December 2004

Curriculum Grant Success Analysis

Alexander Lee White
Worcester Polytechnic Institute

Kelly D. Messier
Worcester Polytechnic Institute

Robert John Caulkins
Worcester Polytechnic Institute

12/17/2004

Dr. Russell L. Pimmel
Division of Undergraduate Education
National Science Foundation
Arlington, Virginia 22230

Dear Dr. Pimmel:

Enclosed is our report entitled Curriculum Grant Success Analysis. It was written at the National Science Foundation during the period October 23 through December 17, 2004. Preliminary work was completed in Worcester, Massachusetts, prior to our arrival in Washington, D.C. Copies of this report are simultaneously being submitted to Professor Dibiasio and Professors Demetry and Petruccelli for evaluation. Upon faculty review, the original copy of this report will be cataloged in the Gordon Library at Worcester Polytechnic Institute. We appreciate the time that you, Dr. Lee Zia, and Dr. Roger Seals have devoted to us.

Sincerely,

Robert J. Caulkins
Kelly A. Driscoll
Alexander L. White

CURRICULUM GRANT SUCCESS ANALYSIS

An Interactive Qualifying Project Report Submitted To:
Professor James Demetry
Professor Joseph Petruccelli
WORCESTER POLYTECHNIC INSTITUTE
Washington, D.C. Project Center

By:

Robert John Caulkins | Kelly A. Driscoll | Alexander L. White

In Cooperation With:
Russell L. Pimmel
National Science Foundation, Arlington, VA

December 17, 2004

Approved by:

Professor James Demetry

Professor Joseph Petruccelli

Abstract

This report, prepared for the National Science Foundation, describes the process we used to develop a tool that predicts a project's outcomes from its proposal. We defined proposal characteristics and outcome extents for 36 completed educational software development awards. We used multiple regression analysis to specify and fit a predictive model, and cross-validation to assess the predictive ability. Finally we developed a software tool that allows program officers to apply this model to help evaluate new proposals.

Authorship Page

Robert J. Caulkins	Conducted 16 PO interviews Conducted 19 PI interviews Developed the Outcome rubric Graded Outcomes for all reports Major contributions to: Executive Summary, Background, Methodology, Results and Discussion, Conclusions, Recommendations, Appendices
Kelly Driscoll	Performed all SAS instances Reached inter-rater reliability with Alex Developed the Predictor Candidate rubric Grade predictor candidates for 20 awards Major contributions to: Introduction, Background, Methodology, Results and Discussion, Recommendations, References, Appendices
Alexander White	Developed the software tool Reached inter-rater reliability with Kelly Developed the Predictor Candidate rubric Grade predictor candidates for 16 awards Major contributions to: Executive Summary, Introduction, Background, Methodology, Results and Discussion, Conclusions, Recommendations, Appendices

Acknowledgements

Prof. James S. Demetry

Prof. Joseph D. Petruccelli

Ms Antoinette T. Allen

Ms Joyce A. Craig

Dr. Rosemary R. Haggett

Dr. R. Corby Hovis

Dr. Russell L. Pimmel

Dr. Roger K. Seals

Ms Melissa F. Squillaro

Dr. Lee L Zia

And all the other DUE staff members who helped us along the way

Table of Contents

Table of Figures

List of Tables

1 Executive Summary

The goal of our project was to develop a tool for the National Science Foundation (NSF) Division of Undergraduate Education (DUE) Program Officers (PO's) that predicts project outcomes. This tool will aid the DUE PO's in making important funding recommendations by providing predictions for each of seven project outcomes based on certain characteristics found in a proposal. The DUE receives thousands of proposals every year from professors and researchers who intend to improve education in science, technology, engineering, and mathematics (DUE, 2004). Due to limited resources, not every proposal can be funded. For this reason, the DUE employs a proven review process and attempts to fund only the best proposals. However this year, 2004, Congress has reduced funding to the NSF (Pear, 2004). It is especially important that the resources available to the NSF are distributed efficiently. By creating a tool that attempts to predict project outcomes, the PO's have an added dimension by which to evaluate proposals.

Because the DUE receives so many proposals each, year we decided to analyze proposals of a very specific nature. We considered those proposals that develop software within the Course, Curriculum, and Laboratory Improvement's Educational Materials Development full-scale development track. These proposals have exclusive characteristics and outcomes, such as the importance of maintenance, which makes them of interest to PO's.

Interviews with division PO's have allowed us to obtain their collective knowledge on important proposal features as well as desirable outcomes. This supplemented our background research and provided a better understanding as to how the PO's make their recommendations. The interview responses were compiled and used to define fifty-two predictor candidates (project characteristics) and seven desirable outcomes (measures of success). The fifty-two predictor candidates represented many different aspects of a proposal such as the project description and information about the submitter or Principal Investigator (PI). Specific characteristics included the proposed software's level of interactivity and whether the PI plans to disseminate his or her project through CD-ROM media. The seven desirable outcomes represented unique measures of success such as project sustainability after grant expiration and how it improves student learning.

By examining 36 completed awards, we collected proposal and outcome data. Each award was scored with respect to the predictor candidates and desirable outcomes. Using multiple linear regressions we obtained predictive models for each outcome. These models showed that each outcome was primarily related to about 12 characteristics. Validation on this model has shown that outcomes can be predicted with a mean error of 0.47. These equations were expressed in a software tool, which was made available to the DUE PO's.

We concluded from this research that both our model and methodology are viable. Our process has produced a tool that can predict the outcomes of a project should it be funded. However, it is important to note that our data set was both small and limited to a specific type of award. The tool may prove useful to DUE PO's but the scope of its use is limited to full-scale EMD proposals that develop software. While this specific model has its limitations, we have proven that predictive relationships do exist between proposal characteristics and project outcomes. We strongly recommend that the NSF pursue this type of research, as a model based on additional data would likely increase its performance. Additionally, it may be possible to write a robust and all-inclusive model for each of its divisions or programs.

2 Introduction

Education has always been a top priority in the United States, receiving $53 billion of federal funds in 2003 (Executive Office of the President, 2004, p. 12, 13). In conjunction with supporting education, there has been a great deal of research on effective learning methods. The advance of technology in the last few decades has offered unique approaches to learning and has given educators the ability for greater interactivity. Many leading professors are trying to take advantage of these new approaches, causing increasing numbers of instructional development projects to be undertaken. (Russ Pimmel, personal communication, 2004). Unfortunately not every institution has enough resources to support its professors' projects. In order to obtain the needed funding, they can apply for grants from sources external to the university to support curriculum development and improvement.

The National Science Foundation's (NSF) Division of Undergraduate Education (DUE) receives project proposals from educators at colleges and universities across the United States; a broad variety of topics and disciplines are represented. The DUE receives thousands of proposals each year, but limited resources allow only 15-20% of proposals to be awarded. As part of the review process, experts in the proposal's discipline read and rate the proposals. Next, after reading the proposal, budget, and peer review comments, the program officers (PO's) make a very important recommendation to the department head of the DUE. This recommendation is very rarely overturned. The outcome of all funded projects would ideally be the achievement of both the goals described in the proposal as well as the DUE's goals, but in reality some projects funded by the DUE do not achieve the desired outcomes as well as expected. If the DUE better understood what information available at the funding decision stage characterizes an ultimately successful project, its funds could be allocated more efficiently.

The DUE receives many proposals aimed at developing software such as web-based laboratories, educational games, electronic homework, and modeling programs. Software development projects are unique because support after completion is very important; software that is not maintained gradually loses its worth and becomes less likely to be used. To understand the implications of a software project, we had to comprehend the outcomes desired by the DUE. We also examined the NSF's review process in detail to gain greater insight into the program officers' evaluation.

There is no systematic approach to proposal evaluation by the program officers. Instead of utilizing a common method, program officers must draw upon their experience and any sense of intuition they have developed by reviewing earlier proposals. Reaching uniformity in proposal review is further complicated by the fact that over half of the program officers are temporary employees who do not have the same wealth of experience to draw upon. By gathering the collective knowledge of the program officers working within the DUE and using it to form a model, we provided an assistive tool to enable a more comprehensive approach to proposal evaluation.

Our project aims at predicting the success of software development proposals if they were to be funded. Specifically, our model is tailored to predict the success of proposals received by the Course, Curriculum, and Laboratory Improvement (CCLI) program within DUE. We analyzed completed software projects and used a statistical model to determine the extent to which certain characteristics contributed to each desirable outcome. Using these relationships we developed a tool that the DUE employees can use to assist in the decision-making process. This, in turn, allows the NSF to distribute funding more efficiently and help ensure that the funded projects meet education's changing needs and "improve the quality of science, technology, engineering, and mathematics education for all students” (NSF, 2004c, DUE's Mission).

3 Background Research

This section describes the process that the NSF uses to make proposal-funding decisions and how our project will be used to assist in this process. The NSF requires that all project applications meet certain standards and move through levels of authorization. The Course, Curriculum, and Laboratory Improvement (CCLI) program, within the DUE, has its own specific rules and regulations regarding the funding of proposals. Currently, the CCLI does not employ a model that predicts award outcomes on the basis of proposal characteristics.

3.1 Proposal Review Process

A proposal must go through several steps before it is awarded or declined. This section will discuss each step in the entire proposal timeline shown below in Figure 1. Once a proposal is received it must meet the NSF's submission standards or those described by the specific program's annual solicitation (if applicable) in order to be accepted for evaluation. After being accepted the merit of the application is evaluated. The first analysis is performed by a peer review panel chosen by Program Officers (PO's). In this review, the proposal is examined by experts in the appropriate discipline. Then, taking careful consideration of the opinion of the panel, the PO makes his or her own recommendation. When a decision to fund a project proposal is made, the proposal and any subsequent information becomes known as an "award."

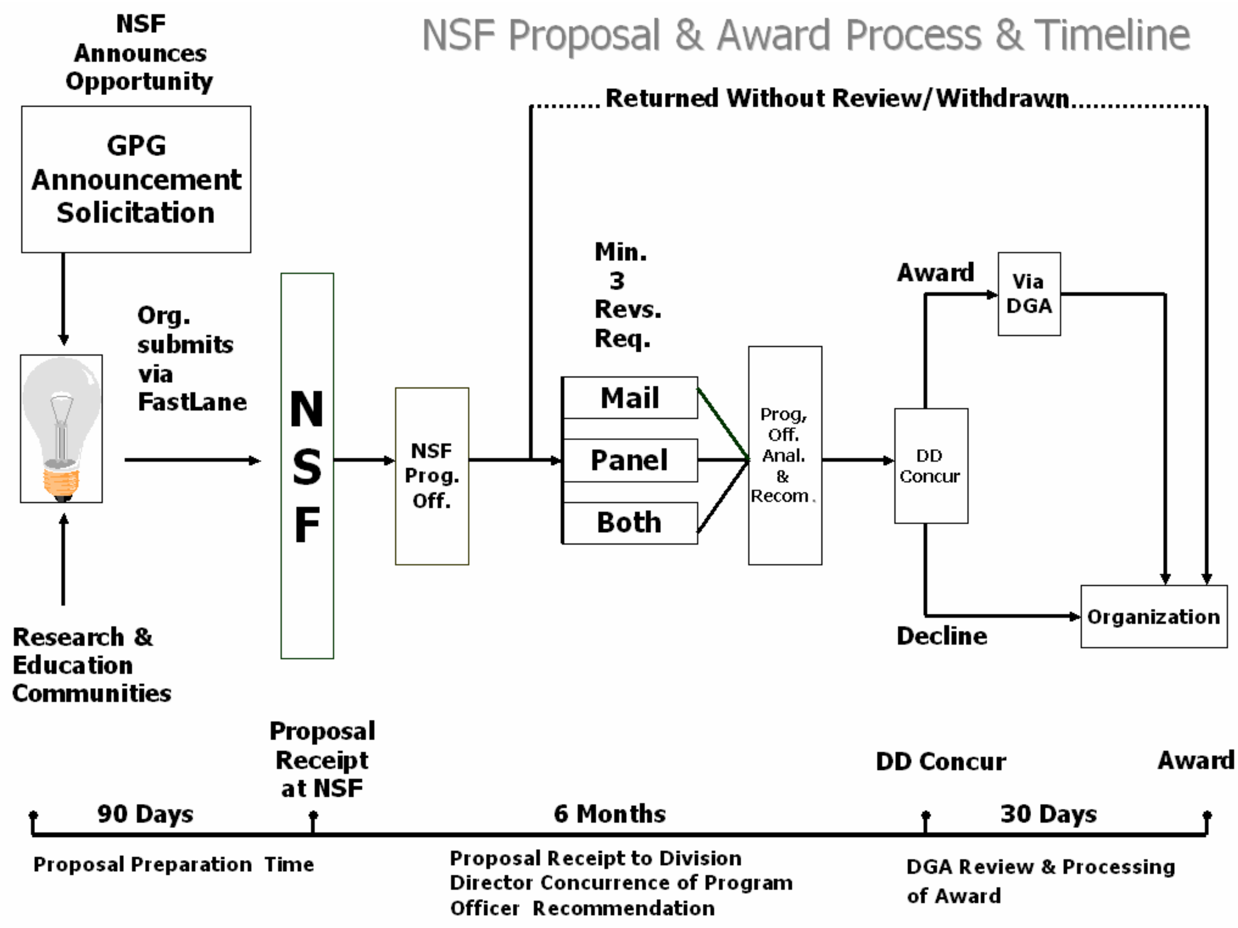


Figure 1: NSF Proposal & Award Process & Timeline
(Source: NSF, 2004a)

3.1.1 Proposal Receipt

Each year, the NSF provides a Grant Proposal Guide that "provides potential proposers with a description of the various categories of funding opportunities utilized by NSF to generate proposals, as well as the appropriate scenarios in which each are used." (NSF, 2004a, "Dear Colleague Letter"). In addition, NSF divisions, such as the DUE, can provide their own guide for proposal writing. Furthermore, each program within these divisions, such as the DUE's CCLI, can have a solicitation that includes specific descriptions for the program and any subdivisions, known as tracks. These solicitations also provide additional criteria for submitted proposals. Programs that do not have solicitations often release grant announcements to educators indicating the availability of funds.

3.1.2 CCLI Submission Standards

All proposals must meet the submission standards in order to be reviewed. Most importantly, the proposal must address both of the merit criteria set by the NSF in a one-page Project Summary (NSF, 2004b, p. 14, 15, 20). These criteria are “What is the intellectual merit of the proposed activity?” and “What are the broader impacts of the proposed activity?”

For the CCLI program, all proposals must also include a cover sheet, a project description, references used, a budget justification, current and pending support, the DUE project data form, and special information and supplementary documents (NSF, 2004b, p. 14-17). The project description must contain the results from prior NSF support (if applicable), the goals and objectives of the activity, a detailed project plan, the experience and capability of the principal investigator (PI) and co-PI(s), the necessary equipment and instrumentation, the evaluation plan, and the dissemination plan. Proposals that do not include all of the above are returned to the PI without review.

3.1.3 Program Officers

Program officers may be permanent employees, Visiting Scientist, Engineer, or Educator (VSEE) employees, temporary employees, or Intergovernmental Personnel Act (IPA) employees (NSF, 2004d, p. 18). (See Figure 2 for a breakdown.) VSEE’s are on sabbatical from their host institutions and work at the NSF for up to three years. IPA employees are employed through grants to their home institutions. This method of employment has increased in recent years. The purpose of non-permanent personnel is to bring in new ideas and viewpoints about science, technology, engineering, and mathematics (STEM), the focus of the DUE.

Program Officers	Total	Percent
Total	**380**	**100%**
Gender		
Male	**220**	**66%**
Female	**114**	**34%**
Race		
Minority	**85**	**22%**
White, Non-Hispanic	**295**	**78%**
Employment		
Permanent	**183**	**48%**
Visiting Scientists, Engineers & Educators (VSEE)	**33**	**9%**
Temporary	**47**	**12%**
Intergovernmental Personnel Act (IPA)	**117**	**31%**
Source: NSF Division of Human Resource Management Notes: VSEE: Individual employed as a Visiting Scientist, Engineer, or Educator (formerly termed "Rotator"). IPA: Individual employed under the Intergovernmental Personnel Act.		

Figure 2: Distribution of NSF Program Officers by Characteristics – October 1, 2003 (Source: NSF, 2004e, p. 18)

Proposals are assigned to program officers based on discipline. Each PO then selects the review panelists for external review and is responsible for making sure the individuals chosen represent varying opinions and are free from conflicts of interest (NSF, 2004e, p. 19-20). Conflicts of interest are not uncommon when working with proposals. This occurs when panelists or PO's may have a stake in the funding of the proposal by having previously taught at the institution in question or by having a close friendship or working relationship with the PI's.

At this stage, the proposal is only looked at for submission standards and enough detail to select effective reviewers.

Program officers at the NSF typically handle about 100 proposals annually (NSF, 2004e, p. 19-20). The DUE, however, handles more proposals. Dr. Russ Pimmel, a lead PO for the DUE, handles around 170 proposals each year (personal communication, 2004). Merit review consumes 40-50% of their time though they are also responsible for award management and oversight, staff oversight, program planning, and other tasks. Since PO's have heavy workloads, it is important that they can evaluate proposals efficiently.

3.1.4 External Review

There are three types of external review used by the NSF: mail-only, panel-only, and mixed (NSF, 2004e, p. 10-16). The CCLI program uses panel-only review exclusively. There are currently two different methods for conducting this type of review, off-site and on-site. For an off-site review, project proposals and review forms are sent to panel members three weeks prior to, and are due back before, the panel meeting. For an on-site review panel members do not see the proposals prior to the panel meeting and all evaluation and discussion is done at the review site. The CCLI program traditionally employed on-site reviews; however DUE PO Dr. Roger Seals indicated that the off-site method is being increasingly used (personal communication, 2004). Normally the NSF requires that at least one PO be present during each panel meeting. The CCLI, however, has a special exemption to this rule. Due to the large number of proposals it receives annually, the CCLI holds multiple panels per discipline and a PO cannot always attend.

It is the policy of the NSF to have at least three external reviewers for each proposal. All reviewers are experts in their disciplines and their service as reviewers is voluntary. The panel rates the proposals on a five-point scale, where 1 is poor and 5 is excellent. The results of the panel are averaged and the proposal is given this score. The largest fraction of proposals falls in the good to very good range, between 3.0 and 4.0 respectively. Proposals receiving scores below 3.0 are referred to as "low declines" and are usually given very little attention and rarely funded. A complete score distribution is shown below in Figure 3 for proposals received in 2003.

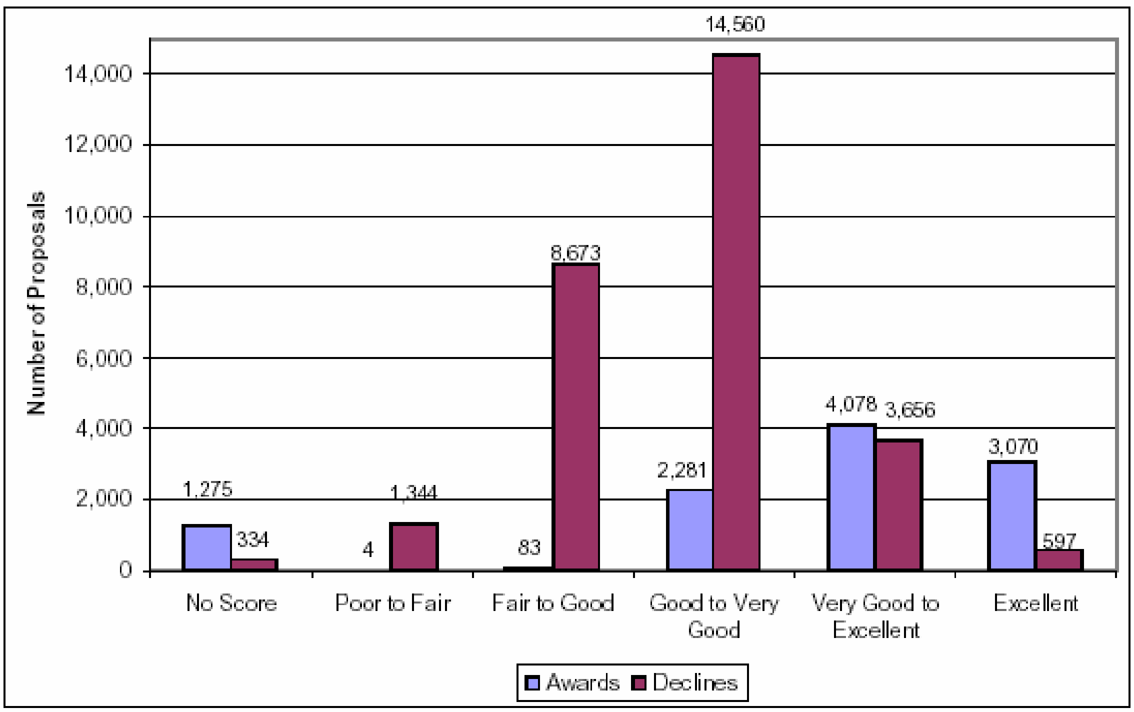


Figure 3: Distribution of Average Reviewer Ratings for 2003 NSF Proposals (Source: NSF, 2004e, p. 16)

3.1.5 Post-Panel Evaluation

After an external review, a PO examines the proposal and the panel reviews. During the panel, a one-page summary is usually created that contains the collective thoughts of the reviewers.

Many of the DUE PO's have indicated that they use the summary and reviews as a guide, weighting reviewer judgments heavily in making the final decision. With the panel reviews in mind, a PO examines the proposal in depth, looking at the intellectual merit and broader impacts of the proposed activity. In addition, for the Educational Materials Development (EMD) track, the idea must be novel. Other ideas could be adaptations of existing solutions and are better suited to the Adaptation and Implementation (A&I) track. Using other criteria defined by the DUE and personal experience, the PO makes a recommendation on whether to decline or award the proposal. He or she often consults other PO's to assist in this recommendation. Essentially, program officers have the duty of predicting which proposals will progress into successful projects.

A PO's recommendation is reviewed by the division director. When programmatic approval has been obtained, the proposal is forwarded to the Division of Grants and Agreement. This division makes the final decision pertaining to the authorization of the NSF and the expenditure of funds. The number of proposals received and number of awards given for each DUE program are shown below in Figure 4.

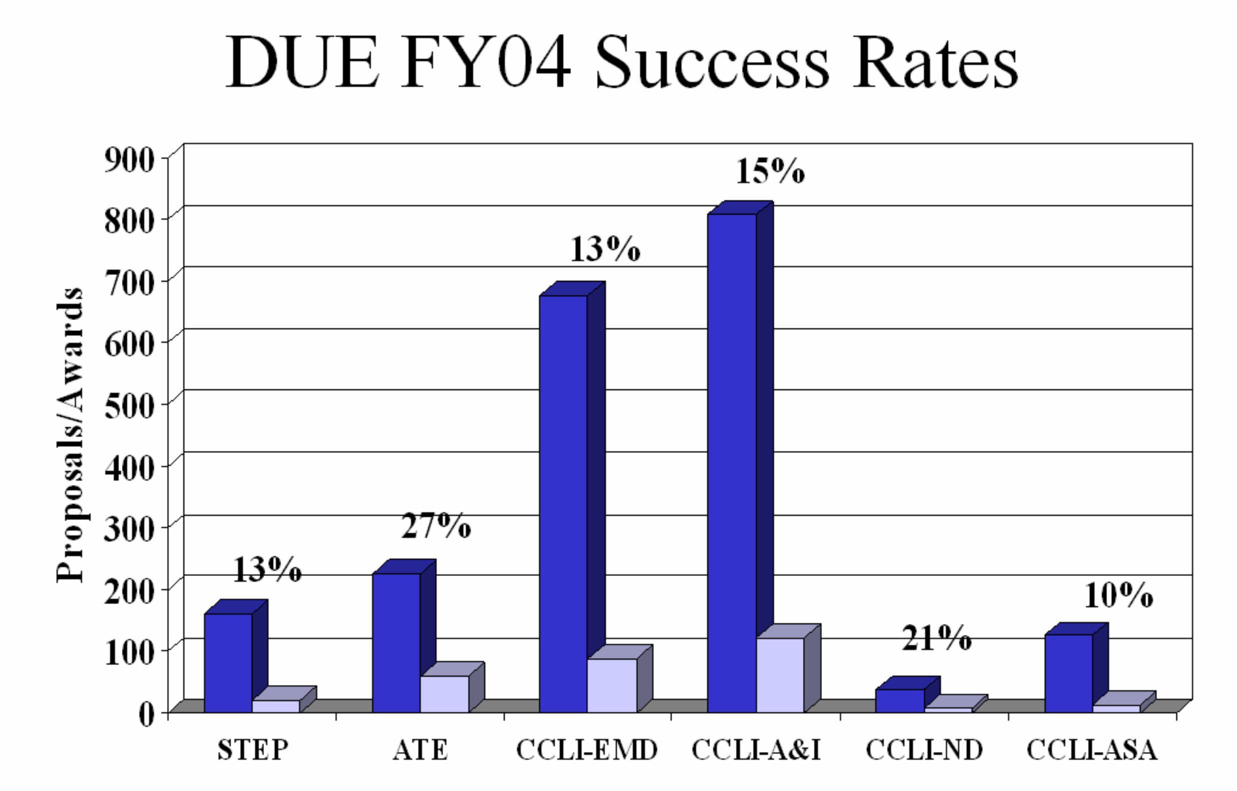


Figure 4: DUE 2004 Proposal to Award Rate

(Source: DUE, 2004)

The EMD track of the CCLI program is one of the most selective programs with an awarding success rate of 13%. This implies that very critical analysis is needed in the review process. Our predictor model provides an additional dimension to consider when making a recommendation.

Permanent PO's can draw upon years of experience evaluating proposals in order to make the awarding recommendation. Temporary PO's, however, do not necessarily have the same level of experience to draw upon. Since personal experience contributes to the decision-making process there may be variations in uniformity. To assist in EMD proposal analysis, a DUE PO has created a number of checklists of things to look for. The full-scale development checklist, provided by Dr. Patrick Carriere, can be seen in Appendix B. It is not sponsored by the DUE but our interviews have indicated that a small portion of the PO's use it during evaluation. Our project has produced a similar assistive tool for use by all DUE PO's, permanent and temporary. Our project goes further by predicting outcomes that are likely to result from this type of proposal. While our predictive tool by no means replaces the PO's decision-making process, it does provide an additional dimension to consider when making a recommendation.

3.2 DUE Project Outcomes

The tool developed by our project is intended to predict how successful a project will be if funded. There are many things to consider when defining success. Success is not a black and white distinction and what the DUE would like to achieve may change over the years. This is reflected by the changing expectations given in the annual CCLI solicitations. Taking this into account we have chosen to define multiple desirable outcomes rather than a single measure of success.

There are many possible outcomes that can characterize a successful award. One such outcome is how well a project achieves the goals and objectives of the CCLI. The EMD track is designed to support the development of new materials that will have national impact. Projects in the EMD track can either be proof-of-concept or full-scale development. The assistive tool developed by our project is designed to predict the success of full-scale development proposals, which have a specific set of outcomes that the DUE wishes to see. Another outcome is the extent to which the award achieved its own goals and objectives. There are additional desirable outcomes and the means by which they were defined are described in our methodology section.

