Title Page:

Measuring the Sustainability of the National Science Foundation's Advanced Technological Education Program

Wayne W. Welch
University of Minnesota
May 2012

Author Note

Wayne W. Welch, Department of Educational Psychology, University of Minnesota (ret.)

Professor Welch is also a consultant for Rainbow Research, Inc., Minneapolis, MN

This material is based upon work supported by the National Science Foundation under Grant No. 1132099. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this report should be addressed to Wayne Welch, 621 W Lake St # 300, Minneapolis, MN 55408. Contact: wwelch@umn.edu

Note to readers of this report

The proposed research includes 11 questions and their attendant tasks. This report addresses questions 1, 2, 3, 4, part of 5 and a bit of seven as they relate to sustainability. The next report will describe these tasks for the impact statements on the survey.

I have written the report following the American Psychological Association (APA) style that is required by educational journals. Eventually, it will be posted on the web and parts of the research may be submitted for publication. Note that APA recommends writing research reports in the first person when there is just one author. I have used this style in this report.

One of the main objectives of the targeted research program is to provide information to the various stakeholders for program improvement. Accordingly, throughout this report I have added "Comments" that address possible program implications of the research findings.

Acknowledgements

I wish to thank my review panel members for their help in preparing this report; Elaine Craft, Director of the South Carolina ATE Center of Excellence, Arlen Gullickson, former director of the ATE Evaluation Program at Western Michigan University, and Elizabeth Teles of Teles Consulting and former co-lead for NSF's ATE program. They were most helpful in reviewing a draft of the report and providing insightful ideas for the "comments" that appear throughout the report. Thanks to my wife, Jerry, for reviewing earlier versions of this report.

The Advanced Technological Education (ATE) program, funded by the National Science Foundation (NSF), is designed to improve the education of technicians in high-technology fields such as biotechnology, advanced manufacturing, information technology, and environmental and energy technologies. The program makes grants to support projects and centers to achieve this goal. It began in 1994 and more than 1,200 awards have been made to two- and four-year colleges and other organizations. The primary focus of the program is on two-year colleges and approximately 74% of the awards have been awarded to these institutions.

The Foundation expects grantees to plan to sustain successful outcomes of their work after NSF funding stops. There are many definitions of sustainability but the one used here is the capability to endure, that is, to last, or continue to exist. Synonyms include endurability, continuation, persistence, and perhaps, survival. The term is not used in the sense of using a resource so that it is not depleted nor permanently damaged as in the currently popular view of the word. In the context of ATE, sustainability means that programs and activities started under Foundation support, and are determined to have value, continue in some form after the grant ends. The purpose of this research was to describe the dimensions of sustainability and develop procedures to assess the persistence or continuation of ATE grants.

> Comment: A brief search of the literature reveals an overwhelming use of the word sustainability to mean using resources so they are not depleted nor permanently damaged. In fact, NSF has a new program seeking proposals to study the topic. It is called "Science, Engineering, and Education for Sustainability NSF-Wide Investment (SEES)". The program solicitation defines sustainability as seeking to meet human needs without harm to the environment, and without sacrificing the ability of future generations to meet their needs.
>
> To avoid confusion and given the growing popularity of the alternative meaning, ATE may want to consider using a different term to describe the continuation of work started under an ATE grant. Perhaps some form of the synonyms mentioned above; endure, persist, or continue. Continuation or "continue the work" seems to work well. On the other hand, maybe the phrase, residual impact, used by some program officers in past years captures the meaning.[1]

Grant sustainability can be viewed in different ways. One approach is to think in terms of the sustainability of the total project or center. For example, the 2002 ATE Program Solicitation stated that proposals for centers should describe a workable plan for sustaining the center after NSF funding ceases (National Science Foundation, 2002). A grant might evolve into a permanent unit of an institution. Projects and centers were expected to continue their work using other sources of funding. This was the view of sustainability during the early years of the program; however, during my initial research on the topic, I suggested a different approach (Welch, 2011). It is based on the assumption that some, but not all, elements of a grant will persist. The task for grantees is to determine those things that should persist and to identify, generate or allocate resources to ensure that this will happen.

