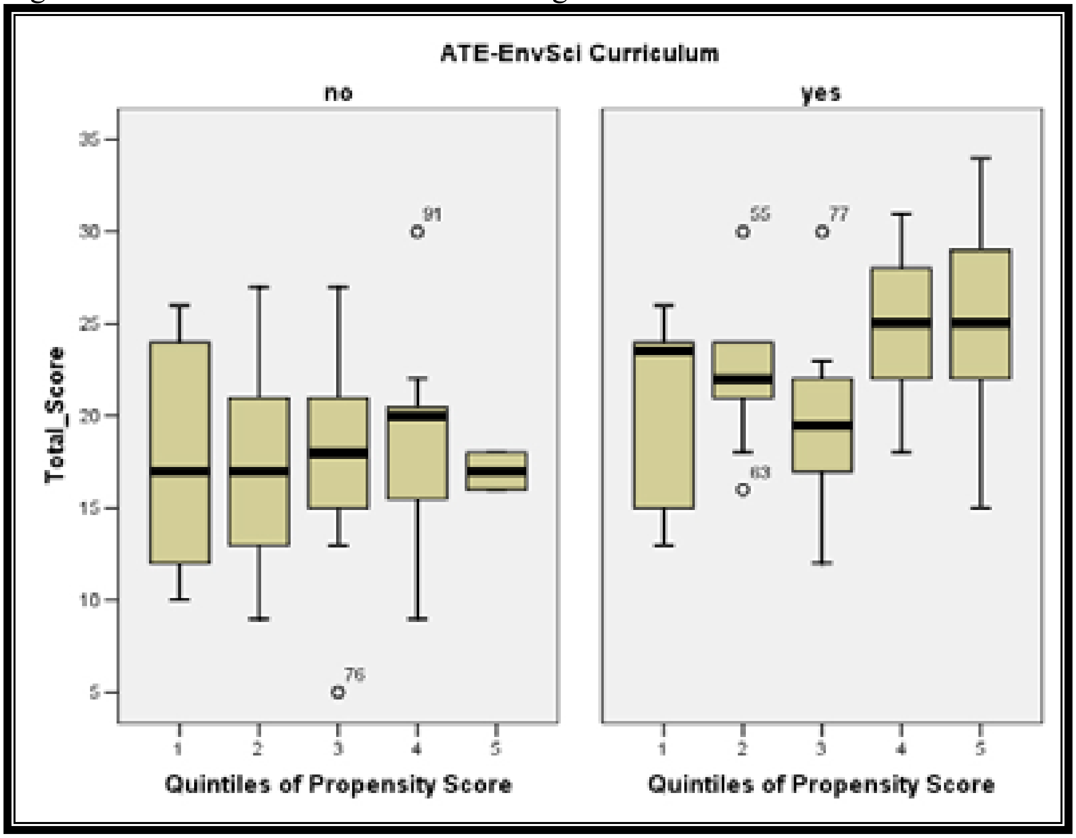


Figure 3: Mean Total ESA Scores and Ranges for ATE-EnvSci and Non-ATE Students



Finally, researchers examined the difference between student scores on the ESA for students receiving the ATE-EnvSci curriculum from different instructors. Here the concern addressed was the possibility that the higher average score for students receiving the ATE-EnvSci curriculum could be the result of one instructor. Instructor means, as shown in Table 10 suggest little instructor-level differences. Additionally, a t-test comparing the scores of the students of the two ATE-EnvSci curriculum instructors was nonsignificant. These results support the conclusion that the ATE-EnvSci curriculum, itself, was responsible for the differences in total ESA score rather than interacting with, or being solely dependent upon, instructor characteristics. The results of the t-test comparing instructors are presented in Table 10.

Table 10: Instructor-Level ESA Score Statistics for the ATE-EnvSci Curriculum

ESA Results by Instructor				
	Instructor	**N**	**Mean**	**SD**
Total ESA Score	*ATE-EnvSci Instructor A*	41	23.29	5.474
	ATE-EnvSci Instructor B	36	22.58	4.487

ATE-EngTech Analyses

The analyses of the data from the ATE-EngTech sites followed much the same procedures as those used in the environmental science analyses. As a first step the data were examined to determine any pre existing differences between the two groups. These included first a determination of the quality of the assessment instruments and their scoring. This was measured through internal consistency and intra- and inter-rater reliability. Then differences between the groups in their background characteristics such as course taking, ethnicity, gender, etc. were examined. Any differences were related to differences in scores on the outcome assessment. Because it was determined that prior differences existed between the groups, analysis of covariance was used to equate the groups. Propensity score analysis such as used with in the environmental science analyses was not possible due to lower sample size. Analysis of covariance allows for the addition of a few variables to the analysis of differences between the means to help adjust for pre-existing differences. The types of analyses, the justification for using them, and the outcomes associated with them are detailed in Table 11.

Table 11: Guide to ATE-EngTech Analyses

Guide to ATE-EngTech Analyses		
Analyses	*Justification*	*Outcome*
• Intra-rater, inter-rater and internal consistency reliability tests	• Examine scoring consistency and test coherence	• High intra-rater (93% and 88%) , high inter-rater (87% to 98%) and adequate internal (α = .70) reliability
• Descriptive analysis of credits accrued	• To determine the comparability of matched (non-randomly assigned) groups	• More ATE than Non-ATE students reported completing the equivalent of 2 semesters of coursework
• Chi-square tests of differences between ATE and Non-ATE students on educational background variables and aspirations	• To check for pre-existing differences between ATE and Non-ATE groups.	• Significantly more Non-ATE aspire to field in science • ATE students had a greater frequency of taking at least three high school science courses.
• T-test analysis of ATE and Non-ATE students for college GPA	• To check for pre-existing differences between ATE and Non-ATE groups.	• No significant differences
• Propensity Score Methodology	• To stratify participants according to probability of receiving ATE materials	• Too few subjects to distribute adequately over five strata and "balance" ATE and Non-ATE groups
• Scatterplot, histogram, and univariate tests with means plot graphs	• To examine the distribution of the data	• Distributions different between groups. Means plots suggest interactions but are not significant
• Inclusion of significantly different background variables in step 3 ANCOVAs	• To model background differences between groups.	• See below
• Bivariate correlations	• To examine the relationship between the AWATA original score, AWATA secondary scale score, ET PSA Form I, and ET PSA Form II.	• High significant correlation between AWATA original and secondary scale scores, moderate correlation between ET PSA Form I and AWATA original score, non significant correlation between AWATA original score and ET PSA Forms II
• ANCOVAs for AWATA Secondary Scale Score, ET PSA Form I and II	• To examine ATE-EngTech materials' effects while controlling for background variables	• No significant differences between ATE and Non-ATE students on any of the three assessments
• Nonparametric analyses for ET PSA Form I and II	• To examine curricular effects with small samples	• No significant differences between ATE and Non-ATE students.

Test Reliability and Item Functioning

One expert (involved in the assessment development) scored all Engineering Technology Problem-Solving Assessments (ET PSAs) and then rescored 10 assessments of each form resulting in a 93% intra-rater reliability on Form I and 88% on Form II. Inter-rater reliability estimates using two additional experts indicated that 87% of the scores of the two raters were within one-half point of each other with 97% within one point of each other on Form I. Results for Form II indicated that 91% of the scores of the two raters were within one-half point of each other with 98% within one point.

The ACT WorkKeys Applied Technology Assessments (AWATAs) were computer-scored via company-based processes. The AWATA has a reliability coefficient of .70, but more meaningful are the reliability of classifications which range from 83% to 100% depending on the level (3-6) to which subjects are classified (ACT, 2001).

Demographic Analyses

Researchers began between group comparisons with descriptive analyses of the number of credits completed by students at each site. Using the number of credits or modules per each individual school expected after two semesters of work suggests that 84% of students receiving the ATE curriculum indicated completing the equivalent of two semesters of coursework while only 24% of students receiving non-ATE curriculum indicated doing so. Table 12 details the comparison.

Table 12: Student Progress in Program by Site

Critical Number of Credits/Modules After the Equivalent of <u>3</u> Semesters:

Non-ATE Curriculum	Credits:
School A:	70
School B:	43
School C:	46
ATE-EngTech Curriculum	Modules:
Schools 1-3	13-16

- If student responses are deemed to represent (3^{rd}) semester credits/modules, 1 of 4 School C, 1 of 2 School B, 0 of 4 School 3, 1 of 18 School 2, 10 of 15 School 1, and 1 of 15 School A students meet criteria (3 of 21 non-ATE (14%) and 11 of 37 ATE-EngTech (30%)).

CRITICAL NUMBER OF CREDITS/MODULES AFTER THE EQUIVALENT OF <u>2</u> SEMESTERS:

Non- ATE-EngTech Curriculum	Credits:
School A:	53
School B:	34
School C:	32
ATE-EngTech Curriculum	Modules:
Schools 1-3	10-12

- If student responses are not deemed to represent current 3^{rd} semester credits/modules (3 of 4 School C, 1 of 2 School B, 4 of 4 School 3, 14 of 18 School 2, 13 of 15 School 1, and 1 of 15 School A meet criteria (5 of 21 non-ATE (24%) and 31 of 37 ATE (84%)).

Alternate explanations such as the integration of several courses into a single module by ATE sites, the recall of 10-15 modules as easier than 20-25 stand alone courses, and the increased recollection of integrated courses may also account for these differences.

Researchers also conducted nonparametric analyses of other demographic variables on which data had been collected. Chi-square tests revealed no significant differences between students who did and did not receive the ATE-EngTech curriculum on variables of engineering course grades, overall course grades, and type of degree sought. A t-test analysis of group differences in college GPA was also nonsignificant.

In contrast, nonparametric analyses of the educational background variables of student intention to pursue a career in a scientific field and the number of high school science courses indicated significant differences ($p<.01$ and $p<.05$, respectively) between students receiving and not receiving the ATE-EngTech curriculum (see Table 13 for further details). More Non-ATE students intended to pursue a career in science (75%) than ATE-EngTech students (32%) yet

ATE-EngTech students had a greater frequency of taking three or more high school science courses (84%) than Non-ATE students (42%). As is apparent, these results are mixed and do not reveal a clear advantage for either group.

Table 13: Nonparametric Examinations of Educational Background Variable Differences

		Career in Science Field**			Total	High School Science Courses Taken*							Total
		No	**Not Sure**	**Yes**		**0**	**1**	**2**	**3**	**4**	**5**	**6**	
ATE-EngTech Curriculum	**No**	2	3	15	20	0	3	8	5	1	1	1	19
	Yes	4	21	12	37	1	2	3	17	10	2	2	37
Total		6	24	27	57	1	5	11	22	11	3	3	56

p<.05* *p*<.01** (based on Chi-Square Statistic)

Given these significant between group results, researchers considered using propensity scores to balance the differences between the ATE-EngTech and comparison students. Researchers calculated propensity scores, but with the smaller number of students (N=58) in the ATE-EngTech propensity score analysis compared to the ATE-EnvSci (N=136) analysis it was difficult to appropriately subdivide the scores. In order to ensure that more than one subject of each group (ATE-EngTech and Non-ATE) was assigned to each subdivision of propensity score, the number of strata had to be reduced to two. This limited number of strata is believed to reduce between group bias by much less (64%) than the 90% expected when using five strata (Cochran, 1968). Since the goal of using propensity scores was to reduce nearly all of the differences in observed variables, the use of two strata was inadequate for achieving that end. Therefore, in this analysis researchers decided to forgo the use of propensity scores and controlled for differences by including the two significant variables as covariates in the initial univariate analyses.

Outcome Comparison Analyses

Preliminary scatterplot, histogram, and means plot graphs (see Appendix F) revealed that the data differed slightly from a normal distribution and that assessment scores on both forms of the Engineering Technology Problem-Solving Assessment (ET PSA) and the ACT WorkKeys Applied Technology Assessment (AWATA) may interact with race for African-American and White students. To account for these possibilities, researchers decided that outcome analyses should involve nonparametric in addition to parametric procedures and also decided to test for curriculum type by race interactions. Tests for curriculum type by race interactions were not significant.

Initially, researchers examined the correlations between the measures of engineering technology problem-solving skills. These results suggested a high correlation between the AWATA original level score and the secondary scale score derived from it. That this correlation is significant, high, and positive (r-square statistic indicates that 92% of the variance in one is explained by the other) is important and necessary in order to use the secondary scale score to differentiate students. Additionally, only Form I of the ET PSA correlates significantly with the AWATA (r = .39) (see Table 14 for complete results).

Table 14: Assessment Measure Intercorrelations

Pearson Correlations Among Assessment Measures

	Total AWATA Secondary Scale Score	**Total AWATA Original Score**	**Total ET PSA Form I Score**	**Total ET PSA Form II Score**
Total AWATA Secondary Scale Score	1			
Total AWATA Original Score	**.961*** **	1		
Total ET PSA Form I Score	.287	**.394***	1	
Total ET PSA Form II Score	.193	.276	.(a)	1

p<.05* *p*<.001***

(a) Cannot be computed (i.e., as there are no students who took both forms)

As both assessments purport to measure an examinee's level of problem-solving skills in applied technology settings, the lack of significant correlation between the AWATA and ET PSA Form II raises some questions. Perhaps this lack of a correlation between the AWATA and ET PSA Form II and lack of a higher correlation between the AWATA and ET PSA Form I (only 16% of variance of one accounted for by the other) results from the response format. For instance, the AWATA consists of an entirely multiple choice format, while the ET PSA consisted entirely of constructed response items. Another potential explanation is the greater breadth of the AWATA assessing applied technology in general, while the ET PSA focuses specifically upon skills within the field of engineering technology.

Researchers then examined both Non- ATE and ATE-EngTech student performance on the AWATA and ET PSA). Scores are portrayed for only the secondary scale of the AWATA as this scale provided a finer level of measurement (25 scale points) and correlated very highly (r = .96) with the five-point level score scale. Scores for the ET PSA will be provided for both Form I and Form II. Since each student completed one form or the other, comparisons are only possible within forms and calculating an overall score across forms would be misleading.

Following these preliminary analyses, researchers statistically tested the scores of students who received the ATE-Eng Tech curriculum with those who did not across each of the measures (i.e. AWATA Secondary Scale Score, ET PSA Form I, and ET PSA Form II). For each of these initial univariate tests, the demographic variables that differed significantly between the ATE-EngTech and Non-ATE groups (i.e., aspirations for a career in science and high school science courses taken) were covaried in the analysis. Each of these analyses of group differences resulted in nonsignificant findings. The means and standard deviations for ATE and Non-ATE groups on the three measures are displayed in Tables 15-17.

Table 15: Group Descriptives for AWATA Secondary Scale Score

Descriptive Statistics Dependent Variable: Total AWATA Secondary Scale Score[a]			
ATE-EngTech Curriculum	**Mean**	**Std. Deviation**	**N**
no	76.63	5.727	19
yes	75.11	3.843	37
Total	75.63	4.575	56

[a]HS Science Courses and Aspirations for Career in Science included as Covariates

Table 16: Group Descriptives for ET PSA Form I

Descriptive Statistics Dependent Variable: Total ET PSA Form I Score [a]			
ATE-EngTech Curriculum	**Mean**	**Std. Deviation**	**N**
no	5.222	3.1929	9
yes	3.889	2.9633	18
Total	4.333	3.0477	27

[a]HS Science Courses and Aspirations for Career in Science included as Covariates

Table 17: Group Descriptives for ET PSA Form II

Descriptive Statistics Dependent Variable: Total ET PSA Form II Score [a]			
ATE-EngTech Curriculum	**Mean**	**Std. Deviation**	**N**
no	7.800	4.3792	10
yes	8.895	3.5457	19
Total	8.517	3.8114	29

[a]HS Science Courses and Aspirations for Career in Science included as Covariates

In addition to these analyses, researchers also used nonparametric tests to conduct comparisons involving smaller sample sizes (i.e., ET PSA Form I: N=27 and ET PSA Form II: N=31) as the distributions of these groups may not be normal. Test results indicated that there were no significant differences between students receiving the ATE-EngTech curriculum and those receiving Non-ATE materials. The number of participants and rank information used in the analysesare depicted in Tables 18-19.

