Supporting Statement B for

National Institutes of Health
National Cross-site Evaluation of the Broadening Experiences in Scientific Training (BEST) Program for the Office of Strategic Coordination, an office of the Division of Program Coordination, Planning, and Strategic Initiatives, within the Office of the Director

March 17, 2015

Patricia Labosky, Ph.D.
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director, NIH
1 Center Drive, MSC 0189
Building 1, Room 214A
Bethesda, MD 20892-0189
Telephone: 301-594-4863
Fax: 301-435-7268
Email: Workforce_Award@mail.nih.gov

Table of Contents

List of Attachments

- Attachment B.2.1.1: Invitation for Graduate Students Surveys – (Data Collection by Awardee)
- Attachment B.2.1.2: Invitation for Postdoctoral Scientists Surveys – (Data Collection by Awardee)
- Attachment B.2.1.3: Invitation for Graduate Students Surveys – (Data Collection by the NIH)
- Attachment B.2.1.4: Invitation for Postdoctoral Scientists Surveys – (Data Collection by the NIH)
- Attachment B.2.1.5: Invitation for Phone Interviews

Supporting Statement for the
Paperwork Reduction Act Submission
B: Justification

National Institutes of Health
National Cross-site Evaluation of the Broadening Experiences in Scientific Training (BEST) Program

This request seeks approval for OMB clearance to conduct a national cross-site evaluation of the Broadening Experiences in Scientific Training (BEST) Program. This request for clearance includes data collection efforts for three populations from the institutions that received the BEST award: graduate students, postdoctoral scientists, and program staff (Principal Investigator (PI), Co-Principal Investigators (co-PIs), Program Director, and local evaluator). The purpose is to identify best practices in the field of biomedical research training. This will be accomplished by assessing two desired outcomes for graduate students and postdoctoral scientists, and one desired outcome for the awardee institutions. The three desired outcomes are: (1) changes in understanding of career opportunities, confidence to make career decisions, and attitudes towards career opportunities; (2) reduced time to desired, non-training, non-terminal career opportunities, and reduced time in postdoctoral positions; and (3) creation/further development of institutional infrastructure to continue BEST-like activities. The third outcome includes actions which will lead to sustainability of BEST programs and the extension of BEST activities within and across multiple graduate programs. Surveys will be used to gather data from graduate students and postdoctoral scientists to assess the first and second outcomes. A Data Form will be used to gather data for all three outcomes. Phone interviews with program staff will be used to gather data for the third outcome. The information gathered from graduate students, postdoctoral scientists, and program staff will document the BEST program operations and activities, and assess its effectiveness.

B. Collection of Information Employing Statistical Methods

B.1. Respondent Universe and Sampling Methods

B.1.1. Respondent Universe

The universe of respondents for which clearance is sought includes: graduate students, postdoctoral scientists, and program staff (PI, co-PIs, Program Director, and local evaluator) from the institutions that received the BEST award. A census will be used to survey graduate students and postdoctoral scientists from the graduate programs/departments participating in the BEST program. Conducting a census is appropriate because sampling would result in a number of respondents from important sub-groups being too small to permit comparative analyses. In addition, given the longitudinal design of the study, and the small population from some of the graduate programs/departments participating in the BEST program, it is necessary to survey the entire BEST program population, rather than a sample.

The evaluation will include 17awardee institutions. Of the 17, 10 institutions received the BEST award in FY2013 and their award will end in FY2018. Seven received the BEST award in FY2014 and their award will end in FY2019. Throughout this document, the term "awardee institution" refers to the 17 institutions that received the NIH BEST award.

A description of the three populations of interest for the national cross-site evaluation of the BEST program is below:

1. Graduate Student Population –This population consists of graduate students from the graduate programs/departments participating in the BEST program at each awardee institution. The following graduate students will be surveyed: (1) graduate students who are participating in the BEST program, and (2) graduate students who are not participating in the BEST program
2. Postdoctoral Scientist Population - This population consists of postdoctoral scientists who have a position within the departments participating in the BEST program at each awardee institution. The following postdoctoral scientists will be surveyed: (1) postdoctoral scientists who are participating in the BEST program, and (2) postdoctoral scientists who are not participating in the BEST program.
3. Program Staff Population – This population consists of PI, co-PIs, Program Director, and local evaluator from each awardee institution.

Table B.1. displays the number of graduate students, postdoctoral scientists, awardee institutions, and program staff entering the evaluation study in 2015. The numbers for graduate students and postdoctoral scientists were used for projections for future years. The total number of projected graduate students is 19,654, which is equal to the sum of the number of existing graduate students who enter the study in 2015 (9,037) and projected cohorts of 2,259 new graduate students who enter later in the year in 2015 and each year 2016 through 2018, with a cohort of 1,581 in the final year of 2019. The total number of postdoctoral scientists is 13,643, which is equal to the sum of the number of existing postdoctoral scientists who enter the study in 2015 (6,273) and projected cohorts of 1,568 new postdoctoral scientists who enter later in 2015 and each year 2016 through 2018, with a cohort of 1,098 in the final year of 2019.