[1] The term was used when I worked with NSF during the 70s and 80s.

The most recent program solicitation, 2011, is consistent with this approach when it described a need for sustainability as follows.

> A project or center is expected to communicate a realistic vision for sustainability and a plan to achieve it. It is expected that at least some aspects of both centers and projects will be sustained or institutionalized past the period of award funding. Being sustainable means that a project or center has developed a product or service that the host institution, its partners, and its target audiences want continued. To be sustainable is to ensure a center or project's products and services have a life beyond ATE funding. (National Science Foundation, 2011, Introduction, para. 5).

There are many indicators of sustainability. An activity might become part of the regular academic curriculum, that is, it becomes *institutionalized*. Perhaps the grant purchases laboratory equipment and it continues to be used after the grant ends. Another example is a college continuing to use the new student recruitment procedures it developed for its grant.

Another possible indicator of sustainability is the *continued presence* of products created during the life of a grant. The grant ends but there are things that remain and continue to be used and shared with others (disseminated). For example, laboratory equipment, educational materials (modules, books, computer programs), or students who join the workforce are examples of this kind of sustainability.

> Comment: When the grant ends or the project stops, well-prepared students are a legacy, surely an important example of sustainability. Project staff may want to consider ways to improve what students learn and place them in positions where they can use their skill in the workforce. Successful students are also likely to stay connected with and "give back" to the college and program, helping with curriculum improvement and student recruitment. A large-scale student follow-up study would provide evidence of their sustainability. Currently few projects and centers systematically follow students after they graduate or leave the programs.

A third indicator is the *prolongation* of things started during a grant. Examples include the continuation of collaborations between a community college and local industries. On the other hand, perhaps an internship program started under an NSF grant persists after the grant ends.

A fourth indicator might be the continued use of *new products or activities* created during the grant. For example, a college works with an industrial partner to develop a new welding program to help them prepare technicians to manufacture rail cars. The program continues after NSF funding stops. I believe this is another indicator of sustainability.

These examples are not meant to be exhaustive but, rather, to describe an elements-based approach to sustainability. It helps projects to plan for sustainability and broadens the places one may look for examples.

> Comment: Under the elements-based approach, sustainment is not a yes/no proposition; there are degrees of sustainability. The challenge for program staff is to determine what is

worthy of sustainment and plan for it to occur. The challenge for the researcher is to find evidence of what was continued.

This report describes procedures used to measure the extent of sustainability of ATE projects and centers. It explains what I did, describes the psychometric properties of the measuring instrument, and presents several research findings.

Methods

Current principal investigators (PIs) and others familiar with the program were asked to describe examples of sustainability for their ATE project or center. I used these statements as items on a survey and mailed them to other ATE PIs. I asked them if the statements were consistent with their sustainability experiences and asked them to respond using a five-point Likert scale that ranged from "strongly agree" to "strongly disagree". There was also an option to circle "not applicable". A few examples of the statements follow.

- "Changes made in our technological education program will keep going after our current grant ends".
- "We have at least one industry partner who is committed to support some of our work after NSF funding has ended."
- "The instrumentation and equipment we secured as part of our ATE grant will have little use by the college after the grant ends".
- "The teaching methods adapted by faculty as part of our ATE project will continue to be used after the grant ends."

The final survey consisted of 23 Likert items. I sent them to 261 current and past ATE principal investigators (PIs).[2] Follow-up contacts were made and eventually 212 completed surveys were returned. The response rate was 81%.

Statisticians differ on how to handle Likert item responses. Most agree that when reporting the results of a single item, it is best to use non-parametric statistics. However, when responses are summed across several items, and meet several other criteria, a summated total scale is appropriate (Uebersax, 2006); (Norman, Published online: February 10, 2010). Norman is particularly forceful in his claims, "Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of ''coming to the wrong conclusion.'' These findings are consistent with empirical literature dating back nearly 80 years." (p. 632).

