PROJECT SUMMARY

At Carnegie Mellon University, the Department of Statistics will build on its successful VIGRE program to (i) train postdoctoral fellows for careers emphasizing research in settings that require versatility, (ii) recruit and retain U.S. graduate students, avoiding excessive time to complete Ph.D.s while providing students with a high probability of success after graduation, and (iii) help increase the numbers of U.S. undergraduates, including women and minorities, with advanced training in statistical science. While maintaining a strong disciplinary foundation for statistical practice, the program emphasizes cross-disciplinary research and understanding the needs of statistical novices. All trainees will be involved in supervised, cross-disciplinary research, where they learn how to translate a research question into well-posed statistical problems, solve these problems, and translate the results back into a product that is accessible to the relevant scientific community. This skill is also central to learning basic statistics and forms a conceptual link between research and education, facilitating their integration. At the graduate level experience in the process includes a year-long project, typically with a faculty member in another domain, while a Statistics faculty member serves as advisor; a series of steps to improve communication skills and teaching effectiveness; and mentoring in professional growth. The graduate curriculum will be modified to make it more effective in building cross-disciplinary skills. Undergraduates will have several new courses available and will be involved in a capstone research project, and we will run a summer program, emphasizing minority students. Postdoctoral fellows will be involved in research projects, and will co-teach courses, with senior faculty; they will also participate in structured mentoring sessions. Undergraduate, graduate, and postdoctoral trainees will be integrated in research teams.
1.A. VISION, OBJECTIVES, AND ANTICIPATED IMPACT

This proposal seeks funding to continue, and improve upon, a training program in Statistics at Carnegie Mellon. Please note that FASTLANE restricts the number of co-investigators, so we chose to include formally only the Department Head and two Associate Heads as PI/co-PIs, but all of the regular faculty in the Department will be involved.

Our major goals are to provide students and postdoctoral fellows with adaptability and breadth, together with good communication skills that include an appreciation for the cognitive needs of the learner: successful trainees should be able to work effectively in teams, across disciplinary boundaries. Our view is that statisticians not only need to command a wide range of disciplinary techniques, but in addition must have the ability to draw out from a colleague the nature of an applied problem, identify potential paths to progress, and iterate this process effectively to make sure the end result will be useful. We believe that young statisticians can gain these skills efficiently through apprenticeship with experts, as long as they are given tasks appropriate to their current level of development and feedback to correct errors and reinforce good practices. Our approach therefore emphasizes this kind of experience throughout. In addition, we are aware that experts have well-developed abilities to “put themselves in the shoes of” their extra-disciplinary collaborators, who are trying to appreciate the way particular statistical methods can help them; the dialogue advances as the expert understands and instructs. Though this collaborative interaction is different from classroom instruction, some of the expertise is similar. Therefore, to improve cross-disciplinary work, as well as teaching, we will help trainees develop their instructional skills and deepen their understanding of the way people learn statistical ideas.

In fact, the interplay of effective cross-disciplinary work and effective teaching is fundamental to our vision of training statisticians. In cross-disciplinary work, trainees will learn how to translate a research question into well-posed statistical problems (A), solve these problems (B), and translate the results back into a product that is accessible to the relevant scientific community. We refer to this process, which is at the heart of statistical practice, as $ABA^{-1}$. We observe that students in basic statistical methods courses are also asked to step through the $ABA^{-1}$ process (though, of course, at a more elementary level). Thus, in developing instructional skills, trainees will again focus much of their attention on “$ABA^{-1}$,” and the ways learners grapple with its many sometimes-subtle aspects. We take this characterization, involving translation, knowledge, and communication, as a foundation for our integration of research and education, and also for our vertical integration among trainees at different levels. This is a vision that has evolved over many years, but as part of our new VIGRE program we aim to refine and improve on it, and to disseminate the results.

**Background** The Departments of Statistics and Mathematical Sciences at Carnegie Mellon University, jointly, received VIGRE funding for 1999-2001. A review of their first two years of activity was critical of Mathematical Sciences, and of some of the efforts of the two departments (which reside in separate Colleges at Carnegie Mellon) to work together. Statistics, however, was reviewed very favorably and we are therefore applying for our own VIGRE grant.

The context for our efforts is the environment at Carnegie Mellon which, as we will explain, makes it relatively easy to engage in cross-disciplinary research, and to integrate key features of the research and educational aspects of the training program.
Impact  Our VIGRE program will, of course, primarily benefit the trainees. We expect to increase (at least a little) the number of students who get advanced training in Statistics. Our outreach to Carnegie Mellon undergraduates is one focus of our efforts, and another is the pool of potential graduate students who will be convinced to enter a Statistics graduate program if they see that it offers a sufficiently hospitable and efficient means of achieving career goals. In addition, we have already attracted one Computer Science Ph.D. as a postdoctoral trainee and hope to find others.

We would also like to have a broader impact on the discipline as a whole and toward this end we plan to publicize various aspects of our program. We will highlight our training in cross-disciplinary research and teaching, spelling out, in particular, some of the details of linking the two by emphasizing communication and the $ABA^{-1}$ process. In addition, we plan to collect visually-appealing images and graphical displays, with supplementary talking points, for faculty to use as a basis for presentations at recruiting visits to undergraduate institutions.

1.B. OUTCOME OF CURRICULUM REVIEW

Leading up to our previous VIGRE proposal we undertook a substantial review of our curriculum, at both the graduate and undergraduate levels.

Graduate Curriculum  When we began working on our previous VIGRE proposal our Master’s program consisted of an intense mix of theoretical and applied courses to be completed in one year. Over the years the program had been updated to include a greater diversity of courses by reducing coverage per topic to “mini” courses (half-semester offerings). This allowed the program to incorporate a greater variety of modern topics. In practice, we discovered that the mini courses were often cramming too much material into the allocated time. The students felt overloaded and discouraged by the experience, and our retention rate (from M.S. to Ph.D.) was only around 50%. To attempt to solve these problems, we revised the schedule of course offerings, and set the “default” time for completion of the M.S. requirements at 1.5 years. This revision of the program increased flexibility for the heterogeneous group of incoming students our program attracts. In addition, student morale improved and retention increased substantially (to just over 75% across three years).

Nevertheless, some serious drawbacks of the current M.S. have appeared: (1) it delays exposure to research experiences; (2) students with a weaker mathematical background (usually U.S. citizens) who take two years to complete the M.S. have their schedules mis-aligned with the Ph.D. courses, which puts them at a disadvantage; and (3) students who discover they did not wish to pursue a Ph.D. often graduate at an awkward time of year and may be swept into the Ph.D. track without making an informed choice about this career path. Prompted by these weaknesses in the new program we examined our basic premises about the M.S. and Ph.D. programs and found that the curriculum could be updated and revised in such a way that we could retain many of the advantages of the new program without incurring the disadvantages. The proposed program returns to a 1 year M.S. for most students, with a 2-year option for those students who need time to bolster their mathematical training. We will require less depth at this stage, favoring exposure to many topics, rather than mathematical detail. We also have found structural ways to emphasize the $ABA^{-1}$ process, and thereby better prepare students for their cross-disciplinary work. In their first semester students will take a seminar called Issues in Statistical Pedagogy which will help them make the link between learning how to do statistics and learning how to teach it. Along the way, they will be asked to analyze data sets in scientific contexts to help them improve their own skills.
First-semester courses will be coordinated to make sure this aspect of statistical problem solving is adequately emphasized, and the students have a steady and manageable work load.

After students obtain their M.S. degree, most (currently, all but a few) go on to the Ph.D. The changes in our M.S. program allow complementary changes to our Ph.D. offerings. In part, this involves a re-organization of topics. One anticipated outcome is that students will be able to obtain a Ph.D. that has a substantial cross-disciplinary emphasis without dramatically lengthening their training. We expect this to occur because many students will become deeply involved in cross-disciplinary research earlier in their graduate training.