3.2.1 Outcomes Specified by DUE

According to the CCLI solicitation (NSF, 2004g, II Section A), a full-scale development project is "expected to produce and evaluate significant new educational materials and pedagogical practices, and to promote their dissemination and effective implementation nationally." (NSF, 2004g, II Section A). The outcomes of full-scale development projects should include:

- Full-scale development of innovative materials that incorporate effective teaching and learning strategies based upon prior experience with the prototype
- Credible evaluation of the effectiveness of the materials or practices on student learning at different types of institutions serving students with diverse backgrounds and career goals
- Faculty at test sites and other potential users who are prepared to use the materials or practice
- Dissemination of information about the developed materials
- Self-sustaining national distribution

The extent to which projects achieve these outcomes can be measured from the final project report that a PI is required to file when the grant period expires.

3.2.2 Outcomes Specified in the Proposal

In addition to the outcomes expected by the DUE, project proposals have a list of goals that the PI wishes to accomplish with the help of NSF funding. The PI is expected to evaluate these accomplishments and submit annual and final reports documenting progress. There is a distinct possibility that an awarded project will go above and beyond its own projected goals. These additional outcomes could, among other things, be in the form of publications, citations, or further development and implementation. Because recognitions usually occur after the final report has been filed, the NSF might not have documentation of them. Therefore it is necessary to contact the PI's personally to accurately measure the success of the awards.

Unfortunately, some funded projects do not achieve a level of success expected by the DUE. By predicting the extent to which outcomes can occur before the project is funded, we can allow the DUE's limited funding to be distributed to other projects that are more likely to succeed.

3.3 Methods of other Organizations

The NSF is only one of many governmental entities that make monetary grants. Furthermore, there are many non-governmental organizations (NGO's) that also fund projects of various natures. Each of these entities must have methods to filter out proposals that are not likely to succeed as awards. This ensures that monetary aid is allocated efficiently. Unfortunately, none of the organizations we contacted was able to provide us with any insight on their decision making processes, aside from the public solicitations.

The United States Department of Education (ED) provides monetary aid, mostly in the form of grants, to states and school districts (ED, 2004, Overview). This aid focuses on improving elementary and secondary schools, meeting the special needs of students, strengthening teaching and learning in colleges, and supporting rehabilitation, research and development, statistics, and assessment. Dr. Lynn Okagaki, the Deputy Directory for Science for the Institute of Education Sciences, described the ED's basic review process via email correspondence, and it contained the same major steps as the NSF process. The most notable difference between these organizations is that ED employs a panel specifically to evaluate proposed budgets, whereas the PO's at the NSF make these recommendations themselves.

The National Institutes of Health (NIH) is another federal funding agency. The mission of the NIH is the pursuit of fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability (NIH, 2004, Mission Statement). Though its review process is similar to the NSF's, the NIH puts a greater deal of emphasis on its Scientific Review Groups (SRG's), which are comparable to the NSF's peer review panels. SRG's opinions are highly valued because this panel of the complex material discussed in NIH proposals; this panel makes the major distinction between feasibility and impracticality. Panel members on an SRG are able to directly correspond with PI's, which is not allowed by the NSF.

Finally, the Ford Foundation is a NGO with an international presence. The main goals of the Ford Foundation include: strengthening democratic values, reducing poverty and injustice, promoting international cooperation, and advancing human achievement (Ford Foundation, 2004, Mission Statement). As a private organization, the Ford Foundation is allowed more freedom in the proposals it chooses to fund. Rather than accepting and reviewing all incoming proposals the way U.S. agencies do, the Ford Foundation defines a set of criteria for a geographical location and only accepts proposals that address these issues. Through consulting experts and performing his or her own extensive studies, a Foundation PO develops a set of criteria. The Foundation uses peer reviewers to analyze the list and if approved, the PO is given the authority to choose any project he or she wishes.

The information concerning these three organizations is not comprehensive enough to affect our project in any significant way. From our correspondence with these organizations, we determined that none employs a predictive model. We can only conclude from this research that the NSF utilizes a widely recognized methodology.

4 Methodology

The goal of this project was to develop a tool that can be used by DUE Program Officers to assist in the evaluation of full-scale development proposals of the Educational Materials Development (EMD) track concerning software development. The primary purpose of this tool is to predict how successful a proposal of this type will be should it be funded. By relating characteristics of awards known to the PO's at the time of the award with the eventual outcomes of the project work, we hoped to obtain a tool that could predict proposal outcomes prior to a funding decision.

We first gathered a list of these past awards. We then determined a set of aspects (proposal characteristics) to evaluate these awards by. We completed the same task with measures of success (award outcomes), evaluating the results of these awards. Next, we used a powerful statistical methodology, multiple regression analysis, to analyze the resulting data and characterize the relationship between these aspects and measures of success. Finally, we represented our findings through a software tool that the DUE PO's could easily use and modify over time.

4.1 Sampling the Awards

The CCLI funds about 250 awards annually across four tracks: EMD, Adaptation and Implementation, National Dissemination, and Assessment of Student Achievement. Since it was not feasible for us to examine such a large and diverse pool, we had the option of selecting a sampling group or choosing a subset of the awards. By choosing a subset of awards (full-scale development EMD) we were able to tailor our analysis to the unique characteristics and outcomes available, Though this narrowed the range of our model's applicability, the trade-off seemed acceptable because we assumed that a narrow scope would help focus our model and increase its predictive accuracy.

The EMD track is comprised of full-scale development projects and proof-of-concept projects. The former produce materials that can be distributed nationally and may involve funding of up to $500,000. Proof-of-concept projects, on the other hand, develop prototypes and new learning methods and therefore require considerably less funding. The DUE is especially interested in full-scale development projects as the division only has enough funding to award a small portion of these larger projects each year. In addition, these projects can have a far greater impact on pedagogy and student learning.

Melissa Squillaro, the Science Education Analyst for the DUE, provided us with an Excel spreadsheet of all EMD awards. Currently, 666 EMD awards have been funded. However, many of these awards were ongoing and it was important that we only studied projects with mature outcomes: projects for which the grant has expired and the final report has been approved by a DUE PO. The spreadsheet included the status of each award and allowed us to separate out completed awards.

We decided early on, with the help of our sponsor, to work exclusively with proposals that plan to develop some type of software. Software projects include virtual teaching aids, educational games, web-based laboratories, electronic homework or lessons, simulation and modeling programs, and toolkits that allow other educators to create their own software materials. This is not an exhaustive list as our definition includes any project that produces materials, both for students and faculty, which are used on computer systems. However, materials that were created in a textbook form and also converted to distribute over the Internet were not considered software development.

The rationale behind this choice was fueled by the fact that software projects have unique outcomes. Projects that develop software require support and maintenance after release. If a product cannot be updated with new information or altered to suit changing needs, then it is less likely to be adopted over time and becomes essentially useless. Another aspect we considered is the relative youth of software-based proposals. Distance learning is becoming more and more popular, according to a DUE PO Russ Pimmel, and as a result web-ready projects are often proposed (personal communication, 2004). As the years go on, the NSF expects to see more projects that deal with these unique characteristics and outcomes. For these reasons we chose to consider these projects exclusively.

Unfortunately, the DUE has does not arrange its awards by the type of material being developed, such as software or textbook-related. Keyword searches of the title and abstract are allowed, but this method will return many awards that are not pertinent but rather contain the word "software" in the abstracts. Furthermore this method might leave out awards that do not specifically mention either, such as those that proposed web-based laboratories. To determine which projects developed software, we read the abstracts of the final reports for the completed awards and identified those having to do with software development.

4.2 Determining the Proposal Predictors

There are two parts to any predictive model, predictors and outcomes. This section describes the approaches we used to determine the characteristics of software development proposals that became the predictor candidates. We refer to these characteristics as "predictor candidates" because not all of them determined in this section were found to correlate with our measures of success.

Due to the time constraints of this project, we were unable to perform content analysis on software development proposals to determine common characteristics. Doing so would have provided an unbiased set of characteristics, but was not possible in the time allotted. As we had no experience in reviewing DUE proposals, our list of predictor candidates was developed exclusively from interviews with DUE program officers. By conducting content analysis on the responses we received, we determined a set of predictor candidates. Using these candidates, we evaluated each of the proposals of the awards chosen by the previous section's methods.

4.2.1 Program Officer Interviews

As there were no specific guidelines for proposal evaluation, we conducted interviews with the DUE PO's to gather the collective knowledge regarding individual evaluation practices. These interviews followed an unstructured format that allowed us to clarify questions further and probe for the types of responses that pertain to our project. The data gained from these interviews provided us with a set of characteristics that PO's deem important. A sampling group was not necessary, as there were only 22 PO's working in the DUE.

Interviews lasted a half hour on average, and the full protocol is provided in Appendix C. This protocol was tested on three individuals at the DUE prior to conducting the official interviews. The individuals involved in this pretest were Russ Pimmel (our liaison), Rosemary Haggett (the division director), and Melissa Squillaro. Dr. Pimmel was not to be included as an official interviewee because of his close connection to our project. Dr. Haggett and Ms Squillaro were not officially interviewed because they are not PO's but are well acquainted with the award process. Dr. Pimmel provided us with a list of topics to use as probes for responses. The pre-testing ensured that the questions asked were clear to the interviewee and addressed all of the issues of importance to us.

We began each official interview with a description of our project and mentioned that we are interested in projects that pertain specifically to software development. Topics we discussed included specific sections of EMD proposals such as dissemination plans, evaluation plans, budgets, and PI experience. John Caulkins conducted each interview while Kelly Driscoll and Alex White took notes and assisted in clarifying interviewee questions. Our intent in using one interviewer for all sessions was to, insofar as possible, assure a consistency and uniformity in the interview results.

4.2.2 Predictor Candidates and Rubrics

After the interviews were completed, the data were organized into a spreadsheet. We individually examined the spreadsheet and each produced a list of unique characteristics. Then together we compiled our three lists into a single list of candidates. Kelly Driscoll and Alex White were in charge of scoring awards by these predictor candidates.

To measure these predictor candidates we decided upon a mix of binary and tertiary responses. Certain candidates, such as “open/closed source,” were better suited to a simple binary response of “yes” or “no”. Others, however, such as the proposed software’s “level of interactivity” required a greater degree of resolution to more accurately represent the range. Binary responses were recorded as either 0 or 1 and tertiary responses recorded as 0, 1, or 2. It became immediately apparent that different people would interpret these scores differently. Our tool’s accuracy depends upon the consistency in values chosen for the predictors. For each predictor candidate, we developed a rubric that would explicitly state what a 0, 1, or 2 represents. The rubrics were designed to minimize, to the greatest extent possible, the element of subjectivity when scoring awards. If a predictor's condition for a 2 is not satisfied, then the condition for a 1 is checked and so on.

4.2.3 Scoring the Award Proposals

Preliminary testing showed that examining each award proposal was a time-consuming process (approximately one hour each) so we decided, with our liaison, that both Kelly and Alex would evaluate half of the entire pool if our inter-rater reliability reach 90%. To test the rubric for inter-rater reliability, two awards that matched our software criterion were evaluated. To protect our limited award pool, ongoing awards were chosen as samples. Outcomes were not important for this test. Both Kelly and Alex evaluated all four awards and compared the results of the scoring. Specific attention was paid to the conflicting scores and the rubric was modified to eliminate confusion. This process was repeated until Kelly and Alex achieved an inter-rater reliability of about 90%. The award pool was then randomly divided between Kelly and Alex to score. Each award was scored with respect to the predictor candidate list and rubric. During inter-rater reliability testing, certain PC's were dropped due to various reasons including complexity and not characteristic of software EMD awards.

Aside from time savings, there is another benefit to having multiple people score awards. Our predictive model and resulting tool is to be used by multiple PO's. It is unlikely that real-world use will remove all subjectivity in scoring, regardless of the detail of the rubric. The few discrepancies Kelly and Alex have in the scoring process more accurately models what will be experienced in real-world use.

4.3 Determining Award Outcomes

During the interviews with the PO's, we determined a set of outcomes that the DUE would like software development projects to accomplish. Since these outcomes can occur to different extents, a range of values was created for scoring. For each award evaluated by Kelly and Alex, the outcomes were evaluated and scored similarly.

Beyond developing the set of desirable outcomes, John took sole responsibility for the outcome domain. It was desirable for John to exclusively deal with outcomes so that Kelly and Alex could remain unbiased toward the awards when scoring the predictor candidates. Furthermore, this produced consistent scoring and eliminated the need for inter-rater reliability in the outcome domain. To properly assign outcome values John read the final reports, searched their websites, and contacted the PI's themselves for phone interviews.

4.3.1 Interviews to Define Outcomes

The outcome-related responses of the PO's were collected during the same interviews as the predictor candidates and thus the same protocol and format applied. Taking into account the extent to which the outcomes can occur, a range of 1 to 5 was decided upon. We felt that this provides a good amount of resolution to distinguish levels of success, without having difficulty in determining the cutoff points. Similar to the predictor candidates, a rubric was developed for each outcome to explain what each value (1 to 5) represents.

4.3.2 Final Report Outcomes

John closely examined the outcomes as they appeared in each award's final report. Content analysis was performed on these reports in a method similar to that performed on the proposals. As no assumptions could be made, the extent to which each outcome was documented reflected the maximum score that the award could receive. To gain a more accurate picture of each award, web searches and projects websites were used.

After reading the final report, John searched their project's website to find further developments in the PI's work. This would help us to determine if the product was widely spread or even still in use. He also checked the project's webpage if one was created. Additionally, some final reports were short and nondescript. Also a final report might not represent an outcome that had occurred and thus receive a lower score than it should have. For these reasons we attempted to obtain as many PI interviews as possible, to further reveal outcomes with as much accuracy as possible. This process was repeated for all awards in our pool.

4.3.3 Principal Investigators Interviews

The easiest and most accurate way to gauge outcomes was through principal investigator interviews. The questions asked were aimed directly at the graded criteria and the answers were accurate and current. Also, John addressed any discrepancies between the final report and the proposal. The PI interview protocol is provided in Appendix D. We were especially interested in the PI's statement concerning any further publications, development, implementation, or awards. Outcomes such as receiving professional recognition cannot usually be determined upon project completion. The PI's however are able to provide us with such information.

The interview scores were kept separate from final report and website scores for several reasons. Mainly we wanted to keep our data in a standard format. We weren't able to schedule interviews with every PI in our pool, and we felt it would give an unfair shift in the scores if we integrated them together. Also we kept the scores separate because they represent somewhat different information. Final reports are available to DUE PO's and represent three to four years of work. Interviews on the other hand are unavailable to PO's and can represent five or more years of work.

4.4 Determining Predictor-Outcome Relationships

Once the award data had been gathered for the entire pool, we proceeded with statistical analysis. This section addresses the methods we used to determine the relationships between the predictor candidates and each outcome. To determine these relationships we applied a published statistical model to the data gathered from the previous sections. Considerations for a statistical model are discussed.

4.4.1 Types of Statistical Models

There are a large number of statistical models available that could be used to relate predictors to outcomes. One of the most basic and robust models is multiple linear regression (SAS Institute Inc., 2004, pp. 3873-3876). A multiple linear regression model characterizes the relation between an outcome variable and a set of predictor variables. In the simplest case, the outcome is taken to be a constant plus a linear combination of the predictor values plus a random error. While this model is easy to use, it predicts outcomes on an unbounded continuous scale, not the 1 to 5 integer scale of the award outcomes we scored. Such predictions could be interpreted as average outcome scores, or they could be rounded if integer scores were desired, though at the cost of lost information. In addition, is it possible that this type of model can produce a result such as “–2.3” or “13.6”. While this model does have its shortcomings it is still a powerful and is easy to represent. Other models exist that can represent the probabilities of the integer outcomes as functions of the set of predictor variables. We did not, however, have an opportunity to use them.

4.4.2 Fitting the Models

It should be noted that our data, discussed in our results section, have indicated that there are a few overly influential predictor candidates. These are candidates having an extreme distribution of values, such as 10% 0’s and 90% 1’s, or vice versa. Preliminary tests showed our model accommodating to these predictor candidates and significantly changing the accuracy of its predictions. Our methodology was modified to fit these models with these influential candidates included as well as removed.

We decided to fit these models using the commercial software SAS. Our data show that the number of predictor candidates outnumbers the awards (observations) in our data set. This allows the model to fit for each specific case, which means that the model is able to exactly predict every outcome in the data set by including an abnormally large number of predictor candidates. However, this does not mean that it will predict data outside of our set well.

To compensate, we used multiple linear regression with a stepwise approach to identify significant predictors. To restrict the candidates to only those significant we adjusted the SAS default entry and exit thresholds to 0.1 from 0.15. This reduced the number of candidates that made it into the final fitting.

To better explain the different fittings that were conducted, see the bulleted process below. Each **bolded** item represents a final model whose accuracies will be compared. Sub-bullets of these bolded items are validation tests, which are described in the following section. This process was completed for each outcome in the two previously described outcome sets.

- Linear Regression Filter on All Predictor Candidates
    - **Multiple Linear Regression**
        - Cross-validation
        - Leave 17% Out
    - Remove Overly-Influential Predictor Candidates
        - **Multiple Linear Regression**
            - Cross-validation
            - Leave 17% Out
- Linear Regression Filter on All Predictor Candidates except Overly-Influential
    - **Multiple Linear Regression**
        - Cross-validation
        - Leave 17% Out

4.4.3 Model Validation

The predictive performance of each fitted model was determined by multiple methods. First, each multiple linear regression model was subjected to "cross-validation" where one observation is left out and the model is fit to the remaining data. The missing value is then predicted from the newly created model. This process is repeated until predictions have been made for every observation. Finally, by examining the mean error in the predictions, we get a measure of predictive performance.

The second validation method was to remove a random set of observations and fit the model once. This fitted model would predict the removed observations. Similar to the cross-validation process, we then examined the mean error in the predictions. This method supplemented our results determined from cross-validation and represented predictions that are less likely to match specific cases. We determined that we would remove about 17% of our observations for this method: of the 36 awards in our data set, we fit with 30 and validated with 6.

4.5 Development of the Predictive Tool

Using the model that produced the most accurate predictions, Alex developed a software program that PO's could use during the proposal evaluation process. This is the final product of this project and allows PO's to interact with the predictive model. This program simply takes the predictor scores as inputs, and when proposal evaluation is complete, the PO hits the "Predict" button and the outcome likelihoods are presented.

Although every PO's workplace computer runs the Windows XP operating System, a few PO's do run Linux on their home PC's or use Macs at home. Since Java is platform independent, it allows the same program to be run on any machine without modifications. In addition, Java provides many Graphical User Interface (GUI) components allowing simple creation of a professional program. The GUI is an important factor to consider when creating a software program as it defines how easy the program is to use.

5 Results and Discussion

This section presents and discusses the results obtained by performing the procedures described in the preceding chapter. This chapter's layout will closely mirror that used in the methodology section where the raw data and any findings are presented in the appropriate sections.

5.1 The Award Pool

As mentioned in the methodology section, Melissa Squillaro provided us with an Excel spreadsheet of all EMD awards. Currently, 666 EMD projects have been funded. The award pool was reduced to 65 by selecting completed full-scale development projects. After reading the final report abstracts of these 65 awards, 46 were preliminarily identified as software development projects. Either during award analysis or through contact with the PI's, 10 projects were no longer considered software development or full-scale development projects. A few of these projects had been modified during the grant period, but most were determined to simply use existing software or publish materials over the Internet. The remaining 36 awards, shown in Appendix E, were evaluated and the resulting data was used to fit our models.

5.2 Predictor Candidates

Our list of predictor candidates was developed exclusively from the interviews with the PO's. Of the 22 PO's in the DUE, 16 agreed to meet with us. We conducted content analysis on the interview data and determined a set of 62 predictor candidates (PC's). The data from these interviews can be found in Appendices E. These PC's fall into six main categories:

- Project Description – details concerning the project such as what it intends to produce and how it intends to be completed
- Assessment/Evaluation Plan – how the project team plans to measure the impact of the project on learning
- Dissemination Plan – how the project team plans to distribute the deliverables and encourage others to adopt the product
- PI Information – information concerning the previous experience of the PI
- Budgetary Information – justifications for funds requested
- Facilities and Support – what equipment is available to the project team as well as resources provided by the host institution

In addition, the average score (1-5) given by the panel review is a predictor candidate.

5.2.1 Predictor Candidates and Rubric

While conducting inter-rater reliability tests, we determined that some of the 62 PC's did not apply to the types of awards we were examining. We also separated a few predictors that represented multiple qualities. The resulting final list consists of 52 PC's and can be found in Appendix F. Each predictor candidate has a label PC1 – PC52 to facilitate data analysis.

For each of the 52 PC's, we developed a rubric to more consistently score the awards. When the rubric was first created, there was a good deal of inter-rater variation in the scores. This rubric underwent many changes and clarifications as our inter-rater reliability testing was conducted. This final rubric is shown in Appendix G.

Using the original rubric, Kelly and Alex were about 75% consistent in their grading. After conducting four trials and making multiple modifications, they achieved an inter-rater reliability of about 90%. Due to time constraints, this value was calculated using only 2 awards; the results of this calibration can be found in Appendix H. Conflicting scores are highlighted on each scored sheet. A reliability rating of 90% is reasonably accurate, but also fairly represents the discrepancy that is likely to be found when PO's use our tool. By having discrepancies in our data we produced a tool that was more in-line with real world situations.

5.2.2 Predictor Candidate Scores

The award pool was randomly divided between Kelly and Alex to score. Each award was scored with respect to the PC's and corresponding rubrics. The proposal portion of each award is roughly 50 pages in length and includes 30 pages of project description and development plans. The remaining 20 pages are mostly dedicated to PI and budgetary information.

The resulting scored sheets were then compiled into one Excel spreadsheet to facilitate statistical analysis. This compiled spreadsheet (raw data) is shown in Appendix I. Bar graphs showing the number of occurrences of 0, 1, or 2 for each predictor candidate are included in Appendix J to convey the raw data more effectively. The distributions from a sample set of PC's are shown below in Figure 5.

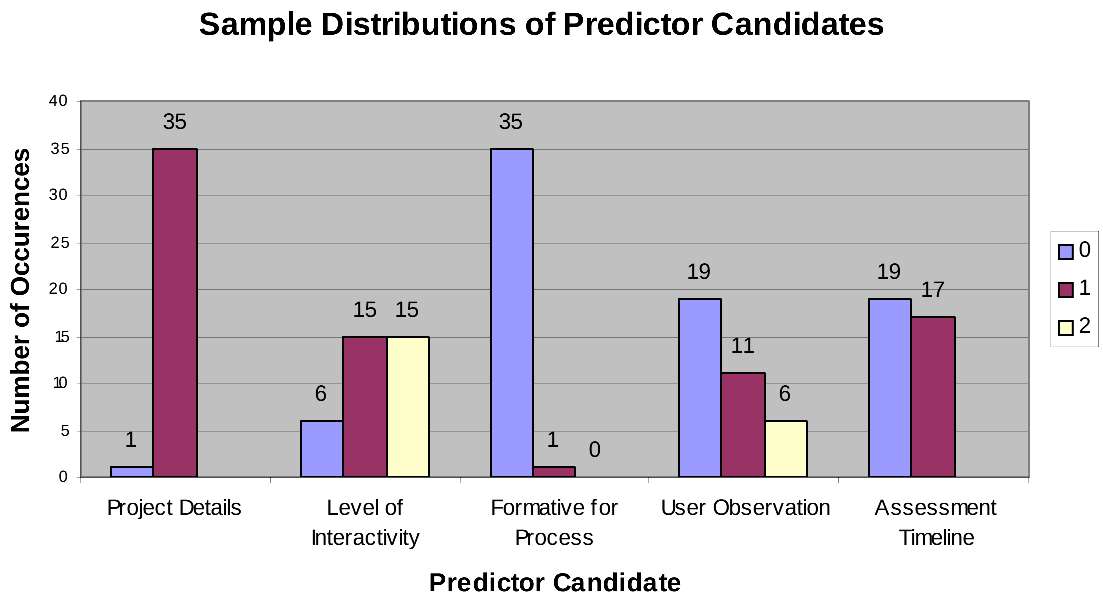


Figure 5: Sample Distributions of Predictor Candidates

The figure above includes a sample set of the 52 total predictor candidates and is not representative. "Project Details" (PC6, does the PI describes the details of what he/she is developing?) is an interesting candidate. The vast majority, 34 of 36, included project details in their proposals. As mentioned previously in our methodology section, this is an example of an overly influential predictor candidate. The number of observations included in the minority is extremely small causing our model to compensate for these cases. Because of the small number of observations, we cannot confirm the validity of predictions including this characteristic. These impacts are discussed in more detail later. There are nine such predictor candidates in our data set (PC6, 12, 20, 22, 39, 40, 43, 46, 47). "Formative for Process" (PC20, did the PI mention that the methodology would be evaluated and changed throughout development?) is another such value, but this was a tertiary choice. The "Level of Interactivity" (PC14, how interactive is the proposed software?) yielded a relatively nice distribution across all three choices. "Assessment

Timeline" (PC32, did the PI provide a timeline for project assessment?) is an example of a binary choice that yielded a nice distribution.

Special attention is given to predictor candidates 1 (Panel Review Score) and 45 (Years as a Higher Education Professor) as these scores are not discrete values in the range of 0, 1, or 2, but are continuous or do not have a specified range. These candidates' distributions are graphed separately at the beginning of Appendix J.

5.2.3 Reviewer-Score Correlation

To analyze Kelly and Alex's inter-rater reliability, we have produced a correlation table between the reviewer and all 52 PC's. Kelly was assigned as 1 and Alex as 0. Scores closer to zero indicate less correlation, which is interpreted as better inter-rater reliability. A score of 1 or -1 would indicate perfect correlation. A negative correlation indicates that Alex gave higher scores for that PC, and vice versa. This correlation table is provided in Appendix K. Alongside the table are the P-Values for each PC. P-Values less than 0.05 were deemed significant.

Of the 52 PC's six showed significant correlations beyond the 0.3 level. Three of these PC's showed correlation beyond 0.4: "Sustaining Plan - Maintenance" (PC18), "Pre-and Post-Testing" (PC29), and "Institutional Support" (PC51). The largest of these, "Pre- and Post-Testing" had a significant correlation of -0.57 with a P-Value of 0.0003. This was a conflicting-score PC during inter-rater reliability testing and indicates that Alex gave consistently higher scores. This can be attributed to the difficulty in determining what constitutes "Pre- and Post-" as was the case during testing. We also determined that "Sustaining Plan - Maintenance" was a difficult PC to characterize. "Institutional Support," on the other hand, was a straightforward PC and we have attributed its correlation to the specific awards reviewed. An interesting thing to note is that all of the six PC's with significant correlation were negative. Additionally, there was a negative correlation with 33 of the 52 PC's. This leads us to believe that Alex was a more generous reviewer, giving higher scores on average. Alex has said that is it possible he "read into the proposals" too much, inferring higher scores. It should be noted that while the award pool was divided randomly, it is possible on a small set of 18 awards that Alex received "better" awards to score. These results indicate the difficulty in providing uniform responses even with a rubric, as personal bias can still influence the scores given. An impractical, but possibly more accurate solution to this problem would be to create individual models tailored to the personal bias of the individual. Such claims have not been validated but might be interesting to explore. This could be done by having a single reviewer score the same award pool and comparing the predictive performance. The amount of time and effort, however, needed by PO's to create the base of each model would likely outweigh its predictive benefits.

5.3 Award Outcomes

The seven desirable outcomes we obtained through the PO interviews are reported below. A rubric for each outcome was written and used during award evaluation. The results of the final report analysis and PI interviews are also contained in this section.