Analysis of Item Responses

The initial analysis computed and reported the responses to each of the 23 items. ATE grantees, in general, reported strong evidence of sustainability (Welch W. W., 2011a). The median affirmation response was 78%. Affirmation means that the subjects agreed or strongly agreed

[2] A description of the data gathering process is available in (Welch W. W., 2011b)

with positively stated items and disagreed with negatively stated ones. Some examples include the following.

"Changes made in our technological education program will keep going after our current grant ends. (90.9% agreed or strongly agreed with the statement)

We have at least one industry partner who is committed to support some of our work after NSF funding has ended. (66.5% agreement)

"The professional development program(s) we developed is/are used at other sites." (63.3% agreement)

The teaching methods adapted by faculty as part of our ATE project will continue to be used after the grant ends. (92.0% agreement)

Comment: The only exceptions to the affirmation findings were three statements about obtaining funds to continue the work of the grant. Only about a third of the respondents to these items reported success in finding new sources of revenue when their ATE grant ended.

Scale Analysis

Single item responses can provide information about a topic but information on validity and reliability is difficult to attain. In order to achieve a more reliable estimate of the construct in question, classical test theory calls for summing the responses to obtain a scale score. This process is appropriate if the items are measuring the same underlying concept, the scores are normally distributed, and they discriminate among the respondents.

Scale score calculation: Normally, one would sum or average the item responses to create a sustainability scale score. However, the survey had an option to mark Not Applicable (NA) to accommodate those grantees who may not have a particular element in their grant. For example, if a statement was about materials development and the scope of work of the grant did not include that activity, it could be marked NA.

Although the use of the NA response provides information on the focus of the grants, it does complicate the calculation of scale scores. How does one code a response of not applicable when computing a scale score? In order to learn more about the frequency of use of NA, I computed the number of such responses and reported them separately for projects and centers. I expected projects to use the option more often because they are smaller in scope than the centers.

The results are shown in Table 1. The statements are grouped by the elements of the domain of content used to develop the survey. For more information on the process, see Welch W. W. (2011a).

Table 1
Number of respondents marking Not Applicable by projects and centers

Statements about sustainability: Organized by elements of the original domain	Number of NA responses	
	Projects (n=170)	Centers (n=42)
Programs *Changes made in our technological education program will keep going after our current grant ends.*	16	0
Most of the programs and activities started during our ATE grant will come to an end when our NSF funding stops.	1	0
The new curriculum was created through the regular institutional approval process so all the new classes are in the regular college catalogue and are part of approved college degrees.	37	6
Collaborations *We have at least one industry partner who is committed to support some of our work after NSF funding has ended.*	20	1
It is doubtful that the relationships we have established with our various partners will continue after our ATE grant has ended.	1	0
Through this grant we have created liaisons with our industry partners and academia that will end when our grant does.	5	0
Educational Materials *The materials we have developed are seldom used by other colleges for technician preparation programs.*	30	0
Revenue earned from the sale of educational materials is used to provide scholarships for students.	121	21
Faculty *Our faculty has become a part of a collegial network that will continue to share program information, work force trends, and cutting-edge instructional technologies.*	13	0
The professional development program(s) we developed is/are used at other sites.	36	2

The teaching methods adapted by faculty as part of our ATE project will continue to be used after the grant ends.	14	0
Facilities *The instrumentation and equipment we secured as part of our ATE grant will have little use by the college after the grant ends.*	51	5
Alumni of our project are ambassadors to the larger technical community and tell potential students about the value of technology education.	35	2
Students *Very few of the graduates prepared under our grant are employed as technicians.*	43	1
The recruitment efforts developed through the grant will be incorporated into W K H F R O O H J H ¶ V J H Q H U D O U H F U X L W P H Q W [illegible]	31	6
Internships, supported by industry, will continue as a way to provide our students with exposure to the real world.	47	6
Institution *Our use of a national review committee has helped make our advisory committees more effective.*	80	3
Our ATE grant experience has caused our administration to encourage other faculty to seek external funding to address workforce needs.	12	1
It is unrealistic to expect that ATE grants will have a long term impact on community colleges.	5	0
The grant has enhanced our reputation as a regional leader in advanced technology education.	7	0
Revenue *Our Center/Project has formed a not-for-profit corporation to help us continue our work beyond NSF funding.*	79	14
We will be able to keep our project/center going by obtaining revenue income for specific education services.	41	2
We would not be able to continue our project/center without continued funding from NSF.	6	0