**Undergraduate Curriculum** When we reviewed our undergraduate curriculum in preparation for the previous VIGRE proposal we found we could create (i) new courses that would stimulate additional interest in Statistics, (ii) a new transition or “bridge” course that would help prepare students drawn in from other areas, and having weaker backgrounds, for advanced undergraduate training, and (iii) additional research experiences for undergraduates. In addition, we were interested in joining the Department of Mathematical Sciences in their long-standing and very successful Summer Undergraduate Institute for Applied Mathematics, in which a substantial percentage of trainees are women or are identified as under-represented minority students.

We have succeeded with all of the first three items and are pleased with the results: we now offer two special Freshman Seminars and two courses aimed at those with interests in computing; we have created a very successful bridge course in mathematical statistics; we have revised our advanced undergraduate data analysis course so that it now emphasizes a research project; and we offer increased opportunities for undergraduates to do projects both within and outside of classes.

As far as the Summer Institute is concerned, very few students were interested in doing statistics in that context and, as a result, we decided to run our own summer program. For the past three summers we have done so, and for the past two we have emphasized inclusion of students from Historically Black Colleges and Universities. This has been going very well and, in fact, will be a starting point for important further efforts toward increasing diversity in our Department and, we hope, in the field of Statistics. We describe the program in more detail below.

Because our undergraduate curriculum revisions under VIGRE are so recent, and appear so successful, we do not at the moment have plans for substantial modifications. We will, however, undertake another review after the 2002–2003 academic year.

**1.C. DESCRIPTION OF THE CORE COMPONENTS**

**1.C.1 OVERVIEW**

To appreciate what we are trying to accomplish, it is important to understand our environment. Carnegie Mellon is a small university (4800 undergraduates, 2700 graduate students, and 1000 faculty), which was created in 1967 by the merger of Carnegie Institute of Technology (founded in 1901) and the Mellon Institute (an industrial research organization). Currently, the Department of Statistics is part of the College of Humanities and Social Sciences, as opposed to the Mellon College of Science, where the Department of Mathematical Sciences resides. This separation of Statistics from Mathematics is unusual, and creates a certain awkwardness in the training of undergraduates in the mathematical sciences. On the other hand, the Department of Statistics was formed in a University environment that emphasized computation and the understanding of human behavior and decision making, and the separate evolution of Statistics within Humanities and
Social Sciences has had its advantages, including a closer relationship with Carnegie Mellon’s outstanding Department of Psychology, with its emphasis on cognitive psychology and computational models of mental processes. Our Department defined a path for itself by focusing on Bayesian statistics, computational statistics, and cross-disciplinary research when none of these was widely recognized as centrally important to a growing department. We now have 16 full-time faculty and 6 Visiting Assistant Professors (postdoctoral fellows) training 49 graduate students and teaching approximately 1500 undergraduates per year. Because of our commonalities of interests, we collaborate frequently, advance our applied research through fertilization of ideas across disciplinary projects, and maintain a vibrant collegial workplace with fluid discussion of exciting new ideas. We believe this provides an unusually stimulating and encouraging setting for undergraduates, graduate students, and postdoctoral fellows.

Carnegie Mellon’s historical emphasis on computing, cognitive science, and artificial intelligence emanated from early work by Alan Newell and Herbert Simon. Initial models of the mind were essentially computer algorithms for processing information in a sequence of ordered steps. Since then, the theories and techniques of cognitive psychology have been advanced and refined so that there is now an understanding not only of the way people decompose difficult problems into simpler subproblems but also (i) the way problem-solving knowledge is stored, (ii) the way different types of practice predict performance, and (iii) the way new skills are learned. These quantitative and statistical theories have produced important results for education by making detailed predictions of learning behavior, with demonstrated success. One consequence is that Carnegie Mellon’s very active Eberly Center for Excellence in Teaching is steeped in the substantial body of cognitive psychology knowledge about human memory, expertise, and learning. In addition, the involvement of cognitive psychologists with statisticians has led to a large collaborative research effort in understanding the way students learn basic statistics. Our colleague Marsha Lovett in Psychology, among others, has shown\(^1\) that learning is improved when the common structure across problems is highlighted, with goals and sub-goals being specified. The research conducted here has influenced our thinking about education generally, and about the training program described in this proposal.

We hope this background helps explain what we see as our distinctive traditions, namely, our emphasis on cross-disciplinary research and understanding the needs of learners in a context of disciplinary advancement. While we do not claim to be proposing revolutionary innovations, we do think we offer a rare ability to implement an integrated environment. We wish to emphasize this point partly because our proposal is, in large part, to continue and extend successful programs already in place. VIGRE funds would (i) allow us to train postdoctoral fellows for careers emphasizing research in settings that require versatility, (ii) help us recruit graduate students and increase the probability they would finish a Ph.D. while both avoiding excessive time to do so and increasing their likelihood of success after graduation, and (iii) help us increase the number of U.S. undergraduates, including women and minorities, who receive advanced training in the statistical sciences.

In outline, our plan involves postdoctoral fellows, graduate students, and undergraduates both separately and together. All trainees will work closely with senior faculty on disciplinary and cross-disciplinary research, often in vertically integrated teams. All will be involved in activities aimed at

---

improving communication skills. To develop and improve teaching ability, graduate students and postdoctoral fellows will be instructed and monitored by the University’s teaching center and by faculty in our Department. Labs and project courses will enrich the undergraduate experience. All trainees will receive formal mentoring, including an introduction to ethical issues in research. We will continue our successful efforts to attract very high quality women into our graduate program, and to retain and guide them on to the Ph.D. A cornerstone of our outreach to a diverse set of undergraduates is our Summer Undergraduate Research Experience in Statistics, which brings a small number of talented minority students into our department and immerses them in research projects.

The vertical integration in our program involves, primarily, the research team. Each team will be led by a senior faculty member and will, depending on the nature of the research problem and the interests of the trainees, include other faculty, postdoctoral researchers, graduate students, and undergraduates. Postdoctoral fellows get the guidance of senior faculty and the opportunity to develop the interpersonal skills needed to engage successfully in cross-disciplinary research. Graduate students receive the same benefits in addition to guidance by postdoctoral fellows near their own professional age. Undergraduates involved in vertically-integrated research teams gain from their interactions with near-peer graduate student colleagues. For undergraduates, it is important to keep their activities at an appropriate level commensurate with their skills. In our limited experience this team activity conveys well the considerable excitement of “doing real science,” and can persuade the undergraduate to go to graduate school or to change fields from a specific engineering or science discipline to statistics.

Cross-Disciplinary Projects

Because cross-disciplinary work plays a relatively large role in our proposal, we would like to stress the benefits of Carnegie Mellon’s environment. These days most active statisticians are involved in projects outside their own discipline, but it can be difficult to find scientific investigators or engineers who are open to the kind of deep and extended discussion it takes to produce a fruitful collaboration: many feel confident in their ability to formulate their problems and believe they can get by with well-established methods. In contrast, those who welcome full collaboration see statisticians as bringing essential skills and viewpoints to the research, thereby increasing the quality of the end product. This is the attitude we continually find here at Carnegie Mellon. As a result, we have many long-term, active projects. Some of these are itemized below. We have also included as our Appendix 2, letters of support from principal researchers in several of these areas. We should emphasize that research in all of these areas has involved graduate students, and in many of them it has involved undergraduates and/or postdoctoral trainees as well.