Table 18: Mean Rank and Sum of Ranks Information for ET PSA Form I

	Curriculum Type	N	Mean Rank	Sum of Ranks
ET PSA Form I Total Score	*Non-ATE*	9	16.28	146.50
	ATE-EngTech	18	12.86	231.50

Table 19: Mean Rank and Sum of Ranks Information for ET PSA Form II

	Curriculum Type	N	Mean Rank	Sum of Ranks
ET PSA Form II Total Score	*Non-ATE*	12	14.04	168.50
	ATE-EngTech	19	17.24	327.50

Power Considerations

An important point to consider with the ATE-EngTech result is the power of these analyses. Such a consideration not only introduces an alternate explanation for findings, but suggests an obstacle in evaluations such as these. According to Cohen (1992) to utilize ANOVA with conventionally desired power of .80 (β = .20) and an alpha of .05 to find a medium effect (.50) with two groups requires at least 64 participants. As is quickly evident, the AWATA comparison neared this number with 58 participants, but the ET PSA comparisons were substantially short of that number (N = 27 and N = 31). The ET PSA analyses would have been able to detect a large effect (.80) which requires an N of 26, but may miss smaller effects. The AWATA sample was also too small to detect a moderate effect of .50 but should have detected slightly larger effects (e.g., .55 or .60).

This issue is being raised for two reasons: 1) the results of the ATE-EngTech analysis are suggestive, but far from conclusive; and 2) despite extensive recruitment efforts by these researchers, the current results highlight the importance of adequate sample size in examining the effectiveness of materials. It is interesting that, using the same criteria, a more than adequate sample was obtained for the ATE-EnvSci comparison in which Cohen (1992) would have recommended 67 participants.

There were not enough subjects for each instructor to conduct between instructor comparisons.

SUMMARY OF OUTCOMES

As mentioned in the introduction this ATE materials development evaluation report is the fourth in a series. It examined the effectiveness of two curriculum materials developed by ATE sites to improve student learning. Achievement of students taught using two of the four highest rated ATE developed curricular materials (i.e., Environmental Science (ATE-EnvSci) and Engineering Technology (ATE-EngTech)) was compared with the achievement of students taught using other materials. In addition to the effectiveness data, the procedures outlined in this report provide examples for other ATE projects to adapt and implement.

Four achievement measures were employed for the study; three were constructed and one standardized achievement measure was purchased. The reliability and validity information about each of the four measures is provided. All measures meet minimal standards for validity and reliability although suggestions about improving the measures are provided. All constructed measures were pilot tested and revised before being used for the study.

One measure was used to assess the ATE-EnvSci materials and three were used for the ATE-EngTech materials. The Environmental Science Assessment (ESA) was constructed by selecting items matched to the AAAS 2061 and NRC Science Education Standards for environmental science from existing national tests. The three measures to assess student understanding of engineering technology used for the ATE-EngTech comparison included: the commercially available ACT WorkKeys Applied Technology Assessment (AWATA), and two forms of the instrument, Engineering Technology Problem-Solving Assessment (ET PSA) developed expressly for this study. The ET PSA was developed based on the biotechnology assessment instrument described in previous reports. The developed instruments were designed to assess STEM content in an integrated fashion consistent with real world experience.

Although the researchers tried to ensure that the characteristics of the two groups were the same, students were not randomly assigned to treatment and comparison groups. Because of this group characteristics were assessed to determine the need to account and adjust for potential selection effects. Group comparisons revealed that students receiving the ATE Environmental Science curriculum (ATE-EnvSci) differed significantly ($p<.001$) from Non-ATE students by having

- a greater number of college science courses,
- higher science-related grades, and
- higher degree aspirations.

To ensure consideration of these and other educational background differences in outcome analyses, researchers created a single number (using propensity score methodology) that reflected these disparities between curricular groups. Even after including pre-existing educational background differences in the analyses, ATE-EnvSci students (N=77) performed significantly better than Non-ATE students (N=68, $p<.001$) on the ESA with what would be considered a medium effect size (r = .328) by Cohen's (1992) criteria.

Results of the engineering technology materials outcome comparison revealed that in comparison to Non-ATE students, ATE Engineering Technology (ATE-EngTech) students

- had taken more of their program-specific coursework,
- had taken a significantly greater number of high school science courses, and
- reported significantly lower scientific career aspirations.

Propensity score methodology was again attempted to adjust for group differences, but the sample size was too small. Instead, researchers used the educational background variables that differed significantly between the groups as covariates instead. There were no significant differences between the groups using the ATE (students=37) and non ATE (students=21) materials on the AWATA Secondary Scale Score or on either form of the ET PSA.

Findings and Recommendations:

Because this is the last report in the series and is a culmination of the materials development evaluation work, the recommendations presented here incorporate the understandings developed over the three years of the evaluation. The findings and recommendations are provided in two sets. The first set is recommendations to the ATE program. The second set is recommendations for improving this type of evaluation.

Recommendations for the ATE Program

Use of the selected ATE developed materials produced students with equal or higher levels of achievement than use of traditional materials. Therefore, it seems reasonable that the ATE program should continue to promote materials development. Given the recent changes to the ATE program solicitation, this type of materials development would most likely be couched within program improvement.

However even within program improvement, the recommendation to continue materials development has some caveats. It must be remembered that the materials tested here were representative not of ATE materials development as a whole but of the "best" materials. Materials development efforts in future ATE projects should be informed by the processes used to develop and identify the exceptional materials reported on here. The ATE Materials Development Processes Report (Lawrenz & Appleton, 2004) provides insight into the processes used to develop the materials rated as excellent (pgs 26-27) as well as an integrated model for guiding the development of technological education materials. Furthermore the Evaluation of Materials Produced by the ATE Program report (Keiser, Lawrenz and Appleton, 2003) outlines a process and provides a rubric for external assessment of the quality of curricular materials.

Students in classes using ATE Environmental science materials had higher achievement than students in classes using other environmental science materials. Although this was only one limited study of the effectiveness of these materials, it has several implications:

- Because these materials appear to help students learn environmental science content, efforts could be made to promote the availability of these materials across the country. The accessibility of these materials via the National Center for Sustainable Resources (NCSR) website provides one means of dissemination. However, although the materials were tested at different sites, all sites were in the northwestern part of the US. Additionally, although some evidence of the portability of the materials is provided by the similar findings of effectiveness across two instructors, more evidence is necessary. Therefore the efficacy of these materials for different areas of the country and with a variety of instructors should be examined. Future ATE projects developing materials should provide plans for future dissemination and for guaranteeing portability should their curriculum be shown to be effective.

- Because the process used to develop the environmental science materials resulted in an effective curriculum, that process might serve as a model for other materials development. The description of the development process provided in the ATE Materials Development Processes Report (Lawrenz & Appleton, 2004) indicates that the key features to the success of the development of the environmental science materials were the expertise and personal commitment to the materials by the single developer and the extensive effort he exerted to utilize the results of several iterations of pilot testing to refine the materials over a long period of time. Having a single person develop the materials was a unique element compared to the processes used to develop the other highly rated materials (Lawrenz & Appleton, 2004). However, the use of expertise, extensive pilot testing and revision over a long period of time were more consistent across the processes used to develop the highly rated materials. Therefore it seems likely that expert attention to all aspects of the materials through a comprehensive process of pilot testing and revision would be most likely to result in effective materials.

- Making the assumption that the success of the environmental science materials could be replicated by other ATE materials development efforts lends support to the recommendation above that ATE continue supporting materials development.

Recommendations for evaluations of ATE developed materials

There are several different implications about the evaluation of materials developed by ATE projects that arose throughout our three year effort. Our evaluation process provides a model for any materials evaluation effort. The suggestions and the findings that lead to them are included below in three sections: before field testing, during field testing and research related to field testing.

Before Field Testing

It is important to recognize that the process of evaluating materials will include several steps. The two comparison studies presented in this report highlight the potential and feasibility for careful student outcome based evaluations of ATE curricula, but this type of study is costly and most relevant for materials in their final stages. Other types of evaluation should be used at different stages in the development of materials.

Involve experts. This is already common in ATE materials development in the form of content experts and industry standards but less common in terms of educational or instructional development expertise.

Conduct iterative pilot testing. The materials field tested in this study had undergone extensive pilot testing where the materials were tried out, student and instructor opinions were gathered and modifications were made. Pilot testing should continue until the materials appear to be meeting the prescribed outcomes. The integrated development process provided in the ATE Materials Development Processes Report, (Lawrenz & Appleton, 2004) can serve as a model to help identify all of the issues to consider when developing materials.

Submit the materials to external review. The rubric provided in the Evaluation of Materials Produced by the ATE Program report (Keiser, Lawrenz, & Appleton, 2003) can serve as a model for conducting this type of multiple expert review and to help identify the issues relevant to high quality technological education materials.

During Field Testing

Once materials have evolved past pilot testing and external review, the next expectation could be field tests. Substantial funds need to be available to conduct these. Although as suggested by the comparison studies presented here, it is possible to create reliable (ESA and ET PSA) and valid (ESA and potentially the ET PSA) assessments as well as recruit adequate comparison groups with which to examine the effectiveness of developed materials, it was a very time consuming, intense, process requiring substantial expertise in research design, sampling, measurement, data analysis and reporting. This type of field testing is designed to showcase how the results produced by using newly developed materials compare to the results produced by other materials. Therefore use of field testing should be restricted to instances where this is an important question.

Plan in advance. By planning for comparative studies at the outset and getting agreement from participating courses, it may be possible to both increase participation and reduce costs by building the comparative testing procedures into regular course expectations.

Ensure comparability of the sites. Extensive efforts were used in this evaluation to ensure that the sites selected were as similar to each other as possible. Even with these careful selection procedures, the students at the sites showed some differences which were adjusted for statistically. Anticipating and collecting data on variables related to the outcomes of interest are critical to obtaining meaningful comparisons.

Use valid and reliable assessment instruments. Three instruments were constructed for use in this evaluation and one was purchased. Despite the careful construction and pilot testing of the developed instruments, they may not have worked in exactly the ways envisioned. Furthermore, commercially available instruments may not fit the exact goals of the courses being examined. Appropriate instrumentation is the key to meaningful interpretation.

Consider outcomes in addition to written achievement tests. The sole use of a written response achievement test to measure curriculum effectiveness:

- may overlook materials' effectiveness at attaining other developer goals (e.g., retaining students)
- may not consider other ways of measuring achievement (e.g., different types of tests or the inclusion of other content) that might have resulted in different findings
- may miss the successes apparent when using different definitions of achievement (e.g., narrowing achievement discrepancies across ethnicities)

Therefore, other indices of achievement and other outcomes besides achievement should be considered in determining materials' effectiveness.

Investigate transferability. The present study investigated the portability of materials across instructors. These results are promising, but future studies should more directly investigate the mechanisms for, and the effects of, transporting and using materials in varied locations as well as with different instructors.

Recruit large samples. The samples included in these studies were sufficient to detect substantial to small main effects for the ATE-EnvSci curriculum comparison and sufficient to detect substantial to moderate main effects with the ATE-EngTech curriculum using the AWATA. The samples were not sufficient to detect moderate effects with the ET PSA. The sample sizes were not sufficient to test other important questions such as special contexts or populations (e.g., ethnic groups or sex) which might yield significant differences. Although increasing sample size is difficult, it would enable a broader array of analyses. Recruiting methods should be carefully considered. In designing studies, it is recommended that researchers consider the tradeoff between the examination of longer term programs, the level of similarity of participants, the value of a longer assessment, and the available advertising avenues on the one hand and sample size on the other.

Research Related to Field Testing

Research factors affecting recruitment. The greater prevalence of introductory environmental science courses, the attempt to recruit subjects from similar geographical areas, the length of the assessment (30 vs. 60 minutes), and the method of advertising the opportunity to participate (in or outside of regular class) each seemed to impact the size of the comparison site sample. For at least these reasons, the ATE-EnvSci sample is much larger than the ATE-EngTech (145 vs. 58, respectively). It should also be noted that compensation, while important, may not offset these other factors. This evaluation recruited more students to complete a 30 minute exam for $10 than a 60 minute exam for $20.

Develop more instruments that are relevant to ATE needs. Only one instrument was available commercially that could be used in this study. Development of instruments is costly and time consuming. Developing and providing more instruments aligned with the goals of ATE projects would facilitate comparison studies.

Improve the measurement tools and sources for gathering background data. Researchers encountered challenges in attempting to gather appropriate data as proxies for educational background and future aspiration differences. For instance, non-ATE students may have misinterpreted a question on courses taken and/or to have forgotten the exact names of courses completed. In addition, self-report methods for obtaining student educational background information (e.g., G.P.A., other test performances, and the rigor of high school and college science courses) may be less objective than sources such as transcripts. Therefore more research about how best to measure background variables either through the use of better survey items or the use of different sources of evidence should be conducted.

REFERENCES

ACT. (2001). WorkKeys Technical Handbook. Available from ACT at (500 ACT Drive, P.O. Box 168,Iowa City, Iowa 52243-0168) or (319) 337-1000.

Bloom, B. S., Englehart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956) Taxonomy of educational objectives: The classification of educational goals: Handbook I, Cognitive domain. New York: Longmans, Green.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24,* 205-213.

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

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112(1),* 155-159.

D'Agostino, R. B., Jr. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265-2281.

Keiser, J., Lawrenz, F., & Appleton, J. (2004). Evaluation of materials produced by the ATE program. Report submitted to NSF.

Lawrenz, F., & Appleton, J. J. (2004). ATE materials development processes report. Report submitted to NSF.

Leow, C., Marcus, S., Zanutto, E., & Boruch, R. (2004). Effects of advanced course-taking on math and science achievement: Addressing selection bias using propensity scores. *American Journal of Evaluation, 25(4),* 461-478.

National Research Council (NRC). (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79,* 516-524.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (revised). Newbury Park, England: Sage Publications.

Standards for Educational and Psychological Testing (1999). Washington, DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

APPENDICES

APPENDIX A: ENVIRONMENTAL SCIENCE ASSESSMENT

ENVIRONMENTAL SCIENCE ASSESSMENT
2004-2005

Directions: Please answer all of the questions on this page and then wait until your instructor tells you to begin before going on to the next page.