The use of "not applicable" by the 212 respondents varied from one to 142 for both groups. The average number per survey was 3.8. However, the distribution of the NA responses is not random. Projects use the response more than centers do. Projects reported 4.3 NAs per survey while centers replied NA an average of 1.7 per survey. Centers are larger on average and have a broader scope of activities.

The area most often marked NA were the items that dealt with obtaining revenue. This may be an area for further investigation because of its importance for sustaining successful grant products and services.

The large number of NA responses for some of the items creates problems when analyzing the data. Surprisingly to me, I found no references during a Google search that addressed this issue. It may be that NA is not often used or that most studies do not have many NA responses.[3]

One might think that such responses should be coded as missing data but this is not the case. Responses were provided; they were not missing. The issue is how to code them.

One method is to exclude those cases that replied to any statement as Not Applicable. This approach works fine for computing item means because SPSS and other analysis programs only include the number of subjects who answer each item. However, there are problems in computing reliability and doing factor analysis because of the serious reduction in the sample size. For example, my sample size dropped from 212 to 38 when I conducted a reliability analysis of the Total Scale score. This is because SPSS uses a listwise exclusion procedure when computing internal consistency indices such as Cronbach's alpha.

Another way to address this issue is to code NAs as zeros. This yields the following values for the Likert responses:

Strongly affirm[4]	5
Affirm	4
Uncertain	3
Deny (do not affirm)	2
Strongly deny	1
NA	0

There are two advantages to this approach. First, all cases are included. Second, when summing the scores, the NA responses are not included in the total. There is one disadvantage and that is computing item means. Including zeros as actual value spuriously reduces the mean values.

One can see this difference for the following item. "The professional development program we developed is used at other sites." Thirty-eight of the grantees marked this NA. The mean item response when those cases with NA are excluded was 3.66 with a standard deviation of

[3] This problem needs further investigation.
[4] Affirmation means the subjects agreed or strongly agreed with positively stated items and disagreed with negatively stated ones

1.01. This result was based on 173 replies. When one assigns a zero for the NA replies, the mean was lowered to 3.00 and the SD was 1.68. The sample size was all 212 cases.

I decided to use both approaches depending on the analysis I was conducting. I coded the NAs as zeros when computing Total Scale scores and for conducting a factor analysis of the responses. When I needed to compute item means, I excluded NA responses from the analysis. Although excluding cases is technically the same as treating them as missing cases, the mindset is different. Missing implies there was no response. I think this psychologically different from choosing to exclude those cases when NA was the response.

Total scorescale I summed the responses across all items to compute a Total Scale score. The response options ranged from zero to five where zero is NA and five indicates maximum affirmation to the statement.[5] This yields a score that is a measure of the scope of sustainability. The maximum score one could receive was 115.

I computed a reliability coefficient for these responses using Cronbach's alpha. This is a measure of the internal consistency of the scale, that is, to what extent are the items measuring the same thing. The alpha value was .89 just below excellent level of .90 recommended by (Mallery & Mallery, 2003).

Parametric tests such as t-tests and correlational analysis are based on the assumption that the underlying population is normally distributed. To test for this, I plotted a histogram of the Total Scale Scores and overlaid a normal curve. The results are shown in Figure 1.

[5] Items that were negatively stated were reverse coded so that strongly disagree was assigned a value of 5, disagree, 4 and so on.