Astrophysics Genovese and Wasserman are founding members of PiCA, the Pittsburgh Computational Astrostatistics group. PiCA is a cross-disciplinary group consisting of astrophysicists, statisticians and computer scientists (http://www.picagroup.org/). There are 7 faculty, 5 post-docs, 8 graduate students and 2 technical staff members. Our primary goal is to make advances in cosmology through the analysis of massive sky survey data. Examples of current and past projects are galaxy clustering, analysis of the Cosmic Microwave Background radiation, estimating the equation of state for dark energy, fitting spikes in galaxy spectra, and estimating n-point correlation functions. Nonparametric statistical methods such as curve fitting play a vital role in much of our work.
Computational Finance and Financial Risk Management  Carnegie Mellon has become a center for research in computational finance and financial risk management. The research effort brings together faculty from the Department of Statistics (Brockwell, Kadane, Lehoczky, Schervish, Seidenfeld, Zhang), the Department of Mathematical Sciences, and the Graduate School of Industrial Administration. These units jointly administer a masters degree program in computational finance, the Department of Mathematical Sciences sponsors a Ph.D. program in mathematical finance, while the Center for Computational Finance sponsors research projects and periodic workshops. Research projects include an NSF Focused Research Grant on the mathematics of financial risk management that addresses the development of coherent measures of risk in dynamic environments, simulation methods for pricing and hedging complex financial instruments including those with early exercise opportunities, fitting heavy tailed distributions to financial data, the modeling of electricity prices, and the modeling of volatility of financial data.

Computational Neuroscience  Functional Magnetic Resonance Imaging (fMRI) is a rapidly developing tool that enables cognitive psychologists and neuroscientists to study the human brain in action. The data from an fMRI experiment consist of a time-series of three-dimensional images; they are a realization of a complex spatio-temporal process with many sources of variation, both biological and technological. In a large, continuing, and well-funded effort, Eddy, Genovese, and Lazar have been working closely with psychologists from Carnegie Mellon and the University of Pittsburgh and faculty in the Departments of Radiology, Neurology, Neurobiology, and Neurosurgery at the University of Pittsburgh Medical School to develop new methods and software for the processing and analysis of fMRI data. Kass, and Ventura have been working for several years on analysis of data from individual and multiply-recorded neurons in the brains of animals while they actively perform some task. (Brockwell has recently joined the team.) Statistically, this involves point process modeling, inference using simulation-based methods including Markov chain Monte Carlo and the Bootstrap, and ideas from the emerging field of functional data analysis. Scientifically, one of the fundamental problems is to determine how neurons carry information; technologically, a goal is to use a small number of neurons to guide a prosthetic robot arm in serving as a replacement for a hand made immobile due to accident or disease.

Computer Science  Several faculty are active collaborators on projects with computer science faculty and students. Eddy, Fienberg, Wasserman, and Seidenfeld all have joint appointments in the Center for Automated Learning and Discovery (CALD) in the School of Computer Science, and others interact with CALD less formally or on specific projects. Recent projects have involved machine learning, document classification, and robotics. We also have students cross-registered for the CALD M.S. degree in Statistical and Computational Learning. Fienberg’s work on disclosure limitation, as part of the project at the National Institute for Statistical Sciences under NSF’s Digital Government Initiative is based in CALD. Kadane’s work on computer intrusion detection also falls in this domain as does a new collaboration, involving several faculty, on computer security.

Educational Statistics  Junker collaborates both locally and internationally on methodology for educational statistics and psychometrics, including characterizations of model features, and practical computational Bayes methodology for hierarchically structured extensions of item response (survey or test question) models. Junker also collaborates with researchers at Carnegie Mellon, the University of Pittsburgh’s Learning Research and Development Center (LRDC), and the Educational Testing Service on various projects, including analyzing outcomes in observational and quasi-experimental studies in education; constructing and validating judging rubrics for instructional quality; and model-building and model-selection issues in applications of cognitive and de-
velopmental psychology in education. Kadane collaborates with workers at Carnegie Mellon and LRDC on detecting illicit answer-copying on examinations; studying the effects of absenteeism and student mobility on standardized test scores; and exploring the applicability of industrial quality control principles to educational outcomes. Fienberg is involved in research on relationships between latent variable models used in psychometrics and medical survey work. Greenhouse and Meyer analyze various designed educational experiments in conjunction with SMARTLAB (see our brief description of Pedagogy and Cognitive Science in Education, below).

**Engineering and Robotics** For several years, faculty in the department have been working with colleagues in Electrical and Computer Engineering, Civil Engineering, Mechanical Engineering, and Robotics. Schervish is developing methods of assessing the distributions of contaminants in drinking water supplies along with the costs associated with proposed regulations. He and colleagues are using pharmicokinetic models to construct relevant hierarchical statistical models. Schervish is also designing strategies based on prior information to help mobile robots search minefields, target ranges, and bomb sites for unexploded ordnance. Brockwell is applying time series methods to understand hard disk traffic. The ultimate goal of this work is to use prediction of read/write requests to a hard disk to design better caching algorithms. Lazar, working with a researcher in Carnegie Mellon’s Robotics Institute, has been developing methods for handling large medical data bases. Combining methods from statistics and computer vision, they have been exploring ways to reduce large feature spaces, and classify and retrieve images, in an automatic and self-consistent framework.

**Genetics and Bioinformatics** Eddy is extending ideas from imaging to some Proteomics data generated via a path-breaking method developed in the Carnegie Mellon Department of Biology. In collaboration with the Carnegie Mellon Philosophy group, Wasserman is applying DAGs to questions arising in the analysis of micro-array data and Genovese and Wasserman are developing theory and procedures for multiple testing and false discovery rates. Roeder and Wasserman are developing branching processes to model massive arrays of overdispersed count data representing RNA splicing patterns, from a novel procedure developed in the Department of Biology. Devlin, Roeder, Seltman, and Wasserman are using haplotype data and graphical models to understand the genetic basis of complex diseases based on studies of linkage disequilibrium. They are also developing methods for incorporating prior information on biological pathways when the number of possible models is vast.

**Pedagogy and Cognitive Science in Education** Meyer collaborates with Marsha Lovett (who was mentioned in our Overview) on applying cognitive psychology to statistics education. Meyer and Lovett, together with Greenhouse, Junker, Kass, and other researchers in Statistics, Psychology and Human-Computer Interaction (HCI), have developed SMARTLAB, a computerized learning environment based on “intelligent tutoring system” technology, in which students solve data-analysis problems on their own, with scaffolding (hints), supervision, and feedback provided by the computer. This teaching software, which plays a key role in our introductory statistics courses, has been shown in designed experiments to significantly reduce student errors in choosing the appropriate analysis for a problem. SMARTLAB is also a research tool that can record student-SMARTLAB interactions at a very fine-grained level. A new collaboration between Statistics, Psychology, HCI and instructional technology communities on campus—including the Vice Provost for Computing Services at Carnegie Mellon—aims to place a more flexible version of SMARTLAB in a broad variety of AP-Statistics and similar classrooms throughout the country, via the World Wide Web. This work is central in pushing our understanding of student learning not only
Mentoring of graduate students and postdoctoral fellows in our Department takes place individually and in groups. Graduate students are mentored by academic advisors, advisors of their research projects, and their thesis advisors. In addition, we plan to repeat a successful mentoring seminar for graduate students, which integrated a postdoctoral fellow as a co-instructor. The postdoctoral fellow helps bridge the age and experience gap between the senior faculty member and the students, and offers a perspective based on current grappling with many of the issues under discussion. Postdoctoral fellows are themselves assigned to an individual who serves as mentor and, in addition, they meet regularly along with all first-term junior faculty for group mentoring sessions with senior faculty. The topics covered in individual and group meetings, for both graduate students and postdoctoral fellows, include Research: theory and methodology; cross-disciplinary work; collaboration; finding problems; writing proposals; attending conferences; Teaching: classroom teaching; advising students; project supervision; Professional activity: refereeing; organizing conferences; other professional service; and Time management.