Table B.1. Graduate Student, Postdoctoral Scientist, Awardee Institution, and Program Staff Counts Entering the Evaluation Study in 2015 and Used for Projections for Future Years					
BEST Recipients	Graduate Students Entering the Study in 2015[1]	Postdoctoral Scientists Entering the Study in 2015[2]	Total Graduate Students and Postdoctoral Scientists Entering the Study in 2015	Number of Awardee Institutions	Program Staff from Awardee Institutions
Recipients of NIH grant in FY2013	5,316	3,690	9,006	10	47
Recipients of NIH grant in FY2014	3,721	2,583	6,304	7	36
Total	9,037	6,273	15,310	17	83

[1,2] Projections are based on anticipated population sizes.

B.1.2. Power Analysis and Estimation Procedure

The goal of this analysis is to ensure that the proposed design has sufficient statistical power to detect meaningful change in graduate students and postdoctoral scientists participating in the intervention relative to those not participating. There are 17 institutions implementing the BEST program. Baseline measures are reassessed after a period in which participation in BEST programming may occur. The outcomes that are being measured include: 1) attitudes toward and perspectives on a range of research and research-related career options, and 2) employment outcomes for graduates and postdoctoral scientists. Two different methods are used to analyze data related to these outcomes. For the first outcome, a two-level repeated measures model is used. For the second outcome, survival analysis techniques will be used to assess employment outcomes. These measures will be compared between BEST program participants and non-participants to ascertain program effectiveness.

The power analyses for outcome 1 focused on the most demanding scenario with the smallest number of repeated measures (2 measures), the highest intercorrelation between measurement points of .5, a small effect size of .2, and expected attrition across the period of study[1]. The Analysis utilized the Hedeker, Gibbons, and Waternaux (1999)[2] method for deriving sample size estimates for longitudinal designs with attrition. In this most conservative scenario outlined above, the power analyses indicate that the study is adequately powered to detect the potential effects (with an expected effect size of .25 for graduate students and .20 for postdoctoral scientists) of BEST programming on participants from both populations (graduate students and postdoctoral scientists).

 A total sample size of 19,654 doctoral students from BEST awardee institutions is expected over a 5-year period. This number consists of 9,037 students who will enter the sample in early 2015, and cohorts of 2,259 newly entering students who will enter the sample in the fall semester of academic years 2015, 2016, 2017, and 2018, and a final cohort of 1,581 entering students will come into the sample in 2019.

Upon entering the sample, each student will complete an Entrance survey with baseline measures on a number of career related items presumed to be affected by participation in the BEST programming. All participants will complete an Exit survey upon graduation. These surveys, along with the two scheduled Interim surveys, yields between two and four measures for those in the graduate student population.

The analyses to determine power were based on a 6-year window for completion of the graduate program, and students entering the study in spring of 2015 are distributed evenly across the six years, resulting in 1,506 graduates per year from 2016 forward. During the 2015 year it is expected that 1,012 students will participate in the BEST program and 6,217 will be available for comparison purposes. By 2020 those numbers decline to 169 in the participant group and 1,036

---

[1]Preliminary reports from participating institutions suggest that fewer than 15% of students in the fields observed for this study depart before completing their degree. Therefore, 15% is used as an upper bound, which was then spread across a 5-year window resulting in a 3% attrition rate between each time point.

[2]Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition. Journal of Educational and Behavioral Statistics, 24, 70-93.

in the non-participant group. Each new cohort between 2015 and 2018 will include 2,259 new graduate students. The power analyses indicate that the study is adequately powered to assess an effect of (.2), and consequently the larger expected BEST programming effect of (.25).

These analyses revealed that a minimum initial sample size of 137 will be required to detect a small effect (.2) under the most conservative scenario outlined above, at a power level of .80. ("Conservative" in this context means that more data points are needed to overcome these assumptions than what is actually expected. Therefore, if the study is powered using these conservative assumptions, it can safely be concluded that the study will be powered under the more realistic, expected, scenarios and assumptions.) The response and attrition calculations of the overall sample shows that the anticipated within-year samples sizes comfortably exceed the minimum of 137 program participants in the years 2016 through 2021. Also, the number of timepoints available for analysis increases as the study moves farther along, enabling greater power to detect BEST programming effects that may exist. Beyond the within-year sample, analysis of more distal effects can be examined by pooling the samples with two time points across time, and similar analyses across pooled data from three or four time points can also be conducted to reveal potential linear or quadratic change across larger numbers of time periods.

The power analyses for graduate students serve as a guide for the assessment of power for the postdoctoral scientist sample, but with three differences. First, a smaller programming effect is expected for the postdoctoral scientists (.20) than for the graduate students (.25). Second, the postdoctoral scientists will not receive Interim surveys, meaning the postdoctoral scientists will have two measurement points—Entry and Exit. And third, a slightly smaller proportion of postdoctoral scientists are expected to participate in the BEST program. These parameters still fit within the conservative scenario outlined previously for the graduate students.