5.3.1 Desirable Outcomes

At first the task of writing a rubric was difficult because the terms we were attempting to measure are strictly qualitative. No single method for obtaining an outcome can be better or worse than any other method. Comparing, for example, dissemination through workshops with dissemination through journals is impossible. We instead had to define a scale of extent for each of these desirable outcomes. This scale begins with one point for non-existence. The values of two, three, and four represent levels of increasing fulfillment of the outcome. Finally, the scale ends with five points for the achievement of excellence. The seven desirable outcomes we have defined are as follows:

- Product (O1) – the achievement of stated goals and objectives
- Dissemination (O2) – the level of distribution attempted or achieved
- Sustainability (O3) – the longevity and robustness of the project's software aspects
- Student Interest (O4) – the extent and diversity of assessment performed
- Student Learning (O5) – the extent and diversity of evaluation performed
- Further Projects (O6) – the importance of this research upon further research
- Professional Awards (O7) – whether or not the project receives professional merit

The five-point rubric for each of these seven outcomes can be found in Appendix L.

5.3.2 Final Report Outcome Scores

Each award was scored for each of the desirable outcomes based on the final report that had been submitted to and approved by the DUE. The final reports in our pool ranged from 3 to 100 pages and averaged 17 pages in length. For each outcome, each award was initially given a score of 1. This score did not increase unless the documentation explicitly stated proof of existence and/or extent. As the dissemination process was described in detail, for example, the grade would increase. If no adoption of the software occurred, a 3 was the maximum possible score. Unfortunately, this process proved harsh against incomplete data. Some awards had three-page long final reports and without a working website or a PI interview, there was virtually no way for these awards to score well. The results of John's scoring can be seen in Appendix M with the distribution of scores in Appendix N.

If there was a website listed in the final report it was taken into account during the evaluation. This score was used to supplement the final report, because the websites were available to PO's at the time of the award closure. These sites rarely made a difference in the grading, however. Websites tended to merely exist as a method of dissemination, rather than a project description. In the event that any of the seven outcomes were addressed on the web site, the official "Final Report" score was adjusted accordingly.

5.3.3 Interview Outcome Scores

We were able to schedule phone interviews with 19 of the 36 PI's. These interviews lasted 10 to 20 minutes apiece, and the PI's were able to elaborate on project growth since the time the final report had been filed. Interview data tended to be more complete and provide a more accurate assessment of project outcomes. We kept these results separate from the final report and website results. The results of evaluation and distribution can also be seen in Appendices M and N respectively.

It was very common for an award to receive higher grades from an interview than from its final report. One reason for this was continued research or development. Another was that the PI was simply able to clarify or add upon the information in the final report. There were, however, rare occurrences that led to a lower interview grade. One example is from a PI who stopped developing his software in order to take on another project. From the interview he was given a 2 for "Sustainability" (O3) to represent static existence. At the time of the final report, however, he had received a 4 due to ongoing development. Interview outcomes are referred to as

IO's while final report outcomes are referred to as FRO's.

5.4 Predictor-Outcome Modeling

This section presents and discusses the relationships between the predictor candidates and each outcome through statistical analysis. The following results were obtained through multiple trials with SAS. We created and cross-validated three different linear models to represent our evaluation data. Furthermore, we validated one trial of "leave 17% out." From these models we used the model with the best overall predictive ability to create our software tool, which is also described in this section.

5.4.1 Fitted Models

For each outcome (two sets of seven) our methodology produced three equations for multiple linear regression. These equations indicate which predictor candidates appear in the final model as well as their weights (coefficients). An example of one of these sets of functions is provided below. All of these function sets are provided in Appendix O. The "all" suffix is the inclusion of *all* PC's into the filter model. The "a" suffix is the elimination of the influential PC's *after* the filter model. The "b" suffix is the elimination of the overly influential PC's *before* the filter model. "Product" (FRO1) is the extent to which the PI produced what he or she had proposed as indicated by the final report.

```
FRO1all = 3.393 – 1.035*PC21 – 0.573*PC49 + 0.514*PC35 –1.770*PC20 + 0.838*PC8 + 0.687*PC1 –
       1.240*PC47 – 1.092*PC48 – 0.289*PC33 + 0.948*PC52 – 0.326*PC25 + 0.475*PC16 – 0.292*PC5 –
       0.602*PC39 – 0.183*PC13 + 0.344*PC17 + 0.314*PC38 + 0.315*PC7 – 0.129*PC41 – 0.117*PC14
```

```
FRO1a = 3.537 – 0.892*PC21 – 0.482*PC49 + 0.322*PC35 +0.552*PC8 + 0.363*PC1 – 0.686*PC48 –
       0.283*PC33 + 0.509*PC52 – 0.180*PC25 + 0.251*PC16 – 0.273*PC5 – 0.129*PC13 + 0.332*PC17 +
       0.319*PC38 + 0.600*PC38 – 0.069*PC41 – 0.262*PC14

FRO1b = 4.399 + 0.494*PC38 – 0.678*PC21 – 0.412*PC49 + 0.352*PC35 + 0.330*PC8
```

It is interesting to note how few predictor candidates were selected in the last case. These functions with the "b" suffix have considerably fewer terms across most outcomes. This occurs due to the nature of the overly influential PC's. Since a small percentage of these awards include them, the "all" model attempts to account for these minority observations by using more PC's to essentially "work around" the overly influential candidates. The "b" model removes these candidates before filtering to determine if this method will produce more accurate or reliable predictive ability. The accuracy of these three models is discussed in the following section.

It is also interesting to note the inclusion of negative coefficients in the equations. Logically, this means that certain PC's are negatively correlated to success as all of our PC's were developed so higher scores were better on a per-candidate basis. However, this does not mean that the model created is incorrect. The models produced from our data are useful for prediction, but none showed any explanatory ability. .

5.4.2 Model Validation

As discussed earlier, we have fitted our data to three different models. To simplify discussion about these three models, each will be referred to by acronym:

- Leaving *ALL* PC's in the filter (ALL)
- Removing the influential PC's *After* Filtering (RAF)
- Removing the influential PC's *Before* Filtering (RBF)

We determined how accurate these models are by examining the mean error in the predictions. Reliability was determined by examining the standard deviation of the prediction errors. The best model would ideally be the most accurate and most reliable.

5.4.2.1 Final Report Outcomes

All three models' cross-validation error distributions for the FRO's can be found in Appendix P. Mean prediction errors and their standard deviations for the different models are shown in Table 1 below. In addition, the number of PC terms (predictors) included in each FRO equation is also shown.

Table 1: Mean Errors and Standard Deviations for Final Report Outcomes

Outcome	Cross-validation Mean Error (36 Observations)			Standard Deviations (Cross-validation)			"Leave 17% Out" Mean Error (6 Observations)			Number of PCs in Function		
	ALL	RAF	RBF	ALL	RAF	RBF	ALL	RAF	RBF	ALL	RAF	RBF
FRO1	0.19	0.70	0.52	0.14	0.51	0.45	1.86	1.80	1.80	20	17	5
FRO2	0.35	0.66	0.86	0.24	0.61	0.53	1.31	1.36	2.5	18	16	6
FRO3	0.57	0.86	0.49	0.36	0.65	0.35	0.92	1.53	0.84	12	9	11
FRO4	0.69	0.93	0.87	0.55	0.74	0.66	0.88	0.83	1.05	8	7	2
FRO5	0.66	0.72	0.68	0.59	0.66	0.55	2.17	1.85	1.93	11	10	13
FRO6	0.48	0.48	0.49	0.33	0.33	0.33	0.65	0.62	1.88	6	6	6
FRO7	0.33	0.33	0.33	0.22	0.22	0.22	1.02	0.93	0.86	10	10	10
Averages	**0.47**	**0.67**	**0.61**	**0.35**	**0.53**	**0.44**	**1.26**	**1.27**	**1.55**	**12.1**	**10.7**	**7.6**
Std. Dev.	**0.19**	**0.21**	**0.20**				**0.56**	**0.49**	**0.64**	**5.11**	**4.23**	**3.87**

Our analysis has shown that for the FRO's, the ALL model produces the most accurate results but the RBF model uses the fewest PC's. Averaging all seven outcomes for the ALL model produced a mean error of 0.47. This is better than the mean errors 0.67 and 0.61 produced by models RAF and RBF respectively. The average number of PC's going into the ALL equations is 12.1, whereas RBF uses substantially fewer PC's (7.6) on average. The removal of the influential PC's cause the predictive ability to decrease in both RAF and RBF models. The standard deviations follow the same trend as the mean errors with the ALL model being the most reliable in its predictions. It should be noted that certain outcomes have been shown to predict better than others. For example the ALL model produces the best mean error of 0.19 for "Product" (FRO1). This is over three times more accurate than the largest mean error exhibited by "Student Interest" (FRO4). These trends are shown across all three models.

In addition to these three validations, we conducted a rougher trial where 17% of the total observations (six for our data set) were removed and model was fit to the remaining observations to predict the removed six. Awards were randomly selected and removed from the fitting process. Due to time constraints, however, we were only able to fit one "leave 17% out" set for all FRO's. Because the results for only one fitting are available, the mean errors are taken lightly in our selection of the best predictive model. According to the mean error in the prediction of the six observations, all three models produced errors significantly greater than those in the cross-validation fittings. This can attributed to the reduced amount of data available to fit the model (30 observations as opposed to 35 in cross-validation fitting). Professor Petruccelli, our advisor, has suggested that "Leaving out six may make it likelier that regions of the predictor space will not be represented in the fitted model, and thus give a bad prediction." (personal communication, 2004). Because we have so few data points, leaving 17% of them out may prevent entire sections of data from being used in fitting the model. Results show that the ALL and RAF models performed slightly better than the RBF (0.29).

5.4.2.2 Interview Outcomes

The Interview Outcomes (IO) were more difficult to model, as only 19 observations were available compared to all 36 used in the FRO fittings. This extremely small data set produced wildly fluctuating results using our predefined filter method. Due to the IO's volatile nature, we reduced the number of PC's that would make it out of the stepwise filter model. We narrowed this selection by using the fewest number of PC's that produced a Model R-Squared of at least 0.900. By reducing the number of PC's available for fitting in the final model, we were able to produce more consistent results. To ensure that the FRO's model is consistent with that of the IO's, we performed the same 0.900 selection narrowing of the FRO model discussed above. This did not affect the PC's selected and the resulting FRO equations remained the same.

Mean prediction errors for the different models for each IO are shown in Table 2 below. In addition, the number of PC terms included in each equation is also shown. The cross-validation distributions for all three models for the IO's can be found in Appendix Q.

Table 2: Mean Errors and Standard Deviations for Interview Outcomes

Outcome	Cross-validation Mean Error (19 Observations)			Standard Deviation (Cross-validation)			"Leave 17% Out" Mean Error (3 Observations)			Number of PCs in Function		
	ALL	RAF	RBF	ALL	RAF	RBF	ALL	RAF	RBF	ALL	RAF	RBF
IO1	0.38	0.41	0.48	0.34	0.43	0.21				8	7	7
IO2	0.39	0.84	0.81	0.36	0.79	0.38				6	5	2
IO3	0.35	0.35	0.35	0.26	0.26	0.26				8	8	8
IO4	0.80	0.80	0.80	0.70	0.70	0.70				2	2	2
IO5	0.49	0.83	0.77	0.27	0.74	0.69				7	6	3
IO6	0.56	0.56	0.42	0.38	0.38	0.31				6	6	8
IO7	0.44	1.31	0.25	0.51	1.52	0.18				6	5	10
Averages	**0.47**	**0.73**	**0.55**	**0.40**	**0.69**	**0.39**	**-----**	**-----**	**-----**	**6.1**	**5.6**	**5.7**
Std. Dev.	**0.16**	**0.33**	**0.23**				**-----**	**-----**	**-----**	**2.04**	**1.90**	**3.30**

The ALL's average mean error for all IO's is 0.47; equal to its average for all FRO's mean errors. Since the IO data set is even smaller than the FRO's, this accuracy level should be taken lightly. The RBF model has produced more accurate predictions for the IO's, but is again outperformed by the ALL model. The standard deviations for these models has shown that both the ALL and RBF models are more reliable than the RAF model. The difference between the ALL and RBF reliabilities is negligible, with the RBF model having the slight advantage. The accuracy and reliability of the IO predictions are on par with those of the FRO's, indicating that IO's and FRO's are similarly predictable. As mentioned with the FRO's, certain IO's show stronger correlations. IO1 has shown to be one of the easiest to predict and IO4 the most difficult. This trend matches what is shown by the FRO's.

5.4.3 The Best Predictive Model

Of the three models we have validated, the ALL model is the most accurate for our both FRO and IO data sets. The cross-validation mean error for both was cases was 0.47 while the second best was IO RBF at 0.55. This shows that regardless of the data set the ALL model has the greatest predictive power. We also looked at the reliability of our models, which is represented by the standard deviation. We have provided both the standard deviations for the absolute predicted errors (the third column) as well as the standard deviation of the mean predicted errors (the bold row below the averages). The first allows us to determine how reliably each model can predict individual outcomes. The second standard deviation determines the reliability of each model across multiple outcomes. The RBF model showed an average standard deviation of 0.39 (for individual outcomes) when used with the IO data. The ALL model however, showed the best average standard deviation for FRO's with 0.35, making it more reliable for that set. Additionally, the ALL model is marginally less reliable than the RBF for the IO outcomes by this average measure. Looking at the standard deviation of the mean errors for all seven outcomes (the bottom row), the ALL model performed marginally better than the other two for the FRO's (0.19 versus 0.21 and 0.20 for RAF and RBF respectively). For the IO's, the ALL model performed substantially better (0.16 versus 0.33 and 0.23). The ALL model proved to be the most accurate and reliable overall and for those reasons we chose it to run our tool.

The only problem with the ALL model is its non-explanatory nature. We had hoped that the RAF or RBF models would help explain the negative values and overly influential candidates. The RBF model however did not offer any clarification and was validated to be less accurate and less reliable than the ALL model. Once more awards are scored and included in the model fitting, the ALL model will likely become more explanatory. Unfortunately we have included all awards available to us and have no additional data to use at this time.

5.4.4 Statistically Important Predictors

After fitting the model for both sets of seven outcomes (FRO and IO), it has been shown that 46 of the 52 PC's made it into the final equations. This indicates that the majority of our PC's determined from the PO interviews have predictive value. The seven FRO equations include 42 of the 52 PC's whereas the IO equations include 30. This shows that certain PC's have predictive value for the IO's but not the FRO's, and vice versa.

In examining the FRO equations, certain trends were apparent. "Panel Score" (PC1), "Attention to Diversity" (PC8), "Attention to User Interface" (PC13), and "Community Building" (PC39) all appeared in four of the seven equations. "Faculty Workshops" (PC35) showed even broader predictive ability by appearing in five equations. However, this does not explain the importance of these characteristics in determining success as the same PC is shown to have a positive coefficient for one outcome and negative for another. These same trends are not readily apparent in the IO's largely due to the small number of PC that were included in the final model. Only "Clearly Defined Goals and Objectives" (PC4) and "Institutional Support" (PC52) appear the maximum of three times in the seven IO equations. Table 3 shows the PC's that appear in the final equations. The rubrics for these PC's can be found in Appendix G.

Table 3: Predictor Candidates Appearing in the Final Model

	Predictor Candidates		
PC1	Panel Score (V)		**Dissemination Plan**
	Project Description	PC33	Textbook/Software Bundle Contract
		PC34	Software Commercialization
PC2	Based on existing Template/Prototype	PC35	Faculty Workshops
PC3	Successful Implementation of Own Prototype	PC36	National Conferences
PC4	Clearly Defined Goals and Objectives (B)	PC37	Local Conferences/Meetings/Presentations
PC5	Detailed Methodology (B)	PC38	Websites
		PC39	Community Building (B)
PC7	Addresses Project Concerns (B)		
PC8	Attention to Diversity	PC41	CD-ROM
PC9	Collaboration of Other Institutions (B)	PC42	Journal Articles/Publications (B)
PC10	Collaboration with Other Experts (B)	PC43	Process Sharing (B)
PC11	Multi-platform or Considerations (B)		
PC12	Open Source (B)		**PI Information**
PC13	Attention to User Interface		
PC14	Level of Interactivity	PC44	Prior Experience as a PI (B)
PC15	Design Allows for Customizability	PC45	Years as Higher Education Professor (V)
PC16	Project URL Given in Proposal (B)		
PC17	Sustaining Plan - Funds	PC47	Project Management Experience (B)
PC18	Sustaining Plan - Maintenance	PC48	Light Work Load (B)
PC19	Timeline		
			Budgetary Information
	Assessment/Evaluation Plan		
		PC49	Detailed Justifications
PC20	Formative for Process		
PC21	Formative for Product		
PC22	Summative for Process (B)		**Facilities and Support**
		PC51	Institutional Support
PC25	User Observation	PC52	Description of Network & Computers (B)
PC26	Qualitative Student/Faculty Interviews		
PC27	Attitude Surveys/Questionnaires		
PC28	Group Discussion/Focus Groups	**(B)**	**Binary Predictor Candidate**
PC29	Pre- and Post- Testing		***All Other Predictor Candidates Tertiary**
PC30	Control Groups	**(V)**	**Quantitative Value**
PC31	Diversity Testing		
PC32	Assessment Timeline (B)		

5.5 Predictive Tool

As a result of fitting our ALL model, a function was produced for each outcome. For a given outcome, a predicted outcome is calculated from the appropriate function using the predictor scores for that proposal. The input interface for the program can be seen in Figure 6. Tool-tips provide the description and rubric for each PC. Additionally, scores can be saved and opened again later. The “Submit Scores” button saves the proposal scores to a text file, which can be used later to supplement the data set once actual outcomes are available (if the proposal is funded). A screenshot of the predictor candidate input can be seen in the screenshot below. Although six PC’s did not appear in any of the FRO or IO equations, we have included these PC’s in the tool. As more data become available, it is possible that these PC’s will also have predictive ability.

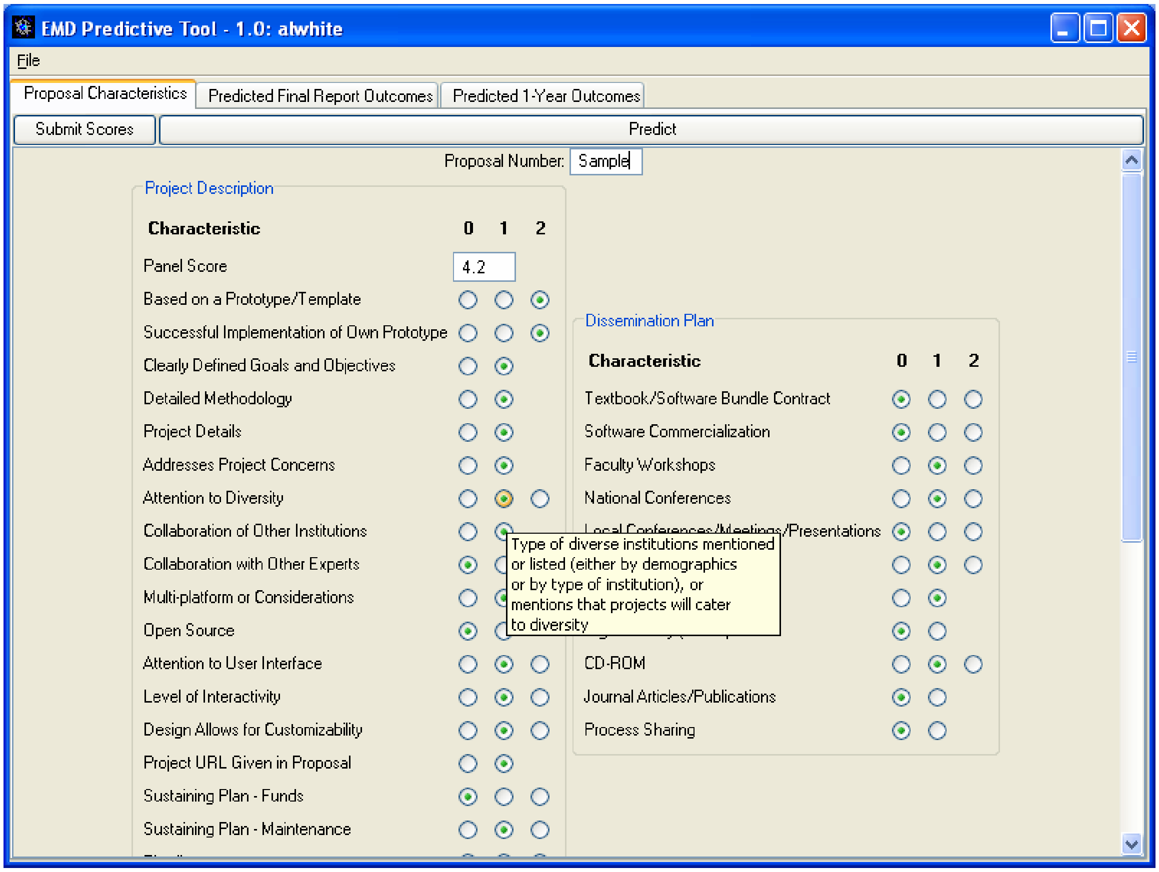


Figure 6: Screenshot of the Software Tool Input

Once the PO hits the "Predict" button a set of graphs are shown, one for each outcome. Each bar shows the value that is to be obtained for that specific outcome, as predicted by our model. It is important to note that outputs of our model are represented as continuous values and not discrete 1 to 5 values. While we could have simply rounded the predicted outcomes, we felt that the information lost was not worth the more intuitive representation. An example output for the interview outcomes (1-Year Outcomes) can be seen in Figure 7. The final report outcomes are displayed in a similar manner on the other program tab. In addition to these basic predictions, the program is able to auto-update itself as newer versions or models become available. It is important that the program use the latest and most accurate predictive model available.

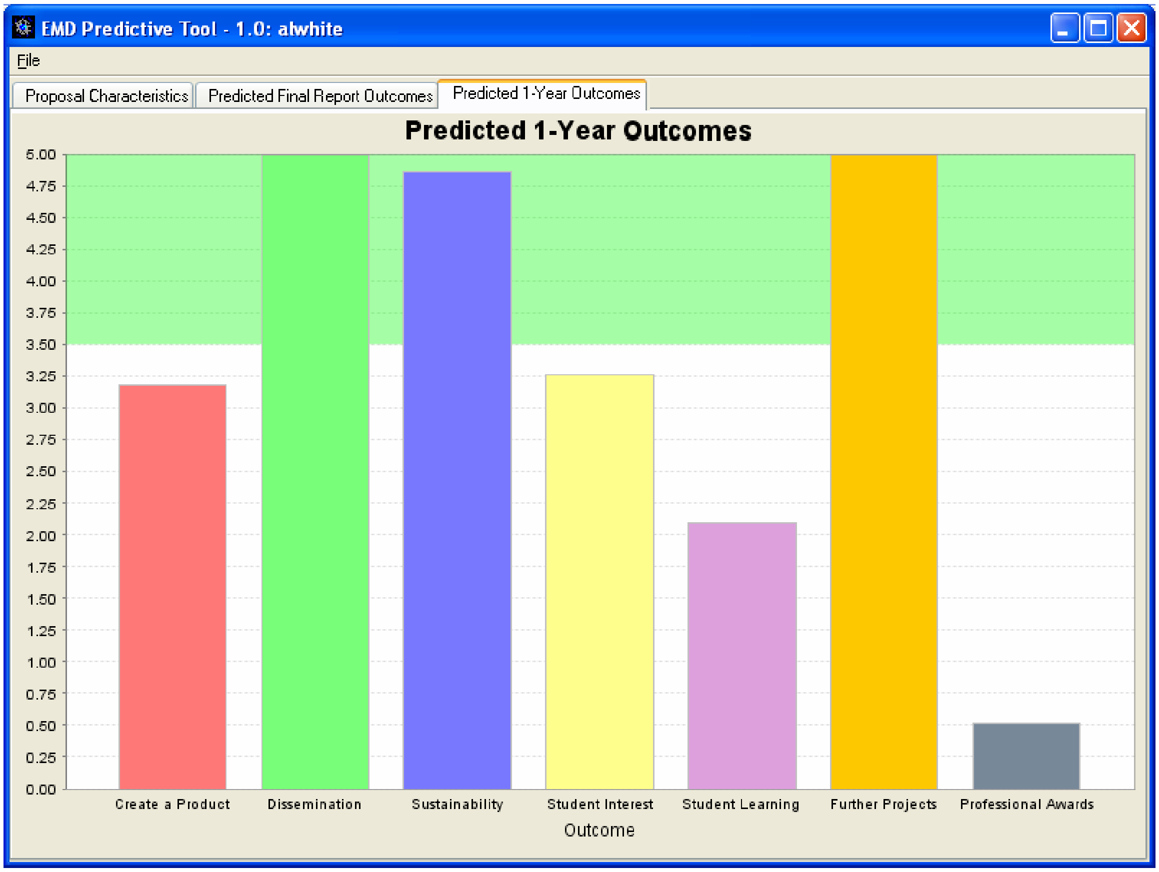


Figure 7: Screenshot of the Software Tool Output

6 Conclusions

The analysis of our results has led us to two major conclusions. The first conclusion is that award outcomes can be predicted from proposal characteristics. This has been confirmed by our model's valid predictive abilities. Our model is capable of predicting final report outcomes (FRO's) for a software EMD award to an average cross-validation error of 0.47 on our 1 to 5 scale. For interview outcomes (IO's) our model predicts with the same average error of 0.47. We have also discovered that some outcomes are more capable of being predicted than others. Both "Product" (FRO1) and "Dissemination" (FRO2) predict very well with mean errors of 0.19 and 0.35 respectively. Conversely, both "Student Interest" (FRO4) and "Student Learning" (FRO5) predict with mean errors greater than 0.65. While this does not indicate that out model predicts these outcomes poorly, it does show that certain outcomes are more strongly correlated to the proposal characteristics we have studied. Similar results are apparent in the IO predictions. While "Professional Awards" (FRO7) predicts with an average error of 0.33, this accuracy is taken lightly as 34 awards received the lowest score and 2 received the highest. On the other hand, IO7 has a much better distribution of scores and predicts with a mean error of 0.44. Our analysis shows that all FRO's and IO's are acceptable for prediction use, with the exception of FRO7. Our software tool does provide a prediction for this outcome but we would not trust its results until the data set the model is based on grows larger.

It is important to note, however, that our results are based on a very specific set of awards (full-scale development EMD awards that developed software) and therefore cannot be expected to apply to other situations such as A&I or non-software proposals. Furthermore, our FRO predictions are based on 36 observations and is not a sufficiently large data set to conclusively confirm our accuracy. The IO predictions are based on 19 observations and these predictions should be taken with less confidence. The average standard deviations indicate that the final report outcome predictions are more reliable, as a whole, than those made for the interview outcomes. Interview outcomes, however, represent a truer assessment of an award's accomplishments than the final reports do. Sometimes final reports are filed before the PI has finished disseminating or applied for a new grant. Other times the final report is short and doesn't portray the actual accomplishments of the award. We feel for this reason that even though we have less IO data, these results are just as important those produced by FRO data.

While we are confident in the model's accuracy over our data set, predicted values have shown errors that are 2.5 off from the actual value. This supports the conclusion that predicting success is hardly a simple, straightforward process. While this is difficult, our model has shown that useful predictions are possible. Incorporating this model into a usable tool has provided an additional consideration for the DUE PO's to use when reviewing proposals.