Mentoring will also explicitly address issues of ethics in research and academic life. Some of these will arise naturally in the interdisciplinary collaboration and publication process, but formal sessions will also be held to discuss various topics, including Written work: working together as a team across seniority levels; determining authorship; plagiarism and acknowledgment of pre-
vious work; responsibility to publish; redundancy or duplication of publications; *Reviewing the work of others:* confidentiality; impartiality; misuse of information; *Oral presentations:* accuracy; acknowledgment of previous results; responsibility to educate; *Grants:* honesty in proposals; accuracy of budgets; responsible use of funds; *Cross-disciplinary work and consulting:* integrity in reporting results; careful attention to statistical assumptions; intellectual property; financial considerations; *Data access:* maintaining confidentiality; making data available; *Scientific misconduct:* kinds of misconduct; complexity of typical scenarios; responsibility and appropriate channels for reporting; and *Job hunting:* honest representation of self.

1.C.3 GRADUATE TRAINING

Our graduate program, which is being revised in concert with this VIGRE proposal, will put beginning students through statistical theory and methods courses while at the same time exposing them to essential ideas of statistics education and preparing them for, then giving them practice in, cross-disciplinary statistical research. At the outset the program is streamlined, allowing students to obtain their M.S. degree quickly and to become acquainted with some of the most exciting tools and applications in the field. The Ph.D. degree provides greater depth of study, with preliminary examinations in statistical theory and probability. A third preliminary examination is based on a year-long cross-disciplinary research project conducted with two faculty members as committee members, one from inside and one outside the Department of Statistics. We also plan to continue, and make more accessible, an alternative path to the Ph.D. degree. This is an option that allows substitution of courses in a substantive field (e.g., genetics or neuroscience) for two advanced theory courses. With this option we anticipate graduating more students capable of engaging in true cross-disciplinary research at an earlier stage in their careers. Moreover, this option makes the Ph.D. degree a realistic possibility for students coming from other fields, who may have considerable mathematical intuition but lack the background or interest in mastering in depth some of the mathematical topics required for our more traditional Ph.D. degree in Statistics.

**Master’s Program** The Master’s program will consist of four year-long sequences: *Theory* (statistics, probability and stochastic processes), *Linear Models* (applied regression analysis, generalized linear models, theory of linear models, design of experiments, and time series), *Computationally-Oriented Methods* (statistical computing, multivariate exploratory data analysis, data mining, dimension reduction techniques, posterior simulation), and *Statistical Practice*. The latter sequence emphasizes training in the $ABA^{-1}$ process identified in our Overview section. It will begin with focused practice in some of the basic skills (described in part again below, under *Training in Teaching*), including the steps involved in translating scientific into statistical problems, and both written and oral communication skills. Students will participate in an applied research project working with scientists in other fields under the supervision of faculty and, sometimes, postdoctoral trainees in our department. In addition, these students will learn by hearing about the progress of their classmates, and by sharing in much of the feedback they are given.

We will shorten our Master’s program by redesigning our core courses so that they expose students to a variety of techniques found in the contemporary world of applied statistics. Students will study many of the methods in greater depth as elective Ph.D. studies.

**Ph.D. Program** The core Ph.D. courses are three year-long sequences, *Advanced Statistical Inference*, *Advanced Probability Theory*, and *Advanced Data Analysis*, but only the first semester of our statistics and probability courses will be required of all students. In this way we will make it easier for students to substitute alternative disciplinary knowledge for the additional theory pro-
vided in the second semesters. Thus, students who wish to pursue a more heavily cross-disciplinary Ph.D. will be able to do so more easily, without dramatically lengthening their time in graduate school. To accomplish this, the probability and statistics courses will be radically restructured so that the first semester of each will cover key results required for a solid understanding of statistical practice.

We have made a pilot run through our new Advanced Probability course. In this new course we aimed to cover a year’s worth of material, from basic measure theory to discrete-time martingales, in a semester. This necessitated the removal of most of the proofs from lecture, which were assigned as out of class reading. Homework assignments did require the students to carry out proofs, at the same level of rigor as in the standard, one-year sequence, providing additional reinforcement of those crucial skills. While the bulk of the students took the traditional course in Advanced Probability, two took the new version: one is pursuing a joint PhD with the Public Policy School at Carnegie Mellon, the other is focusing on Data Mining; both students are American, and female. Although there was concern at the outset, both among faculty members and the students themselves, about whether the overview course would give sufficient exposure to concepts and tools, and about whether sacrificing depth would result in a “watered down” version of the material, the outcomes have been favorable. Although it is impossible to draw firm inferences based on two students, and in the presence of many confounding variables, we were extremely pleased to observe that this pair performed very well in the Advanced Statistical Theory course and passed the Ph.D. qualifying exam at a level comparable to the students who took the year-long probability sequence.

The Advanced Data Analysis sequence is taken after the M.S. program is completed. Like Statistical Practice, this course stresses training in the ABA⁻¹ process. It will include a discussion of data quality, manipulation of large data sets, and advanced data analysis techniques that arise in the context of the research projects, but its main component is a very substantial cross-disciplinary research project, with both a Statistics faculty member and a faculty member from another department (who supplies the data) as advisors. The student must write a substantial report (which we hope will lead to a research publication in the other discipline), and present their work orally to the entire Department.

Training in Teaching One of the fundamental goals of the VIGRE program has been to prepare graduate students for the many demands they may face as faculty members. Initially, the VIGRE management team insisted that we require graduate trainees to teach their own courses during the academic year. We appealed, however, arguing successfully that not only would this be difficult at Carnegie Mellon (where the use of graduate students as instructors is tightly controlled), but such a move would do more harm than good in our environment. We believe that what Ph.D. students destined for academic careers in statistics need most in order to adjust to teaching demands in their new environments is teaching ability. Teaching a course does not guarantee knowledge of how to teach, nor would it necessarily provide any insight into the way to succeed while juggling diverse activities. Indeed, nothing in our experience suggests that merely by having a course under their belt will a junior faculty member be better able to manage their time and teach well. We have found that junior faculty benefit most from an understanding of how students learn, and how to structure classes to facilitate the learning process. These ideas must be practiced in a supervised situation where novice instructors are given feedback on the effectiveness of their efforts. Our view on the best way to make sure Ph.D. students learn how to teach is no different than our view on the
most effective way to make sure undergraduates learn basic elements of statistics: in either case, students must master well-defined skills, and this can be achieved over time by instruction in basic principles together with practice and feedback.

We have designed our program in order to help students build their teaching skills gradually, with expert supervision. It includes the following elements: (i) Instruction on what cognitive science has, through careful experimentation, demonstrated about the ways people learn; (ii) Discussion of the relationship between what it takes to be effective as a teacher and what it takes to be effective as a cross-disciplinary team member, together with extensive mentored practice in cross-disciplinary statistical work; (iii) Discussion of both general and course-specific teaching issues; (iv) Supervised classroom practice; and (v) Discussion of time management.