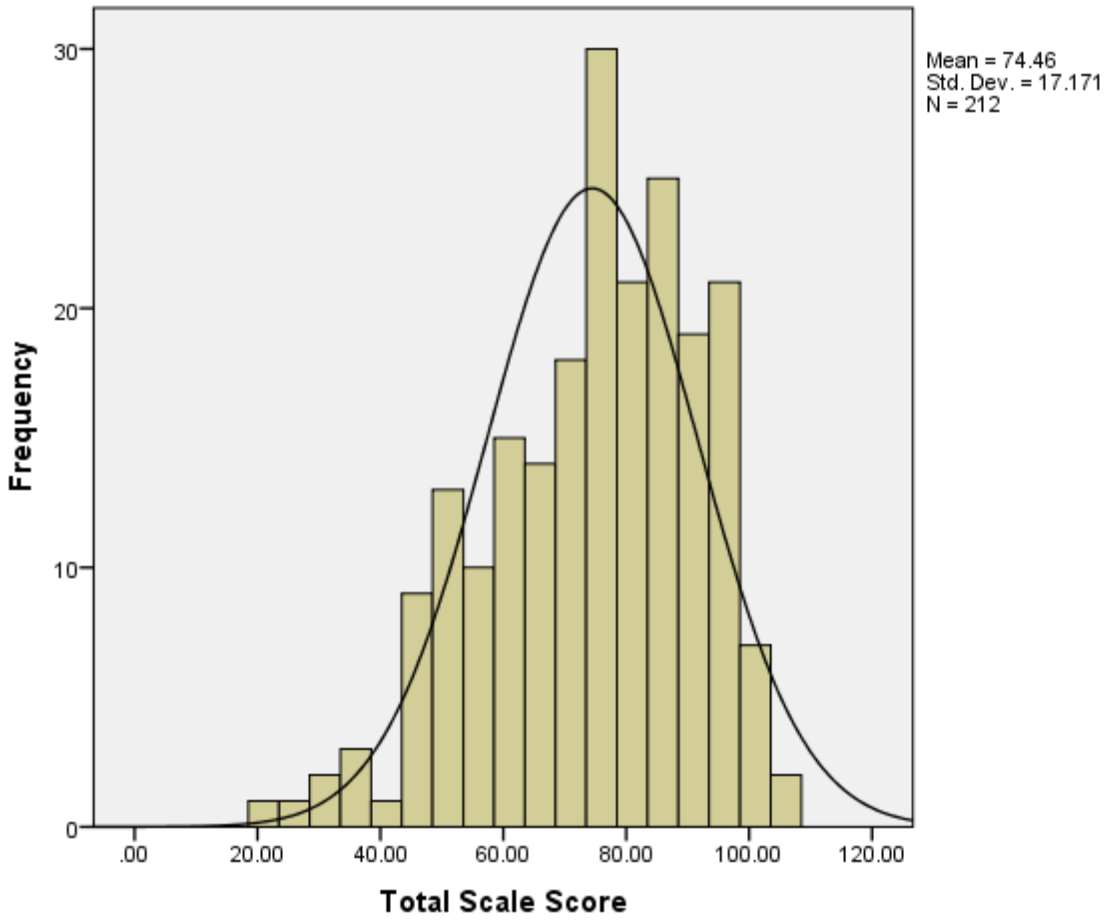

Figure 1: Histogram of total scale scores on the sustainability survey

The scores ranged from 21 to 107 with a mean of 74.46 (SD = 17.17). The median was 77.

I computed the skewness and kurtosis of the scores. Skewness is a measure of how symmetrical the data are; a skewed variable is one whose mean is not in the middle of the distribution (that is, the mean and median are different). "Kurtosis" has to do with how peaked the distribution is, either too peaked or too flat. The skewness was -.60 (SE = .17) and almost no kurtosis -.06 (SE = .33). These values are within generally accepted levels of plus or minus two (http://dss.princeton.edu/online_help/analysis/regression_intro.htm). The population estimate based on the distribution of the sample's scores suggests there should be no problems in using parametric tests such as t-tests.