The second conclusion that we've reached is that a predictive tool of this nature is worth pursuing. This fact is almost as important as the model itself. This year, 2004, Congress has reduced funding to the NSF (Pear, 2004). Additionally, recent studies have shown that U.S. students have been surpassed by other nations in the fields of science and mathematics (Grimm, 2004), a key focus of the DUE. It is especially important that the resources available to the NSF be distributed efficiently. Our research has shown that proposal evaluation is a time-consuming and difficult process. By having a predictive tool at a PO's disposal an additional dimension for proposal review is provided that can be taken into consideration when making a funding recommendation. We strongly believe that this predictive research be continued. Additional development and research could make this tool more accurate and better able to predict the outcomes of these proposals. Through further development it is also possible to produce more sophisticated models for predicting other types of proposals, rather than exclusively software. We have been a pilot group mainly to study the viability of making such predictions. Our resulting tool has become a prototype and the NSF now has the option of pursuing its development. By thoroughly documenting our process, others can follow similar steps and expand upon our research. Our recommendations for additional research and the use of our software tool are discussed in the following section.

7 Recommendations

In this age of increasing applications of technology, more and more DUE projects will propose the development of some sort of software. We recommend that our model, through our software tool, be applied as necessary to these proposals. Furthermore, if the NSF continues to sponsor this research a robust and all-inclusive tool can be created. We recommend that either our model is maintained over the years or that a new model be developed to serve the same purpose. Finally, during our research we ran into a few challenges. We make a few recommendations to the NSF and to anyone who builds on our research so that this process will be easier in the future.

7.1 Recommendations for the NSF

Recommendation 1: That the NSF consider using the software tool's predictions to assist in proposal evaluation.

The data in our results section show that certain predictive characteristics do exist within project proposals of this type. Our model is non-explanatory, however, so it is not possible to see which characteristics are most important. This limits the use of these relationships to within the tool we developed. We recommend that this tool be used as an added dimension for proposal evaluation. Our experience has shown that scoring proposals takes little effort beyond reading them. The prediction can be taken with moderate confidence as our validation has shown that the average predicted error is 0.47 on a 1 to 5 scale. On occasion, however, our model has been known to predict with an error up to 2.

Recommendation 2: That the NSF use this tool to record proposal scores.

The model would have the potential to be far more accurate with an increased data set. We strongly recommend that the software tool be used to record the scores given to proposals. Once the funded proposals are completed, the project outcomes can be scored. These outcome scores, coupled with the predictor candidate scores from the software tool, would provide additional observations to our data set. By refitting our documented model, accuracy could be improved at marginal time and effort.

Recommendation 3: That the NSF consider modifying its final report criteria to include "Plans for the Future."

We found that scoring proposals by the predictor candidates was a straightforward task. They were all very descriptive and specific in the proposal sections. On the other hand, we noticed it was difficult to accurately gauge the seven outcomes for several awards in our pool. This was due to short nondescript final reports. In the event that no interview was scheduled and no useful information was found online, the award received minimal scores. It is quite possible that some of our scores do not reflect reality in terms of outcomes. While that is unfortunate, it is important to note that this information is all that is available to PO's at the time an award is closed.

The outcomes we used for our analysis were given to us through PO interviews and thus are known to be desired by PO's. Specifically, important to all software awards is the notion of sustainability, which is not asked for in the final report documentation. Furthermore, scores for further research tended to spike drastically upwards during the interviews showing this topic is underrepresented in the final report. It might be useful for the NSF to consider modifying its final report criteria to include “Plans for the Future.” Such a category would help PO’s evaluate an award more fully from its final report.

7.2 Recommendations for the Software Tool

Recommendation 4: That the software tool be modified to automatically refit the model when new data are available.

The software tool developed has been designed in a robust manner so that changes can be incorporated easily. To simplify the refitting of the model when additional data are added, an automated process could be employed. This would reduce the amount of maintenance needed to a bare minimum while providing models that are always based upon the latest data available. This can be done through the open source statistical program R. With the help of scripting and macros, the model could be refitted and validated automatically.

Recommendation 5: That the software tool’s usability be evaluated by program officers.

We recommend that the software tool be provided to at least four PO’s and ask them to evaluate the proposal of one award from our pool. Using our tool, with the detailed descriptions and rubric of the predictor candidates, they would score the proposal and have the tool predict its outcomes. After this testing, we recommend that a focus group be held with the PO’s, discussing the tool and its use. This testing would ensure that the software tool caters to the needs of the PO's.

7.3 Recommendations for Future Projects

Recommendation 6: That future project groups solicit the advice and feedback of the program officers early and often.

Our project is intended to be a pilot and that if the results look promising, future projects are likely to expand upon our model. Our first recommendation is that groups solicit the advice and feedback of the program officers early and often. We had originally planned to present our preliminary list of predictor candidates and outcomes to the PO's during one of their weekly staff meetings. Due to the scheduling of these meetings and the timing of our own project, however, this did not occur.

Recommendation 7: That future project groups work with the NSF to obtain sample awards prior to arriving in Washington.

Another recommendation is that future groups work with the NSF to obtain awards, which will not be in the award pool, prior to arriving in Washington. Examples of these could be projects that have been awarded but not yet completed. This would allow the group to become familiar with the layout of a proposal and gain an idea of what is typical early on, while maintaining the size of the data set available for modeling. It would also be possible to conduct content analysis on these awards to identify common characteristics. These could be used to supplement characteristics given by PO's. As we scored the awards, ideas of potentially viable characteristics emerged too late to be incorporated into our study.

Recommendation 8: That inter-rater reliability be obtained early.

Finally we highly recommend that inter-rater reliability be obtained early. It is extremely important and also helpful in that each member may not have the time to read every award. We did not schedule time specifically for this task and ended up rearranging our deadlines. Planning for these test runs will allow more efficient use of time as well as help prevent the problem of differing results. These tests also help ensure that the rubric is comprehensive and well defined.

Recommendation 9: That our model be further validated and other models also explored.

We also propose the following improvements to be considered. The multiple linear regression model used could be further validated. Other models could also be looked into as they may be better able to predict outcomes. For example a multinomial model can predict discrete values for each outcome. In our rubric, a higher score equates to a better outcome of that type. To take advantage of this ordering, it is possible to use a type of ordinal multinomial model. Our outcomes can be affected by any combination of predictors, as they are not likely to be independent of one another. We had looked into this model extensively and had hoped to use it for our tool, however time constraints prevented us from validating this model. The SAS scripts we used as well as a write up for multinomial models by Professor Petruccelli have been provided in a supplementary CD-ROM.

It may be possible to find a model which would serve as explanative as well as predictive. Our validated model, multiple linear regression, works as a predictive tool but is not explanative. For example, in the equation for FRO6, PC1 (panel review score) has a negative coefficient. This would seem to imply that a higher score is detrimental to further research, which does not make sense. Our small award pool contributes to this issue.

References

Executive Office of the President. (2004.) *Department of Education.* Retrieved November 9, 2004, from The Executive Office of the President, Office of Management and Budget web site: http://www.whitehouse.gov/omb/budget/fy2005/pdf/budget/education.pdf

Ford Foundation. (2004). *Mission Statement*. Retrieved October 2, 2004, from http://www.fordfound.org/about/mission.cfm

Grimm, David. (2004, December 7.) U.S. Students Trail Peers in Math and Science. *Science Now*. Retrieved December 14, 2004 from http://sciencenow.sciencemag.org/cgi/content/full/2004/1207/1

Kulik, J. (1994). Meta-analytic studies of findings on computer-based instruction. In Baker, E. L. and O'Neil, H. F. Jr. (Eds.), *Technology assessment in education and training.* (pp. 9-33) Hillsdale, NJ: Lawrence Erlbaum.

National Institutes of Health. (2004). *Mission Statement*. Retrieved October 2, 2004, from http://www.nih.gov/about/almanac/index.html

National Science Foundation. (2003). *NSF by the Numbers.* Retrieved October 11, 2004, from http://www.nsf.gov/od/lpa/nsf50/nsfnumbers.htm

National Science Foundation. (2004a). *Course, Curriculum, and Laboratory Improvement (CCLI) Program*. Retrieved September 26, 2004, from National Science Foundation Division of Undergraduate Education Web site: http://www.ehr.nsf.gov/ehr/DUE/programs/ccli/

National Science Foundation. (2004b). *Course, Curriculum, and Laboratory Improvement Program Solicitation NSF 04-565.* Arlington, VA: Author.

National Science Foundation. (2004c). *DUE's Mission.* Retrieved October 2, 2004, from National Science Foundation Department of Undergraduate Education Web site: http://www.ehr.nsf.gov/ehr/due/about/mission.asp

National Science Foundation. (2004d). *A Guide for Proposal Writing.* Arlington, VA: Author.

National Science Foundation. (2004e). *FY 2003 Report on the NSF Merit Review System NSB 04-43.* Arlington, VA: Author.

National Science Foundation. (2004f). *National Science Foundation (NSF)*. Retrieved September 26, 2004.

National Science Foundation. (2004g). *NSF - Division of Undergraduate Education*. Retrieved September 26, 2004 from http://www.ehr.nsf.gov/due/

National Science Foundation. (2004h). *NSF – Organizational Chart.* Retrieved September 26, 2004 from http://www.nsf.gov/home/chart/orgchart.htm

Pear, Robert. (2004, November 30.) Congress Trims Money for Science Agency. *The New York Times.* Retrieved December 12, 2004 from http://www.nytimes.com

SAS Institute Inc. (2004.) *SAS/STAT 9.1 User's Guide.* Retrieved November 29, 2004 from http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_ug_7313.pdf

Seals, R. K. (2003). *Course, Curriculum and Laboratory Improvement Funding Opportunities at NSF*. Retrieved September 26, 2004, from University of Minnesota, Electrical and Computer Engineering Department Web site: www.ece.umn.edu/groups/PowerElectronics_Drives/workshop2003/workshop_jul03/seals.pdf

Schaffter, D. (1969). *The National Science Foundation.* New York: Fredrick A. Praeger, Inc.

U.S. Department of Education. (2004). *Overview.* Retrieved October 2, 2004, from: http://www.ed.gov/fund/landing.jhtml?src=rt

Appendix A: Sponsor Description

The National Science Foundation (NSF) is a government agency that was created by the National Science Foundation Act of 1950 (Schaffter, 1969, p. 3). The NSF is a very large organization. It is divided into the following seven directorates and three offices under the Director of the NSF: Office of the Director, Office of Budget, Finance and Management, Office of Information and Resource Management, Directorate for Biological Sciences, Directorate for Computer and Information Sciences and Engineering, Directorate for Engineering, Directorate for Mathematical and Physical Sciences, Directorate for Geosciences, Directorate for Social, Behavioral, and Economic Sciences, and Directorate for Education and Human Resources (NSF, 2004h, Organizational Chart). Each of these is divided into a number of parts, as shown on the following chart.

We are working with the Course, Curriculum, and Laboratory Improvement (CCLI) Program. The CCLI Program is part of the Division of Undergraduate Education (DUE), which is under the Directorate for Education and Human Resources (EHR). The primary goal of this program is to "support efforts in colleges and universities to develop the capacity to meet the learning needs of all undergraduate students in [science, technology, engineering and mathematics]" (NSF, 2004a, Overview). Specifically, the CCLI program is geared toward the activities affecting learning environments, course content, curricula, and educational practices. A visual representation of the organization of the DUE is on the second chart.

There are four tracks within the CCLI program: Educational Materials Development (EMD), Adaptation and Implementation (A&I), National Dissemination (ND) and Assessment of Student Achievement (ASA) (NSF, 2004b). The ASA track deals with the assessment of student achievement, including research on assessment and the development of assessment tools and practices. The ND track is aimed at providing institutional faculty with information about new materials and processes that further science, technology, and mathematics (STEM) education. The EMD track promotes the development of new educational materials for national dissemination, whereas the A&I track supports the adaptation of existing successful projects. Figure 8 below shows the relative nature of the EMD and A&I tracks in terms of innovation/adoption and the target audience. We will further narrow our focus and concentrate on the EMD track within the CCLI. In a 2004 CCLI solicitation (NSF, 2004a), the program estimates that 250 awards will be given by the program. Of these 250 awards, 115 are estimated to be given to projects from the EMD track.

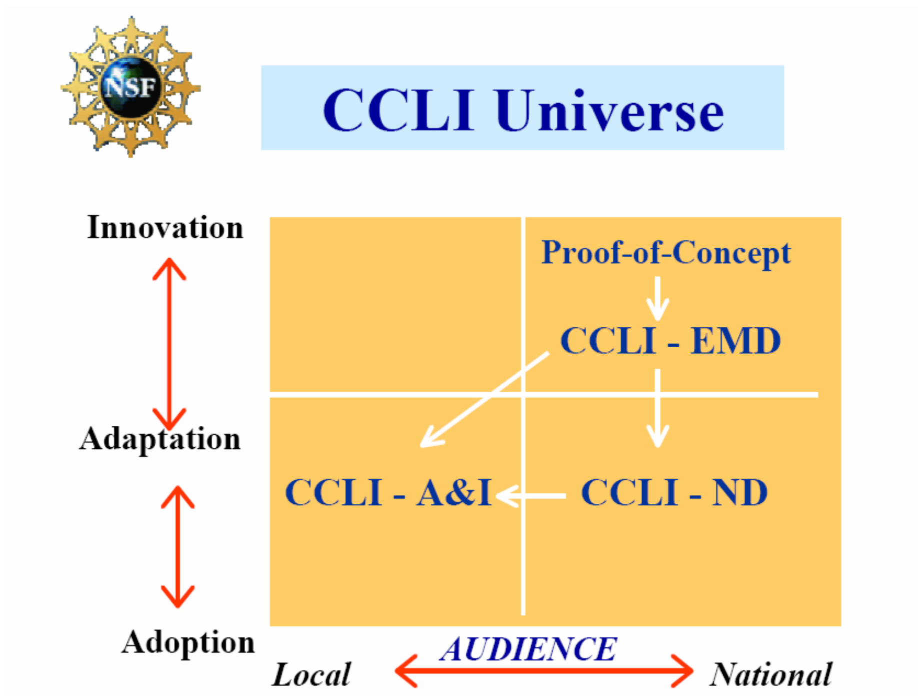


Figure 8: Relative nature of the CCLI tracks
(Source: Seals, 2003, CCLI Universe)

EMD projects are at the forefront of educational techniques. The projects are innovative and hope to discover new ways to approach educational material. One such method is through educational software. James Kulik spent more than a decade studying the effects of using computers for instruction (Kulik, 1994, pp. 9-33). He drew the conclusion that providing a solid interaction layer between the students and materials allows students to learn the material better and with less time. It is due to this innovation that the NSF puts a strong emphasis on the EMD track.

Part of the National Science Foundation Act of 1950 was a provision for an annual appropriation of funds from Congress (Schaffter, 1969, pp. 43-45). The current budget is approximately four billion dollars (NSF, 2003). Of this, the CCLI Program receives approximately $40 million dollars annually (Seals, 2003, Approximate FY 04 Funding). A breakdown of DUE funding can be seen in Figure 9 below. The CCLI dedicates approximately $18 million in funds to EMD projects.

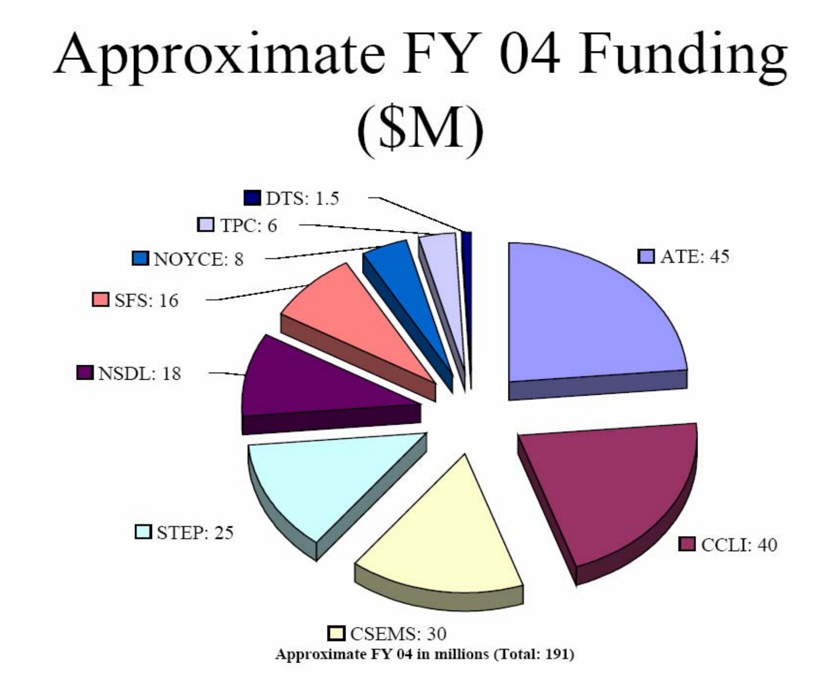


Figure 9: DUE FY 2004 Funding
(Source: Seals, 2003, Approximate FY 04 Funding)

The following pages show the original letter from NSF identifying our project topic as well as organizational charts of the NSF and the DUE.

<u>National Science Foundation</u>

Division of Undergraduate Education

Curriculum Grant Success Analysis

The mission of NSF's Division of Undergraduate Education (DUE) is to promote excellence in undergraduate science, technology, engineering, and mathematics (STEM) education for all students. The division accomplishes its mission through several strategies including supporting curriculum development, stimulating and funding research on learning, and promoting development of exemplary materials and strategies for education. The primary mechanism is the funding of instructional development projects at colleges and universities throughout the US.

An important instructional development area within DUE is the Course, Curriculum, and Laboratory Improvement (CCLI) program. This program seeks to improve the quality of STEM education for all students, based on educational research and empirical data concerning needs and opportunities in undergraduate education and effective ways to address them. The program targets activities affecting learning environments, course content, curricula, and educational practices. It aims to improve learning that contributes to the knowledge base supporting future efforts to enhance STEM education.

The CCLI program funds many projects that do one of the following: (1) purchase instruments and equipment, (2) develop software, or (3) develop laboratory materials. Resources could be much more efficiently directed if the division better understood what characterizes a potentially successful project. The WPI project would develop a "success predictor" that would use qualitative and quantitative project characteristics that could be evaluated at the start of the project to predict several outcomes. Project characteristics would be features of the investigators, the context where the project will be implemented (i. e., the department, the college, and the institution), as well as those capturing the nature of the project itself as described in the proposal. The outcomes would be observed at project completion and describe how well the project achieved its goals and those of the CCLI program. The student team, working with NSF personnel, would select one of the three types of projects, identify a suitable set of project characteristics, establish a set of appropriate outcomes, select a predictor model from the published literature, and use data from a set of projects to parameterize and validate the model. The model could be pilot tested on a different set of projects and refined during follow-on projects.

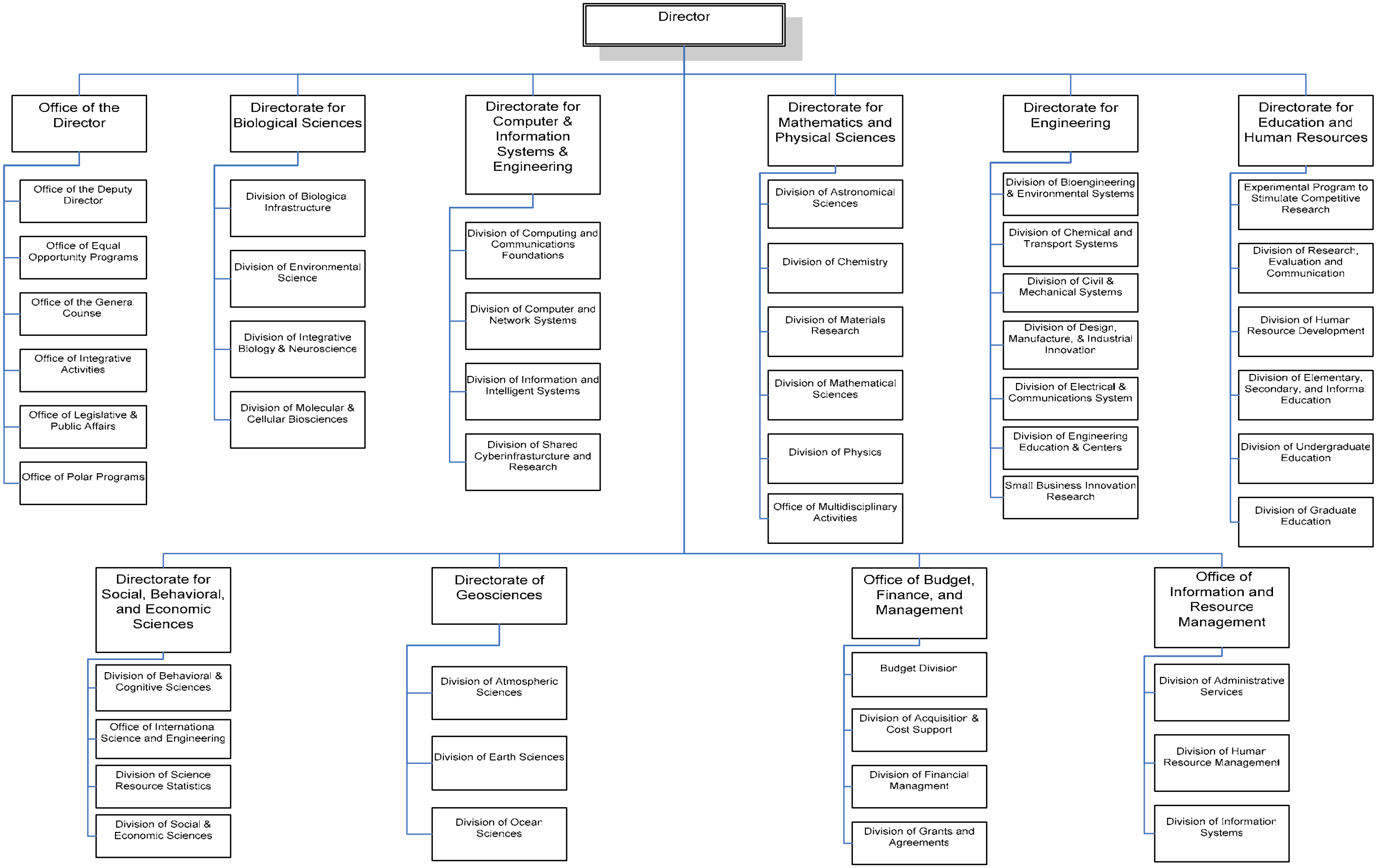


Figure 10: Organizational Chart of the NSF
(Source: NSF, 2004f)

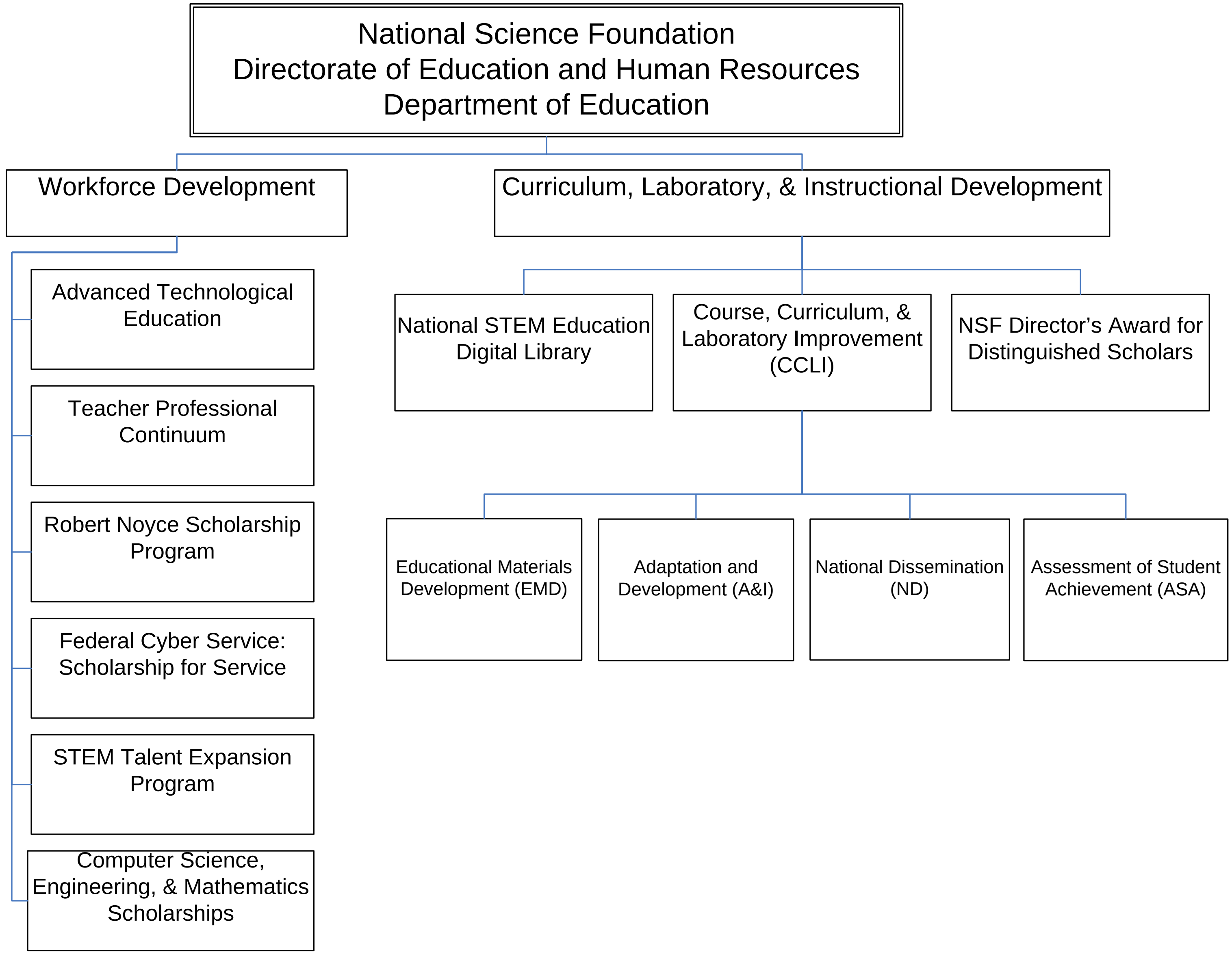


Figure 11: Organizational Chart of the DUE
(Source: NSF, 2004g)

Appendix B: EMD Full-Scale Checklist

What to Look for in Proposals, & to Address in PD's Review Analysis

Item	Comment
I. Development and Testing	
A. Development of a substantive <u>product</u> is planned, based on prior experience with a prototype?	
B. Product is based on sound, <u>effective pedagogy</u>?	
C. Plan to <u>assess</u> effectiveness of product is credible?	
D. Plan to <u>pilot test</u> at developer's institution?	
E. Plan to prepare <u>documentation</u> so others can test product?	
F. Plan to <u>beta test</u> at <u>diverse</u> types of institutions and with diverse student populations? (if diversity missing, it should be negotiated in and funds may be added to cover cost)	
G. The PI has a plan to contact commercial publishers, and has been told by DUE PD to provide documentation of contact with publishers within 6 months of receiving the award; or describes other credible plans for self-sustained distribution of product.	
II. Dissemination	
A. Plan to <u>publish about</u> the development?	
B. Plan to <u>orally present</u> about the development?	
III. Faculty Development	
A. Plan to <u>prepare faculty</u> at test sites and other potential adapters to use product? Supplemental funding for more extensive faculty development may be requested in later stages of project, to make more widespread the implementation of the developed materials.	
IV. Cost Effectiveness	
A. What: Which tasks contribute to achievement of objectives? B. How: Time devoted to tasks reasonable? C. Who: Appropriate people doing tasks? D. How much: Cost of doing task reasonable?	

Appendix C: PO Interview Protocol

General:

How many years have you worked with the NSF?

What specific discipline to you work with? Do you handle any interdisciplinary panels?