1. Are you? ○ Male ○ Female

2. Do you usually speak a language other than English in your home? ○ Yes ○ No

3. Which of the following describes you? *(Check all that apply)*
   - o **African-American**
   - o **American Indian** or **Alaskan Native** *("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)*
   - o **Asian** or **Pacific Islander** *("Asian or Pacific Islander" means someone who is Chinese, Japanese, Korean, Filipino, South East Asian, East Indian, Asian American or from some other Asian or Pacific Island background.)*
   - o **Hispanic/Latino** *("Hispanic/Latino" means someone who is South American, Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or from some other Spanish or Hispanic background.)*
   - o **White** *(Not Hispanic or Latino)*
   - o **Other** (Please Specify): ____

How would you describe the grades that you usually get in school?	A's	B's	C's	D's	Below D's
4. In my science-related classes, I get mostly: *(please choose only one answer)*	○	○	○	○	○
5. In all of my classes, I get mostly: *(please choose only one answer)*	○	○	○	○	○

6. Which of the following science courses have you taken in **high school**? *(Choose all that apply)*
   - o Physical Science
   - o Chemistry
   - o Earth Science
   - o Biology
   - o Physics
   - o Environmental Science
   - o Other *(please describe)*: ____

7. Which of the following science courses have you taken in **college**? *(Choose all that apply)*
   - o Physical Science
   - o Chemistry
   - o Earth Science
   - o Biology
   - o Physics
   - o Environmental Science
   - o Other *(please describe)*: ____

8. Do you plan to work in a scientific field after you graduate?
   - o Yes
   - o No
   - o Not Sure

9. Which choice best represents the highest educational degree you plan to pursue?
   - o An Associate's Degree from a two-year school
   - o A Bachelor's Degree from a four-year school
   - o A degree for graduate school studies

1) An ecosystem will most likely remain stable if

a) it has more predators than prey
b) it has a high level of biodiversity
c) biotic factors decrease
d) finite resources decrease

2) Which human activity would have the most direct impact on the oxygen-carbon dioxide cycle?

a) reducing the rate of ecological succession
b) decreasing the use of water
c) destroying large forest areas
d) enforcing laws that prevent the use of leaded gasoline

3) Decomposers are important in the environment because they

a) convert large molecules into simpler molecules that can then be recycled
b) release heat from large molecules so that the heat can be recycled through the ecosystem
c) can take in carbon dioxide and convert it into oxygen
d) convert molecules of dead organisms into permanent biotic parts of an ecosystem

4) Self-nourishing organisms are called:

a) exotrophs
b) endotrophs
c) autotrophs
d) heterotrophs
e) chemoautotrophs

5) Some factories have a negative impact on the Earth's ecosystems because they

a) have high energy demands that require the use of fossil fuels and nuclear fuels
b) utilize agricultural technology that decreases soil erosion
c) decrease the need for finite resources
d) limit the amount of emissions produced each year

6) For a natural ecosystem to be self-sustaining, many essential chemical elements must be

a) converted to energy
b) changed into fossil fuels such as oil and coal
c) permanently removed from the environment
d) cycled between organisms and the environment

The data table below contains information on the growth of eight white pine trees, planted in eight different locations, after a period of time.

Data Table

Tree Number	Trunk Diameter 1.2 Meters Above Soil Surface (m)	Soil pH	Elevation Above Sea Level (ft)
1	0.54	4.0	1,200
2	0.79	6.5	1,650
3	0.64	4.5	1,400
4	1.04	5.0	1,350
5	0.96	5.0	1,350
6	0.82	4.5	1,250
7	0.80	5.5	1,400
8	0.52	5.0	1,600

7) Which statement is best supported by the data in the table?

a) White pines grow best at higher elevations.
b) White pines are not found at elevations below 1,000 feet.
c) White pines have a long life span.
d) White pines can grow in acidic soil.

8) Which population of predators is potentially most stable?

a) a population that specializes on one type of prey
b) a population that is composed primarily of juveniles
c) a population at its maximum sustainable yield
d) a population that feeds on many types of prey
e) a population that is composed of fertile adults

9) In an area in Africa, temporary pools form where rivers flow during the rainy months. Some fish have developed the ability to use their ventral fins as “feet” to travel on land from one of these temporary pools to another. Other fish in these pools die when the pools dry up. What can be expected to happen in this area after many years?

a) The fish using ventral fins as “feet” will be present in increasing numbers.
b) “Feet” in the form of ventral fins will develop on all fish.
c) The fish using ventral fins as “feet” will develop real feet.
d) All of the varieties of fish will survive and produce many offspring.

10) Which of the following factors would tend to <u>increase</u> the biodiversity of a community?

a) physically diverse habitat
b) environmental stress
c) catastrophic disturbances
d) geographic isolation
e) introduction of an exotic species

11) The source of energy for the Earth’s water cycle is the

a) wind
b) sun’s radiation
c) earth’s radiation
d) sun’s gravity

12) The nitrogen cycle, one of the most important biochemical cycles, may cause environmental problems because too much nitrogen can:

a) result in acid rain
b) deplete the ozone shield
c) contribute to the greenhouse effect
d) reduce earthshine
e) cause eutrophication in bodies of water

Ecosystem	Events
A	A severe ice storm occurs during the winter, damaging trees and shrubs. No ice storms occur for the next 20 years.
B	A severe drought causes most of the leaves to fall from the trees during a single summer. There are no serious droughts for the next 20 years.
C	An island with a dense shrub population becomes submerged for 3 years. When the river water lowers, the island does not become submerged for the next 20 years.
D	A fire burns through a large grassy area. Fires do not occur in the area for the next 20 years.

13) Which ecosystem would most likely require the most time for ecological succession to restore it to its original state?

a) A
b) B
c) C
d) D

14) The three basic kinds of interaction between species are:

a) competition, symbiosis and predation-parasitism
b) competition, parasitism and adaptive radiation
c) symbiosis, predation-parasitism and migration
d) migration, symbiosis and adaptive radiation
e) there are more than three basic kinds of interaction

15) Human demography suggests that an improving economy in a country correlates with:

a) decreased birth rate, increased population growth rate
b) decreased death rate, increased population growth rate
c) decreased birth rate, decreased population growth rate
d) increased birth rate, decreased population growth rate
e) increased birth rate, increased population growth rate

16) Which population is most likely to exhibit an evolutionary response to a change in its environment?

a) a population in which all organisms are genetically identical and which has a high rate of reproduction
b) a population which has genetic variability but with a low rate of reproduction
c) a population in which the effect of intraspecific competition is reduced by behavioral adaptations
d) a population undergoing genetic drift
e) a population highly susceptible to predators

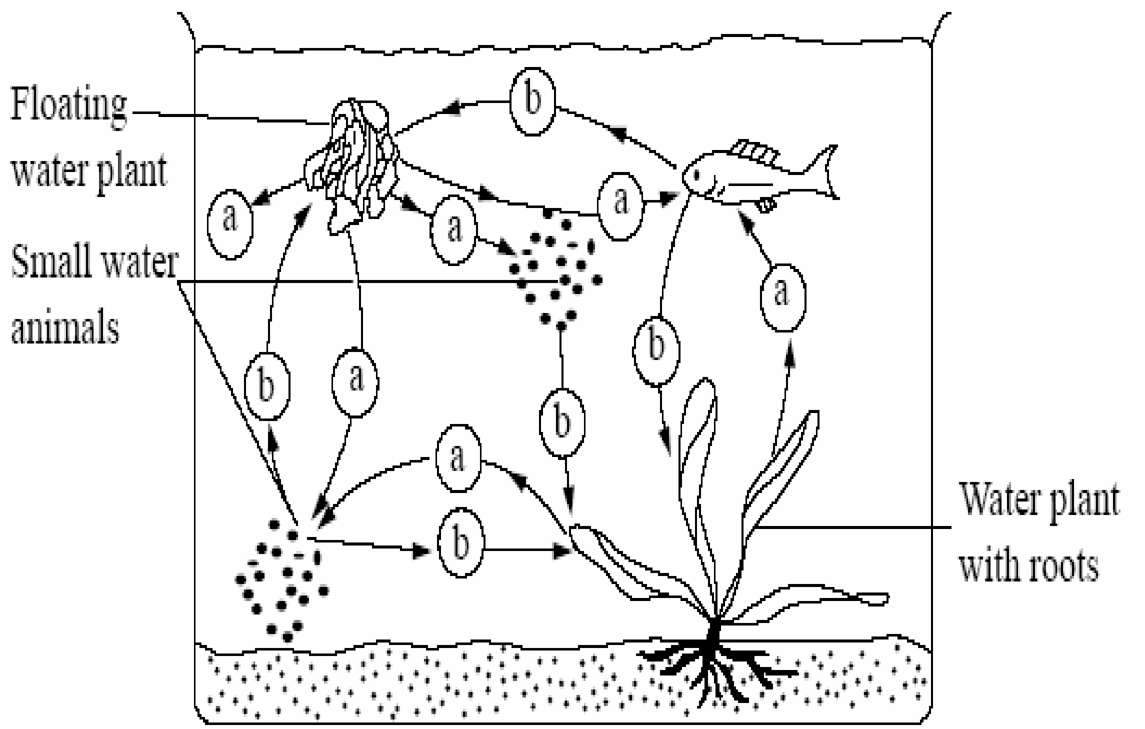
17) "Carrying capacity" refers to:

a) the maximum weight that can be put on a vehicle or machine
b) the nutrient value of a food source
c) the amount of a mineral resource that can be recovered economically from a mine
d) the average life-expectancy of an individual in a population
e) the maximum number of individuals that can be supported by an ecosystem

18) A scientist tested a hypothesis that white-tailed deer would prefer apples over corn as a primary food source. The findings of the test, in which the scientist claimed that the deer preferred apples, were published. Which research technique, if used by the scientist, might result in this claim being questioned?

a) The scientist observed four deer in different locations at various times of the day.
b) The scientist observed a total of 500 deer in 20 different locations at various times of the day.
c) The scientist observed 200 deer in various natural settings, but none in captivity.
d) The scientist observed 300 deer in various locations in captivity, but none in natural settings

The diagram below shows an example of interdependence among aquatic organisms. During the day the organisms either use up or give off (a) or (b) as shown by the arrows.



19) Choose the right answer for (a) and (b) from the alternatives given.

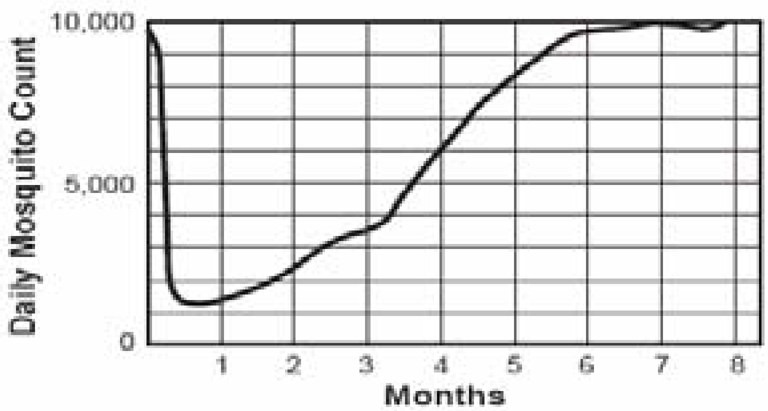
a) (a) is oxygen and (b) is carbon dioxide
b) (a) is oxygen and (b) is carbohydrate
c) (a) is nitrogen and (b) is carbon dioxide
d) (a) is carbon dioxide and (b) is oxygen
e) (a) is carbon dioxide and (b) is carbohydrate

20) Tomato plants in a garden are not growing well. The gardener hypothesizes that the soil is too acidic. To test this hypothesis accurately, the gardener could

a) plant seeds of a different kind of plant
b) move the tomato plants to an area with less sunlight
c) change the pH of the soil
d) reduce the amount of water available to the plant

*Base your answer to question **21** on the information and graph below and on your knowledge of environmental science.*

A small community that is heavily infested with mosquitoes was sprayed weekly with the insecticide DDT for several months. Daily counts providing information on mosquito population size are represented in the graph below.



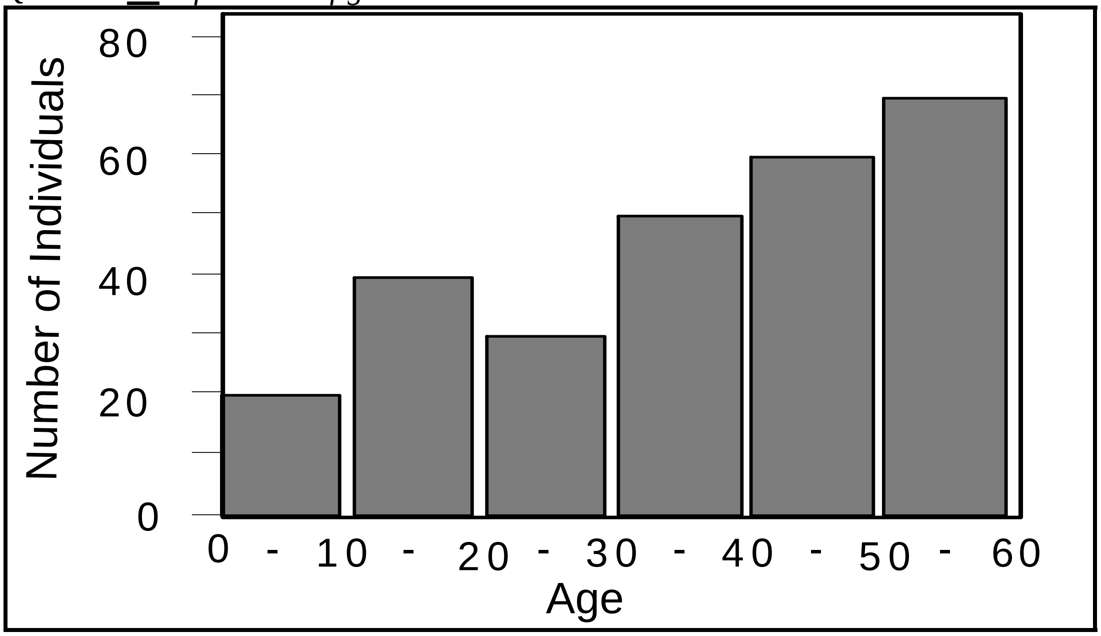
21) Which statement best explains why some mosquitoes survived the first spraying?

a) The weather in early summer was probably cool.
b) Most of the mosquitoes were of reproductive age.
c) Environmental factors varied slightly as the summer progressed.
d) Natural variation existed within the population.

22) Why might a stream or river, especially one in a humid area, have a much higher proportion of heterotrophs to autotrophs?

a) CO_2 dissolved in water is less readily available for photosynthesis in streams
b) aquatic food chains are longer and therefore support more heterotrophs
c) high pH in humid-area waterways inhibits photosynthesis
d) less sunlight penetrates into the water
e) a lot of dead organic matter from the land is washed into the water and is a potential food source for heterotrophs

Question ***23*** *refers to the figure below:*



23) Examine the graph above illustrating the age distribution of people in a tribe in the Amazon basin. What might be the significance, from a demographic standpoint, of this distribution for the future of this population?

a) the population is growing toward the maximum human population sustainable by the basin
b) the total population size is increasing exponentially and they will therefore run out of resources
c) the population is growing logistically and is balanced with its resources
d) the population has type I survivorship curve.
e) the population is not replacing itself and may go extinct

24) On cold days, snakes usually lie very still and eat little or nothing, while birds usually move around and eat a lot of food. Which statement best explains this?

a) Both animals are cold-blooded, but without feathers to keep warm, snakes get too cold to move
b) Unlike birds, snakes are warm-blooded; they must hibernate during cold weather.
c) Unlike snakes, birds are cold-blooded; they are less affected by the cold than snakes
d) Unlike snakes, birds are warm-blooded; they must eat food to maintain a constant temperature.