Mean item scale score I computed the average response to the 23 items for each subject. The possible range of means could be from 1.0 to 5.0. Means were computed only for the items that each person answered. Those items left blank (this seldom occurred) or those marked NA were excluded from the averaging process. The number of valid cases varies from item to item depending on the frequency of excluded cases. This approach yields a measure of

the relative success rather than a total sustainability score. The Item Mean score is an indicator of success for those activities and products promised by the grantee, it is not an indicator of total sustainability.

Calculating the reliability of this approach is a bit problematic because there are a different number of cases for each item mean. However, one can use an approach based on the correlations between each pair of items. One computes the average inter-correlation of the items and then applies the Spearman-Brown prophecy formula to calculate the reliability for a 23-item scale. The procedure is similar to the single measures intraclass correlation coefficient available in the SPSS program (IBM SPSS Statistics 20, 2011). However, they use it for comparing the reliability of one rater (judge) to that of multiple raters. The single measures correlation is the reliability you would get with just one item. The "average (or sum) measures" is the reliability you would expect with 23 items.

The sample sizes of the various pairwise values ranged from 70 to 211. The average item inter-correlation was .242. Applying the Spearman-Brown prophecy formula[6] yielded a value of 0.88, which is considered very good.

I plotted a histogram of the Mean Item Scale and overlaid a normal curve to examine the distribution of these scores. This is shown in Figure 2.

[6] $r_{nn} = nr_{ii} / (1 + (n - 1) r_{ii})$ where n equals the number of items and i refers to individual items. r_{ii} is the correlation between each pair of items.