Predictors:

What characteristics do you look for in the following sections? What criteria do you use to evaluate the following sections?

Review Panel Comments

Project Proposals

Biographical Sketches

Facilities, Equipment, and Other Resources (Context)

Project Summary, References Cited, Current and Pending Support

What do you consider as a good Evaluation Plan? Can you give an example?

What do you consider as a good Assessment Plan? Can you give an example?

What do you consider as a good Dissemination Plan? Can you give an example?

Outcomes:

What outcomes do you expect to see from all of the DUE's awards? Software specific?

Where would you gather evidence of these successes?

What do you consider to be successful dissemination?

Appendix D: PI Interview Protocol

Product:

Is the final product being implemented/used in the manner described in your proposal?

Are you still using this program/product? Is the class still being taught?

How did you evaluate your award?

Did you accomplish your goals and objectives?

Sustainability:

Do you still maintain this product? Bug fixes, support, upgrades?

If so, who supports this maintenance?

If there is a website, is it maintained regularly?

Student Interest:

Has enrollment increased as a result of using your program/product?

Has it impacted the retention rate of students?

How do students tend to feel about the product?

Student Learning:

Did your curriculum affect student learning in any noticeable way?

Did grades increase on average?

Did understanding of the material increase on average?

Has your program changed the way professors teach this course?

Dissemination:

Did you hold any workshops? How many were in attendance? How long did your workshops last?

How many professors use your program at your institution? Do you know of any other professors using it at other institutions?

What types of active dissemination (face-to-face, conferences, workshops, speeches) did you employ?

What types of passive dissemination (literature, textbooks, journals, CD-ROMs, websites) did you employ?

Further Development:

Have you received additional support because of or based on this project?

Has anybody seriously approached you about adapting your product/process for another institution?

Professional Nominations and Awards:

Was your project nominated for or did it receive a professional award?

Appendix E: Award Pool

Award	Project Title	PI Last Name	Institution	Effective	Expired
0088304	PsychExperiments: Expanding and Training the User-Developer Community	McGraw	University of Mississippi	04/15/01	03/31/03
0088657	Collaborative Research--Visualizing and Exploring United States Urban and Rural Social Change, 1790-2000: Interactive Multimedia and Web Based Tools	Halle	University of California-Los Angeles	05/01/01	04/30/03
0088704	Collaborative Research -- Visualizing and Exploring United States Urban and Rural Social Change, 1790 - 2000 -- Interactive Multimedia and Web Based Tools	Beveridge	CUNY Queens College	05/01/01	04/30/04
0088709	Developing a Technology Enhanced Guided Inquiry Workbook for General Chemistry	Greenbowe	Iowa State University	03/01/01	02/29/04
0088758	Guided Discovery and Intelligent Tutoring Materials for Calculus and their Electronic Delivery on the World Wide Web	Eisenberg	University of Massachusetts Amherst	01/01/01	12/31/03
9950239	ANT: A Coherent Framework for Computer Science Education	Seltzer	Harvard University	07/01/99	06/30/03
9950301	Industrial Systems Design and Analysis: High-Fidelity Learning Environments for Engineering Education	McGinnis	Georgia Tech Research Corporation - GA Institute of Technology	07/01/99	06/30/03
9950356	Science, Mathematics, Engineering and Technology Learning Modules for an Electronics Curriculum	Millard	Rensselaer Polytechnic Institute	07/01/99	12/31/01
9950480	Interactive Spatial Science: Multimedia Geography Education	Wisniewski	Cambridge Studios Inc.	10/01/99	03/31/02
9950506	Development of Experential Learning Modules for Environment Systems Analysis	Paterson	Michigan Technological University	09/01/99	12/31/03
9950566	Project CLEA: Contemporary Laboratory Experiences in Astronomy	Marschall	Gettysburg College	08/01/99	01/31/02
9950567	An Internet Based System for Mathematics Homework Problems	Pizer	University of Rochester	07/15/99	06/30/03
9950568	Interactive Software to Improve Student Success in Developmental Mathematics	Manville	Bunker Hill Community College	08/15/99	07/31/02
9950603	Neurobiology Software Package for Teaching Through Interactive Laboratory	Moody	University of Washington	07/15/99	06/30/02
9950612	LUCID-A New Model for Computer-Assisted Instruction in Chemistry	Hanson	SUNY at Stony Brook	09/01/99	08/31/03
9950613	BIOSCAPES: Biological Informatics Office-Studies of Complex Anatomy and Physiology within Education Simulations	Tashiro	Northern Arizona University	09/01/99	08/31/01

9950673	A Flexible Networked Laboratory Simulation for Use in Introductory Chemistry Courses	Yaron	Carnegie-Mellon University	09/01/99	12/31/02
9950689	Biology Student Workbench: Inquiry Tools for the Use of Molecular Data in Undergraduate Biology	Jakobsson	University of Illinois at Urbana-Champaign	07/15/99	06/30/02
9950714	Mathematical Activities Using JavaBeans	Yanik	Emporia State University	07/01/99	06/30/02
9950740	BIRDD Digital Library: Enhancing Evolution Education	Jungck	Beloit College	07/01/99	12/31/01
9950746	An Animation-based Tutor for Algebra-Word Problems	Reed	San Diego State University Foundation	07/01/99	06/30/03
9950762	Improving Learning in Undergraduate Engineering: The Homework Laboratory	Henderson	Tennessee Technological University	07/15/99	12/31/02
9950829	Computer Science Laboratory Projects: Breadth Through Depth	Rasala	Northeastern University	07/01/99	12/31/03
9950906	Development of Virtual Soil Laboratory Tests Courseware	Budhu	University of Arizona	07/15/99	06/30/02
9950948	Effective Internet Education for Everyone	Kurtz	Appalachian State University	09/01/00	05/31/03
9952246	Strengthening Undergraduate Education through Research in Radio Astronomy	Salah	Northeast Radio Observatory Corp	05/01/00	04/30/03
9952509	Concurrent Computing in an Upper-Level Computer Science Curriculum	Carr	Michigan Technological University	06/01/00	06/30/03
9952525	Life Lines Online: Accessible Investigative Biology for Community Colleges	Waterman	Southeast Missouri State University	01/01/00	12/31/03
9952540	Hardware-Software Co-Design in an Undergraduate Microcontroller Laboratory	Beetner	University of Missouri-Rolla	01/15/00	12/31/03
9952630	Courses and Capabilities for Asynchronous Learning in Engineering	Culver	SUNY at Binghamton	03/15/00	08/31/03
9952703	Intelligent Tutoring System for Visual Reasoning in Solid Modeling	McRoy	University of Wisconsin-Milwaukee	01/01/00	12/31/03
9972394	Improving the Linkage Between Applied Mechanics and Materials Science in the Engineering Curriculum: Model Curricula and Multi-Media Courseware	Jenkins	South Dakota School of Mines and Technology	07/15/99	06/30/02
9972437	STEP - A System for Teaching Experimental Psychology	MacWhinney	Carnegie-Mellon University	09/15/99	08/31/03
9972486	An Inquiry-Based Simulation Learning Environment for the Ecology of Forest Growth	Murray	Hampshire College	09/01/99	08/31/02
9980802	Creating a Community of Peer Instruction Users: Dissemination and Electronic Resources	Mazur	Harvard University	05/15/00	04/30/02
9980935	Student Authoring of Three-Dimensional Illustrations in Undergraduate Biochemistry	Bateman	University of Southern Mississippi	02/15/00	01/31/04

Appendix F: Predictor Candidate Score-Sheet

			Project:				**Reviewer:**
PC1			**Panel Score**				
	pg				pg		
			Project Description				**Dissemination Plan**
PC2			Based on existing Template/Prototype	PC33			Textbook/Software Bundle Contract
PC3			Successful Implementation of Own Prototype	PC34			Software Commercialization
PC4			Clearly Defined Goals and Objectives (B)	PC35			Faculty Workshops
PC5			Detailed Methodology (B)	PC36			National Conferences
PC6			Project Details (B)	PC37			Local Conferences/Meetings/Presentations
PC7			Addresses Project Concerns (B)	PC38			Websites
PC8			Attention to Diversity	PC39			Community Building (B)
PC9			Collaboration of Other Institutions (B)	PC40			Digital Library (NSDL) (B)
PC10			Collaboration with Other Experts (B)	PC41			CD-ROM
PC11			Multi-platform or Considerations (B)	PC42			Journal Articles/Publications (B)
PC12			Open Source (B)	PC43			Process Sharing (B)
PC13			Attention to User Interface				
PC14			Level of Interactivity				**Biographical Sketches**
PC15			Design Allows for Customizability				
PC16			Project URL Given in Proposal (B)	PC44			Prior Experience as a PI (B)
PC17			Sustaining Plan - Funds	PC45			Years as Higher Education Professor (V)
PC18			Sustaining Plan - Maintenance	PC46			Technical Expertise on Project (B)
PC19			Timeline	PC47			Project Management Experience (B)
				PC48			Light Work Load (B)
			Assessment/Evaluation Plan				
							Budget
PC20			Formative for Process				
PC21			Formative for Product	PC49			Detailed Justifications
PC22			Summative for Process (B)	PC50			Student Involvement
PC23			Summative for Product (B)				
PC24			Outside/Impartial Evaluation (B)				**Context**
PC25			User Observation				
PC26			Qualitative Student/Faculty Interviews	PC51			Institutional Support
PC27			Attitude Surveys/Questionnaires	PC52			Description of Network & Computers (B)
PC28			Group Discussion/Focus Groups				
PC29			Pre- and Post- Testing			**(B)**	**Binary Predictor Candidate**
PC30			Control Groups				***Other Predictor Candidates Tertiary**
PC31			Diversity Testing			**(V)**	**Quantitative Value**
PC32			Assessment Timeline (B)				

Appendix G: Predictor Candidate Rubrics

	Predictor Candidates	Scoring Rubric		
PC1	Panel Score	Average value of panel review scores	------------	------------
	Project Description	**0**	**1**	**2**
PC2	Based on existing Template/Prototype	No Prototype/template (POC project or other to work off of)	Prototype/template summarized/described	Prototype described in detail with impact and evaluation, problems will be addressed
PC3	Successful Implementation of Own Prototype	Not in use at home institution or elsewhere/Not tested	Prototype being used currently	Prototype tested/evaluated and being used currently
PC4	Clearly Defined Goals and Objectives (B)	Not clearly stated, no objectives given	Clear details of goals/objectives (bulleted or short thought-out sentences)	------------
PC5	Detailed Methodology (B)	Listed/minor details about process	Description of each step and how it will be created/done (example: by whom, etc.)	------------
PC6	Project Details (B)	Minor details about content of project (educational material)	Specific examples of what project will produce in terms of content (description of modules to be created), what	------------
PC7	Addresses Project Concerns (B)	No or no concerns given	Explains concerns (about pedagogy or usage, etc) or and how they will be taken into account	------------
PC8	Attention to Diversity	No attention	Type of diverse institutions mentioned or listed (either by demographics or by type of institution), or mentions that projects will cater to diversity	Details about how project will cater to diverse institutions and include names where project will be developed at (both by demographics and by type of institution)
PC9	Collaboration of Other Institutions (B)	No	Yes, other institutions involved with development/ Beta-Testing	------------

PC10	Collaboration with Other Experts (B)	No	Yes, other experts not included in the project team will provide input to the content aspects of the project	------------
PC11	Multi-platform or Considerations (B)	No mention of OS/platform support	Yes (web-based, java or separate builds for different OS's)	------------
PC12	Open Source (B)	No	Yes, source to project released as well	------------
PC13	Attention to User Interface (UI)	None/not a focus	Mentioned that UI will be addressed or is important	Described how UI will be focused on (past research or UI, usability testing, testing during development)
PC14	Level of Interactivity	Static/Navigation Only	Predetermined inputs such as tutorials or examples simulations, etc.	User solves own problems, creates/experiments
PC15	Design Allows for Customizability	No, cannot modify project to meet needs (set number of modules or lectures or lessons, cannot be added to by users)	Some, can modify existing material (change homework questions	Can redesign and create additional modules/materials as needed (provides tools or toolkit to do so)
PC16	Project URL Given in Proposal (B)	No	Yes	------------
PC17	Sustaining Plan - Funds	No/not mentioned	Plan proposed ways to obtain funding needed (commercialization, institutional support to keep webpage going, etc)	Approved or a letter of support
PC18	Sustaining Plan - Maintenance	No/not mentioned	Plan proposes that resulting materials will be supported/more produced/ someone in charge of maintenance	Details about said plan -> careful thought given about support
PC19	Timeline	No	Yes, provides a timeline of events or order in which development will occur	Timeline provide details about when different stages of development take place and justifies/explains the time requirements

	Assessment/Evaluation Plan	0	1	2
PC20	Formative for Process	Not mentioned	Mentions that feedback will be used to modify development process (or adjust timeline accordingly)	Detailed Planning or examples as to how the feedback will be used
PC21	Formative for Product	Not mentioned	Mentions that feedback will be used to modify product	Detailed Planning as to how feedback will be gathered and what information it will provide
PC22	Summative for Process (B)	Not mentioned	Mentions that the development process will be evaluated at the end	Detailed Planning as to how/by who it will be evaluated
PC23	Summative for Product (B)	No mention or mentions that the product will be evaluated at the end	Detailed Planning as to how/by who it will be evaluated (questions to ask, who will be asked, etc.)	------------
PC24	Outside/Impartial Evaluation (B)	No	Yes, someone not on the project team has been contacted to evaluate the product	------------
PC25	User Observation	Not used	Mentioned that students will be observed using product	Detailed planning as to whom by to observed, by who, why, etc.
PC26	Qualitative Student/Faculty Interviews	Not used	Mentions that students/faculty will be interviewed about product and impact	Detailed Planning as to who will be interviewed, and what questions will be asked/why
PC27	Attitude Surveys/Questionnaires	Not used	Mentions that student/faculty opinion about product and its impact will be surveyed	Detailed Planning as to who will be surveyed and what questions will be asked/why, etc.
PC28	Group Discussion/Focus Groups	Not used	Mentions that student/faculty will be gathered together to discuss the product and its impact	Detailed Planning as to who will participate and what questions will be asked/why, etc.
PC29	Pre- and Post- Testing	Not used	Mentions that testing will be done before and after product used	Detailed Planning as to who will be tested and when and what results will show
PC30	Control Groups	Not used	Mentions that control groups (not using the product) will be used to determine how effective the product is	Detailed Planning as to who the control groups will be and how the product will be used and testing done at the end

PC31	Diversity Testing	Not used	Mentions that diverse institutions will be included in the testing/assessment	Detailed Planning as to who will participate and how diversity will be addresses
PC32	Assessment Timeline (B)	No	Yes, gives timeframes (years, quarters, or sequential) when different testing will take pace	------------
	Dissemination Plan	**0**	**1**	**2**
PC33	Textbook/Software Bundle Contract	No plans for a textbook bundle	Bundle proposed/looking for publishers or authors to bundle with	Publisher/author found with letter of support
PC34	Software Commercialization	No plans to commercialize	Commercialization proposed/ looking for publisher	Publisher found with letter of support
PC35	Faculty Workshops	No	Mentioned, can list possible locations	Includes where they will be held and detailed planning/benefits
PC36	National Conferences	No	Mentioned, may mention specific conferences	Mentions specific conferences and describes topics of discussion
PC37	Local Conferences/Meetings/Presentations	No	Mentioned, may mention specific conferences	Mentions specific conferences and describes topics of discussion
PC38	Websites	No	Mentioned, may mention user's sites, not much thought given	Major specific websites and details about how they will be used, or advertising or unique methods (placed on searches or directories), considerable thought given
PC39	Community Building (B)	No	Mentions that a community of faculty or developers will be organized to contribute to project	------------
PC40	Digital Library (NSDL) (B)	No	Yes, plans for	------------
PC41	CD-ROM	No	Distributed by PI or available	Distributed by many (publisher or many other collaborators or larger companies)
PC42	Journal Articles/Publications (B)	No, not mentioned	Yes, plans to produce	------------

PC43	Process Sharing (B)	No	Plans to sharing the development experience to other developers (NOT the final product)	------------
	PI Information	**0**	**1**	**2**
PC44	Prior Experience as a PI (B)	No	Yes, has received an NSF grant before (not a co-PI)	------------
PC45	Years as Higher Education Professor (V)	Number of years teaching college students	------------	------------
PC46	Technical Expertise on Project (B)	No	Yes, team has the capability to produce product (web developers, programmers, etc.) without doing technical learning during project	------------
PC47	Project Management Experience (B)	No	Yes, has either PI'ed a project before or managed a project	------------
PC48	Light Work Load (B)	Currently working on another project (not completed with final report)	Other	------------
	Budgetary Information	**0**	**1**	**2**
PC49	Detailed Justifications	List	Explain the amount or why needed	Explain both and justify why needed
PC50	Student Involvement	No	Yes	Yes with justification/explanation as to their purpose on the project and what they will accomplish
	Facilities and Support	**0**	**1**	**2**
PC51	Institutional Support	No	Yes, institution provides support beyond cost-sharing either in terms of equipment or funds or maintenance or other benefits	Yes with letter as proof of support as well
PC52	Description of Network & Computers (B)	No	Yes, describes equipment available to team/explain	------------

Appendix H: Inter-rater Reliability Results

		Project: #######			
		K = Kelly			
		A = Alex			
K	**A**				
5	5	**Panel Score**			
		Project Description	**K**	**A**	**Dissemination Plan**
1	1	Based on existing Template/Prototype	0	0	Textbook Contract
1	1	Successful Implementation of own Prototype	0	0	Software Commercialization
1	1	Clearly defined Goals and Objectives	0	0	Faculty Workshops
0	0	Detailed Methodology	1	0	National Conferences
0	0	Project details	0	0	Local Conferences/Meetings/Presentations
1	1	Addresses Concerns	2	1	Websites
0	0	Attention to Diversity	0	0	Community Building
1	1	Collaboration of other Institutions	1	1	Digital Library (NSDL)
1	1	Collaboration with other experts	0	0	CD-ROM
1	1	Adaptable Programming Language or Considerations	1	1	Journal Articles/ Publications
0	0	Open source	0	0	Process Sharing
0	1	Attention to User Interface			
1	1	Level of Interactivity			**PI Information**
0	0	Design Allows for Customizability			
1	1	Project URL listed in proposal	1	1	Prior experience as PI
0	0	Sustaining Plan - Funds		---	Years as Professor
0	0	Sustaining Plan - Maintenance	1	1	Technical expertise on the project
1	1	Timeline	1	1	Project Management Experience
			0	0	Light Work Load
		Assessment/Evaluation Plan			
					Budgetary Information
0	0	Formative for Process			
2	2	Formative for Product	1	1	Detailed Justifications
1	1	Beta-Testing	2	2	Student Involvement
0	0	Summative for Process			
1	1	Summative for Product			**Facilities and Support**
0	0	Outside/Impartial Evaluation			
2	2	Qualitative Student/Faculty Interviews	0	0	Institutional Support
1	1	Attitude Surveys	1	1	Description of Network & Computers
0	0	Group Discussion			
1	1	Pre- and Post- Testing			
0	0	Control Groups			
0	0	Diversity Testing			
1	0	Assessment Timeline			

		Project: #######			
		K = Kelly			
		A = Alex			
K	**A**				
4.8	4.8	**Panel Score**			
		Project Description	**K**	**A**	**Dissemination Plan**
2	2	Based on existing Template/Prototype	0	0	Textbook Contract
2	2	Successful Implementation of own Prototype	0	0	Software Commercialization
1	1	Clearly defined Goals and Objectives	1	1	Faculty Workshops
1	1	Detailed Methodology	1	1	National Conferences
1	1	Project details	0	0	Local Conferences/Meetings/Presentations
1	1	Addresses Concerns	2	2	Websites
0	1	Attention to Diversity	0	0	Community Building
1	1	Collaboration of other Institutions	0	0	Digital Library (NSDL)
1	1	Collaboration with other experts	1	1	CD-ROM
1	1	Adaptable Programming Language or Considerations	1	1	Journal Articles/ Publications
1	1	Open source	0	0	Process Sharing
1	0	Attention to User Interface			
1	2	Level of Interactivity			**PI Information**
2	2	Design Allows for Customizability			
1	1	Project URL listed in proposal	1	1	Prior experience as PI
0	0	Sustaining Plan - Funds	24	24	Years as Professor
0	0	Sustaining Plan - Maintenance	1	1	Technical expertise on the project
1	1	Timeline	1	1	Project Management Experience
			1	1	Light Work Load
		Assessment/Evaluation Plan			
					Budgetary Information
0	0	Formative for Process			
2	2	Formative for Product	1	1	Detailed Justifications
1	1	Beta-Testing	2	2	Student Involvement
0	0	Summative for Process			
1	1	Summative for Product			**Facilities and Support**
0	0	Outside/Impartial Evaluation			
1	1	Qualitative Student/Faculty Interviews	0	0	Institutional Support
1	1	Attitude Surveys	1	1	Description of Network & Computers
0	0	Group Discussion			
2	2	Pre- and Post- Testing			
0	0	Control Groups			
0	0	Diversity Testing			
1	1	Assessment Timeline			

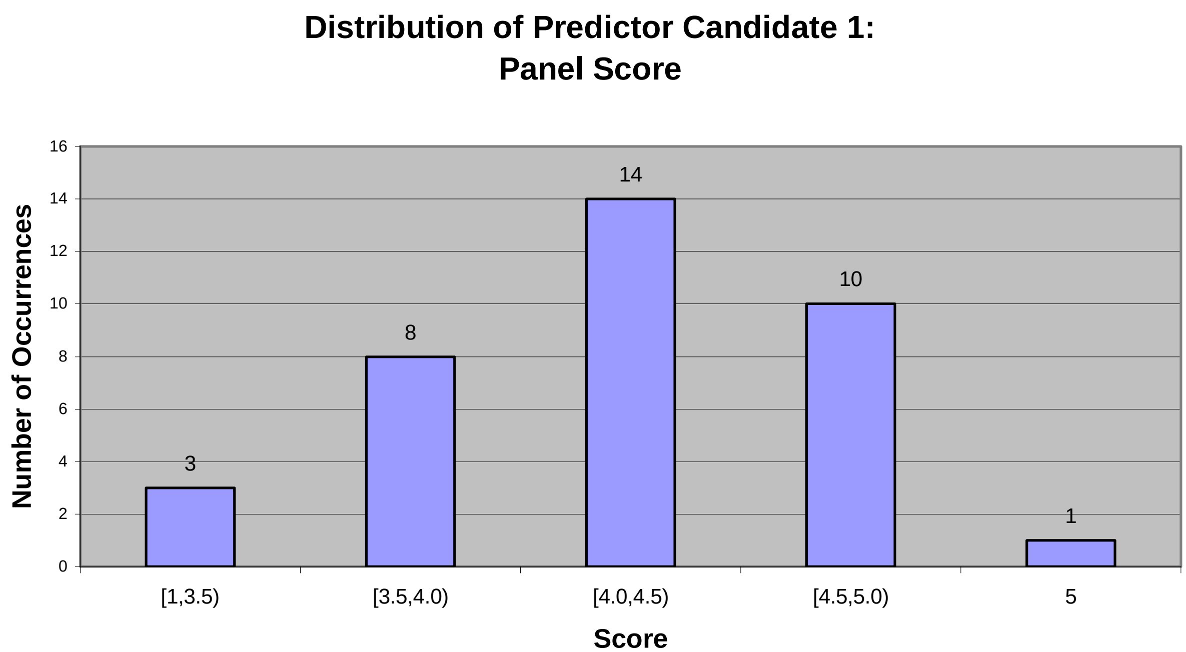
Appendix I: Predictor Candidate Scores

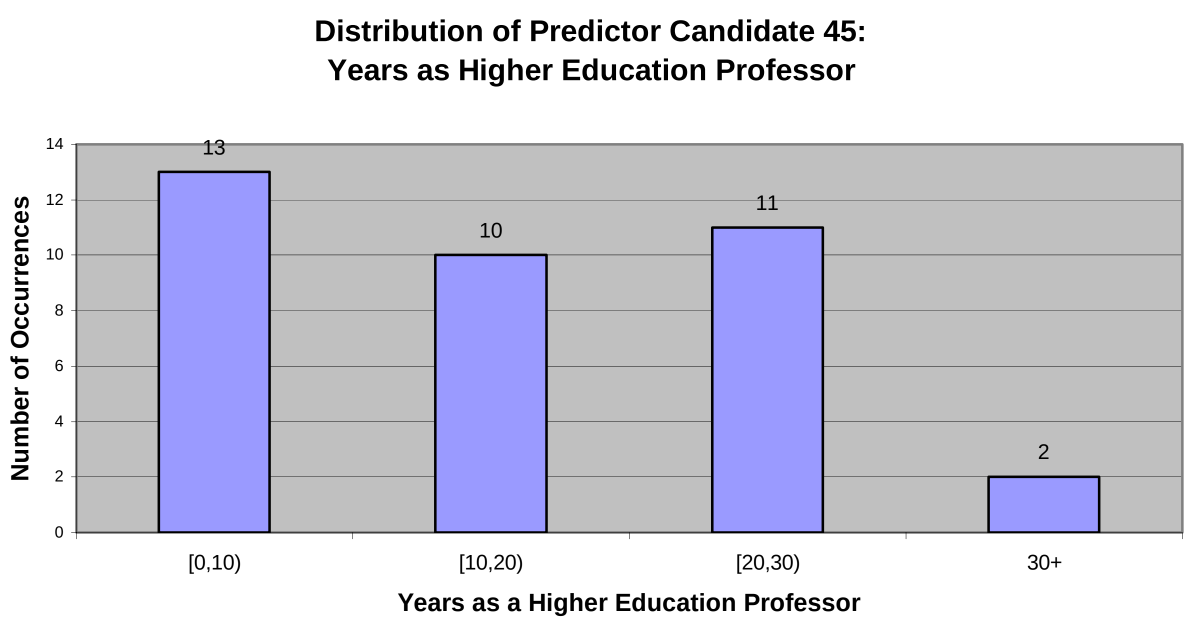
Award	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
1	3.3	1	1	1	0	1	0	0	0	0	1	0	2	2	2	0	2	2	0	0
2	4.2	1	1	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0
3	4.2	1	1	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0
4	4.6	2	0	0	0	1	0	1	1	0	0	0	0	0	0	1	2	1	1	0
5	4.4	1	1	0	0	1	0	1	1	0	1	0	1	1	1	0	1	0	1	0
6	3.9	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	0	2	0
7	4.0	0	0	1	1	1	1	0	1	0	0	0	2	2	0	0	0	0	1	0
8	4.5	1	1	1	0	1	1	0	1	0	1	0	0	2	1	1	1	0	1	0
9	3.0	2	2	0	1	1	1	2	1	0	1	0	1	0	1	0	0	1	2	0
10	3.6	1	2	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	1	0
11	4.4	2	2	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1	2	0
12	4.9	1	1	1	1	1	1	1	0	0	1	0	2	1	2	1	0	0	1	0
13	3.8	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	0
14	4.8	1	1	1	1	1	1	1	1	1	1	0	2	2	2	0	2	2	1	0
15	3.7	2	2	1	1	1	1	1	1	0	1	0	0	2	2	0	0	1	1	0
16	3.6	1	0	1	0	1	1	1	0	0	0	0	0	2	0	0	0	0	1	0
17	4.5	1	0	0	0	1	1	0	1	0	1	0	2	2	2	1	1	1	2	0
18	4.6	0	0	1	0	1	0	1	1	1	1	0	2	0	0	1	0	1	1	0
19	4.2	2	1	0	1	1	1	1	1	1	1	0	0	1	2	0	1	2	2	0
20	4.5	1	1	1	1	1	0	0	1	1	1	0	2	0	0	1	0	0	0	0
21	4.3	1	0	0	0	1	1	1	0	0	0	0	1	1	0	1	0	0	0	0
22	4.2	2	2	1	1	1	1	1	1	1	1	0	1	1	1	0	2	0	1	0
23	3.9	1	0	1	0	1	1	0	0	0	1	0	0	2	0	0	1	1	1	0
24	4.0	1	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0
25	4.3	2	2	0	1	1	1	1	0	0	1	0	1	2	2	0	1	2	1	0
26	4.3	2	1	1	1	1	0	0	1	0	0	0	0	1	1	0	1	0	0	0
27	4.7	1	0	1	0	1	1	0	1	0	1	1	0	2	2	0	0	1	1	0
28	5.0	1	1	1	0	1	1	0	1	1	1	0	0	0	2	1	0	0	2	0
29	4.0	0	0	1	0	1	1	1	1	0	0	1	1	2	2	1	0	2	1	0
30	3.0	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0
31	3.6	2	0	0	0	1	0	1	1	0	0	0	1	2	1	0	1	0	2	0
32	4.4	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0
33	4.4	2	2	0	1	1	1	1	1	0	0	0	2	2	2	0	2	2	0	0
34	4.6	0	0	1	0	1	1	1	1	0	1	0	2	2	1	0	1	1	1	0
35	3.8	2	2	1	1	1	1	2	1	0	1	0	1	1	0	0	0	1	1	0
36	4.7	0	0	0	0	1	0	1	1	1	1	0	2	2	0	1	0	0	0	1