25) The second law of thermodynamics states that when energy is transformed from one form to another, it always goes from a more useful form to a less useful form. We also know that living organisms create order from disorder and create useful forms of energy. Is life a violation of the second law of thermodynamics?

a) No, because biological energy is fundamentally different from physical energy.
b) No, because organisms create greater overall disorder in the process of creating order locally.
c) Yes, but organisms do not obey the laws of thermodynamics.
d) No, because organisms transfer energy with near 100% efficiency, while energy transfer in physical systems is almost always much lower.
e) Yes, and the thermodynamic police are on their way.

26) A student designed an investigation to determine the effect of temperature on the rate of seed germination. The student placed moist filter paper in each of four culture dishes. Ten bean seeds were placed on the filter paper in each dish. The four dishes were numbered and placed in the dark at different temperatures as follows: Dish 1: 10°C, Dish 2:15°C, Dish 3:20°C, Dish 4:25°C. The total number of germinated seeds in each culture dish was counted each day for two weeks.

What data table is best for recording the results of this investigation?

Petri Dish	Day	Temperature	Amount of Light
1			
2			
3			
4			

a)

Day	Temperature			
	Dish 1	Dish 2	Dish 3	Dish 4

b)

Petri Dish	Amount of Water	Number of Germinated Seeds	Amount of Light
1			
2			
3			
4			

c)

Day	Number of Germinated Seeds			
	10°C	15°C	20°C	25°C

d)

*Base your answer to question **27** on the cartoon below, which refers to certain concepts of natural selection, and on your knowledge of environmental science.*

Cartoon 2

"Listen... I'm fed up with this 'weeding out the sick and the old' business... I want something in its prime."

27) Identify one concept represented in the cartoon above and explain how this concept supports the theory of natural selection. Your answer must:

- identify *one* concept represented in the cartoon [1]
- briefly explain the concept you identified [1]
- explain the relationship between this concept and the process of natural selection [1]

When an animal or plant species is introduced to an area where it has never previously existed, it frequently creates a problem by multiplying out of control and displacing established species. One way of fighting introduced species is to poison them. This may be impractical, be very costly or carry heavy risks. Another method, called *biological control*, involves the use of living organisms, other than human beings, to control the pest species.

28) Give an actual example of biological control.

29) For a long-term ecology study, a meadow in a large forest is divided into two plots. One plot is mowed once a year, while the other plot is not. Describe what each plot will look like after 40 years and justify your answer.

APPENDIX B: ENGINEERING TECHNOLOGY PROBLEM SOLVING ASSESSMENT (FORM I)

Case A: Your company will be installing an automatic just-in-time assembly line to assemble electrical switches. The assembly process requires ten operations all of which will be automatic except one that requires an operator. Parts will be moved from one station to another by a conveyer belt. Each day the line will assemble switches to fill the previous day's orders. Based on the number of daily orders, the assembly line will not need to run more than seven hours a day to complete production. Before installing the line, your team has tested assembly stations for each operation and determined the production rate and maximum repair time for each station. Station two has a production rate of 10 units per minute and a maximum repair time for a technician of 45 minutes. Station three has a production rate of 10 units per minute and requires less than 5 minutes to repair. All other stations require 5 minutes or less to repair. Since this is a just-in-time line with one operator, the line must be designed to run continuously during the time of operation. How should your team design the configuration of station two and three and the conveyer system between two and three to ensure continuous operations during a single 8-hour shift in the event that station two needs repair?

1. What are four possible ways to solve this problem?

    1.____

    2.____

    3.____

    4.____

2. a. Which solution to the problem would you use? ________________________

____

b. Why is this solution better than the others?

____

____

____

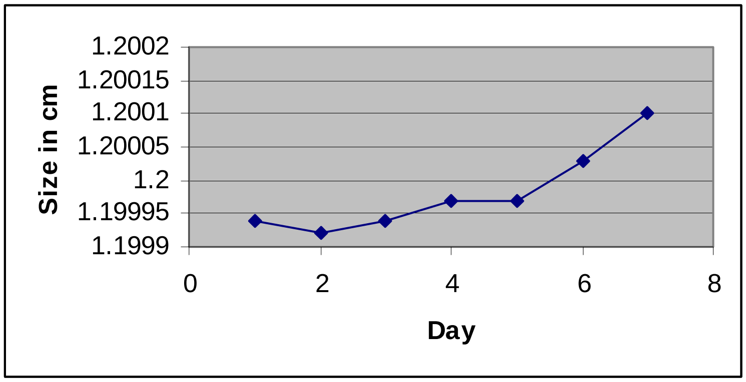
____

____

____

Case B: You work as a technician responsible for quality control in the production of automotive parts. A part is to be machined by a computer numerical control (CNC) lathe to a length of 1.20000 cm +/- .00015 cm. As a quality check each day, measurements of 25 random parts are averaged. Results of seven days are:

Day 1	1.19994 cm
Day 2	1.19992 cm
Day 3	1.19994 cm
Day 4	1.19997 cm
Day 5	1.19997 cm
Day 6	1.20003 cm
Day 7	1.20010 cm



3. a. Do you have a problem? Yes, No, or Maybe (circle one).

Why or why not?

____

____

____

b. If there is a problem, what is it?

____

____

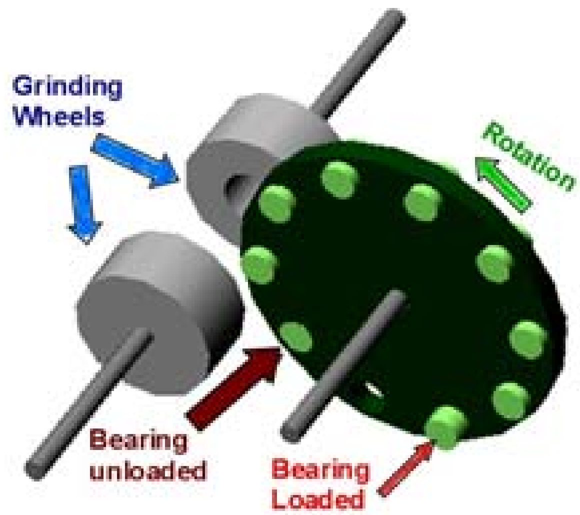
____

____

____

____

Case C: Below is a sketch of a machine that grinds the ends of cylindrical bearings to proper length and perpendicular to the cylindrical axis. The bearings are ground round and ground cut (oversize) in operations prior to this grinding station. The precut bearings are loaded into the holding disk in the lower right hand side and rotated into the grinding position by rotating the disk counter clockwise. After grinding, the bearings are ejected into a collection bin. Twenty percent (20%) of the bearings ground by this machine are out of specifications. You have eliminated problems with the grinding wheels and need to determine if the holding disk is warped.



4.a. What do you do next?

____

____

____

b. Why?

____

____

____

____

____

____

Case D: Your engineering consulting team is designing a sprinkler system for a new five-floor office building to be built in a small town. The building site has one of the highest elevations of the town. The town water is supplied from a well, with pressure maintained through a pumping system. Your team has recommended that the sprinkler system for the office building be supplied with water from a reservoir placed on the roof.

5. a. List two significant advantages of your solution:

1.____

2.____

b. List two significant disadvantages of your solution:

1.____

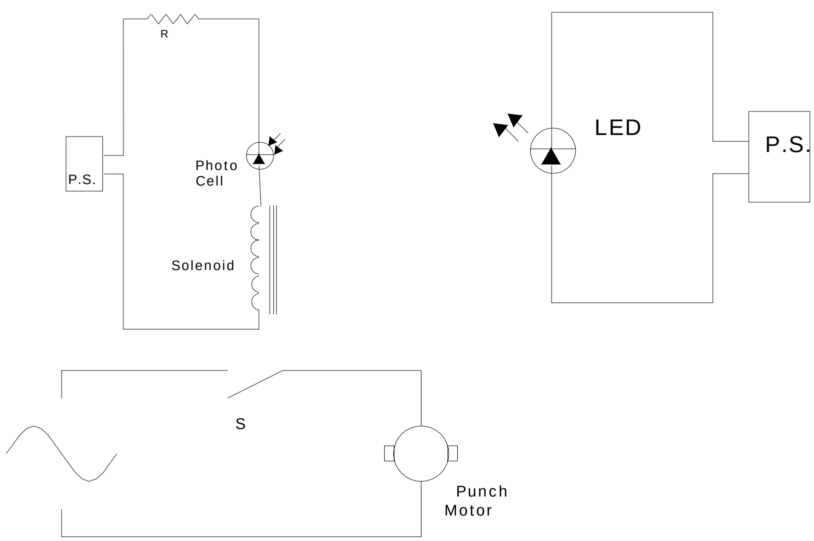
2.____

6. What information will you gather and how will you gather it to find out if your solution will work?

APPENDIX C: ENGINEERING TECHNOLOGY PROBLEM SOLVING ASSESSMENT (FORM II)

Case A: In your plant is an assembly operation that requires manual insertion of a metal strap and metal plate into a press to attach the strap to the plate. The strap and plate are aligned in the press and tabs on the plate are pressed around the strap when the operator removes his hands from the press. To prevent the press from operating while the operator's hands are in the press, there is a safety lockout that detects the operator's hands in the press. The safety lockout on this press uses an optical system at the opening to the press. When the operator breaks the light beam at the opening to the press, the power to the press is shut-off. The light source is a LED that shines across the opening. The LED illuminates a photocell that is in series with a resistor (R), power supply (PS), and a solenoid. When the cell is illuminated there is sufficient current flowing in the detector circuit to cause the solenoid to hold the switch (S) closed to supply power to the press.

The schematics below show the circuits for the LED, light detection, and the punch motor. The operator has informed you that the press will not operate.



1. What are four possible causes for the press not to operate?

    1.____

    2.____

    3.____

    4.____

2. Explain what two measurements you would gather that would help you determine the actual cause of the problem.

3. What will the measurements tell you about the operation of the system?

Case B

Directions: Read the following case and write an e-mail that contains your answers in the space below. You may use the back of the page as well, if necessary.

4. Your company is a high volume (one unit every three seconds) manufacturer of an electronic unit that attaches to a submersible pump. Your company supplies units to ten major customers. The casing for the units is a sealed metal box except for a round electrical port one and a half (1.5) inches in diameter. A round neoprene cap is placed over the port by a robotic assembly process. The cap is used to protect the unit during shipping and it makes a water-tight seal.

One of your major customers, whose assembly plant is located in Canada, has called in late January to tell you that they are now experiencing a water leak around some of the caps when they are tested in a water bath. This is a recent occurrence and other customers in more temperate climates have not reported a problem with the units. These units are shipped to the Canadian plant by truck and held in the truck at the loading dock until transferred to the assembly plant. The time it takes for the units to leave your plant and be held at the loading dock until they are taken into the assembly area may be up to two weeks. You have been sent to the Canadian plant to inspect the units they have received, to determine the cause for the water leakage, and to recommend actions that should be taken to satisfy the customer.

You are to write an e-mail to your supervisor explaining the following.

a. the problem(s)
b. possible causes of the problem
c. how you will know which one is the real cause
d. possible solutions both immediate and long term
e. the solution you recommend and why it is best
f. & g. your plan to monitor the solution (In other words, what information will you gather and how will you gather it to find out if this solution is working? What criteria will you use to determine if this solution is successful?)
h. what you will do if the solution doesn't work

Your supervisor will judge your email and solution to the problem on accuracy, cost-effectiveness, efficiency and ability to convey the information she needs to know.

____

____

____

____

____

APPENDIX D: ACT WORKKEYS ASSESSMENT DESCRIPTION

Article I. Applied Technology

The WorkKeys *Applied Technology* test measures the skill people use when they solve problems with machines and equipment found in the workplace. This skill includes four areas of technology: electricity, mechanics, fluid dynamics, and thermodynamics. Individuals need to know the basic principles of each area.

The Applied Technology skill focuses on reasoning, not math. Therefore, individuals do not need to make calculations or use formulas to solve problems. When individuals use the Applied Technology skill, they can:

- Analyze a problem by identifying the problem and its parts.
- Decide which parts of a problem are important.
- Decide on the order to follow when dealing with the parts of the problem.
- Apply existing tools, materials, or methods to new situations.

Applied Technology Terms

Electricity involves the flow of electrons. Individuals need to know how electrical current moves through a circuit or a system and how electricity affects a circuit or a system. For example, they need to know how to control current and resistance.

Mechanics involves the way solid things move and how leverage, force, friction, and momentum affect that motion. Individuals need to solve problems with simple machines, complex machines, and mechanical systems.

Fluid dynamics involves the way fluids (liquids and gases such as water and air) move through systems. Individuals need to use this knowledge to solve problems with plumbing, hydraulics, or pneumatics (compressed gas).

Thermodynamics involves the movement of heat. Individuals need to know which substances warm up quickly when heated and which ones warm up more slowly. For example, air heats faster than water. They also need to know how specific heat works. That is, they need to know how different materials hold heat for different amounts of time. They need to solve problems with refrigeration, heating, air conditioning, and phase changes.

Article II. Applied Technology

Characteristics/Skills

There are four levels of difficulty. Level 3 is the least complex and Level 6 is the most complex. The levels build on each other, each incorporating the skills assessed at the previous levels. For example, Level 5 includes the skills used at Levels 3, 4, and 5. Individual problems may involve only one area of technology, but each skill level requires individuals to know the basic principles of all four areas at that skill level.