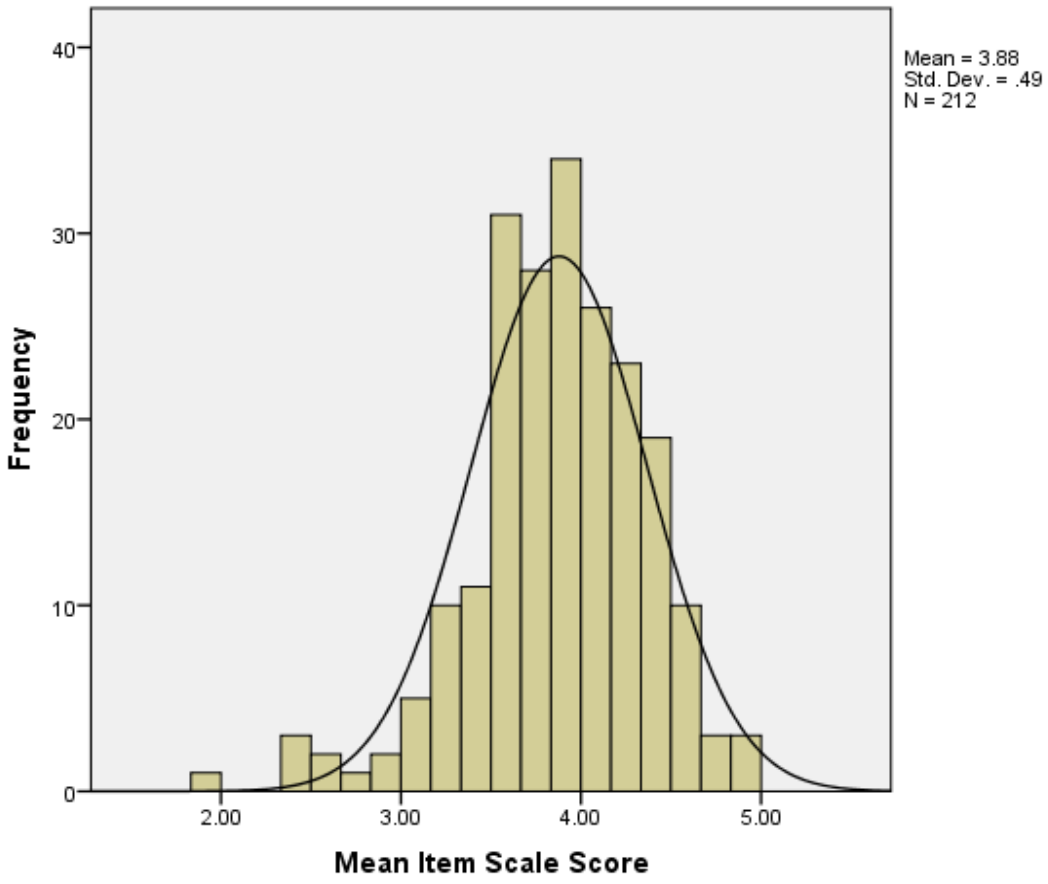

Figure 2: Histogram of mean item scale score on a sustainability survey

The range of scores was 1.96 to 5.00. The mean was 3.88 (SD = .50). The skewness was -.66 (SE = 1.17) and the kurtosis was .19 (SE = .33). Neither of these values is significant and these values are within plus or minus two limits. There should be no problem in using parametric statistics for additional analysis.

When I computed the Total Scale score, I assigned zeros to NA responses. When I computed the Item Mean scores, cases with NA responses were excluded from the analysis. Even though they are calculated in different ways, one would expect to find a relationship between these two measures of sustainability. I tested this assumption by computing the correlation between them. The Pearson r between the total scale scores and mean item scores was .63 which was significant at the p = .00 level. Effect sizes can be measured using correlation coefficients or Cohen's d for the difference between means. An effect size of .63 is in the medium range (Cohen, 1988) as reported in (Howell, 2011). Cohen uses the following rule of thumb to interpret effect sizes: d = .20 (Small), d = .50 (Medium), d = .80 (Large).

In the sections that follow, I compare the sustainability scores of various sub-groups to determine if differences exist and to investigate the causes of such differences.

Comparisons across sub-groups

Projects versus centers: I compared the responses of the 166 projects and the 45 centers on the Total Sustainability Scale. The centers (M = 84.83, SD = 11.67) reported significantly higher levels of sustainability than did the projects (M = 71.89, SD = 11.67), t (91) = 5.78, p = .00, two-tailed, d=.79. Levene's test for equality of variance showed the variances for the two groups were not equal so an adjustment was made that reduced the degrees of freedom from 210 to 91.[7]

The statistic "d" above is the effect size (Cohen, 1988). It is the mean difference divided by the pooled standard deviation. This is a standardized estimate of the size of the differences between groups. Cohen considers an effect size of .78 to be very large. Because ATE centers are funded to implement a larger number of activities and the scale measures the scope of activities and products that endure, this finding seems consistent with expectations.

There is a possible confounding factor in the above analysis and that is size of the grants. Centers generally received larger funding levels. The average center grant was about $1.76 million while projects averaged about $575,000. Grant size and the grant type (project or center) are related (r = .64) which indicates a strong relationship. I did a regression analysis predicting Total Test score using grant type and grant size as predictor variables.

Grant type was regressed on Total Score. The overall regression (correlation) was significant, (F (1, 210) = 18.77, p = .00) and R = .29. When I added size of the grant to the regression, the R stayed the same. The regression was still significant (F (2, 209) = 9.79, p = .00) but the size of the grant explained very little additional variance. This suggests that the project/center distinction was a better predictor of sustainability score than was the size of the grant.

I reversed the order of entry in a multiple regression analysis and included size of grant first. The R was .23 and statistically significant, F (2, 210) = 11.64, p = .001. When I added in the project/center difference, R increased to .29 [F (2, 209) = 9.79, p = .000]. However, the beta weights were .24 (t = 2.75, p = .006) for the project/center variable and size of grant dropped to .08 (t = .91, p =. 37). The weight for size of grant was not significant.