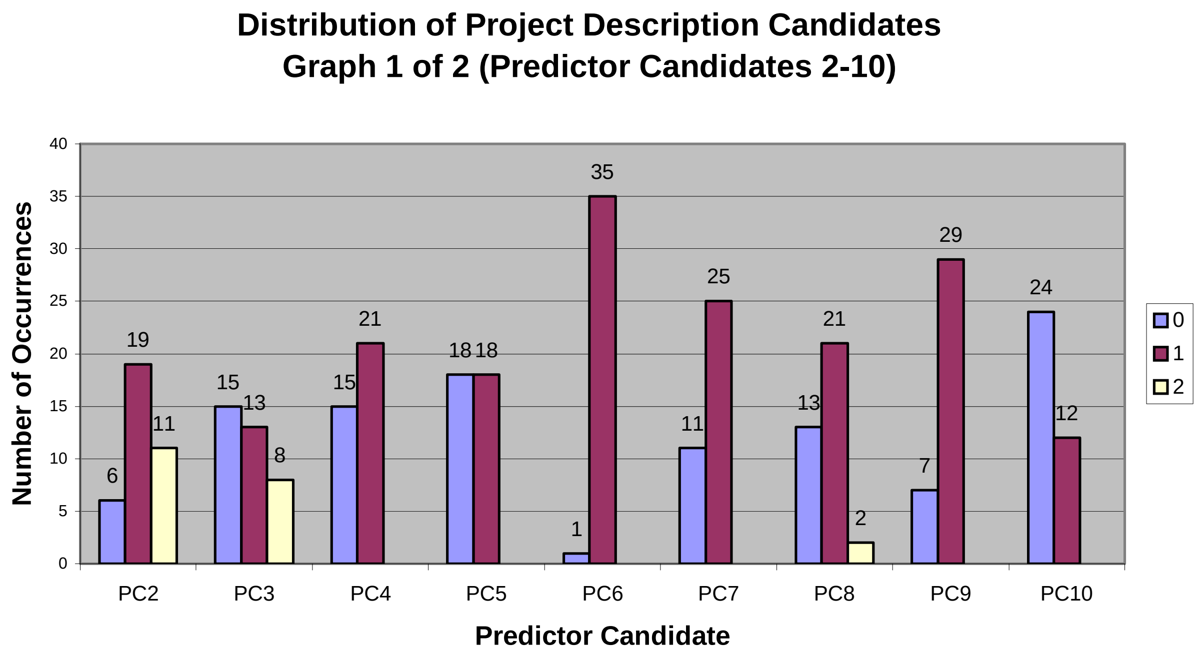
Award	PC21	PC22	PC23	PC24	PC25	PC26	PC27	PC28	PC29	PC30	PC31	PC32	PC33	PC34	PC35	PC36	PC37	PC38	PC39	PC40
1	2	0	1	1	0	0	1	0	0	0	0	0	1	0	2	2	2	0	0	0
2	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1	1	0	0
3	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1	1	0	0
4	2	0	0	1	2	1	1	2	2	2	0	1	2	2	1	1	1	1	0	0
5	2	0	1	1	0	1	1	1	0	0	0	1	1	0	0	1	0	0	0	0
6	1	0	1	1	0	1	2	0	0	1	0	1	1	1	0	0	0	2	0	0
7	1	0	1	1	2	1	0	1	0	0	0	0	1	1	0	1	0	1	0	0
8	2	0	1	1	2	1	0	0	0	0	0	0	2	2	2	0	0	1	0	0
9	2	0	1	1	0	1	0	0	0	0	1	1	1	2	0	0	0	0	0	0
10	2	0	1	0	0	2	1	0	1	1	1	1	0	0	0	2	0	2	0	0
11	2	0	1	0	1	1	1	0	1	2	0	1	0	0	2	1	1	1	0	0
12	2	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	1	0	0
13	1	0	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	0	0
14	2	0	1	1	0	1	1	0	1	0	2	1	1	2	0	0	1	2	0	0
15	1	0	1	1	0	1	1	1	1	0	0	0	1	0	2	0	1	0	0	0
16	1	0	1	0	1	1	1	1	0	0	0	0	2	2	0	0	1	1	0	0
17	1	0	1	0	1	1	0	0	2	2	0	1	0	1	0	0	0	1	0	0
18	2	0	1	1	1	1	1	1	0	0	0	0	0	2	2	1	1	2	1	0
19	1	0	1	0	1	1	2	2	0	0	0	0	0	0	2	1	2	1	0	0
20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
21	1	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0
22	2	0	1	1	0	0	1	1	1	2	0	1	2	2	1	1	0	0	0	0
23	2	0	1	0	1	0	2	1	0	0	0	0	1	0	1	1	1	1	0	0
24	1	0	0	1	0	0	1	0	0	0	0	0	0	2	0	0	0	2	0	0
25	2	0	1	1	1	0	2	0	1	1	1	1	2	0	1	1	1	1	0	0
26	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0
27	1	0	1	1	0	1	2	0	1	0	0	1	1	0	1	1	0	1	1	0
28	2	0	1	1	0	1	1	0	1	1	0	1	0	0	2	1	1	2	1	0
29	2	0	1	0	2	0	2	0	2	0	0	0	0	0	0	1	0	2	0	0
30	1	0	1	1	0	0	2	0	0	2	1	1	0	1	1	0	0	0	0	0
31	1	0	1	1	0	0	0	0	1	1	0	0	0	1	2	2	1	1	0	0
32	1	0	0	0	0	0	0	2	1	2	0	1	2	0	2	1	0	1	0	0
33	2	0	0	1	2	1	1	1	1	2	1	1	0	2	0	1	0	1	1	0
34	2	1	1	1	2	0	1	0	1	0	0	1	0	1	2	0	1	1	0	0
35	1	0	0	0	0	0	2	0	0	0	0	0	2	2	0	0	2	2	0	0
36	1	0	0	0	0	1	2	0	0	2	0	0	1	0	1	0	0	2	0	0

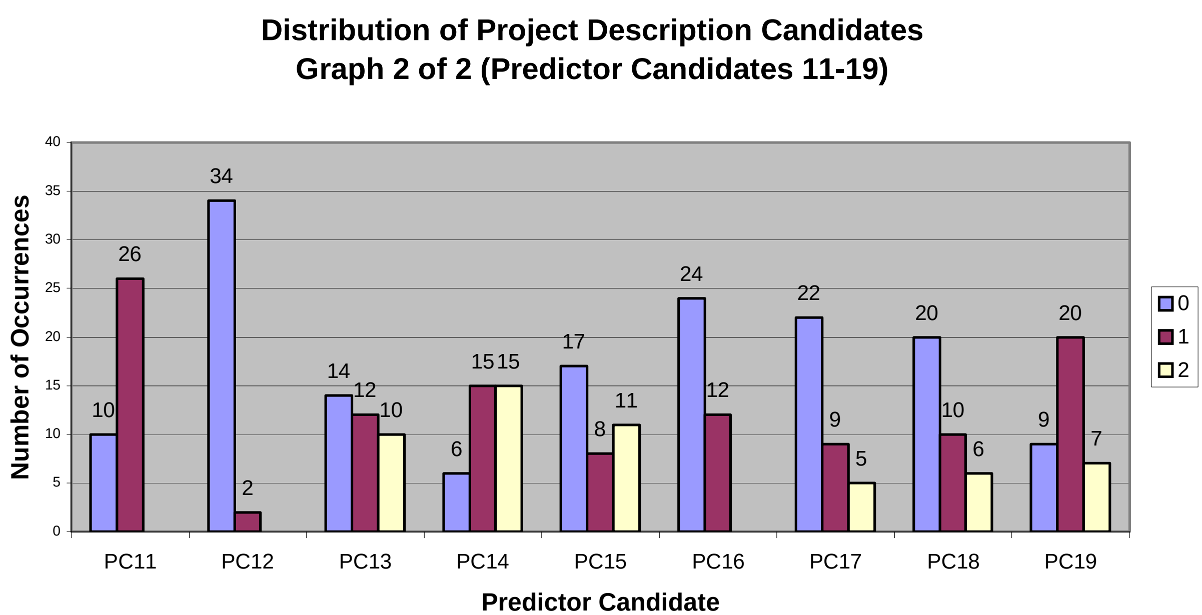
Award	PC41	PC42	PC43	PC44	PC45	PC46	PC47	PC48	PC49	PC50	PC51	PC52
1	0	1	0	1	24	1	1	0	2	1	0	1
2	1	0	0	0	5	1	1	0	1	1	0	1
3	1	0	0	1	27	1	1	0	0	2	0	1
4	2	1	0	1	17	1	1	0	1	2	0	0
5	0	1	0	1	44	1	1	1	1	2	1	1
6	2	1	0	0	7	1	0	1	0	2	0	1
7	1	1	0	0	24	1	1	0	1	1	1	1
8	2	1	0	1	25	1	1	0	0	1	0	0
9	2	0	0	1	11	1	1	1	0	2	0	1
10	1	0	0	0	6	1	1	0	1	2	1	1
11	1	1	0	1	27	1	1	0	1	2	0	1
12	0	0	0	1	29	1	1	1	0	1	1	1
13	1	0	0	0	21	1	1	1	1	0	2	1
14	2	1	0	1	17	1	1	0	1	1	0	0
15	0	1	1	1	30	1	1	0	2	2	2	0
16	2	0	0	1	10	1	1	0	1	0	0	1
17	0	0	0	0	7	1	1	0	1	2	2	0
18	2	1	0	1	8	1	1	0	1	1	0	0
19	0	0	0	1	19	1	1	1	2	2	2	1
20	2	1	0	0	0	1	1	0	1	0	0	1
21	0	1	0	1	27	1	1	1	0	1	0	1
22	2	1	0	0	4	1	0	1	0	1	2	0
23	1	0	0	1	21	1	1	1	2	2	0	1
24	2	0	0	1	15	1	1	0	1	1	0	0
25	2	1	1	1	18	1	1	1	2	2	1	1
26	0	1	0	1	0	1	1	0	2	0	0	0
27	1	1	0	0	6	1	1	0	2	1	2	1
28	1	1	0	0	13	1	1	1	2	1	1	1
29	0	1	0	0	2	1	0	1	2	2	2	1
30	0	1	0	1	25	1	1	0	1	0	1	0
31	0	0	0	1	9	1	1	1	1	1	0	0
32	1	0	0	1	9	1	1	0	0	1	0	0
33	1	0	0	1	23	1	1	0	0	0	0	1
34	1	0	1	1	2	1	1	0	1	2	0	1
35	0	1	0	1	15	1	1	0	2	2	2	1
36	0	1	0	1	11	1	1	1	0	0	0	1

Appendix J: Predictor Candidate Score Distributions

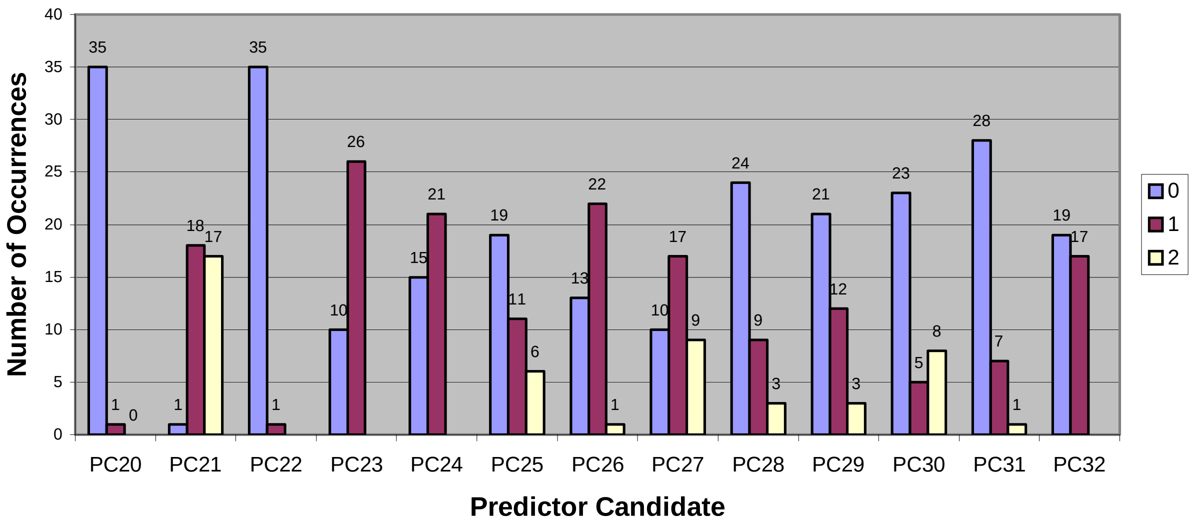




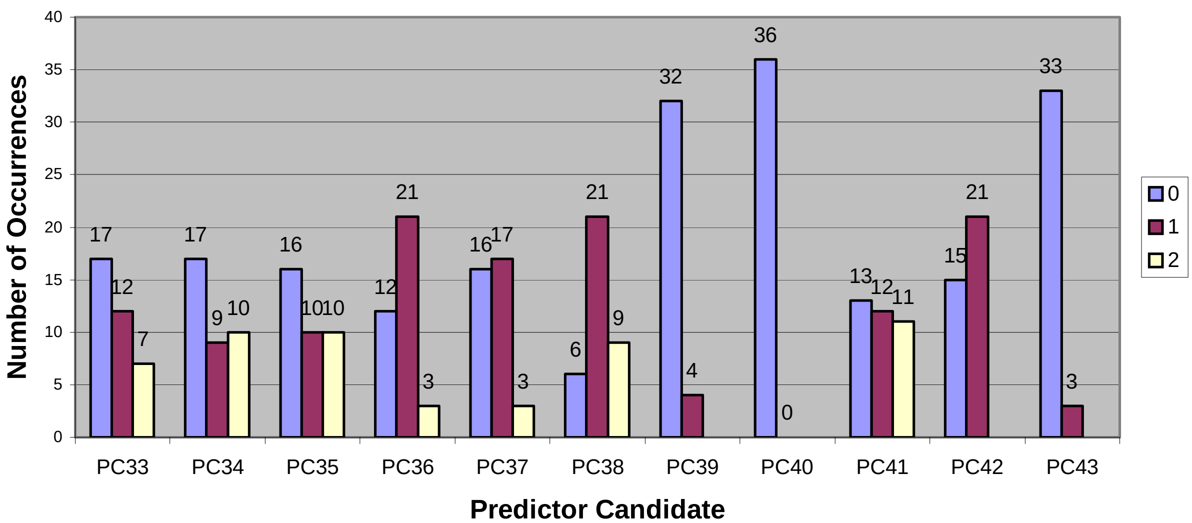

Distribution of Project Description Candidates
Graph 1 of 2 (Predictor Candidates 2-10)
Number of Occurrences
Predictor Candidate
0
1
2
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10

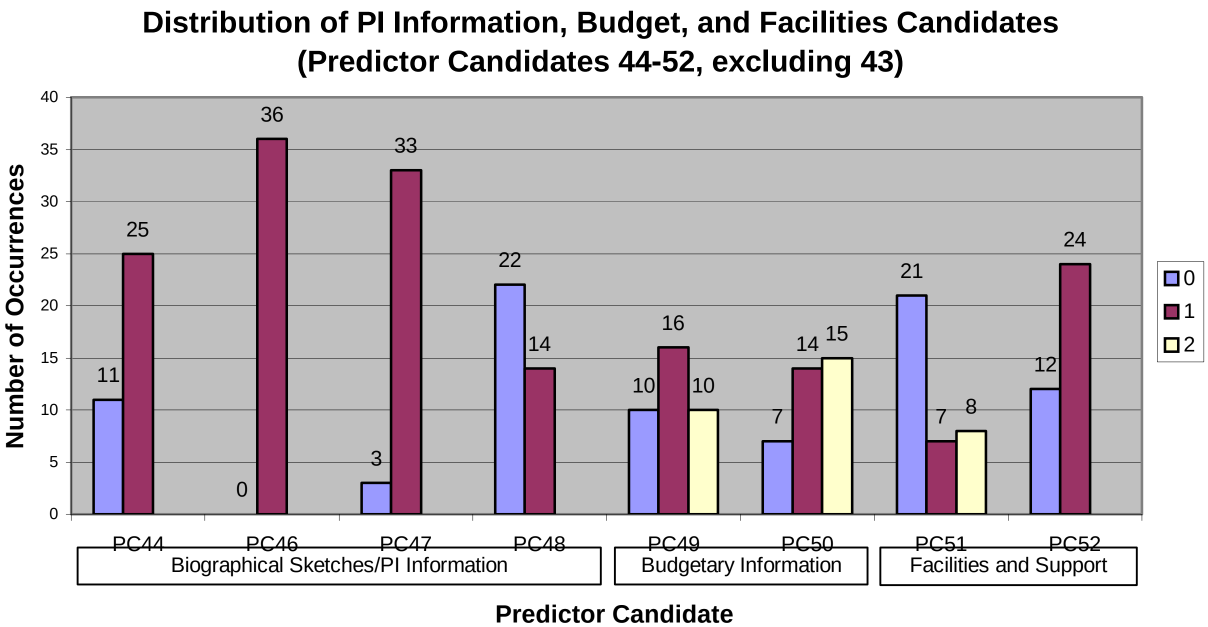

Distribution of Project Description Candidates
Graph 2 of 2 (Predictor Candidates 11-19)
Number of Occurrences
Predictor Candidate
0
1
2
PC11
PC12
PC13
PC14
PC15
PC16
PC17
PC18
PC19

Distribution of Evaluation/Assessment Plan Candidates (Predictor Candidates 20-32)



Distribution of Dissemination Plan Candidates (Predictor Candidates 33-43)





Distribution of PI Information, Budget, and Facilities Candidates
(Predictor Candidates 44-52, excluding 43)
Number of Occurrences
40
35
30
25
20
15
10
5
0
11
25
0
36
3
33
22
14
10
16
10
7
14
15
21
7
8
12
24
PC44
PC46
PC47
PC48
PC49
PC50
PC51
PC52
Biographical Sketches/PI Information
Budgetary Information
Facilities and Support
0
1
2
Predictor Candidate

Appendix K: Predictor Candidate-Reviewer Correlation

For this correlation, a 1 was assigned to Kelly and a 0 to Alex. P-Values less than 0.05 were deemed to have significant correlation. These rows are highlighted.

Reviewer Correlation with P-Values

PC	Correlation	P-Value
1	0.1573	0.3596
2	-0.3279	0.0509
3	-0.0666	0.6995
4	0.0857	0.6192
5	-0.0563	0.7441
6	-0.1429	0.4059
7	-0.3610	0.0605
8	0.2560	0.1318
9	0.0119	0.9453
10	0.2390	0.1603
11	-0.0210	0.9043
12	-0.2870	0.0879
13	0.0929	0.5901
14	-0.2537	0.1354
15	-0.3578	0.0321
16	0.1195	0.4875
17	-0.3169	0.0597
18	-0.4350	0.0080
19	-0.3251	0.0530
20	0.1429	0.4059
21	-0.2390	0.1603
22	-0.2000	0.2422
23	-0.1468	0.3930
24	-0.1429	0.4059
25	-0.2563	0.1313
26	0.0000	1.0000
27	-0.1876	0.2733
28	-0.1540	0.3697
29	-0.5674	0.0003
30	-0.1529	0.3733
31	-0.1429	0.4059
32	-0.2163	0.2051
33	-0.0122	0.9436
34	0.0056	0.9743
35	-0.1014	0.5565
36	-0.1657	0.3342
37	0.0521	0.7627
38	0.1981	0.2469
39	-0.0598	0.7292
40	0.0000	1.0000
41	0.0115	0.9468
42	0.3143	0.0619
43	-0.3568	0.0327
44	0.1733	0.3122
45	0.2029	0.2352
46	0.0000	1.0000
47	0.1529	0.3733
48	-0.2504	0.1407
49	-0.2268	0.1835
50	-0.3508	0.0359
51	-0.4400	0.0072
52	0.0000	1.0000

Appendix L: Outcome Rubrics

Product – O1

1 – No product is created by the time the award closes

2 – The product has been completed however it did not work as intended. An example might be if the students did not take to the software.

3 – Due to time constraints or complications, the software was created at the expense of a few features. If for example, the original plan was for 5 simulations and only 4 were constructed. This would be different from a grade of 2 because the 4 simulations that were created work as intended.

4 – All goals are completed as originally stated in the project proposal.

5 – For whatever reason, if the project achieves further development than originally planned, it receives a grade of 5. As in the previous example, this grade is achieved if 5 simulations were planned but 6 were ultimately constructed.

Dissemination – O2

1 – Either no dissemination was attempted, or very little. This means either passive or active attempts were made exclusively and only several times at most.

2 – Both active and passive attempts are present, as well as different types of each. Multiple attempts are made for example: 5 journal publications, a website, and 4 conferences.

3 – The PI must show that he or she tried a variety of methods. Also he or she should explain which methods worked or did not work. Repetition of working methods (such as adding a third workshop after two successful ones) is also looked for.

4 – Despite how aggressive the dissemination attempts were, a project automatically receives a 4 if the free product is proven to be used elsewhere. The rationale is that no matter how hard you try to disseminate, the ultimate purpose is to achieve it.

5 – If shown that third party distribution is achieved, the project receives a 5. An example of this might be a CD-ROM packaged with a textbook.

Sustainability – O3

1 – If the product cannot be shown as active, even in the PI's own class, it receives a 1 and is considered obsolete.

2 – If the product is being used, however is not compatible with multiple operating systems or evolving software it receives a 2.

3 – The product is shown to fit multiple operating systems and/or if the software becomes more user friendly or aesthetically pleasing over time.

4 – The product is actively being developed. This is different from a grade of 3 in the sense that adding an entire module or simulation is different from fixing a typographical error.

5 – The product is robust enough and popular enough that other developers make their own modifications, either for their own purposes or for public use.

Student Interest – O4

1 – No assessment was made.

2 – All assessment attempts were inconclusive

3 – If the only control is a group of students taught by the PI himself or herself, the assessment receives this grade.

4 – Multiple assessments must be taken, whether they are across a single department or across multiple universities.

5 – Multiple assessments must be taken, however they must be taken specifically with a diverse student population in mind.

Student Learning – O5

1 – No evaluation attempt was made.

2 – All evaluation attempts were inconclusive.

3 – If the only control is a group of students taught by the PI himself or herself, the evaluation receives this grade.

4 – Multiple evaluations must be taken, whether they are across a single department or across multiple universities.

5 – Multiple evaluations must be taken, however they must be taken specifically with a diverse student population in mind.

Spawns Further Projects – O6

1 – No further development has occurred

2 – Further development occurs without funding

3 – An original PI or co-PI has attempted or is attempting to seek funding based on the research of this project.

4 – A different developer has attempted or is attempting to seek funding based on the research of this project. This would be primarily in the form of an A&I grant based on the original EMD project. If a project receives a 3 and a 4, it is given a score of 4.

5 – If anyone obtains funding, whether or not they were on the original EMD grant, the project receives a 5.

Professional Awards – O7

1 – No professional recognition is received.

2 – Professional recognition is received, however not in the form of an award. This might be the acceptance into a competitive database.

3 – Any number of nominations is received for professional awards.

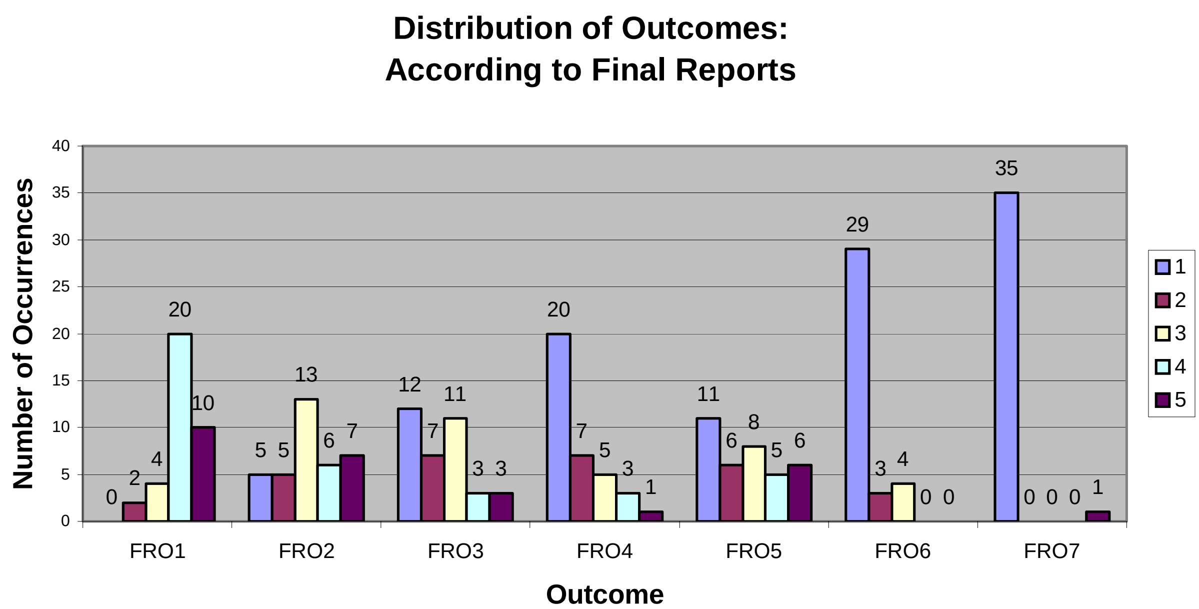
4 – The project wins a professional award.