Level	Characteristics of Items	Skills
3	• Straightforward • One simple system that generally has two to five components • Situation exhibits clear physical symptoms • Situation usually has only one variable • All needed information is present • Only elementary technical terms are used	• Identify how basic tools work • Identify how simple machine parts work • Apply basic principles to solve problems involving a simple system • Solve basic problems • Identify the clear physical symptom that points to the potential source of a problem • Identify the best solution after eliminating clearly unsuitable possibilities

Level	Characteristics of Items	Skills
4	• Moderately complex because they can involve two or more simple systems that work together or one moderately complex system • Systems may have up to ten components • Situation can have one or two variables • All needed information is present • Extraneous information may be included • Less common technical terms are defined	• Understand the operation of moderately complex tools and diagnostic equipment • Understand the operation of moderately complex machines and systems • Apply less obvious basic principles to solve problems within physical systems • Solve moderate problems • Eliminate physical symptoms that do not point to the source of a problem, disregarding extraneous information • Identify the best solution after eliminating other unsuitable possibilities

Level	Characteristics of Items	Skills
5	• Moderately complex or advanced, involving two or more simple tools or systems that affect each other or a complex system that includes several components • Systems perform somewhat complex operations and generally have more than ten components • May involve two or three variables and may require use of technical knowledge • Extraneous information is often included • Technical terms may be explicitly defined or their meaning can be implicit in context and illustrations	• Understand the operation of moderately complex tools and diagnostic equipment, choosing the best tool for the task • Understand the operation of complex machines and systems • Apply two or more principles of technology as they interact in moderately complex systems • Solve moderate and advanced problems • Eliminate physical symptoms that do not lead to the source of a problem by disregarding extraneous information; use clues to find the source of a problem • Identify the best solution after eliminating unsuitable possibilities

Level	Characteristics of Items	Skills
6	• Advanced, involving complex tools or systems with more than ten components • Include large amounts of information and present a variety of possible problem sources that are subtle and difficult to diagnose • Require the use of technical knowledge • Contain considerable extraneous information • Technical terms may be explicitly defined or their meaning may be implicit in complex context and illustrations	• Understand the operation of complex tools and diagnostic equipment, choosing the best tool for the task • Understand the operation of complex machines and their components • Apply two or more principles of technology as they interact in complex systems • Solve advanced problems where a variety of mechanical, electrical, thermal, or fluid faults could be the reason for the problem • Eliminate physical symptoms that do not lead to the source of a problem by disregarding extraneous information; use less obvious clues to find the source of a problem • Test possible hypotheses to ensure the problem is diagnosed correctly and the best solution is found

Appendix E: NRC Standards and AAAS 2061 Benchmarks For ESA Construction

Table 20: Consolidated Standards and Original Benchmarks and Standards

CONSOLIDATED STANDARD (CS)	SOURCE(S)
Biodiversity/Natural selection	***Benchmark(s):*** 5A1, 5F6, 10H6; ***NSES****:* Biological Evolution 2;
Ecosystems	***Benchmark(s):*** 5D: 1-2, 5E2, 7F1; ***NSES****:* Interdependence of 0rganisms: 2-4; Matter Energy and Organization in Living Systems 5, Population Growth 3
Energy and the environment	***Benchmark(s):*** 5E3; ***NSES****:* Matter Energy and Organization in Living Systems: 1-2; EES4
Geochemical cycles	***Benchmark(s):*** 4C1; ***NSES****:* Geochemical Cycles: 1-2
Human effects on the environment	***Benchmark(s):*** 3C: 4,5; 5D3; 5E1; 7C2; ***NSES****:* Interdependence of Organisms 5; Natural Resources: 1,2,3; Environmental Quality: 1,2,3; Natural and Human-Induced Hazards: 2,3
Population size and rate of growth (influences and results)	***Benchmark(s):*** 7C1; ***NSES****:* Population Growth: 1-2
Scientific method/scientific inquiry	***Benchmark(s):***1A: 1-3; 1B: 1-7; 1C7; ***NSES****:* Abilities Necessary to Do Scientific Inquiry: 1-6, Understanding Scientific Inquiry: 1-6

CS #	Original benchmarks and standards (**Benchmarks in bold type;** *NSES in italic type)*	Description
1	**The Living Environment (Diversity of Life)...5A1**	Variation(organisms & species) ensuring survival
1	**The Living Environment (Evolution of Life)…5F6**	Natural selection/fluctuating usefulness of traits
1	**Historical Perspectives (Explaining the Diversity of Life)...10H6**	General scientific acceptance of evolution
1	*Biological Evolution...2*	Diversity of organisms
2	**The Living Environment (Interdependence of Life)...5D1**	Ecosystem stability
2	**The Living Environment (Interdependence of Life)…5D2**	Ecosystem fluctuation with climate or species change
2	**The Living Environment (Flow of Matter and Energy)…5E2 /** *Population Growth…3*	Carrying capacity
2	**Human Society (Social Conflict)…7F1**	Sources of group conflict (e.g., resources)
2	*Interdependence of Organisms...2*	Ecosystem energy flow is unidirectional
2	*Interdependence of Organisms…3*	Organism cooperation and competition w/in ecosystems
2	*Interdependence of Organisms…4*	Carry capacity vs. organism reproductive ability
2	*Matter, Energy, and Organization in Living Systems…5*	Organisms and populations limited by matter, energy, and recycling ability of the ecosystem
3	**The Living Environment (Flow of Matter and Energy)…5E3**	Continuous energy from the sun is required and passes through food web or dissipates
3	*Matter, Energy, and Organization in Living Systems...1*	Continuous energy required for matter to maintain organized state
3	*Matter, Energy, and Organization in Living Systems…2*	Solar energy as processed by plants provides energy for life processes

3	*Energy in the Earth System...4*	Solar energy (as impacted by dynamic and static processes) determines global climate
4	**The Physical Setting (Processes That Shape the Earth)...4C1**	Plant alterations of atmosphere
4	*Geochemical Cycles...1*	Fixed amounts of elements move as part of geochemical cycles
4	*Geochemical Cycles…2*	Matter movement and change in chemical and physical properties
5	**The Nature of Technology (Issues in Technology)…3C4**	Human impact on other species
5	**The Nature of Technology (Issues in Technology)…3C5**	Effects of human inventiveness
5	**The Living Environment (Interdependence of Life)…5D3 /** *Interdependence of Organisms…5*	Human alteration of ecosystems
5	**The Living Environment (Flow of Matter and Energy)…5E1**	Creation and burning of fossil fuels
5	**Human Society (Social Change)…7C2**	Impact of the decisions of one generation upon another
5	*Natural Resources...1*	Role of environmental resources in sustaining human existence
5	*Natural Resources…2*	Stress of human consumption on the earth's finite resources
5	*Natural Resources…3*	Natural systems change can be beyond human capacity to adapt
5	*Environmental Quality...1*	Human alterations of ecosystem processes may be detrimental to humans
5	*Environmental Quality…2*	Earth physical and chemical cycles are affected by materials from human societies
5	*Environmental Quality…3*	Numerous factors influence environmental quality
5	*Natural and Human-Induced Hazards...2*	Human activities can enhance hazard potential by accelerating rates of natural change
5	*Natural and Human-Induced Hazards…3*	Hazard rates vary from rapid to slow, but all negatively affect society
6	**Human Society (Social Change)...7C1** / *Population Growth...2*	Many factors affect population size and rate of growth

6	*Population Growth...1*	Factors influencing and results of...population growth
7	**The Nature of Science (The Scientific World View)…1A1**	Scientists assume that the universe is a vast single system in which the basic rules are the same everywhere and that these rules, while varying from very simple to extremely complex, can be discovered by careful, systematic study.
7	**The Nature of Science (The Scientific World View)…1A2**	Major shifts in the scientific view of how the world works occur infrequently and most changes are small modifications of prior knowledge. Change and continuity are persistent features of science.
7	**The Nature of Science (The Scientific World View)…1A3**	There is always the chance that a new theory will fit observations as well or better than a current theory. Science is marked by the continuous testing, revising, and occasionally discarding of theories. While this process leads to an increasingly better understanding of how things work in the world, it does not lead to absolute truth. Evidence of the usefulness of this process is in scientists improving ability to offer reliable explanations and accurate predictions.
7	**The Nature of Science (Scientific Inquiry)…1B1**	Investigations are conducted for many reasons (e.g., to explore new phenomena, check on previous results, test how well a theory predicts, and compare different theories).
7	**The Nature of Science (Scientific Inquiry)…1B2**	Hypotheses are used for choosing what data to attend to, choosing what additional data to seek, and for guiding the interpretations of the data.
7	**The Nature of Science (Scientific Inquiry)…1B3**	When scientists cannot control conditions in order to obtain evidence (e.g., for practical or ethical reasons) they try to observe as wide a range of natural occurrences as possible to be able to discern patterns.
7	**The Nature of Science (Scientific Inquiry)…1B4**	Despite different traditions in science about what is investigated and how, they all value evidence, logic, and good arguments. Agreement prevails that progress in all fields of science depends on intelligence, hard work, imagination, and even chance.
7	**The Nature of Science (Scientific Inquiry)…1B5**	Even groups of scientists may have trouble being entirely objective about their methods and findings leading to the expectation that scientific teams seek out possible sources of bias in both the design of their investigations and their data analysis. Checking each other's results and explanations helps, but does not guarantee objectivity.
7	**The Nature of Science (Scientific Inquiry)…1B6**	Initially new ideas that do not fit well with mainstream ideas in science often meet with vigorous criticism, yet in the long run theories are judged by how well they fit with other theories, the range of observations they explain, how well they explain observations, and how effective they are in predicting new findings.

7	**The Nature of Science (Scientific Inquiry)…1B7**	New ideas in science are limited by the context within which they are conceived; are often rejected by the scientific establishment; may spring from unexpected findings; and usually grow slowly, through contributions from many investigators.
7	**The Nature of Science (The Scientific Enterprise)…1C7**	The commitment of science to peer review and publication keep the vast majority of scientists well within the bounds of ethical professional behavior. Deliberate deceit is rare and likely to be exposed leading to condemnation by the scientific community accompanied by difficulty regaining the respect of this community.
7	*Abilities Necessary to Do Scientific Inquiry…1*	Identify questions and concepts that guide scientific investigations
7	*Abilities Necessary to Do Scientific Inquiry…2*	Design and conduct scientific investigations
7	*Abilities Necessary to Do Scientific Inquiry…3*	Use technology and mathematics to improve investigations
7	*Abilities Necessary to Do Scientific Inquiry…4*	Formulate and revise scientific explanations and models using logic and evidence
7	*Abilities Necessary to Do Scientific Inquiry…5*	Recognize and analyze alternative explanations and models
7	*Abilities Necessary to Do Scientific Inquiry…6*	Communicate and defend a scientific argument
7	*Understanding Scientific Inquiry…1*	Scientists usually inquire into the functioning of systems and do so based on the historical and current knowledge base as well as proposed explanations by other scientists.
7	*Understanding Scientific Inquiry…2*	Investigations are conducted for many reasons including discovering new aspects of the natural world, explaining recently observed phenomena, or testing the conclusions of prior investigations/predictions of current theories.
7	*Understanding Scientific Inquiry…3-4*	Technology and mathematics are essential to scientific inquiry and impact the questions posed, accuracy, precision, and type of data gathered, and the explanations constructed and communicated.
7	*Understanding Scientific Inquiry…5*	Scientific explanations adhere to criteria such as a logically consistent proposed explanation that follows the rules of evidence, is open to questions and possible modification, and based on historical and current scientific knowledge
7	*Understanding Scientific Inquiry…6*	New knowledge and methods result from different types of investigations and public communication between scientists. Arguments must be logical and demonstrate connections between natural phenomena, investigations, and the historical body of scientific knowledge. Clear reporting of methods and procedures is needed to enhance opportunities for further investigation.

APPENDIX F: METHODOLOGICAL AND STATISTICAL DETAILS

<u>COHEN'S D CALCULATION:</u>

Cohen's *d* was calculated using the means of each of the demographic variables by group (ATE-EnvSci vs. Non-ATE) and a pooled measure of the standard deviation.

The formula used was: $d = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(SD_1^2 + SD_2^2)/2}}$ and the means and standard deviations used were from the following output with ATE-EnvSci group mean used as $\overline{X}_1$. The *d* calculations were .36, .68, and .56 for science-related grades, college science courses, and type of degree, respectively.

Table 21: Descriptive Statistics for ATE-EnvSci Groups

	ATE-EnvSci Curriculum	N	Mean ($\overline{X}$)	Std. Deviation (SD)
HS Science Courses	**no**	26	2.69	.884
	yes	8	3.00	.756
Science Related Grades	**no**	65	2.86	.682
	yes	74	3.11*	.713
All Grades	**no**	26	3.35	.562
	yes	8	3.25	.707
College Science Courses	**no**	67	1.36	.595
	yes	76	1.87***	.869
Science Field	**no**	67	1.58	.819
	yes	76	1.83	.870
Type of Degree	**no**	67	2.06	.600
	yes	76	2.39**	.568

$p<.05$* $p=.001$** $p<.001$***

NON-PARAMETRIC EXAMINATION OF DEMOGRAPHIC DIFFERENCES:

Table 22: Proportion of Selected Subgroups Experiencing ATE-EnvSci Curriculum

Crosstabs

		Gender**		*Total*	Other Language in the Home		*Total*	Hispanic /Latino		*Total*
		male	**female**		**yes**	**no**		**yes**	**no**	
ATE-EnvSci Curriculum	**no**	18	48	*66*	16	50	*66*	14	54	*68*
	yes	38	37	*75*	10	61	*71*	8	69	*77*