Trainees go through the following:

- Introductory lectures on teaching and learning during our orientation program.
- The year-long sequence in *Statistical Practice*, including a segment on *Issues in Statistical Pedagogy* in the Fall of their first year. In the seminar-style segment on pedagogy we will discuss the essential tasks any data analyst must face, which we have itemized as follows: (1) Understand the problem (check data format; consider study design), (2) Reflect on the question (state expected findings; identify relevant variables), (3) Analyze the data (classify the variables; perform exploratory analyses; conduct formal analyses; report results), (4) Draw conclusions (recall expectations; consider what results mean), (5) Summarize (choose key results to report; place conclusions in substantive context). Trainees will have to carry out these tasks themselves in working on a data-analysis problem, and we will discuss with them the steps that most people find difficult. We will also discuss communication skills and the ways to make productive the interactions between instructors/assistants and students, and we will examine a few concepts that students typically find particularly difficult. Trainees will be given feedback on their communications with students.
- Mentored feedback on cross-disciplinary and oral communication skills. This occurs in the cross-disciplinary project courses, and in supervised cross-disciplinary research. In addition, students are required to make two formal presentations: their first project is presented to the first cross-disciplinary project class (*Statistical Practice*) and its instructors, then the second (*Advanced Data Analysis*) is given to the entire department. The latter particularly is taken very seriously; it involves extensive discussion and practice with an advisor and we have been very impressed with the high quality of the results. We believe our repeated review of the elements of successful presentation, together with our emphasis on careful preparation, help students develop and improve this crucial teaching skill.
- TA activities, beginning with holding office hours and running labs, grading, and tending to administrative matters in course delivery. Instructors will be conscious of their supervisory role in helping the trainees with their apprenticeship.
- Associate Instructor activities, including team teaching. Here the trainee not only participates in planning the course and, where appropriate, developing the syllabus, but also has substantial responsibility for it, including sometimes selecting the homework, improving the lab sessions, and occasionally giving lectures. In this process, the Associate Instructor will
be mentored by the faculty member teaching the course, particularly with regard to the lectures they give. The trainee will take significant administrative responsibility for the course, including supervising the other TAs, overseeing all grading activities, and posting solutions for homework assignments.

- Mentoring discussions (see the section on mentoring, above), some of which focus on teaching and time-management issues.

- Teaching teas, which are informal discussions of teaching issues and run roughly 2-3 times per semester.

- A series of seminars by our Teaching Center (in which one of the three senior staff members has a Ph.D. in cognitive science, and another has a Ph.D. in Statistics), aimed at preparing students for teaching their own courses. The seminars go over topics such as planning and syllabus construction, lecturing skills (including watching videotapes of good statistics graduate student and faculty lecturers), motivating students, grading and student evaluation, etc.

- Teaching, during the summer. The classes we offer during the summer are the most heavily-populated basic statistics courses in our curriculum. Enrollments are generally between about 12 and 30 students. Several of our faculty take turns meeting with the summer instructors, beginning shortly before the summer session begins, to provide advice and supervision, help with planning and anticipating difficulties, and assistance with evaluating effectiveness. The instructors are often observed and videotaped, and they are critiqued by teaching center personnel on a confidential basis.

   A final point concerning our VIGRE graduate training is that we will attempt to minimize the discrepancies between trainees and their non-U.S. counterparts. There will be some inequalities in work load and compensation, but the main intellectual benefits of the training program will be available to all students. For example, aside from students with severe language difficulties (whom we try to avoid on admission), we expect all first-year students to be in the Statistical Practice sequence, including Issues in Statistical Pedagogy. Similarly, we expect foreign students with strong English language skills will go through the full training in teaching.

1.C.4 UNDERGRADUATE TRAINING

The Department of Statistics sees undergraduates in classrooms and in research projects at every level and in every college and school at Carnegie Mellon. Under our previous VIGRE grant, we have expanded the number of majors through which students may study statistics, and developed new courses intended to appeal to more students. Supervised practice, collaborative learning, and timely feedback are now staples in all of our regular freshman/sophomore courses, as well as many junior/senior ones.

Freshmen are exposed to Statistical Reasoning and Research Methods in lively introductory courses that incorporate computer-aided data analysis (often supervised by SMARTLAB), in-class discussion and demonstration, and consideration of case studies. We also offer smaller Freshman Seminars exposing students to statistics and visualization techniques through text analysis, and exploring statistics though the policy and demography issues in the U.S. Census. Intermediate
level courses in *Experimental Design, Survey Methodology, Statistical Graphics,* and *Data Mining* are available to Statistics minors and majors and also to other Bachelor of Science majors who need additional methodology in their programs. The latter two are new, and capture students who would otherwise not take an intermediate-level course in Statistics. We hope this will become a good source of additional Statistics majors and minors. Junior/Senior level courses in *Probability, Mathematical Statistics* and *Modern Regression* are taken by Statistics majors and minors, as well as students in related fields such as Computer Science and Mathematics. In our capstone course *Advanced Data Analysis,* majors in Statistics, Information Systems, Mathematics, Economics, Psychology, and other fields work in teams supervised by a statistics faculty member or graduate student, and often a non-statistics researcher, on a substantial data analysis project. In order to help undergraduates move more quickly into the capstone research course, a research team, or an independent research project with a faculty member, we have collapsed the introductory two-semester freshman statistics course into a single semester, and similarly created a single-semester “bridge” course as an optional alternative to the standard two-semester sophomore/junior level mathematical statistics course. We have also created a new version of our honors-level mathematical statistics course, which now emphasizes connections between Statistics and Machine Learning.

Most undergraduate research projects arise in our intermediate-level *Survey Methodology* and senior-level capstone *Advanced Data Analysis* courses. The structure of these courses mimics the vertically integrated research teams that carry out much of the interdisciplinary work outside of classes. Thus, the same principles—and the same training of teacher/mentors—that work in the more controlled classroom setting should also be applicable to research teams, where many undergraduate students eventually find themselves as they look for independent study projects. We will maintain the basic structure of these courses, but with some additional mentoring of the TA’s and junior instructors, aimed at helping undergraduates with this transition from classroom to “live” research teams. This mentoring will occur through regular meetings among the instructional staff for the courses, and also through the Teaching Training mentioned earlier.

**Summer Undergraduate Research Experience in Statistics** With partial support from the previous VIGRE training grant, we have offered for the past three summers (2000–2002) a program for undergraduates intended to show them the excitement and utility of statistical analysis in an interesting scientific context, and to help them develop their statistical thinking. Although these students come from varied backgrounds, we aim to draw some of them into careers in the statistical sciences. Undergraduate interns in the program work in various research teams, typically composed of a statistics faculty member, at least one graduate student, and often a clinical investigator, on a substantial research project. Interns also participate in colloquia facilitated by faculty and graduate students.

Starting in the summer of 2001, we have partnered with a program sponsored by Morehouse College in Atlanta called Project IMHOTEP which places students from Historically Black Colleges and Universities in internships in biostatistics and epidemiology. Four of the seven 2001 interns, and four of the seven 2002 interns, were women of color from the Morehouse program. The remaining three interns in both years were undergraduates of mixed gender and race from Carnegie Mellon.

At the core of our summer program is experiential learning through the supervised practice of research. Research topics have included directional tuning of neurons, evaluation of an intelligent

---

2See [http://www.morehouse.edu/publichealth/imhotep/index.htm](http://www.morehouse.edu/publichealth/imhotep/index.htm).
cognitive tutor, the temporal-spatial distribution of sexually transmitted diseases, statistical models for biological rhythm data, and analysis of national health survey data to investigate questions such as factors related to perceived BMI, relationship between socio-economic status and hypertension, and factors related to testing for HIV in OB/GYN ambulatory care settings.

For the IMHOTEP students, the program has started with two weeks of training at Morehouse College. All students participate in a two-week introduction at Carnegie Mellon to statistical methods and practice, computing, and to the research projects, followed by six weeks of focused hands-on research. At the end of the summer, students write a paper describing their project and results and present their work at a public poster session.

In addition to hands-on research experience, the program also provides summer interns with a broad exposure to the statistical sciences. Approximately twice a week for an hour, faculty and graduate students meet with interns to discuss topics such as alternative approaches to data analysis, computing, epidemiology, academic life, the research process, and ethics. These meetings also include informal “Perspectives on Statistics” presentations given by faculty and graduate students, with personal reflections on how they became interested in statistics, what they like about their chosen field, and usually an introductory overview to the presenter’s research. These interactions solidify relationships between interns, graduate students, and faculty. They also provide undergraduates with invaluable pointers on applying to graduate schools and professional positions.