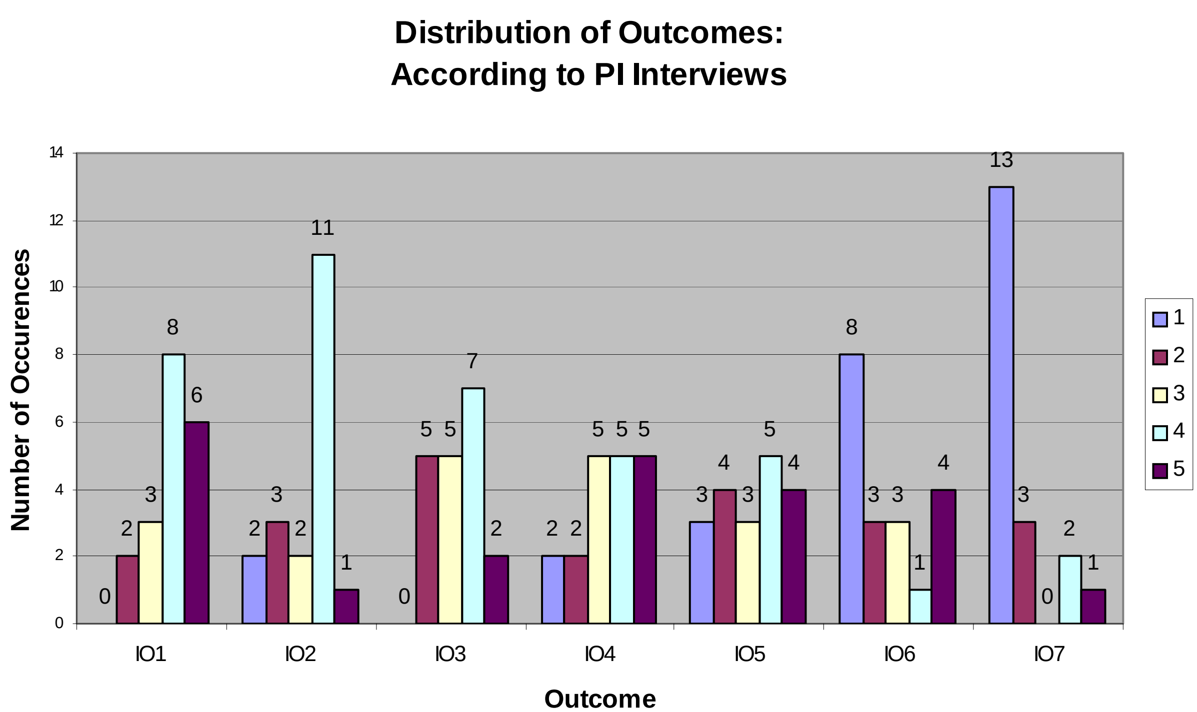
5 – The project wins two or more professional awards

Appendix M: Award Outcome Scores

Award	FRO1	FRO2	FRO3	FRO4	FRO5	FRO6	FRO7	IO1	IO2	IO3	IO4	IO5	IO6	IO7
1	3	4	4	1	1	1	1	3	4	2	5	2	5	1
2	4	5	1	1	1	1	1							
3	4	5	1	1	1	1	1							
4	4	3	2	1	5	1	1	4	4	3	5	5	1	1
5	3	2	3	4	5	1	1	3	2	4	4	5	2	1
6	5	4	2	1	1	1	1	5	2	2	3	3	1	1
7	3	3	1	4	4	1	1							
8	4	5	1	5	5	3	5	4	4	4	5	5	5	5
9	3	5	3	1	1	3	1	3	5	3	1	1	3	1
10	4	2	3	3	3	1	1							
11	5	4	3	2	2	1	1							
12	5	4	5	3	3	1	1	5	4	5	5	2	5	1
13	4	3	2	1	1	1	1	4	4	2	4	3	1	1
14	4	1	1	1	3	1	1							
15	4	2	2	1	1	2	1							
16	4	3	1	1	3	1	1							
17	4	4	3	2	2	1	1							
18	4	3	3	1	1	2	1							
19	5	4	5	1	1	1	1	5	4	3	2	3	1	2
20	5	5	3	2	2	3	1							
21	5	3	3	1	5	1	1	5	2	4	3	4	3	1
22	4	1	2	4	4	1	1							
23	2	3	4	1	1	1	1	2	4	4	1	1	1	2
24	5	5	1	1	4	1	1							
25	2	1	1	3	2	1	1	2	1	4	3	2	5	1
26	4	5	2	2	5	1	1	4	4	5	5	5	2	1
27	4	3	3	1	3	1	1	4	4	3	4	4	1	4
28	4	3	5	1	4	1	1							
29	4	3	1	2	4	1	1	4	4	4	4	4	4	4
30	4	2	1	1	3	3	1							
31	4	1	4	3	2	2	1							
32	5	2	3	3	3	1	1	5	4	3	3	4	1	1
33	4	1	1	1	1	1	1	4	1	2	2	1	2	1
34	5	3	1	2	3	1	1	5	3	2	4	4	1	2
35	5	3	2	1	5	1	1							
36	4	3	3	2	2	1	1	4	3	4	3	2	3	1

Appendix N: Outcome Distributions





Appendix O: Outcome Prediction Equation Set

FRO1all = 3.393 – 1.035*PC21 – 0.573*PC49 + 0.514*PC35 –1.770*PC20 + 0.838*PC8 + 0.687*PC1 – 1.240*PC47 – 1.092*PC48 – 0.289*PC33 + 0.948*PC52 – 0.326*PC25 + 0.475*PC16 – 0.292*PC5 – 0.602*PC39 – 0.183*PC13 + 0.344*PC17 + 0.314*PC38 + 0.315*PC7 – 0.129*PC41 – 0.117*PC14

FRO1a = 3.537 – 0.892*PC21 – 0.482*PC49 + 0.322*PC35 +0.552*PC8 + 0.363*PC1 – 0.686*PC48 – 0.283*PC33 + 0.509*PC52 – 0.180*PC25 + 0.251*PC16 – 0.273*PC5 – 0.129*PC13 + 0.332*PC17 + 0.319*PC38 + 0.600*PC38 – 0.069*PC41 – 0.262*PC14

FRO1b = 4.399 + 0.494*PC38 – 0.678*PC21 – 0.412*PC49 + 0.352*PC35 + 0.330*PC8

FRO2all = 6.333 – 1.468*PC29 – 1.617*PC31 – 0.809*PC28 – 0.634*PC14 + 0.782*PC26 – 1.570*PC39 + 1.310*PC18 – 0.553*PC8 – 0.569*PC42 + 0.975*PC16 – 0.414*PC36 – 0.649*PC27 + 1.016*PC12 – 0.384*PC13 – 0.404*PC11 + 0.712*PC32 – 0.280*PC19 – 0.015*PC45

FRO2a = 6.317 – 1.240*PC29 – 1.584*PC31 – 0.881*PC28 – 0.513*PC14 + 0.560*PC26 + 1.129*PC18 – 0.437*PC8 – 0.583*PC42 + 0.804*PC 16 – 0.493*PC36 – 0.628*PC27 – 0.429*PC13 – 0.352*PC11 + 0.404*PC32 – 0.247*PC19 – 0.012*PC45

FRO2b = 4.257 – 0.713*PC29 – 1.320*PC21 – 0.864*PC28 – 0.564*PC14 + 0.473*PC25 + 0.497*PC26

FRO3all = 2.205 – 0.917*PC34 + 0.593*PC19 – 2.781*PC43 + 0.492*PC35 – 0.541*PC9 + 0.672*PC11 + 0.627*PC8 – 0.521*PC7 + 0.859*PC39 – 1.527*PC12 – 0.431*PC10 + 0.230*PC15

FRO3a = 2.645 – 0.584*PC34 + 0.983*PC19 + 0.267*PC35 – 0.996*PC9 + 0.406*PC11 + 0.164*PC8 – 0.996*PC7 + 0.235*PC10 + 0.096*PC15

FRO3b = 0.167 + 1.029*PC19 – 1.006*PC25 + 0.711*PC36 + 1.426*PC16 + 1.046*PC44 + 1.234*PC52 + 0.552*PC28 – 0.592*PC8 – 0.353*PC41 + 0.273*PC17 – 0.019*PC45

FRO4all = -2.916 – 1.211*PC37 – 2.254*PC39 + 0.444*PC35 + 0.819*PC36 + 0.941*PC1 + 0.369*PC3 + 0.528*PC4 + 0.329*PC34

FRO4a = -0.285 – 0.988*PC37 + 0.283*PC35 + 0.498*PC36 + 0.377*PC1 + 0.297*PC3 + 0.442*PC4 + 0.091*PC34

FRO4b = 2.039 – 0.825*PC37 + 0.385*PC35

FRO5all = -1.362 + 1.426*PC42 – 0.803*PC10 – 1.035*PC15 + 1.264*PC1 + 1.293*PC8 – 0.798*PC13 + 0.923*PC17 – 0.880*PC2 – 0.460*PC35 – 2.439*PC20 – 0.382*PC41

FRO5a = -1.016 + 1.324*PC42 – 0.971*PC10 – 0.932*PC18 + 1.151*PC1 + 1.100*PC8 – 0.510*PC13 + 0.881*PC17 – 0.708*PC2 – 0.460*PC35 – 0.268*PC41

FRO5b = -2.092 + 1.334*PC42 – 0.778*PC10 – 1.125*PC18 + 1.347*PC1 + 1.409*PC8 – 0.829*PC13 + 0.961*PC17 – 0.811*PC2 – 0.599*PC35 – 0.520*PC26 – 0.639*PC27 + 0.655*PC49 + 0.480*PC7

FRO6all = 4.363 - 0.611*PC1 – 0.237*PC14 + 0.293*PC35 – 0.400*PC36 – 0.357*PC27 + 0.318*PC42

FRO6a = 4.363 - 0.611*PC1 – 0.237*PC14 + 0.293*PC35 – 0.400*PC36 – 0.357*PC27 + 0.318*PC42

FRO6b = 4.363 - 0.611*PC1 – 0.237*PC14 + 0.293*PC35 – 0.400*PC36 – 0.357*PC27 + 0.318*PC42

FRO7all = 1.375 + 0.626*PC25 – 0.584*PC52 – 0.483*PC29 + 0.350*PC11 – 0.436*PC28 – 0.371*PC13 + 0.188*PC33 + 0.321*PC15 – 0.354*PC49 + 0.303*PC4

FRO7a = 1.375 + 0.626*PC25 – 0.584*PC52 – 0.483*PC29 + 0.350* PC11 – 0.436*PC28 – 0.371*PC13 + 0.188*PC33 + 0.321*PC15 – 0.354*PC49 + 0.303*PC4

FRO7b = 1.375 + 0.626*PC25 – 0.584*PC52 – 0.483* PC29 + 0.350*PC11 – 0.436*PC28 – 0.371*PC13 + 0.188*PC33 + 0.321*PC15 – 0.354*PC49 + 0.303*PC4

IO1all = 5.901 – 1.088*PC49 – 0.696*PC21 + 1.419*PC22 + 0.553*PC51 – 1.076*PC31 – 0.981*PC11 + 0.571*PC37 + 0.570*PC7

IO1a = 5.713 – 1.059*PC49 – 0.628*PC21 + 0.431*PC51 – 1.226*PC31 – 0.951*PC11 + 0.651*PC37 + 0.804*PC7

IO1b = 5.813 – 0.885*PC49 – 1.444*PC21 – 0.852*PC33 + 0.621*PC35 + 0.801*PC29 + 0.449*PC50 + 0.026*PC45

IO2all = 4.130 + 2.162*PC4 – 1.548*PC14 + 3.553*PC20 + 0.688*PC18 – 0.397*PC13 – 0.397*PC30

IO2a = 3.835 + 1.680*PC4 – 1.019*PC14 + 0.276*PC18 – 0.197*PC13 – 0.094*PC30

IO2b = 3.744 + 1.707*PC4 – 0.970*PC14

IO3all = 2.851 – 0.750*PC34 + 1.181*PC16 + 1.812*PC44 – 0.754*PC35 – 0.612*PC13 + 0.463*PC15 – 0.722*PC36 + 0.643*PC4

IO3a = 2.851 – 0.750*PC34 + 1.181*PC16 + 1.812*PC44 – 0.754*PC35 – 0.612*PC13 + 0.463*PC15 – 0.722*PC36 – 0.643*PC4

IO3b = 2.851 – 0.750*PC34 + 1.181*PC16 + 1.812*PC44 – 0.754*PC35 – 0.612*PC13 + 0.463*PC15 – 0.722*PC36 + 0.643*PC4

IO4all = 1.901 + 1.481*PC42 + 1.358*PC4

IO4a = 1.901 + 1.481*PC42 + 1.358*PC4

IO4b = 1.901 + 1.481*PC42 + 1.358*PC4

IO5all = 2.972 – 1.400*PC52 + 2.086*PC42 + 0.987*PC51 + 3.423*PC22 – 0.465*PC18 + 0.595*PC28 – 0.530*PC27

IO5a = 3.808 – 1.560*PC52 + 1.367*PC42 + 0.770*PC51 – 0.378*PC18 + 0.214*PC28 – 0.406*PC27

IO5b = 5.064 – 2.428*PC52 + 0.818*PC51 – 1.255*PC5

IO6all = 2.814 – 1.517*PC9 + 2.366*PC16 + 2.721*PC3 – 1.702*PC2 – 0.954*PC52 + 0.716*PC49

IO6a = 2.814 – 1.517*PC9 + 2.366*PC16 + 2.721*PC3 – 1.702*PC2 – 0.954*PC52 + 0.716*PC49

IO6b = 1.611 – 1.217*PC9 + 2.188*PC16 + 2.515*PC3 – 1.447*PC2 + 0.714*PC49 + 0.791*PC42 – 0.474*PC27 + 0.275*PC35

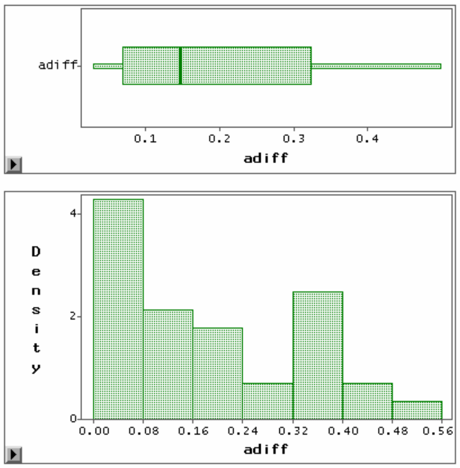
IO7all = 1.849 + 4.317*PC12 + 0.448*PC35 – 0.462*PC36 + 0.852*PC25 – 1.220*PC29 – 0.950*PC52

IO7a = 1.342 + 0.329*PC35 – 0.436*PC36 + 0.433*PC25 + 0.188*PC29 – 0.049*PC52

IO7b = -1.056 + 1.825*PC14 – 0.515*PC13 + 1.869*PC16 + 1.282*PC19 – 1.343*PC48 + 0.874*PC23 + 0.820*PC9 – 0.780*PC38 – 0.501*PC35 – 0.592*PC24

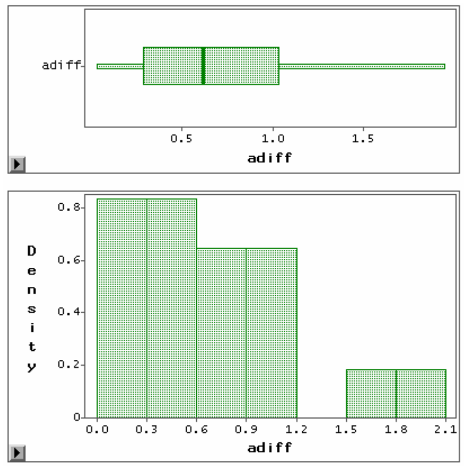
Appendix P: Error Distributions for FRO Cross-Validation

The following distributions show the absolute value of the error in the predicted values (adiff) of the tested models for each final report outcome. Appendix Q shows the same distributions for the interview outcomes. Important features are the density graph, the mean value, and the standard deviation.



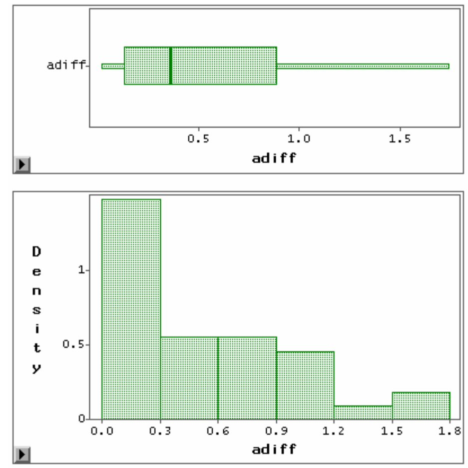
Moments			
N	35.0000	Sum Wgts	35.0000
Mean	0.1934	Sum	6.7700
Std Dev	0.1406	Variance	0.0198
Skewness	0.5609	Kurtosis	-1.0088
USS	1.9819	CSS	0.6724
CV	72.7038	Std Mean	0.0238

FRO1 ALL



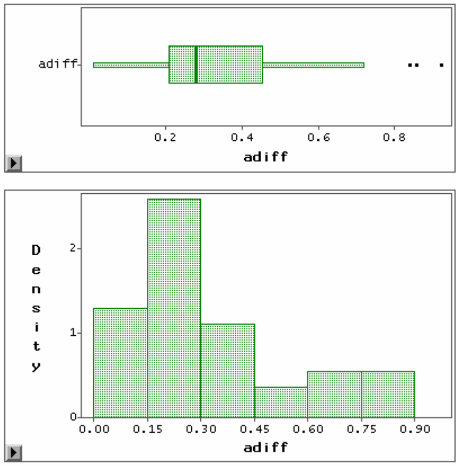
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.7029	Sum	25.3057
Std Dev	0.5109	Variance	0.2611
Skewness	0.7706	Kurtosis	0.1658
USS	26.9251	CSS	9.1368
CV	72.6853	Std Mean	0.0852

FRO1 RAF



Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.5230	Sum	18.8275
Std Dev	0.4505	Variance	0.2029
Skewness	1.0021	Kurtosis	0.4189
USS	16.9483	CSS	7.1017
CV	86.1305	Std Mean	0.0751

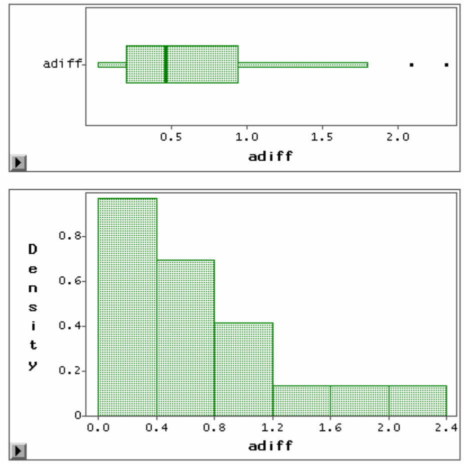
FRO1 RBF



Moments

N	36.0000	Sum Wgts	36.0000
Mean	0.3526	Sum	12.6952
Std Dev	0.2465	Variance	0.0608
Skewness	0.9690	Kurtosis	0.0652
USS	6.6042	CSS	2.1272
CV	69.9095	Std Mean	0.0411

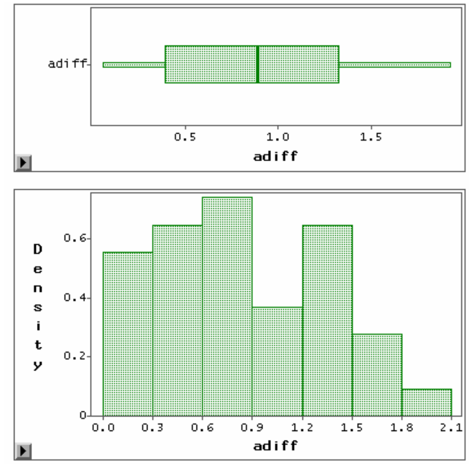
FRO2 ALL



Moments

N	36.0000	Sum Wgts	36.0000
Mean	0.6613	Sum	23.8083
Std Dev	0.6098	Variance	0.3718
Skewness	1.1667	Kurtosis	0.7239
USS	28.7600	CSS	13.0145
CV	92.2049	Std Mean	0.1016

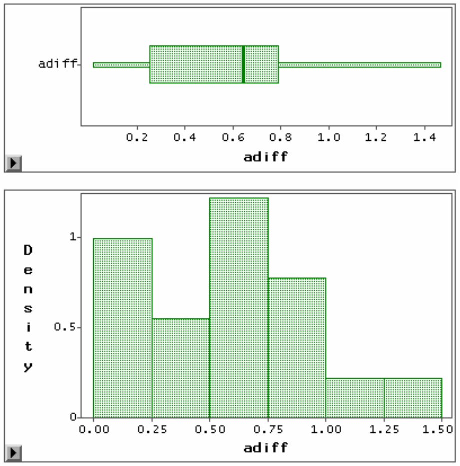
FRO2 RAF



Moments

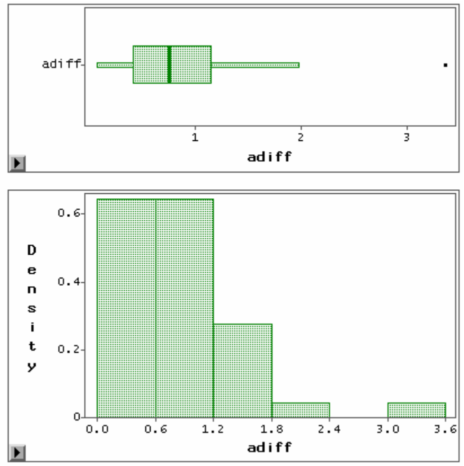
N	36.0000	Sum Wgts	36.0000
Mean	0.8617	Sum	31.0228
Std Dev	0.5269	Variance	0.2776
Skewness	0.2267	Kurtosis	-0.9433
USS	36.4507	CSS	9.7170
CV	61.1441	Std Mean	0.0878

FRO2 RBF



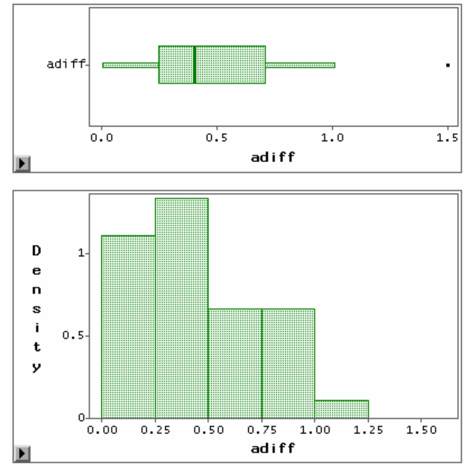
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.5722	Sum	20.5975
Std Dev	0.3641	Variance	0.1326
Skewness	0.5114	Kurtosis	0.1490
USS	16.4254	CSS	4.6404
CV	63.6403	Std Mean	0.0607

FRO3 ALL



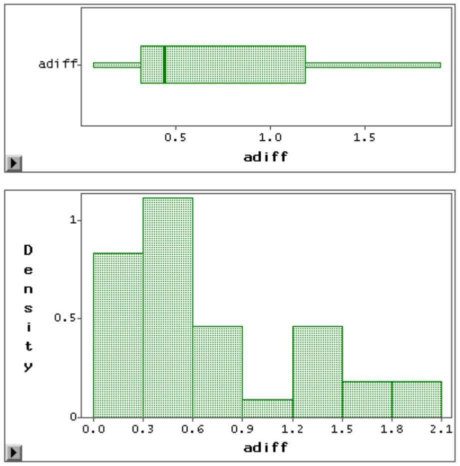
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.8645	Sum	31.1230
Std Dev	0.6527	Variance	0.4260
Skewness	1.7453	Kurtosis	4.8830
USS	41.8167	CSS	14.9100
CV	75.4964	Std Mean	0.1088

FRO3 RAF



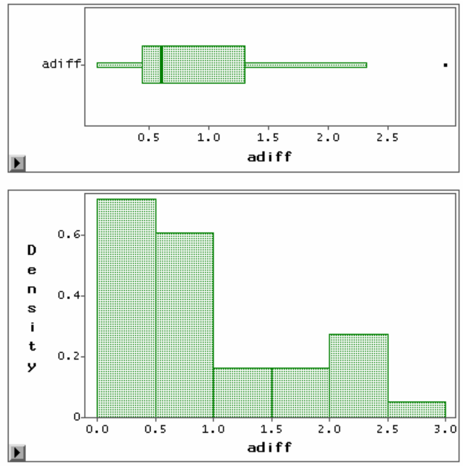
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.4854	Sum	17.4746
Std Dev	0.3457	Variance	0.1195
Skewness	0.7925	Kurtosis	0.5124
USS	12.6660	CSS	4.1838
CV	71.2272	Std Mean	0.0576

FRO3 RBF



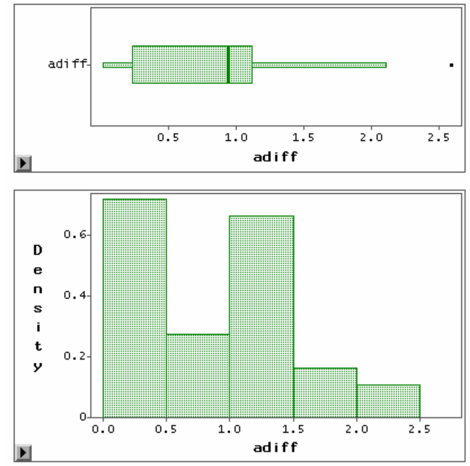
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.6942	Sum	24.9900
Std Dev	0.5542	Variance	0.3072
Skewness	0.9270	Kurtosis	-0.3480
USS	28.0977	CSS	10.7505
CV	79.8390	Std Mean	0.0924

FRO4 ALL



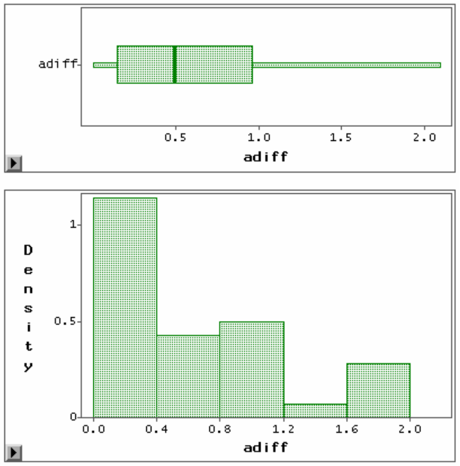
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.9287	Sum	33.4343
Std Dev	0.7375	Variance	0.5438
Skewness	1.1164	Kurtosis	0.4164
USS	50.0862	CSS	19.0347
CV	79.4053	Std Mean	0.1229

FRO4 RAF



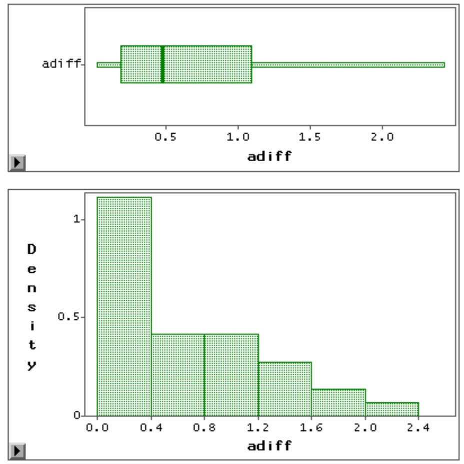
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.8744	Sum	31.4779
Std Dev	0.6568	Variance	0.4314
Skewness	0.6224	Kurtosis	-0.0255
USS	42.6228	CSS	15.0989
CV	75.1166	Std Mean	0.1095