> **Comment:** This finding was somewhat surprising. It will require additional analysis. What is it about being a center that would yield higher sustainability scores? It may be the broad scope of activities that centers are expected to implement. Furthermore, some large project grants had a single focus. For example, the largest project grant was to produce a series of videos. The products were sustained but the overall sustainability score was low because there were several NA responses.

I have a measure of the age of a grant but it is just the difference in months between the start date and the expiration date. It does not reflect the fact that most centers receive more than one grant.

---

[7] As mentioned in the author notes, I am reporting the results in APA format because the reports will be published on web sites and possibly in journals. M is the symbol it recommends for the mean and SD is the standard deviation.

Some have been in existence for a dozen years or more and have received multiple grants. The correlation between my age measure and the Total Scale score was not significant; r (212) = .01, p = .81).

I also compared projects and centers on the Mean Item Scale. Projects (M = 3.86, SD = .49) were not significantly different from centers (M = 3.94, SD = .49), t (210) = 0.89, p = .37, two-tailed. This suggests that while projects did not implement as many things as centers, the things they did do were sustained at about the same level.

> Comment: Recall that the Item Mean score excludes NA responses. The score is computed just for those who responded 1 through 5 on the Likert scale. Thus, a grantee who promised and delivered two activities could obtain the same score as a site that did seven activities.

Active versus expired The sample of grants included both active grants (n = 131, 62%) and those that had expired during the three years prior to the study (n = 81, 38%). Previous research (Gullickson & Welch, 2006) had shown that after completion of ATE funding, productivity declines were reported in several areas including the number of collaborations, amount of funding, professional development, and materials development. Steady state findings occurred for course and curriculum use and articulation agreements.

In this analysis, I compared the responses of active and expired grants on the Total Scale score. No significant differences were found between active (M = 75.27, SD = 17.03) grants and expired grants (M = 73.14, SD = 17.42) t (210) = .88, p = .38, two-tailed.

The Gullickson and Welch (2006) approach was to ask PIs the number of collaborations they had during the year after the grant ended. The sustainability survey asks them to judge statements made by their peers. It is a measure of their perception rather than an estimate of the number of collaborations.

> Comment: Contrary to the earlier finding, this study found no differences on the sustainability scales between active and expired grants. It might be useful for future research to compare the findings for projects that had ended one, two, and three years prior.

Grantee institution: NSF made ATE grants to different institutions. About 74% (n = 156) went to two-year colleges, 18% (n = 39) to four-year institutions and 8% (n = 17) to other, for example, organizations or secondary schools.

I compared the Total Scale scores between two-year and four-year colleges. I found that two-year colleges had significantly higher scores (M = 77.2, SD = 15.5) than did four-year colleges (M = 70.8, SD = 18.3) t (193) = 2.23, p = .03 two-tailed, d =.41). The effect size of the differences was .41, just below the standard for a large effect, d = 50 (Cohen, 1988). Given the focus of the program on two-year colleges, this would appear to be a positive outcome. Greater change was reported in those institutions than in the four-year colleges.

> Comment: One might speculate that receiving grants from NSF is a new thing for community colleges and they would be more motivated to succeed in the goals of their grant

than the four-year colleges who have received NSF funds for many years. Maybe two-year colleges are often doing things for their own students, while four-year institutions are doing it for other institutions. Perhaps some PIs at four-year institutions often receive funding and may be less passionate about the subject. These questions and other possible reasons for the differences will be explored in the series of site visits to be held during the second year of this research grant.

Although I did not include the "other" grantees in the statistical comparison, the mean value for the 17 cases in this group was 57.6 with a standard deviation of 17.2. This is considerably lower than the scores for colleges. I examined the grants in the "other" category and they span a wide range of activities. These include secondary schools, educational development organizations, large business organizations, media production agencies, non-profit educational technology laboratories and others. They are generally quite large and funded to do specific tasks, e.g. develop media presentations. When their ATE work was completed, they moved on to other projects. Given that nature of the organizations, the fact that there was some sustainability is interesting. Further research is needed to determine just what kind of things were sustained and why.

Factor Analysis