All the students have provided very favorable feedback about the program and their experience at Carnegie Mellon.

As we indicated, VIGRE has supplied some initial partial support for this program. To be successful, it is essential that we generate support for faculty time, and we will seek such funding. In addition, we expect to modify our participation in IMHOTEP so that we can get additional applications from students in their junior year who may be oriented less toward public health and more toward the sciences, including computer science and mathematical and statistical sciences.

1.C.5 POSTDOCTORAL TRAINING

We aim to help postdoctoral trainees develop the diverse set of skills they will need to succeed as new faculty members, or as research workers in cross-disciplinary environments. Specifically, they will be involved in disciplinary research, cross-disciplinary research, and teaching with active mentoring and assistance; throughout their training they will have a reduced course load. We request VIGRE support for four postdoctoral fellows per year; we will supplement this with University support for teaching and, where possible, additional external funding from cross-disciplinary projects. We appoint postdoctoral fellows as Visiting Assistant Professors. They are sheltered from some aspects of running the department, they are given a reduced teaching load with some co-teaching and are expected to join at least one cross-disciplinary research team, but in most other respects they are treated like regular faculty members.

Research Post-doctoral fellows will become involved in several projects, some disciplinary, others cross-disciplinary. Cross-disciplinary research is a major strength of our Department, and we have itemized several of the currently-available problem areas in our Overview. Fellows will be mentored partners with one or more regular faculty. There is no specific mechanism for matching visitors to research projects; we have found that it suffices for faculty and fellows to give presentations of their work, and for there to be an expectation by both visitors and the permanent faculty that collaboration will take place. In this environment, visitors find projects and faculty with whom they are comfortable. The Department Head will monitor the situation to make sure that trainees
do, in fact, find suitable research partners and mentors.

**Teaching** As we mentioned in our introduction, one of the defining characteristics of Carnegie Mellon’s evolution has been the prominent role of research in cognitive science and artificial intelligence, and its direct descendant, the strong and visible research in the theory of learning. Postdoctoral fellows have regularly received an orientation to the role of cognitive psychology in understanding learning at the beginning of their stay here. They also participate in discussions on this topic as part of our “teaching tea” series, which meets two or three times per semester. This is an outgrowth of the Eberly Teaching Center’s lunch series, in which faculty join focused discussions over lunch on several occasions during each semester.

Courses taught by postdoctoral trainees are monitored by senior faculty as well as the Teaching Center. Postdoctoral fellows receive assistance from the Teaching Center in the form of confidential discussion, class visitation, and optional videotaping. Because these consultations are confidential, they are supportive and generally productive. We will also attempt to have each postdoctoral trainee co-teach a course with a senior faculty member. We have done this under our previous VIGRE support and believe it is very beneficial to the trainees.

In addition to getting support and assistance with their teaching fellows will be able to participate in supervision of undergraduate student projects, and to talk with undergraduates about research within the cross-disciplinary research teams. This is considerably more difficult than teaching from a standard text, and at Carnegie Mellon we have some non-threatening points of entry for young postdocs into this arena. Short projects are completed in our Summer Undergraduate Research program and semester-long projects will be conducted by undergraduates in the existing course “Advanced Data Analysis II.” In these settings, regular faculty have primary responsibility and post-doctoral fellows will assist. Because these activities may also involve graduate students in junior mentoring roles, they afford an opportunity for full vertically integrated learning.

**Formal Mentoring** As we described above, in our section on Mentoring, postdoctoral fellows will have an individual mentor and will also participate in group mentoring sessions together with all first-term junior faculty. In addition, the Department Head will meet with the fellows to make sure the process is working properly.

### 2. RECRUITMENT AND RETENTION

We understand VIGRE to be promoting both the recruitment of talented young U.S. citizens into the mathematical sciences, and, in particular, the recruitment of women and under-represented minorities. Our approach to these twin recruitment efforts involves (i) a program that is geared toward the interests and needs of these students, (ii) an environment that is engaging and supportive, and (iii) a focused publicity campaign that will attempt to showcase the excitement of statistical work here at Carnegie Mellon and the wide-ranging demand for well-trained statisticians.

We have already discussed above (under Outcome of Curriculum Review) our efforts and successes in modifying both the M.S. and Ph.D. programs so that they will be both more attractive and more manageable for the diverse set of U.S. citizens we have been and will continue to be aiming at under the VIGRE program. Historically, we have done well in bringing U.S. citizens into our program, and the modifications seem to be especially important for women. We have for many years placed great emphasis on maintaining a collegial, relatively non-competitive working environment, where we do everything we can to help each student get the training they need to meet their career goals. This involves the way we structure our requirements, assess our students, and give them feedback and advice. As one particular example, we have chosen to run our own
small course in Real Analysis, taught by a senior graduate student during the summer and taken by selected graduate students after their first year in our program. This is essentially an extended series of group-tutorial sessions, building up skills needed for the advanced probability and statistical theory courses. It has been very successful. We have also worked hard to expand the presence of women and minorities on our faculty, and now have 7 women on our faculty, one of whom, a Visiting Assistant Professor, Kimberly Sellers, is an African-American woman. In 1995 just under 40% of our entering graduate students (averaged across several years) were women, but there was somewhat greater attrition among women than men. Over the past three years 16 out of 33 entering graduate students (48%) have been women and currently 22 out of 49 students (45%) are women.

While we believe we are doing reasonably well with U.S. citizens, and women in particular, and have taken steps to make further improvements, we are very conscious of the national need for greater numbers of U.S. citizens, women, and minorities in graduate programs in Statistics, and we have ourselves had only a small number of Hispanic and African-American students, at any level. We propose to take an additional step involving more active recruitment. This comes, in part, from our experience with our new Summer Program in Undergraduate Research, which has been designed specifically to bring minority students from Historically Black Colleges and Universities to the campus to learn about opportunities in statistics and to carry out research projects. As we have considered how to reach additional students, we have felt that some of the best opportunities will come from outside mathematics: we would like to do our part to enlighten more quantitatively and computationally-oriented undergraduates about the exciting opportunities in this field.

Our plan is to construct a self-contained highly visual and entertaining presentation, and to have it given by various faculty on recruiting visits to selected institutions, especially (though not exclusively) Historically Black Colleges and Universities. We envision a presentation featuring perhaps four selected application areas, such as those itemized in our Overview, where we highlight the role of data analysis in important problems of science, technology, and business. We would avoid the use of the word “Statistics” until near the end, when we would draw together the applications, explain their commonalities, and discuss what it is that statisticians do. We would title the presentation something like *Turning Information into Knowledge*. In our first iteration we will aim to make the presentation effective, but we will keep it modest in production cost, using one or two graduate students who will collect the graphics from relevant faculty and help draft the talking points. Once we have a version we can try out, we will go to a small number of institutions, and then iterate. We would also expect to interest the much wider Statistics community (e.g., the American Statistical Association) in producing a high-quality, serious (and much more expensive) version of this presentation.