p<.01**

Figure 4: Gender by Curriculum Type Data

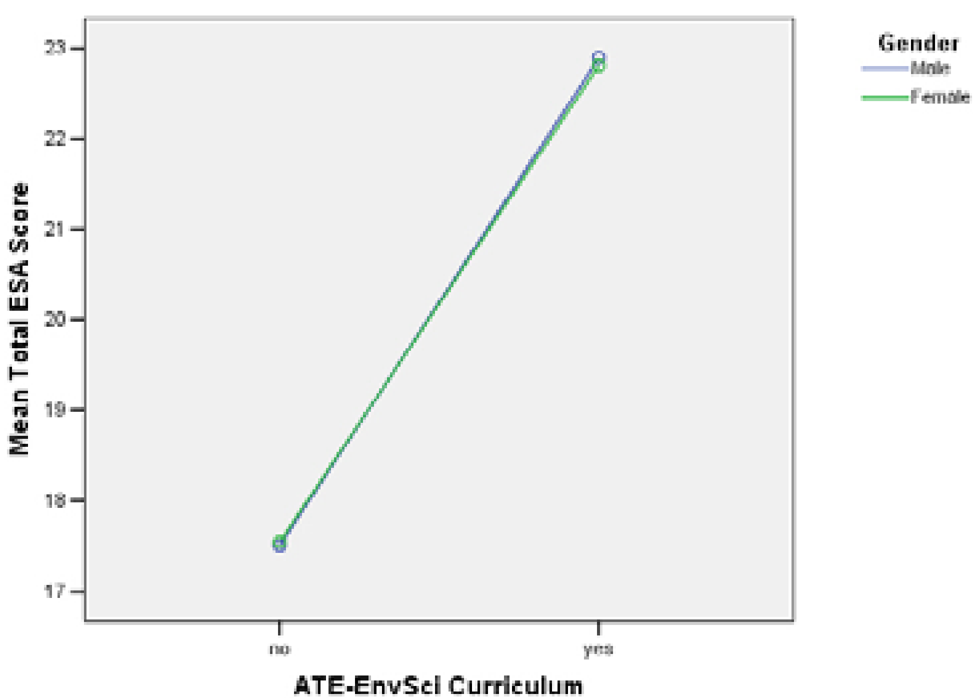


Figure 5: Other Home Language by Curriculum Type Data

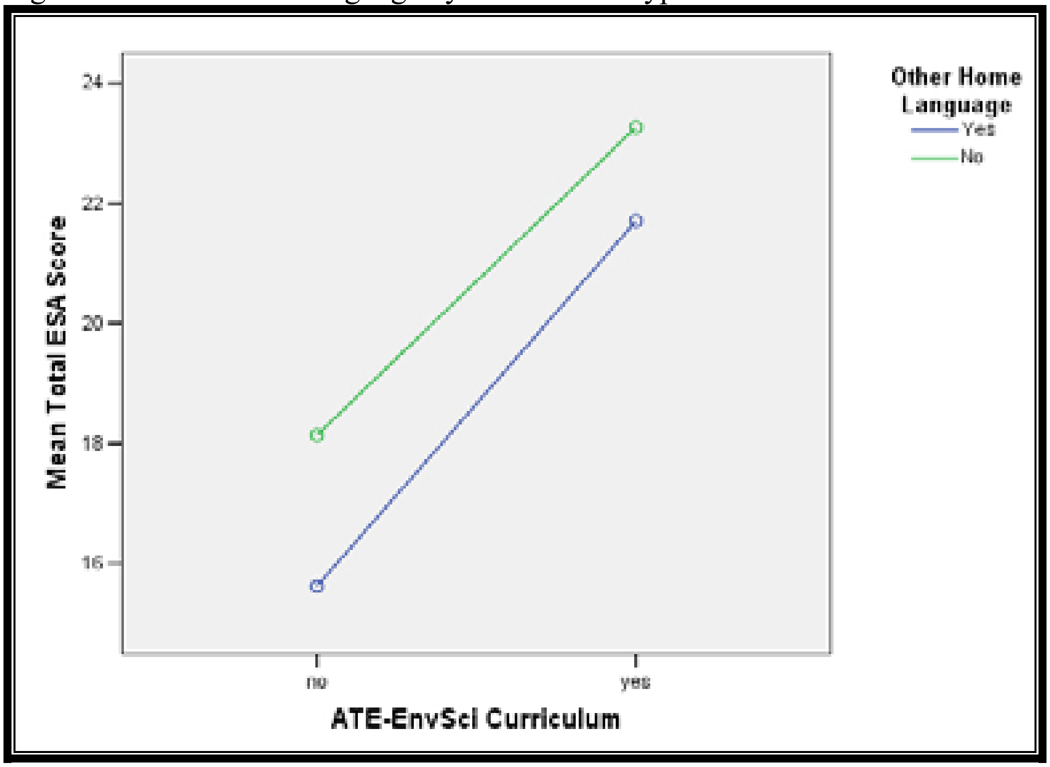
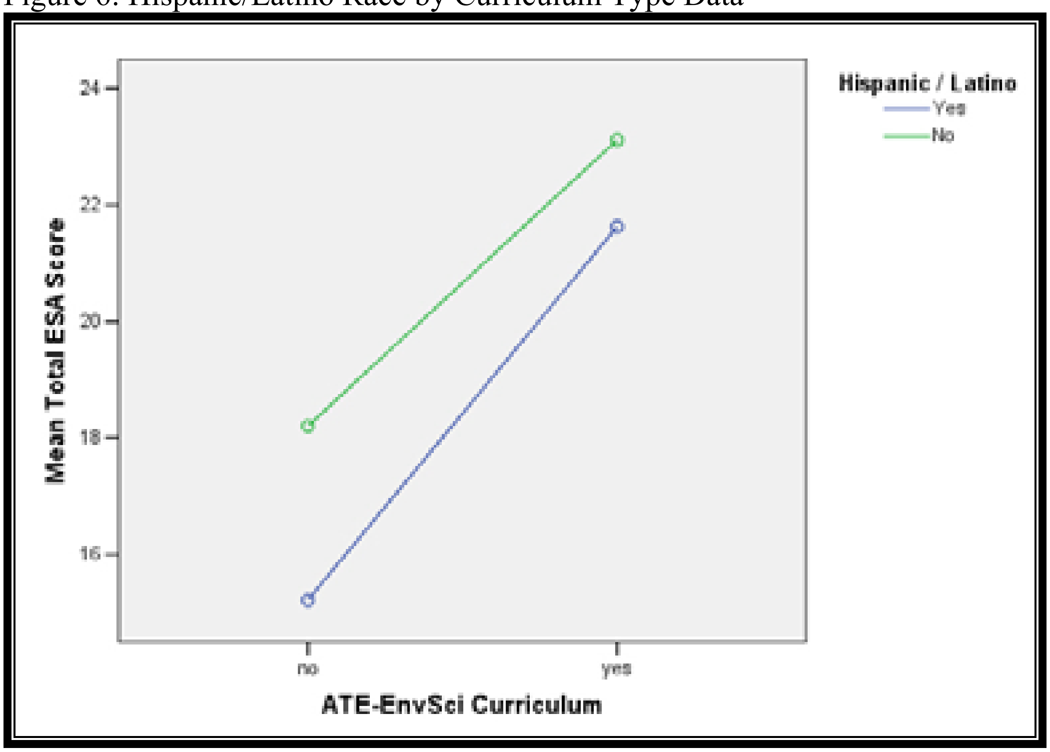


Figure 6: Hispanic/Latino Race by Curriculum Type Data



PROPENSITY SCORE CALCULATION AND STRATA ASSIGNMENT DETAILS:

Propensity score analysis often involves logistic regression and is useful in that it maps a large number of variables onto a single scalar. This scalar represents the probability of receiving treatment given the subject's scores on the variables included in the analysis. Another benefit of propensity scores is that the analyses take place outside of the data analytic model. Therefore researchers are able to conduct numerous iterations of the analyses to determine which variables and terms (e.g., squared terms, interactions) best capture the probability of receiving treatment.

Researchers in medical fields (where investigators frequently lack control over treatment assignment) as well as in economics have been using propensity scores for some time (D'Agostino, 1998). In defining a propensity score, D'Agostino cites Rosenbaum and Rubin (1983) who described it as the conditional probability of a subject being treated given that subject's covariates or

$$e(x_i) = pr(Z_i = 1 | X_i = x_i)$$

where $e(x_i)$ represents the propensity score and is based upon the conditional probability of receiving a specific treatment ($Z_i = 1$) as opposed to control ($Z_i = 0$) given some vector of covariates (x_i) (D'Agostino).

In calculating the propensity score many variables can be included, yet the score itself remains a "scalar summary" (D'Agostino, 1998, p. 2268) of those variables chosen. The benefit (when matching or using these scores as a covariate) is that the score is a single number for each subject rather than several scores or numerous covariates. The goal in using propensity scores is to render the assignment to treatment group "ignorable" (D'Agostino, p. 2266). Retaining Z to represent treatment assignment and utilizing Y to symbolize the response variable(s), the assignment to treatment would be thought essentially ignorable if Z is independent of Y given covariates X ($Y \perp Z \mid X$.) (D'Agostino).

Given a thorough and well-conceived set of propensity scores, the differences between treatment and control subjects on variables specified in the propensity score should be "strongly ignorable." An important consideration is that propensity scores can only account for differences explained by the observed covariates included in the calculation of the score.

One use of propensity scores relevant for the current analyses is stratification of participants based upon their propensity score. Typically 4 to 5 strata are adequate for grouping similar subjects and the range of values for the strata are based upon the distribution of propensity scores (Leow, Marcus, Zanutto, & Boruch, 2004). One final advantage in using propensity scores in this manner is the ability to include differing numbers of control and treatment subjects in the same stratum. That is, the matching need not be a one-to-one type of matching (Leow et al.) Upon separating subjects into the stratum assigned to them based on their propensity score, follow-up analyses can be conducted to determine if the propensity score accounts for the variance that previously would have explained differences in the variable examined. That is, does including a strata specifying variable in the model reduce (and render insignificant) the variance of the treatment factor in differentiating levels of the different educational background variables (e.g., aspirations to work in a scientific field). Given the advantages and relevance of propensity scores for attempting to account for several educational background variables, researchers used that methodology in the current set of analyses.

To create the propensity scores, researchers used six t-test analyses to determine any significant differences between treatment and control groups on the variables of science-related course grades, grades in all courses, high school science courses taken, college science courses taken, aspirations to work in a scientific field, and the highest educational degree students intended to earn. While conducting six t-tests may inflate alpha and indicate more than 5% of differences significant by chance, this type of error is on the side of caution and beneficial for the analysis. In other words, finding differences between background variables merely informs the propensity score model which is later tested for its usefulness anyway. It is better at this point to err on the side of caution rather than overlooking a variable that differs between ATE and Non-ATE groups. The results of the t-tests indicated that science-related grades, college science courses taken, and highest educational degree were significantly different between groups.

Understanding that students receiving the ATE curriculum had achieved significantly higher science-related grades, had taken significantly more college science courses and had significantly higher educational aspirations, researchers computed propensity scores to enable matched comparisons between subjects within similar strata of propensity scores. The first step in computing propensity scores was to include all relevant variables (i.e., those that may be different across levels of curriculum type). Certainly those variables found significantly different (i.e., science-related grades, number of college science courses and highest educational degree the student intends to pursue) were different across levels of curriculum type.

In addition, since propensity score estimation occurs outside of the data analytic model (and therefore does not affect family-wise error levels), one is able to include any variables potentially relevant and engage in as many iterations of estimation as necessary until the scores seem well conceived. To this end, researchers included each of the variables tested (i.e., typical science-related course grades, typical grades in all courses, high school science courses taken, college science courses taken, aspirations to work in a scientific field, and the highest educational degree they intended to pursue) in the initial calculation of scores.

In specifying the logistic regression model to generate propensity scores, it is common to initially include all potentially related variables (Rosenbaum & Rubin, 1984) as well as interaction and quadratic terms as suggested by a correlation matrix divided by the levels of group one is attempting to render equivalent. Correlations that differ substantially across levels of the group suggest an interaction and extremely influential variables may suggest a quadratic term.

Since the correlation matrix is too large to reproduce here, it will have to suffice to say that based upon significant differences in the correlations across levels of the curriculum type variable (ATE-EnvSci vs. Non-ATE) *science-related grades X type of degree (SRG*TOD), all grades X type of degree (AG*TOD), and high school science courses X science field (HSC*SF)* were included in the propensity score model as interaction terms. In addition, quadratic terms for each of the significantly different demographic variables were included (i.e., *college science courses* (CSC^2), *type of degree* (TOD^2), and *science-related grades* (SRG^2).

These variables were entered as one block into a logistic regression analysis with "curriculum type" as the dependent variable. Due to missing data, the logistic regression analysis utilized information from 136 of the 145 subjects. The resulting propensity scores incorporated information from each of the variables included, based upon the usefulness of these variables in explaining a student's inclusion in their respective "curriculum type" group.

Following the calculation of these scores, subjects were initially assigned to one of five strata given their predicted probability of receiving the ATE curriculum (as suggested by the propensity score computed from the 6 educational background variables and the interaction and quadratic terms created from them). Researchers then evaluated the effectiveness of the propensity scores for reducing biases between treatment and control groups.

In order to examine evidence of the effectiveness of the propensity scores at reducing between group bias given the reduced sample size (due to stratification of subjects), researchers conducted 30 non-parametric analyses (5 strata containing 6 variables each) between ATE-EnvSci and Non-ATE groups with the demographic variable as the criterion. Even with the inflated level of alpha, no between-group differences were found on the demographic variables.

These results were considered evidence for the effectiveness of the propensity scores in reducing pre-existing educational background and future aspiration differences between ATE-EnvSci and Non-ATE groups.

AWATA Results:

Figure 7: Range of Student AWATA Scores by Curriculum Type

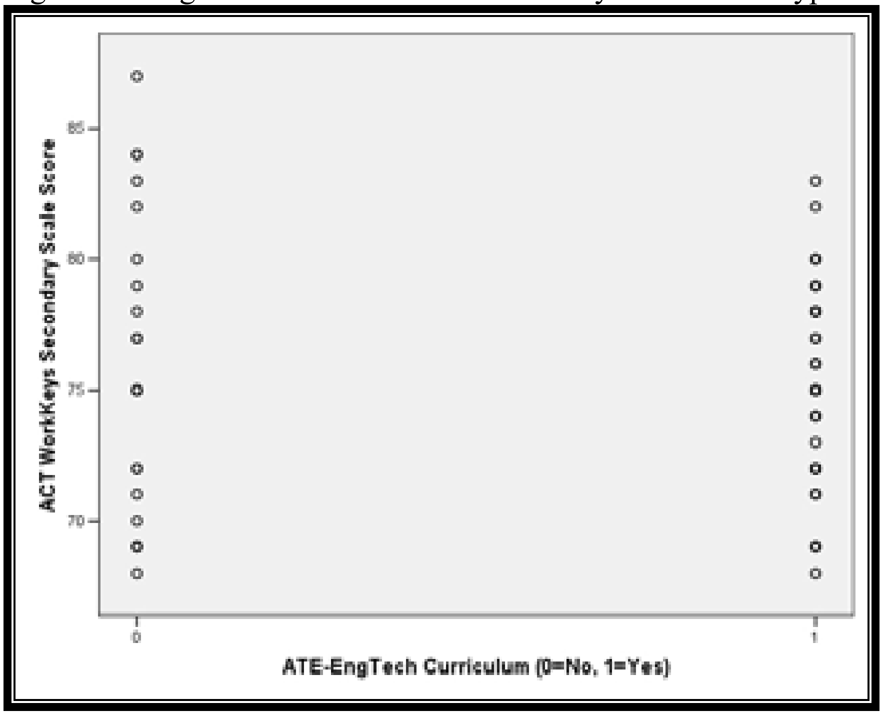


Figure 8: Frequency and Distribution of Non-ATE Student AWATA Scores

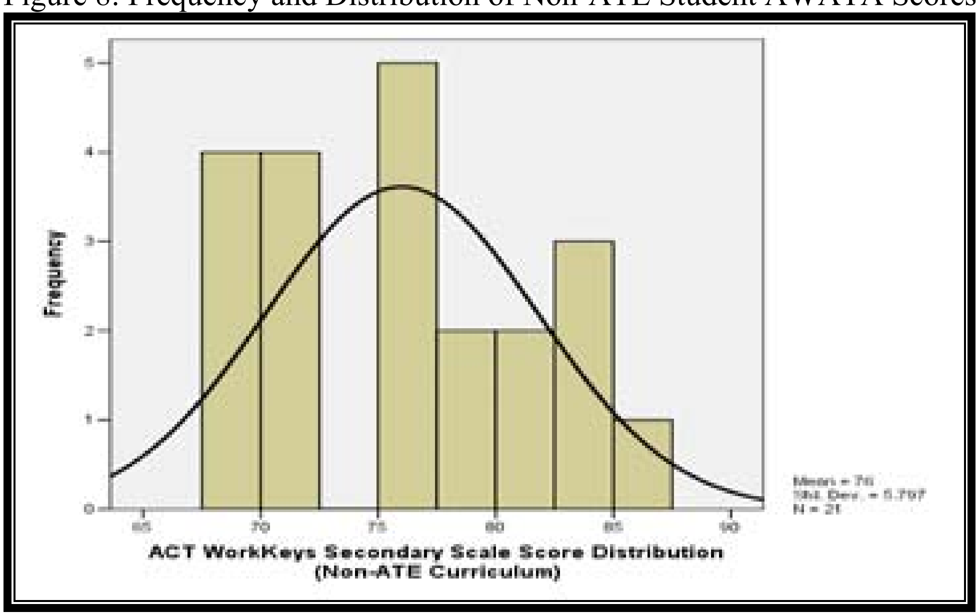
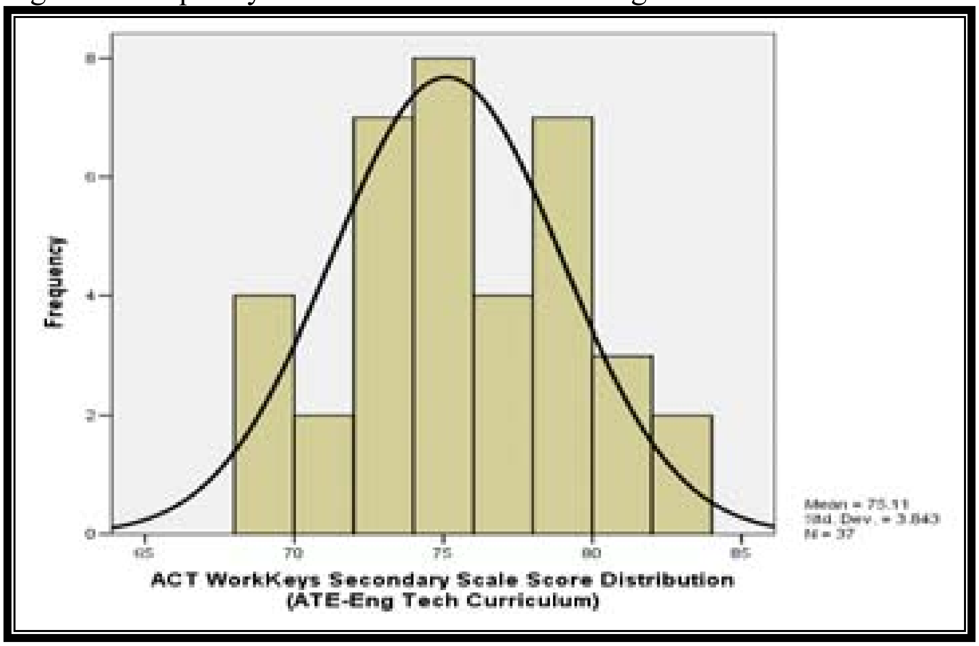
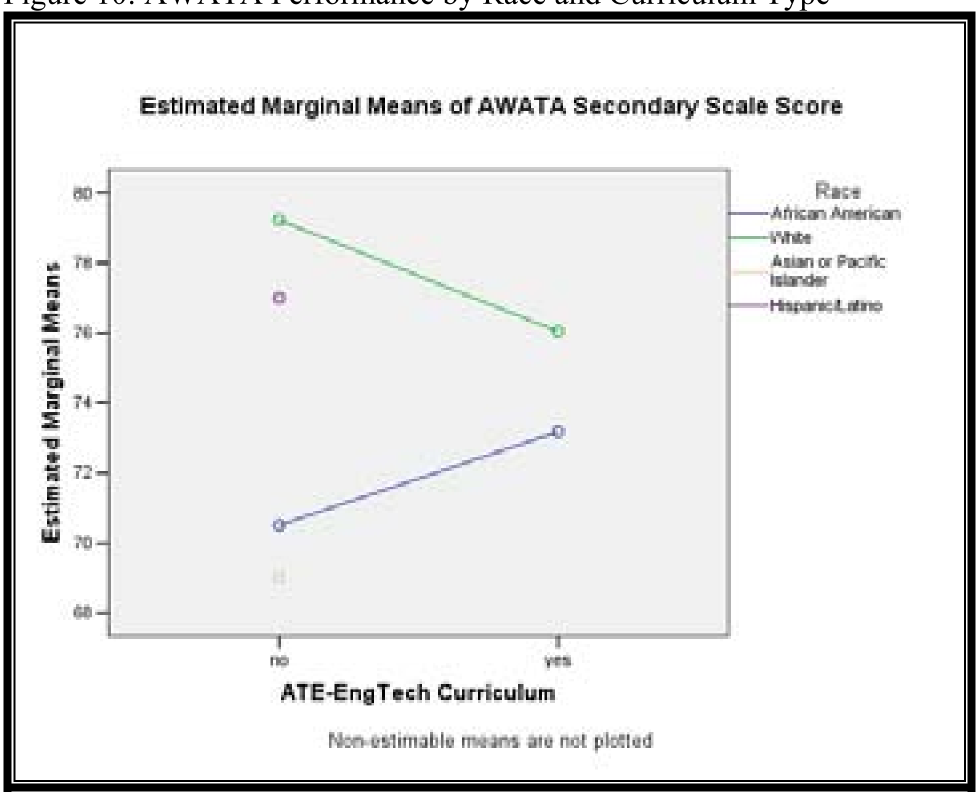