FRO4 RBF



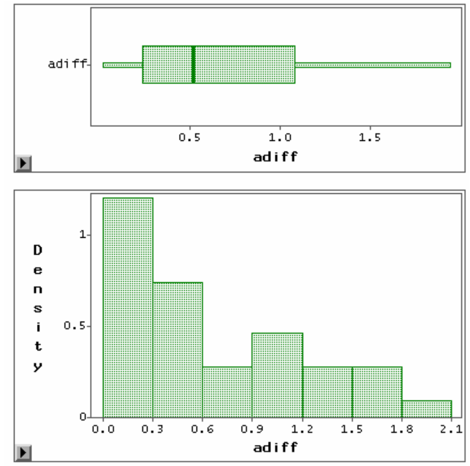
Moments			
N	35.0000	Sum Wgts	35.0000
Mean	0.6582	Sum	23.0357
Std Dev	0.5863	Variance	0.3438
Skewness	0.9320	Kurtosis	-0.0956
USS	26.8490	CSS	11.6878
CV	89.0827	Std Mean	0.0991

FRO5 ALL



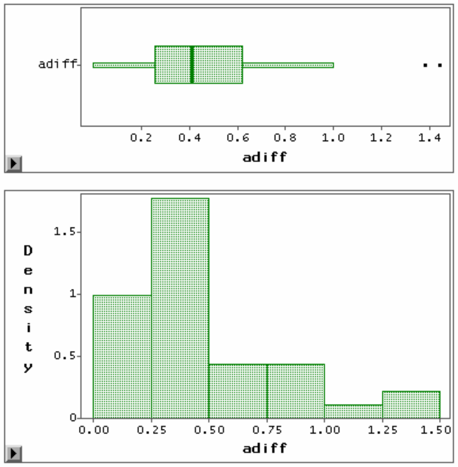
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.7247	Sum	26.0897
Std Dev	0.6627	Variance	0.4392
Skewness	1.0613	Kurtosis	0.4339
USS	34.2794	CSS	15.3719
CV	91.4458	Std Mean	0.1105

FRO5 RAF



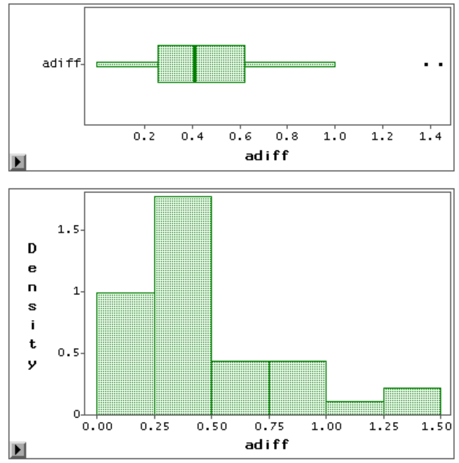
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.6835	Sum	24.6072
Std Dev	0.5546	Variance	0.3076
Skewness	0.7563	Kurtosis	-0.5926
USS	27.5862	CSS	10.7663
CV	81.1405	Std Mean	0.0924

FRO5 RBF



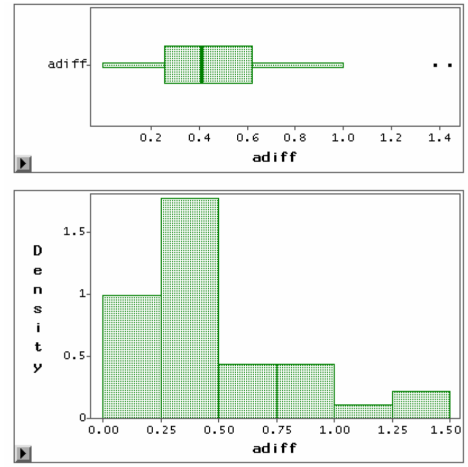
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.4763	Sum	17.1461
Std Dev	0.3334	Variance	0.1111
Skewness	1.2337	Kurtosis	1.8845
USS	12.0559	CSS	3.8896
CV	69.9930	Std Mean	0.0556

FRO6 ALL



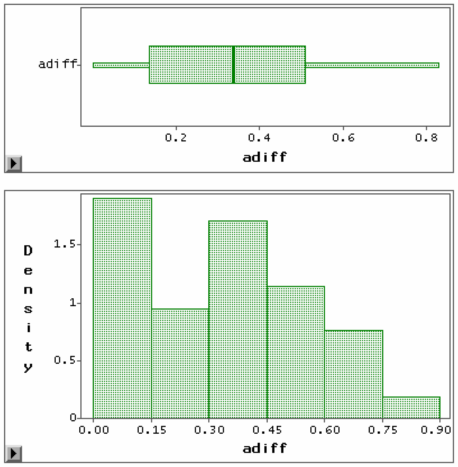
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.4763	Sum	17.1461
Std Dev	0.3334	Variance	0.1111
Skewness	1.2337	Kurtosis	1.8845
USS	12.0559	CSS	3.8896
CV	69.9930	Std Mean	0.0556

FRO6 RAF



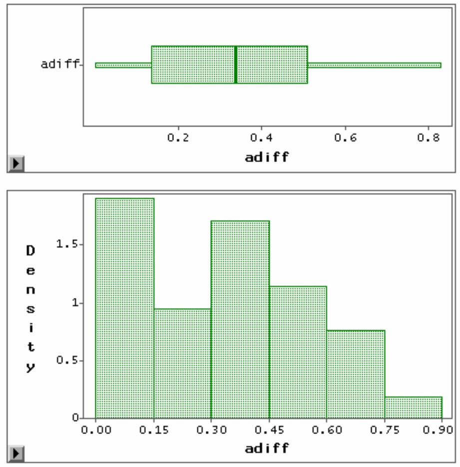
Moments			
N	36.0000	Sum Wgts	36.0000
Mean	0.4763	Sum	17.1461
Std Dev	0.3334	Variance	0.1111
Skewness	1.2337	Kurtosis	1.8845
USS	12.0559	CSS	3.8896
CV	69.9930	Std Mean	0.0556

FRO6 RBF



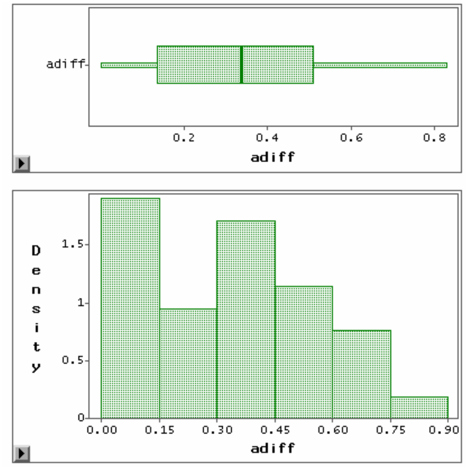
Moments			
N	35.0000	Sum Wgts	35.0000
Mean	0.3349	Sum	11.7208
Std Dev	0.2196	Variance	0.0482
Skewness	0.2767	Kurtosis	-0.8355
USS	5.5652	CSS	1.6402
CV	65.5862	Std Mean	0.0371

FRO7 ALL



Moments			
N	35.0000	Sum Wgts	35.0000
Mean	0.3349	Sum	11.7208
Std Dev	0.2196	Variance	0.0482
Skewness	0.2767	Kurtosis	-0.8355
USS	5.5652	CSS	1.6402
CV	65.5862	Std Mean	0.0371

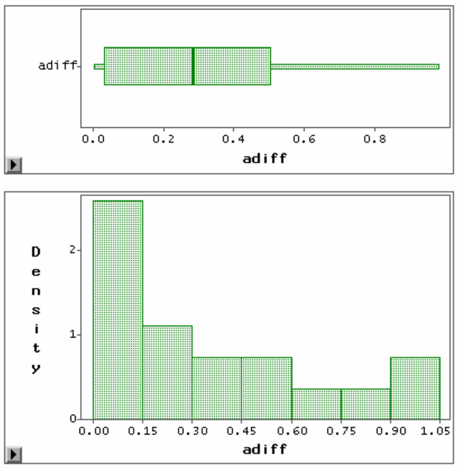
FRO7 RAF



Moments			
N	35.0000	Sum Wgts	35.0000
Mean	0.3349	Sum	11.7208
Std Dev	0.2196	Variance	0.0482
Skewness	0.2767	Kurtosis	-0.8355
USS	5.5652	CSS	1.6402
CV	65.5862	Std Mean	0.0371

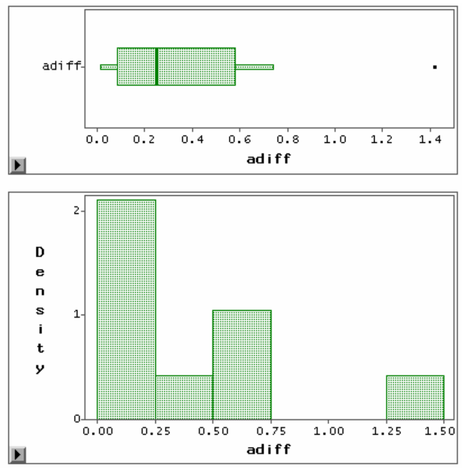
FRO7 RBF

Appendix Q: Error Distributions for IO Cross-Validation



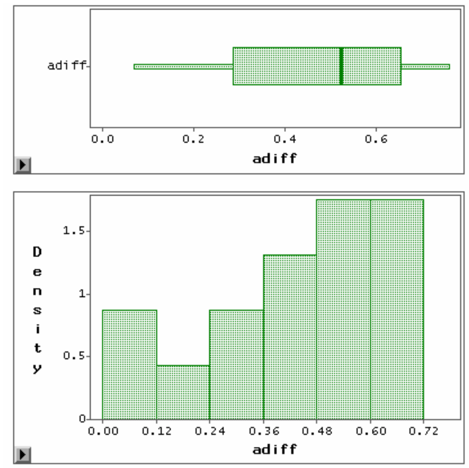
Moments			
N	18.0000	Sum Wgts	18.0000
Mean	0.3576	Sum	6.4368
Std Dev	0.3377	Variance	0.1140
Skewness	0.7631	Kurtosis	-0.5139
USS	4.2406	CSS	1.9388
CV	94.4378	Std Mean	0.0796

IO1 ALL



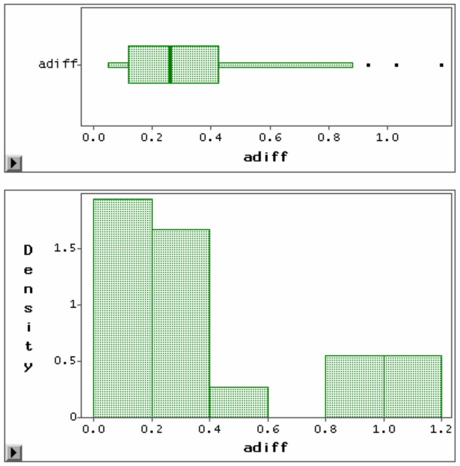
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.4125	Sum	7.8366
Std Dev	0.4339	Variance	0.1883
Skewness	1.4467	Kurtosis	1.5936
USS	6.6210	CSS	3.3888
CV	105.1983	Std Mean	0.0995

IO1 RAF



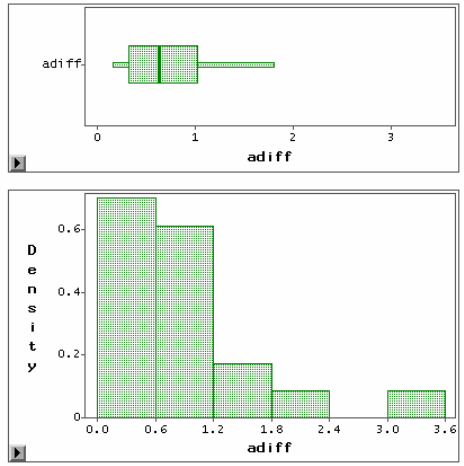
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.4771	Sum	9.0654
Std Dev	0.2179	Variance	0.0475
Skewness	-0.5048	Kurtosis	-0.8294
USS	5.1801	CSS	0.8548
CV	45.6728	Std Mean	0.0500

IO1 RBF



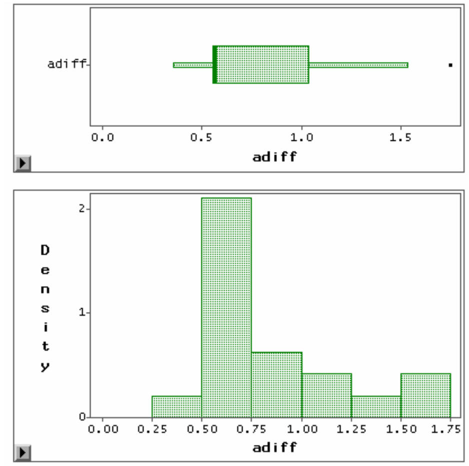
Moments			
N	18.0000	Sum Wgts	18.0000
Mean	0.3867	Sum	6.9601
Std Dev	0.3621	Variance	0.1311
Skewness	1.2330	Kurtosis	0.1626
USS	4.9204	CSS	2.2292
CV	93.6490	Std Mean	0.0854

IO2 ALL



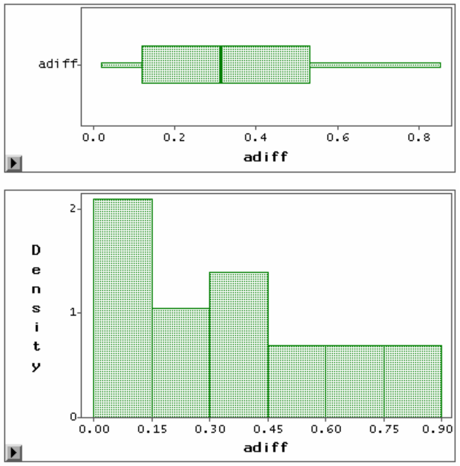
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8425	Sum	16.0081
Std Dev	0.7930	Variance	0.6289
Skewness	2.4570	Kurtosis	7.3773
USS	24.8079	CSS	11.3205
CV	94.1257	Std Mean	0.1819

IO2 RAF



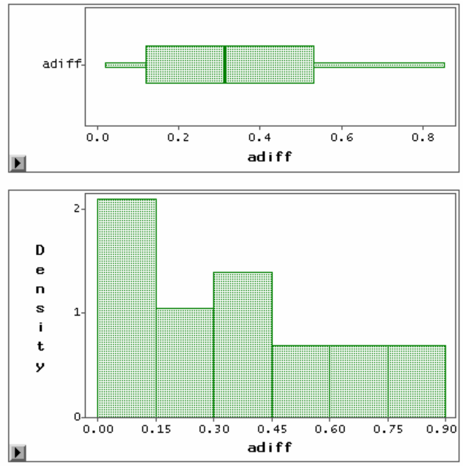
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8064	Sum	15.3218
Std Dev	0.3837	Variance	0.1473
Skewness	1.3038	Kurtosis	0.8855
USS	15.0064	CSS	2.6507
CV	47.5871	Std Mean	0.0880

IO2 RBF



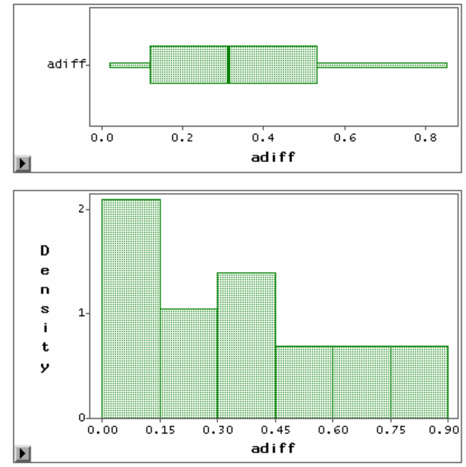
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.3496	Sum	6.6424
Std Dev	0.2560	Variance	0.0655
Skewness	0.7282	Kurtosis	-0.5202
USS	3.5021	CSS	1.1799
CV	73.2333	Std Mean	0.0587

IO3 ALL



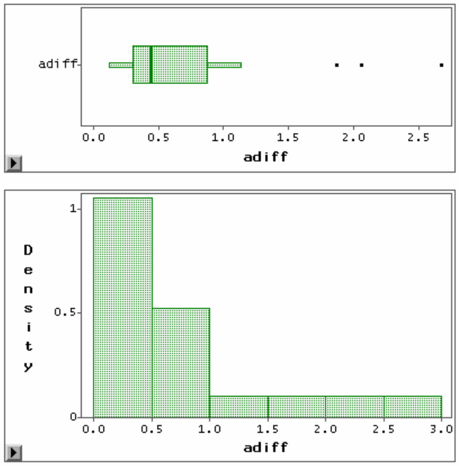
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.3496	Sum	6.6424
Std Dev	0.2560	Variance	0.0655
Skewness	0.7282	Kurtosis	-0.5202
USS	3.5021	CSS	1.1799
CV	73.2333	Std Mean	0.0587

IO3 RAF



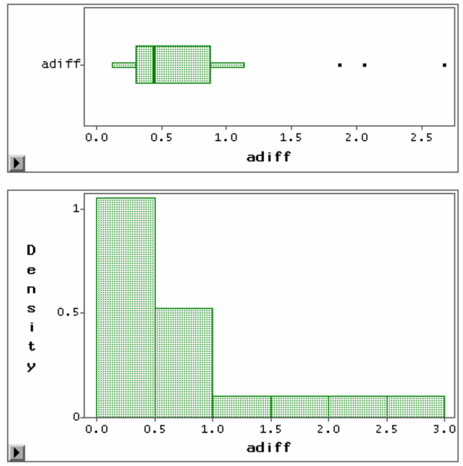
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.3496	Sum	6.6424
Std Dev	0.2560	Variance	0.0655
Skewness	0.7282	Kurtosis	-0.5202
USS	3.5021	CSS	1.1799
CV	73.2333	Std Mean	0.0587

IO3 RBF



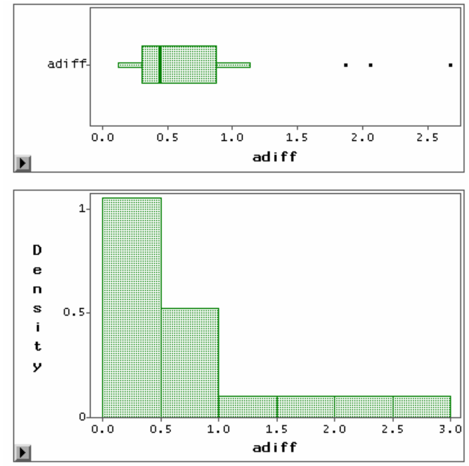
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8011	Sum	15.2215
Std Dev	0.6991	Variance	0.4887
Skewness	1.5570	Kurtosis	1.9219
USS	20.9915	CSS	8.7971
CV	87.2633	Std Mean	0.1604

IO4 ALL



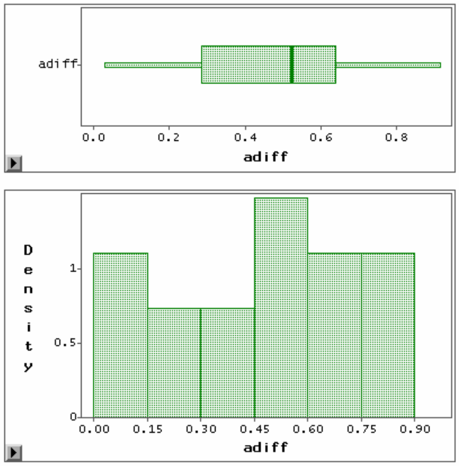
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8011	Sum	15.2215
Std Dev	0.6991	Variance	0.4887
Skewness	1.5570	Kurtosis	1.9219
USS	20.9915	CSS	8.7971
CV	87.2633	Std Mean	0.1604

IO4 RAF



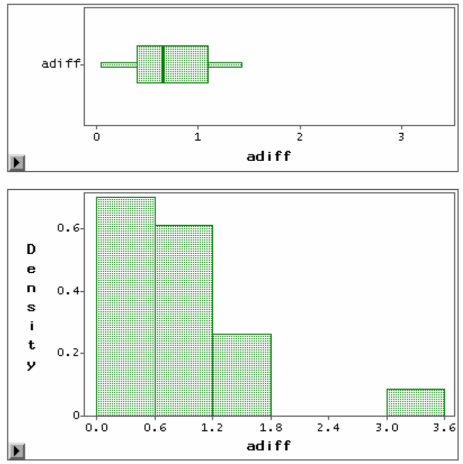
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8011	Sum	15.2215
Std Dev	0.6991	Variance	0.4887
Skewness	1.5570	Kurtosis	1.9219
USS	20.9915	CSS	8.7971
CV	87.2633	Std Mean	0.1604

IO4 RBF



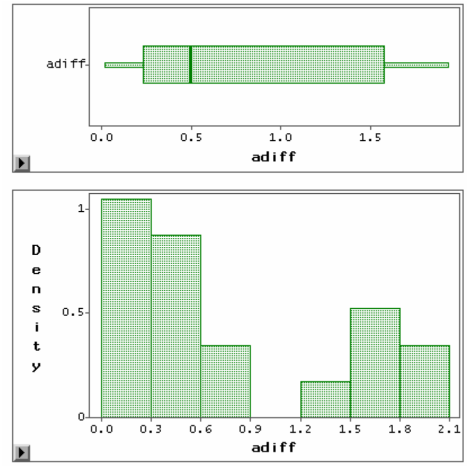
Moments			
N	18.0000	Sum Wgts	18.0000
Mean	0.4933	Sum	8.8786
Std Dev	0.2735	Variance	0.0748
Skewness	-0.1247	Kurtosis	-0.9719
USS	5.6514	CSS	1.2720
CV	55.4547	Std Mean	0.0645

IO5 ALL



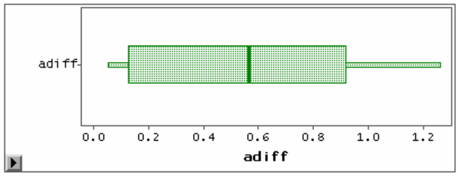
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.8266	Sum	15.7048
Std Dev	0.7396	Variance	0.5470
Skewness	2.5775	Kurtosis	8.4626
USS	22.8262	CSS	9.8452
CV	89.4740	Std Mean	0.1697

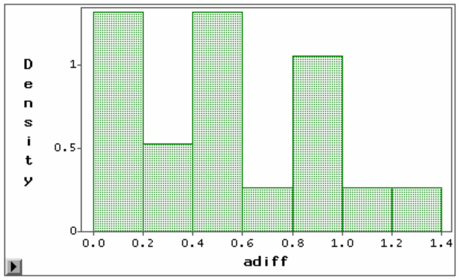
IO5 RAF



Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.7658	Sum	14.5505
Std Dev	0.6904	Variance	0.4767
Skewness	0.6600	Kurtosis	-1.2285
USS	19.7234	CSS	8.5804
CV	90.1554	Std Mean	0.1584

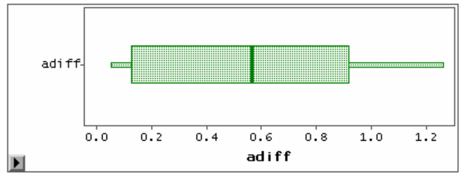
IO5 RBF

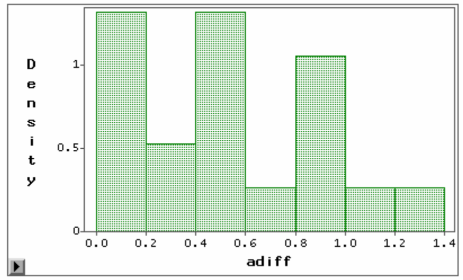




Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.5600	Sum	10.6402
Std Dev	0.3811	Variance	0.1452
Skewness	0.1623	Kurtosis	-1.1774
USS	8.5730	CSS	2.6143
CV	68.0533	Std Mean	0.0874

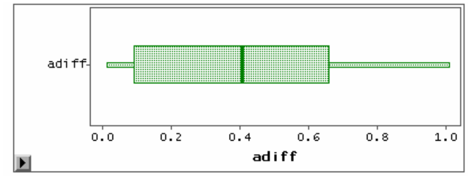
IO6 ALL

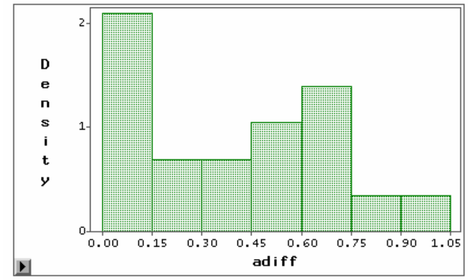




Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.5600	Sum	10.6402
Std Dev	0.3811	Variance	0.1452
Skewness	0.1623	Kurtosis	-1.1774
USS	8.5730	CSS	2.6143
CV	68.0533	Std Mean	0.0874

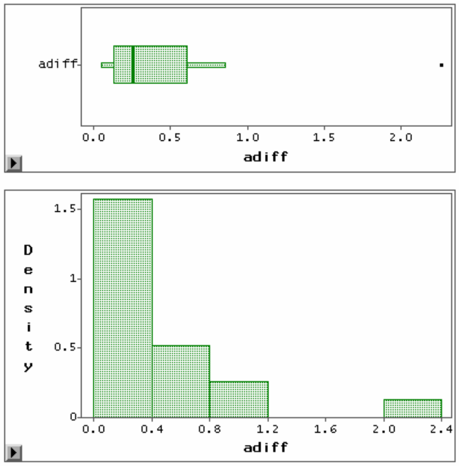
IO6 RAF





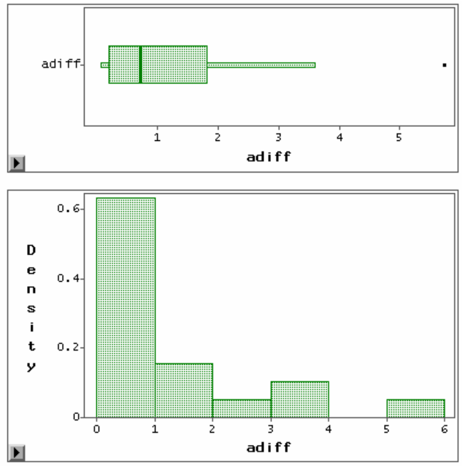
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.4163	Sum	7.9102
Std Dev	0.3096	Variance	0.0958
Skewness	0.1715	Kurtosis	-1.1636
USS	5.0180	CSS	1.7248
CV	74.3536	Std Mean	0.0710

IO6 RBF



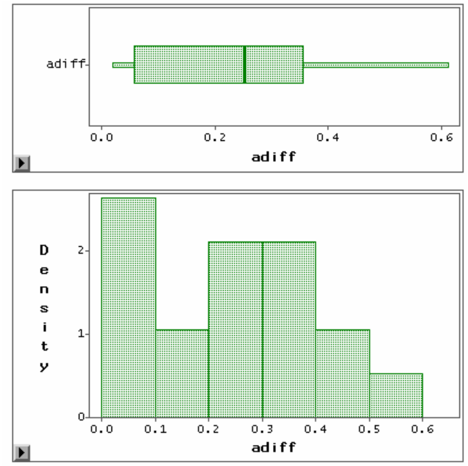
Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.4416	Sum	8.3910
Std Dev	0.5077	Variance	0.2578
Skewness	2.7798	Kurtosis	9.3789
USS	8.3459	CSS	4.6401
CV	114.9659	Std Mean	0.1165

IO7 ALL



Moments			
N	19.0000	Sum Wgts	19.0000
Mean	1.3074	Sum	24.8410
Std Dev	1.5208	Variance	2.3128
Skewness	1.7468	Kurtosis	2.8702
USS	74.1077	CSS	41.6301
CV	116.3195	Std Mean	0.3489

IO7 RAF



Moments			
N	19.0000	Sum Wgts	19.0000
Mean	0.2588	Sum	4.9176
Std Dev	0.1808	Variance	0.0327
Skewness	0.3606	Kurtosis	-0.6754
USS	1.8610	CSS	0.5882
CV	69.8456	Std Mean	0.0415

IO7 RBF