**3. ORGANIZATION AND MANAGEMENT PLAN**

The primary responsibility for managing VIGRE will lie with the Department Head. The Associate Heads will supervise the graduate and undergraduate aspects of VIGRE, and the Department Head will supervise the postdoctoral training program. We expect the Summer Program to be supervised by Joel Greenhouse, and mentoring activities to be supervised by Jay Kadane together with the Department Head. We have been, and intend to continue committing 10% time of a staff member, Margie Smykla, to assisting with the oversight, management, and data collection. Because all of our regular faculty members will be involved with VIGRE, we plan to discuss the VIGRE training program during several faculty meetings throughout the year.
The lower-level management of our VIGRE program is summarized in the accompanying table, where Sr. Fac. refers to senior faculty, Sr. Fac. Leader refers to the cross-disciplinary research team leader, and TA Supervisor refers to the graduate student Associate Instructor or head TA. We consider the redundancy in the system (e.g., a postdoctoral trainee may be guided by different faculty on disciplinary and cross-disciplinary projects) a very good thing, though it remains essential that we have whole-faculty discussions about all our trainees, and the program, at least once per semester.

<table>
<thead>
<tr>
<th>Activity Group</th>
<th>Research</th>
<th>Teaching</th>
<th>Mentoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergrad.</td>
<td>Sr. Fac. Leader</td>
<td>Course Instructor</td>
<td>Sr. Fac. Leader</td>
</tr>
<tr>
<td></td>
<td>Course Instructor</td>
<td>TA Supervisor</td>
<td>Course Instructor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teaching Center</td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>Sr. Fac. Leader</td>
<td>Course Instructor</td>
<td>Sr. Fac. Leader</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TA Supervisor</td>
<td>Academic Advisor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teaching Center</td>
<td>Course Instructor</td>
</tr>
<tr>
<td>Postdoc</td>
<td>Sr. Fac. Leader</td>
<td>Sr. Fac. Co-Instructor</td>
<td>Sr. Fac. Leader</td>
</tr>
<tr>
<td></td>
<td>Sr. Fac.</td>
<td>Teaching Center</td>
<td>Sr. Fac. Co-Instructor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mentor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Group Mentoring Fac.</td>
</tr>
</tbody>
</table>

4. PERFORMANCE ASSESSMENT

We are keenly aware that in a formal sense it will be impossible to determine definitively the effectiveness of our program: we cannot run a randomized experiment in which some (unidentified) students and postdoctoral fellows go through the program while others do not, nor can we wait the many years it would take to find out how successful they ultimately become. Instead, we must rely on expert judgment. We previously asked Professor Mary Ellen Bock (and a mathematician, because previously the effort was joint with the Department of Mathematical Sciences) to serve as an external evaluator, which was very helpful. This provides not only another perspective, but also a filter that allows us to get anonymous comments. The evaluators will come to our department for a site visit, and have access to confidential correspondence with trainees and former trainees.

The general issues we will ask our evaluators to examine include the following:
• Are we attracting good postdoctoral fellows and providing them with well-rounded and important additional training?

• Does the postdoctoral training we provide help fellows become contributors to areas where there are national needs in mathematics and statistics, through disciplinary and cross-disciplinary research and teaching? Which aspects of the training are most (or least) helpful?

• Are we attracting high-quality U.S. students to our graduate programs, and do we maintain a diverse graduate environment?

• Are we providing graduate students with the training they need to succeed in academic, government, or industrial research settings? Which aspects of the training are most helpful?

• Are we successfully shepherding U.S. students, including women, through our graduate programs?

• Are we doing a good job in encouraging undergraduates, including women and minorities, to continue with careers in mathematics and statistics? Which parts of our program and curriculum make the most positive impression, or are of greatest lasting value?

• How well does our model of vertical integration work, and how much is it valued by participants?

• How successful have we been in producing acceptable average time-to-degree for graduate students?

To assess the program we will continue to collect data (number of students, length of time to completion of degree, publications, etc.) and use questionnaires, and we will provide the evaluators with a variety of materials.

Research Evaluation of disciplinary and cross-disciplinary research must be based in part on the quality of the reports and publications that are produced. We will provide summaries of these, including a statement of significance of contribution. We will also ask all collaborators to complete a short survey:

1. Please describe the status of the person, the activities in which she or he is engaged, and summarize any papers, including working papers, which she or he has co-authored with you.

2. Does the person have or is the person developing the ability to ask appropriate questions and independently learn relevant disciplinary material?

3. Does the person have or is the person developing the ability to identify places where additional mathematical and/or statistical analysis would be helpful?

4. Does the person have or is the person developing good communication skills?

5. What is your opinion of the quality of work produced by this person?
6. If possible, would you wish to continue your collaboration with this person?

7. Please provide any additional comments you wish to make based on your impressions of our cross-disciplinary training program.

This will be applied at all levels of the training program.

**Teaching**

Teaching is difficult to evaluate. We believe a particularly important outcome is downstream learning, and we will work on developing relevant assessment tools, but this is a large endeavor of its own and will only apply to a limited set of courses. We do have teaching evaluations administered by the University. One procedure we have found very useful is to organize a group discussion of students in a particular class, with a graduate student serving as an impartial facilitator and reporter. We find undergraduates are generally quite willing to speak freely in this context, and graduate students can be adept at constructively summarizing their comments. Sometimes senior faculty may be involved in this process. Because we will be actively supervising them, we will have quite a bit of information about the teaching efforts of trainees. We will assess our program in light of our knowledge about their successes and difficulties.

**Mentoring**

Because we are unable to collect hard evidence, we will rely primarily on the opinions of trainees. We will survey them, asking these questions:

1. What is your research area? What are your career goals?
2. Have you been engaged in joint research with anyone at Carnegie Mellon? If so, with whom?
3. Has your time at Carnegie Mellon enhanced your abilities and advanced your career? Please be specific.
4. How would you rate the quality of the mentoring and advice you have received concerning research, professional development, and teaching?
5. Have you received adequate assistance with and supervision of your teaching? Please be specific.
6. Describe any activities in which you have been engaged which included graduate students and/or undergraduate students? Do you think the students benefitted from these activities? Did you?
7. What are the best features of the Carnegie Mellon environment for post-doctoral fellows?
8. What would you most like to change about the environment for post-doctoral fellows?

Questions 6-8 will be suitably modified for graduate students.

5. **DISSEMINATION**
We believe we have been very successful so far with many aspects of our VIGRE program, and we expect our additional modifications will lead to further improvements. Many of the things we are doing, and will do, should be transferable to other institutions. We plan to disseminate our knowledge and experiences primarily through articles and discussions in professional meetings, and also by posting materials on our VIGRE web site.

The articles we plan to write would be submitted to journals such as *The American Statistician*, *The Journal of Educational and Behavioral Statistics*, and *The Journal of Statistical Education*. We expect to be part of panel discussions, or special sessions, on graduate and undergraduate education, and will actively seek these out through the professional organizations. The self-contained presentation we described above, under Recruitment and Retention, will also serve as a vehicle for dissemination.

One topic we will likely write and talk about concerns the core link between learning to do statistics and learning to teach it. A paper on this subject might include the following: Our essential thesis on the centrality and importance of $ABA^{-1}$; the evidence from Cognitive Science, including the need to break each problem into its components; our approach to teaching basic statistics; the effectiveness of SMARTLAB; breaking $ABA^{-1}$ into its components; our Statistical Practice sequence, including *Issues in Statistical Pedagogy*; some specifics on both cross-disciplinary research and teaching.

Additional topics we could discuss include the whole of our first-year program; our new Advanced Probability course, our Advanced Data Analysis project as a preliminary exam for the Ph.D., and, more generally, our method of managing research project courses; our vertically-integrated cross-disciplinary research teams; our summer program; our undergraduate curriculum; and our undergraduate training in research.

We will improve substantially our VIGRE web site, so that it contains descriptions of various aspects of our program. Currently, we have only a very primitive site. Our plan is to give additional details, including links to the papers we will write, and information on the innovative parts of our curriculum.