Figure 9: Frequency and Distribution of ATE-EngTech Student AWATA Scores



- *(Note scale differences between Figure 8 and Figure 9)*

Figure 10: AWATA Performance by Race and Curriculum Type



ET PSA Form I Results:

Figure 11: ET PSA Form I Range of Student Scores by Curriculum Type

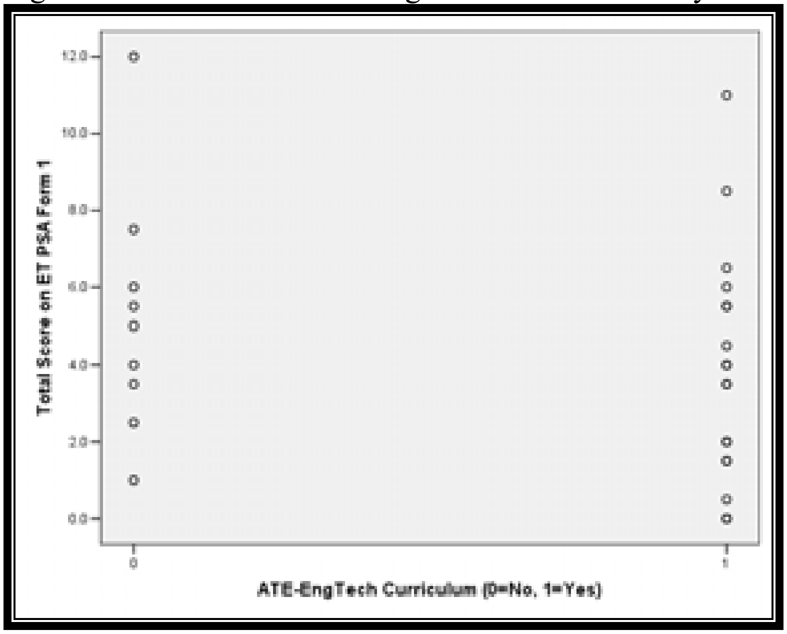


Figure 12: Frequency and Distribution of ET PSA Form I Non-ATE Student Scores

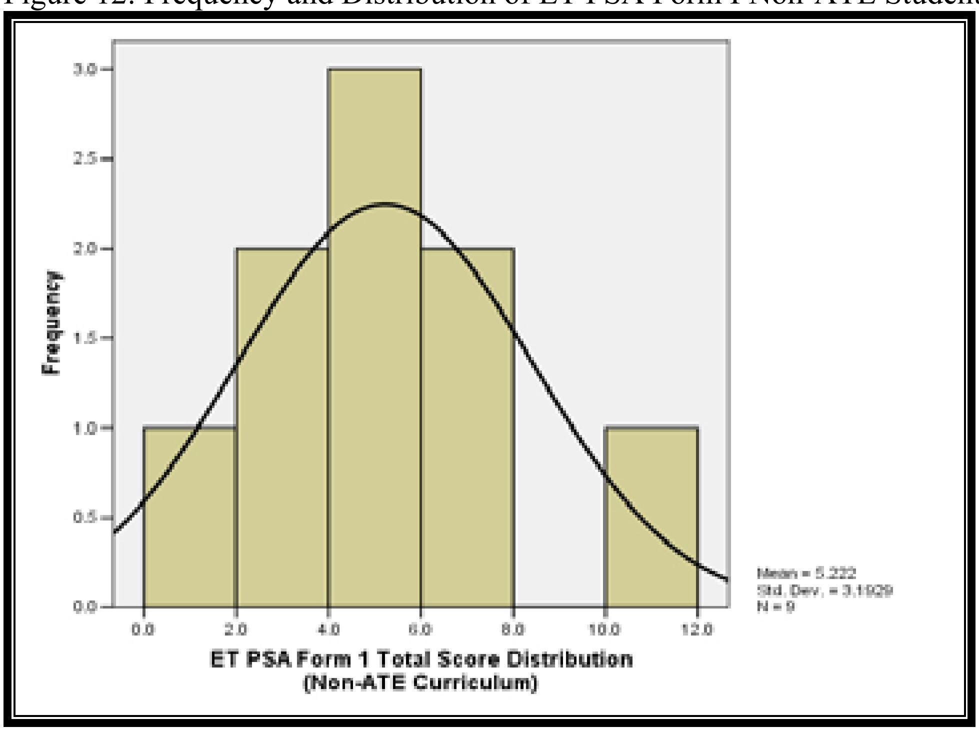


Figure 13: Frequency and Distribution of ET PSA Form I ATE-EngTech Scores

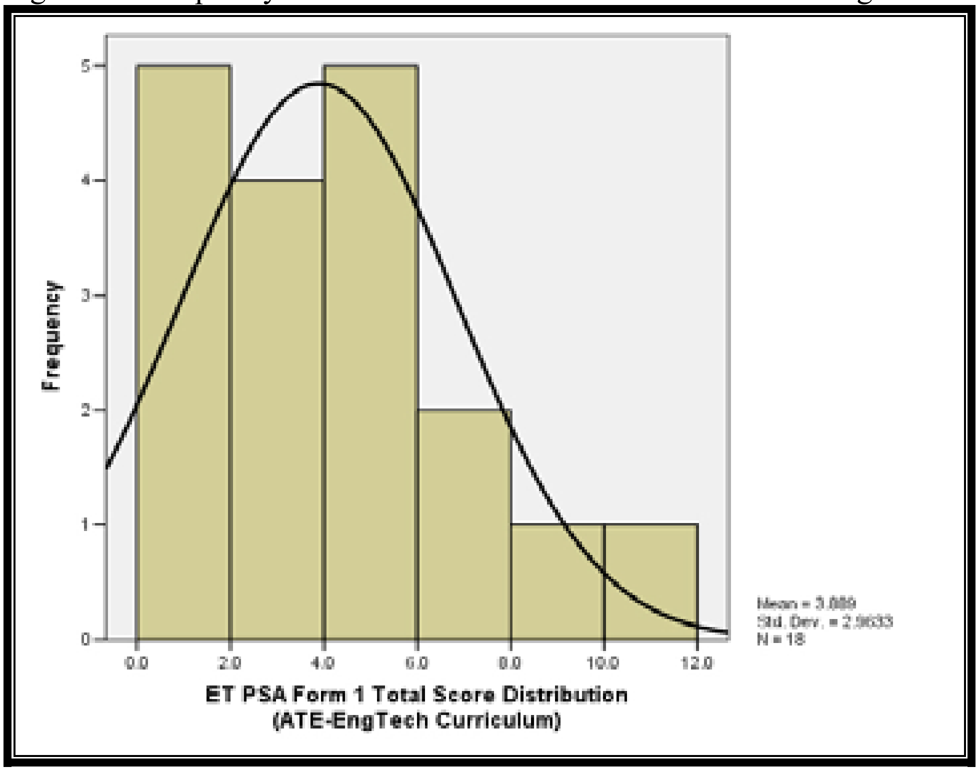
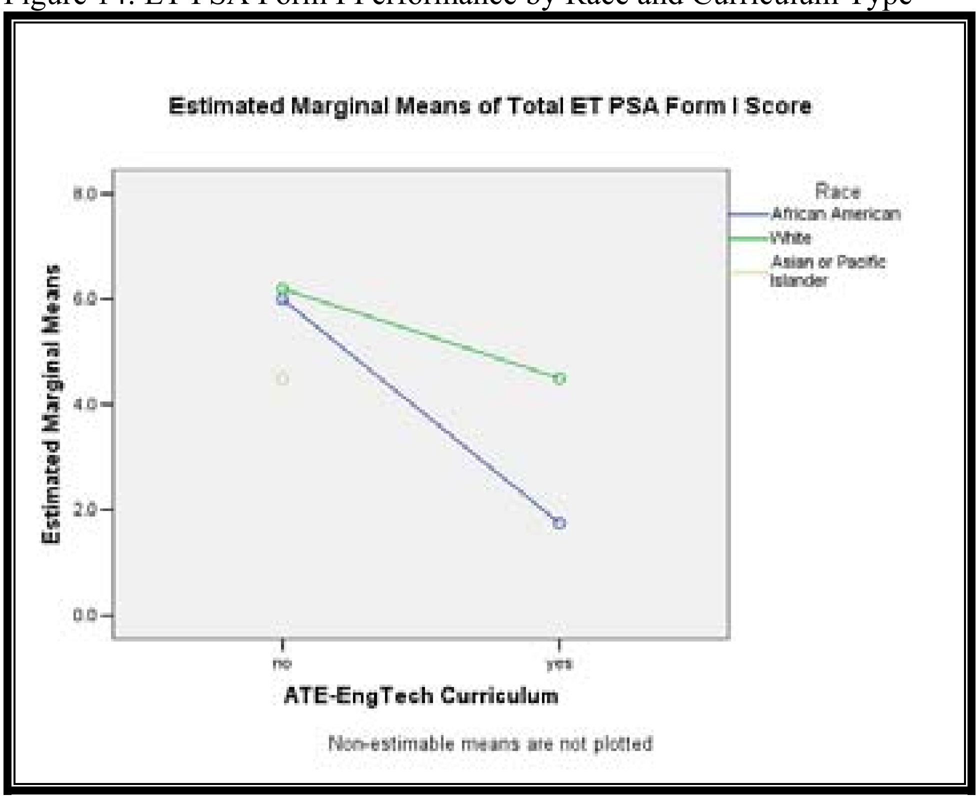


Figure 14: ET PSA Form I Performance by Race and Curriculum Type



ET PSA Form II Results:

Figure 15: ET PSA Form II Range of Student Scores by Curriculum Type

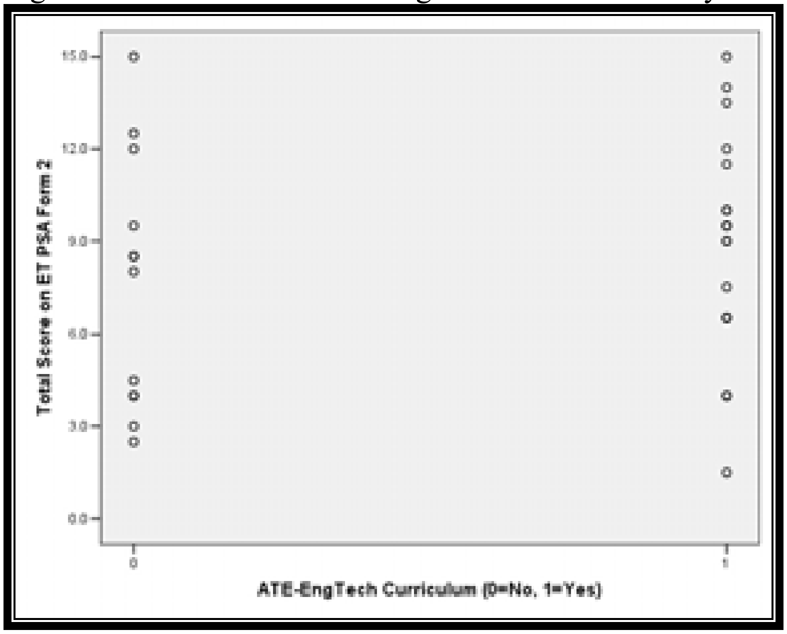


Figure 16: Frequency and Distribution of ET PSA Form II Non-ATE Student Scores

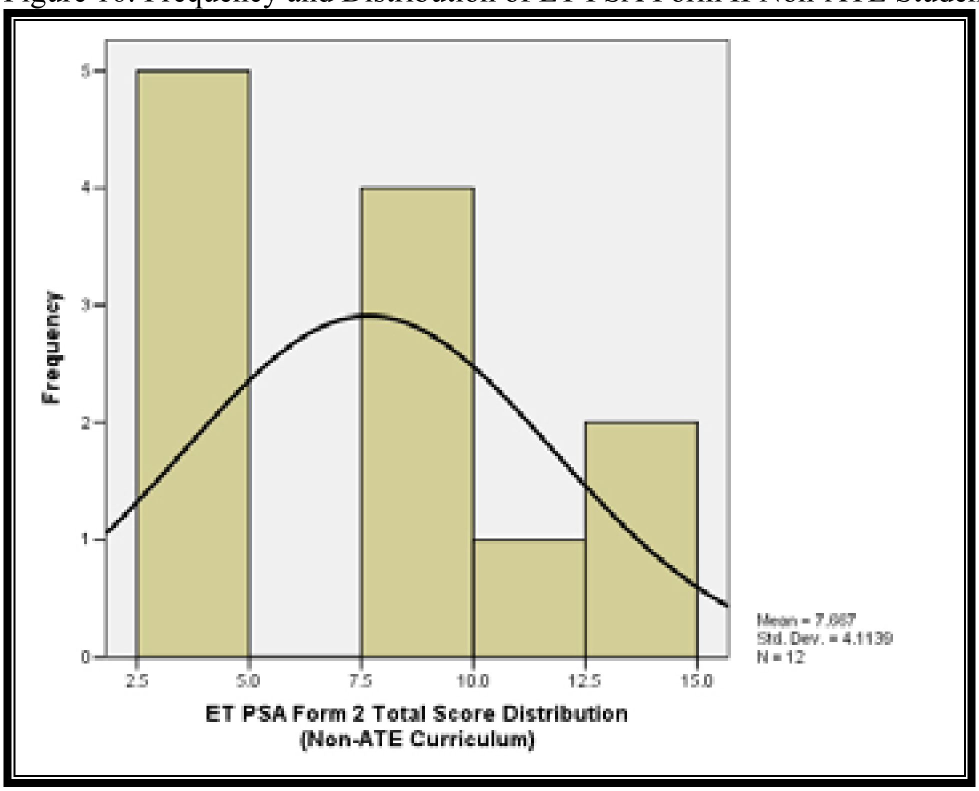


Figure 17: Frequency and Distribution of ET PSA Form II ATE-EngTech Scores

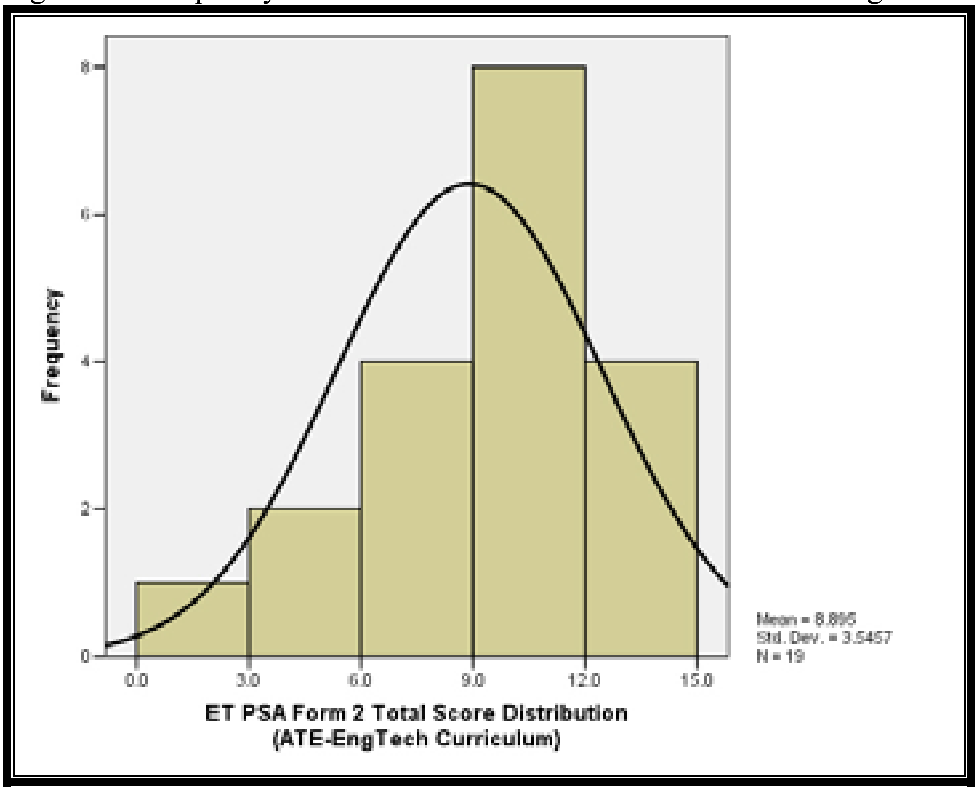
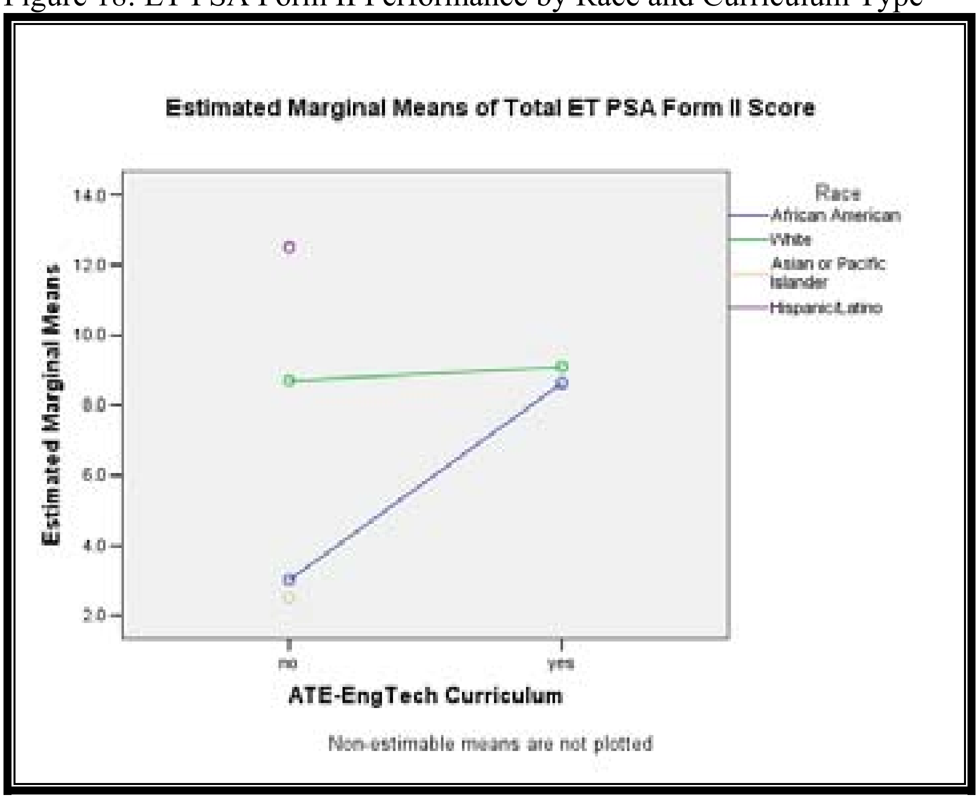


Figure 18: ET PSA Form II Performance by Race and Curriculum Type



N=136 STEPWISE HIERARCHICAL REGRESSION (WITH PROPENSITY SCORE QUNITILES) MODEL FITTING DETAILS:

Figure 19: Q-Q Plot of Unstandardized Residual Values

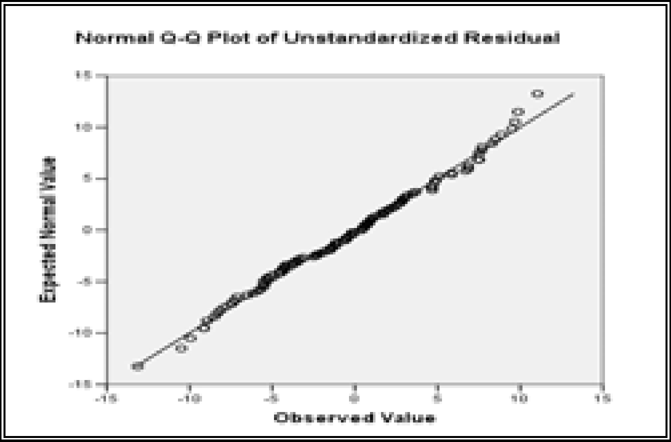


Figure 20: Predicted vs. Residual Scatterplot

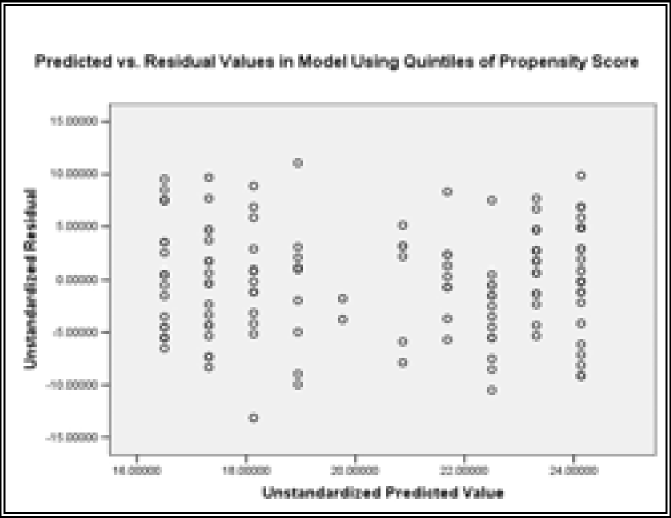
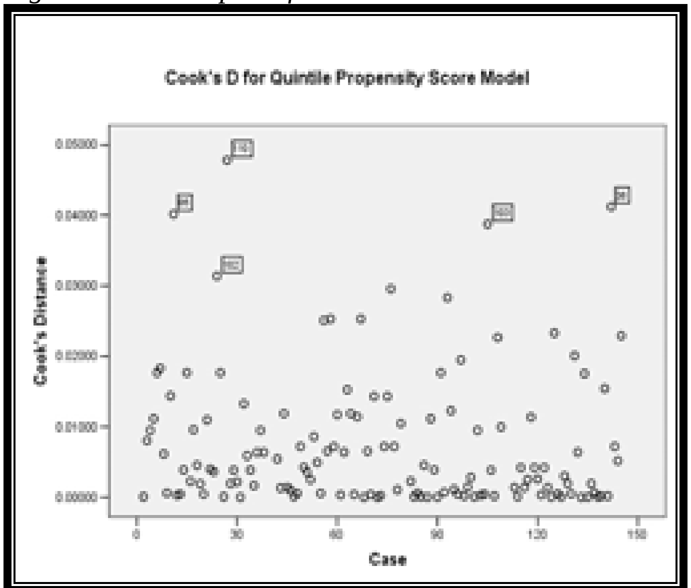


Figure 21: Scatterplot of Cook's D



Model Fit Indices:

- Figure 19 plots the unstandardized residuals of the regression that are observed with the fitted values of the model. Adherence to the plot line provides evidence of normality. Also, given the N of 136 in this analysis, the Central Limit Theorem also supports these data being normally distributed.

- Figure 20 plots the unstandardized predicted values with the unstandardized residual values. A random pattern suggesting no relationship is evidence of homoscedasticity. Since these data were grouped according to quintile and curriculum type the categorization along the x-axis is to be expected. In terms of random distribution along the y-axis, the values seem to cover the area fairly well with little in the way of patterns. This appearance seems to support the assumption of homoscedasticity.

- Figure 21 graphs Cook's D which provides information on potential outliers among the predictor variables and evidence of model-fit. Given the five marked data points, there is some potential for outliers or suspect model fit, yet the majority of the points cluster near the bottom which bodes well for the model. The numbered points should be noted as exerting more influence than the others.

- Leverage values (outliers among predictor variables) were saved and the equations (p+1)/n and 3(p+1)/n (with p = # of groups) were used to calculate the mean and suspect leverage values, respectively.
  - **Mean Lev = .022**
  - **Suspect Lev = > .066**
  - **Largest Lev in analysis = .042**

APPENDIX G: ET PSA ITEM AND SCALE INFORMATION

Table 23: ET PSA Item Discrimination and Scale Reliability Information

Score Received for ET PSA Form I and Form II Responses by Percentile of Total Score

		FORM I						*FORM II*									
		Q1F1	**Q2abF1**	**Q3abF1**	**Q4abF1**	**Q5abF1**	**Q6F1**	**Q1F2**	**Q2 and 3F2**	**Q4aF2**	**Q4bF2**	**Q4cF2**	**Q4dF2**	**Q4eF2**	**Q4fF2**	**Q4gF2**	**Q4hF2**
25th Percentile and Lower	*Mean*	.143	.000	.250	.143	.583	.167	1.563	.750	.333	.500	.083	.500	.000	.000	.000	.083
	Median	.000	.000	.000	.000	.500	.000	1.750	.750	.000	.500	.000	.000	.000	.000	.000	.000
75th Percentile and Higher	*Mean*	.643	1.214	1.500	.857	2.643	1.357	2.000	2.000	.500	.875	.813	1.875	2.375	1.375	.750	.714
	Median	.500	1.000	2.000	1.000	3.000	2.000	2.000	2.000	.500	1.000	1.000	2.000	2.500	2.000	1.000	1.000

Form I: $\alpha = .76$
Percentiles: >=75th: (6.0); =25th: (2.0)

Form II: $\alpha = .77$
Percentiles: >=75th: (11.5); =25th: (4.5)

APPENDIX H: ESA ITEM AND SCALE INFORMATION

Table 24: ESA Item Discrimination and Scale Reliability Information

		All	**>75% (top students)**	**<25% (low students)**	**Males (N=56)**	**Females (N=85)**		**R of SciC & Score**
ST-Level	***Item***	*% answering correctly*	*% answering correctly*	*% answering correctly*	*% answering correctly*	*% answering correctly*	***Item***	Coll_SciCourses =
2-1	**1**	95.2	100.0	91.4	94.6	95.3	**1**	.254** (p<.001)
5-1	**2**	79.3	88.9	65.7	83.9	76.5	**2**	Type of Degree =
3-(1-2)	**3**	38.6	66.7	20.0	44.6	32.9	**3**	.107* (p<.05)
3-1	**4**	77.2	88.9	65.7	82.1	74.1	**4**	
5-1	**5**	91.7	100.0	77.1	89.3	92.9	**5**	**Mean Score/SD**
3-1	**6**	80.7	100.0	60.0	85.7	77.6	**6**	16.5/4.0
7-2	**7**	56.6	86.1	34.3	53.6	57.6	**7**	
6-1	**8**	83.4	97.2	62.9	78.6	85.9	**8**	**Percentiles**
1-2	**9**	75.2	97.2	40.0	80.4	71.8	**9**	75 (19.5)
1-2	**10**	77.9	100.0	54.3	80.4	75.3	**10**	25 (14)
3-1	**11**	71.7	100.0	42.9	76.8	68.2	**11**	
4-2	**12**	35.2	72.2	14.3	44.6	29.4	**12**	**Reliability**
2-2	**13**	53.8	91.7	37.1	53.6	52.9	**13**	α = .77
2-1	**14**	55.9	80.6	28.6	51.8	57.6	**14**	
6-2	**15**	49.7	61.1	37.1	53.6	47.1	**15**	
1-2	**16**	16.6	19.4	17.1	12.5	20.0	**16**	
2-1	**17**	93.8	100.0	77.1	98.2	90.6	**17**	
7-2	**18**	49.0	69.4	28.6	57.1	44.7	**18**	
4-1	**19**	64.1	100.0	25.7	64.3	63.5	**19**	
7-2	**20**	79.3	94.4	45.7	82.1	77.6	**20**	
5-2	**21**	68.3	94.4	34.3	75.0	63.5	**21**	
3-2	**22**	31.0	58.3	20.0	41.1	23.5	**22**	
6-2	**23**	52.4	75.0	40.0	44.6	55.3	**23**	
3-1	**24**	77.9	94.4	54.3	78.6	76.5	**24**	
3-2	**25**	33.8	50.0	17.1	30.4	34.1	**25**	
7-3	**26**	57.9	80.6	22.9	55.4	60.0	**26**	
				Avg. % Correct	65.1	61.7		