### 6. POST-VIGRE PLAN

VIGRE does two things. It supplies much-needed financial resources, and it forces departments to revise, defend, and assess their training programs. When VIGRE support ends, we will of course, lose a key mechanism for bringing trainees into our Department. However, if we can somehow find alternative resources, the programs themselves will survive. Specifically, we expect to keep the modifications we plan to make in our graduate program having to do with statistical methodology, research, and preparation for teaching. These are very important structural changes that would not be reversed due to termination of funding. Similar statements apply to the undergraduate program, and at the postdoctoral level we would expect to maintain much of the training program using alternative sources of support.

On the other hand we must note that NSF VIGRE funding provides a crucial degree of freedom to postdoctoral trainees: they are not hired full-time into a single research project, nor is their time split solely between a single pre-specified project and their teaching. The trainees may act to a substantial extent like other junior faculty and select projects to work on; and a similar situation occurs at graduate level. This gives trainees a greater ability to work on projects that are scientifically meaningful to them. Ultimately, we feel, this is a deep benefit that NSF VIGRE funding
provides. Thus, while many aspects of our program will continue unchanged, when VIGRE terminates we will need to find new ways to support trainees that continue to give them freedom to define their own path.
Appendix 1:

A. Number of baccalaureate degrees in Statistics awarded in last 5 years:

<table>
<thead>
<tr>
<th></th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>17</td>
<td>10</td>
</tr>
</tbody>
</table>

B. Number of full-time Ph.D. students for each of the past 5 years:

<table>
<thead>
<tr>
<th></th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35</td>
<td>39</td>
<td>41</td>
<td>47</td>
<td>49</td>
</tr>
</tbody>
</table>

C. List of Ph.D. recipients during past 5 years, along with each individual’s a) citizenship status, b) baccalaureate institution, c) time-to-degree, d) post-Ph.D. placement, and e) thesis advisor:

1998

Chris Andrews
a. U.S.A.
b. Oberlin College
c. 4.5 years
d. NSF postdoc, University of California, Berkeley
e. Kathryn Roeder

Ashish Sanil
a. India
b. Indian Institute of Technology
c. 4.5 years
d. Junior Fellow, National Institute of Statistical Sciences, Research Triangle Park, NC
e. William F. Eddy

1999

Xu Fan
a. P.R. China
b. University of Science and Technology of China
c. 4.5 years
d. AlphaMetrics, LLC, West Roxbury, MA
Christopher Genovese

Herbie Lee III
  a. U.S.A.
  b. Yale University
  c. 4.5 years
  d. Visiting assistant professor, Institute of Statistics and Decision Sciences, Duke University
  e. Larry Wasserman

Terra McKinnish
  a. U.S.A.
  b. University of Richmond
  c. 4.5 years
  d. Assistant professor, University of Colorado, Boulder
  e. Stephen Fienberg and Seth Sanders (CMU’s Heinz School)

2000

Dan Cork
  a. U.S.A.
  b. George Washington University
  c. 5.5 years
  d. Committee on National Statistics, Washington, D.C.
  e. Steve Fienberg and Al Blumstein (CMU’s Heinz School)

Alix I. Gitelman
  a. U.S.A.
  b. Columbia College
  c. 5.5 years
  d. Assistant professor, Dept. of Statistics, Oregon State University
  e. Brian Junker

Howard Seltman
  a. U.S.A.
  b. Oberlin College
  c. 4.5 years
  d. Visiting assistant professor, Dept. of Statistics, Carnegie Mellon University
  e. Joel Greenhouse
2001

Feng Tang
a. P.R. China
b. Shanghai Jiao Tong University
c. 5.5 years
d. Statistician, modeling and analysis, Cendant Corp.
e. Mark Schervish

Ilaria DiMatteo
a. Italy
b. University of Rome ‘‘La Sapienza’’
c. 5.5 years
d. Associate statistician, statistics division, United Nations
e. Rob Kass

Matthew Johnson
a. U.S.A.
b. Indiana University
c. 4.5 years
d. Research scientist, Educational Testing Service
e. Brian Junker

John R. Lockwood
a. U.S.A.
b. Duke University
c. 4.5 years
d. Associate statistician, The Rand Corporation
e. Mark Schervish

Tzee-Ming Huang
a. Taiwan
b. National Taiwan University
c. 4.5 years
d. Assistant professor, Dept. of Statistics, Iowa State University
e. Larry Wasserman

Bobby Jones
a. U.S.A.
b. Miami University
c. 5.5 years
d. Postdoctoral trainee, Dept. of Human Genetics, University of Pittsburgh
e. Kathryn Roeder
Thomas Nichols
   a. U.S.A.
   b. Carnegie Mellon University
   c. 5.5 years
   d. Assistant professor of biostatistics, University of Michigan
   e. William Eddy

2002

Michele DiPietro
   a. Italy
   b. Universita D’Annunzio
   c. 4.5 years
   d. Assistant director, Eberly Center for Teaching Excellence, Carnegie Mellon University
   e. John Lehoczky

Adrian Dobra
   a. Romania
   b. University of Bucharest
   c. 4.5 years
   d. Researcher, Duke University and the National Institute of Statistical Sciences
   e. Stephen Fienberg

Cristian Ghiuvea
   a. Romania
   b. University of Bucharest
   c. 4.5 years
   d. Associate, quantitative research group, Zurich Capital Markets, New York City
   e. John Lehoczky

Iuliana Ianus
   a. Romania
   b. University of Bucharest
   c. 4.5 years
   d. Biometrician, Merck Research Labs, Blue Bell, Pa.
   e. Larry Wasserman and Joel Greenhouse

Stella Salvatierra
   a. Argentina
   b. University of Tucuman
c. 4.5 years
d. Assistant professor, School of Economics, University of Navarra, Spain
e. Stephen Fienberg

Shu-Ngai Yeung
a. Hong Kong
b. Hong Kong University of Science and Technology
c. 4.5 years
d. Data analyst, AT&T Laboratories, Florham, NJ
e. John Lehoczky

Xiaohua Zhang
a. P.R. China
b. Beijing Normal University
c. 3.5 years
d. Biometrician, Merck Research Labs, West Point, Pa.
e. Kathryn Roeder

D. The names of postdoctoral associates (visiting assistant professors) during the past 5 years, a) their Ph.D. institutions, b) postdoctoral mentors, c) post-appointment placements:

Anthony Brockwell
a. University of Melbourne
b. Jay Kadane and John Lehoczky
c. Still at CMU

Thomas Minka
a. Massachusetts Institute of Technology
b. Jay Kadane and Stephen Fienberg
c. Still at CMU

Kimberly Sellers
a. The George Washington University
b. Stephen Fienberg
c. Still at CMU

Howard Seltman
a. Carnegie Mellon University
b. Joel Greenhouse and Kathryn Roeder
c. Still at CMU
Galit Shmueli  
a. Israel Institute of Technology  
b. Stephen Fienberg  
c. University of Maryland  

Garrick Wallstrom  
a. University of Minnesota  
b. Robert Kass  
c. Still at CMU  

E. The dollar amount of non-teaching support of graduate students supplied by the university for each of the past five years and the anticipated changes in university support of this kind in the event of an award:

<table>
<thead>
<tr>
<th>Years</th>
<th>Support Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997-1998</td>
<td>$599,383</td>
</tr>
<tr>
<td>1998-1999</td>
<td>$782,858</td>
</tr>
<tr>
<td>1999-2000</td>
<td>$736,101</td>
</tr>
<tr>
<td>2000-2001</td>
<td>$913,823</td>
</tr>
<tr>
<td>2001-2002</td>
<td>$1,000,427</td>
</tr>
</tbody>
</table>

No change is anticipated in university support of this kind in the event of an award.

F. The dollar amount of funding by federal agencies for REUs, for graduate students and for postdoctoral associates in each of the past five years:

This information is in process of being gathered and will be forwarded as available.

G. The anticipated size of the graduate program should this award be received:

49 students