The analyses for postdoctoral scientists assumes a 5-year window for completion of the post-doc training. Trainees entering the study in spring of 2015 are distributed evenly across the five years it takes to complete the average term. Five years is just beyond a median term completion time and therefore will result in a conservative number of postdoctoral scientists (an underestimate) continuing each year. To model a reasonable rate of progression for continuing postdoctoral scientists, the postdoctoral scientists entering the study in the spring of 2015 (6,273) is divided by five years to yield 1,255 completers per year from 2016 forward. Each cohort between 2015 and 2018 will include 1,568 new postdoctoral scientists, and a final cohort of 1,098 in the final year.

The analyses revealed that a minimum sample size of 132 will be required to detect a small effect (.2) under the most conservative scenario outlined above, at a power level of .80. The response and attrition calculations of the overall sample shows that the anticipated within-year samples sizes comfortably exceed this minimum in the years 2016 through 2021. Also, the number of timepoints available for analysis increases as the study moves farther along, enabling greater power to detect BEST programming effects that may exist.

Therefore, in conclusion, the analyses demonstrate that for outcome 1, sufficient sample sizes in both populations can be expected to achieve adequate power (.80).

Post Graduate Employment Outcomes

To gauge the degree to which the analysis is sufficiently powered for the post graduate employment outcomes, the nQuery (2014)[3] statistical software package was used to derive preliminary power estimates based on minimum expected sample sizes. The simulation was run based on the assumption that 80% of graduates will be unemployed at the end of the first quarter after graduating. The survival is reduced by 10% each quarter thereafter, so that in the final period, 10% of students are still seeking employment.

Simulations were run for post graduate employment measures for both graduate students and postdoctoral scientists. The simulations assumed a possible 25% difference in time to employment (i.e. the program effect), and a more conservative 20%. At anticipated sample sizes, the study was found to be powered under both assumptions and for both populations. For graduate students, the estimate of power was based on 10,000 simulations of the study with a seed for the random number generator of 536333. At a 25% effect, the estimated power was found .96 and at a 20% effect, the estimated power was found to be .85. Similarly, for postdoctoral scientists, the estimate of power was based on 10,000 simulations of the study with a seed for the random number generator of 563747. For this population, at a 25% effect, the estimated power was found .99 and at a 20% effect, the estimated power was found to be .97.

B.2. Procedures for the Collection of Information

B.2.1. Data Collection Procedures

The data collection activities include online surveys for graduate students and postdoctoral scientists, phone interviews with program staff from the awardee institutions, and a Data Form for PIs to provide information requested in the RFAs for the BEST program.

Procedures for Online surveys

Data will be gathered from graduate students and postdoctoral scientists using a common set of survey questions, which the NIH has developed with input from awardee institutions. Based on the awardee institution' preference and capacity, two approaches will be used:

Approach 1: Some surveys will be administered by the awardee institution and some surveys will be administered by the NIH.
- The Awardee, on behalf of NIH, will administer the Entrance, Interim, and Exit surveys to graduate students while they are enrolled at the awardee institution, and the Entrance survey to postdoctoral scientists while they are employed by the awardee institution during the grant period. The invitations when the awardee institutions administer the surveys are in Attachments B.2.1.1 and B.2.1.2. B.2.1.1 contains the invitation for graduate students, and Attachment B.2.1.2 contains the invitation for postdoctoral scientists.
- The NIH will administer:
  o Exit surveys for up to four years after the grant ends to graduate students who receive Entrance surveys during the grant period, but graduate after the grant ends;

---

[3] nQuery Statistical Software (2007). Version 7 downloaded from http://www.statsols.com/products/nquery-advisor-nterim/ on September 1, 2014.

- o Exit surveys to postdoctoral scientists;
- o Post-exit surveys at 2, 6, 10, and 15 years after graduate students and postdoctoral scientists complete the Exit survey.

Approach 2: All surveys will be administered by the NIH. The NIH will administer:
- Entrance, Interim, and Exit surveys to graduate students while they are enrolled at the awardee institution;
- Entrance survey to the postdoctoral scientists while they are employed by the awardee institution during the grant period;
- Exit surveys for up to four years after the grant ends to graduate students who receive Entrance surveys during the grant period, but graduate after the grant ends;
- Exit surveys to postdoctoral scientists; and
- Post-Exit surveys at 2, 6, 10, and 15 years after graduate students and postdoctoral scientists complete the Exit survey.

Note: Regarding the Post-Exit surveys, we are requesting clearance for the 2 year Post-Exit surveys. In the future, NIH will seek clearance to administer the Post-Exit surveys at 6, 10, and 15 years after the Exit survey.

The invitations when the NIH administers the surveys are in Attachments B.2.1.3 and B.2.1.4. Attachment B.2.1.3 contains the invitation for graduate students, and Attachment B.2.1.4 contains the invitation for postdoctoral scientists.

The awardee institutions and the NIH will enter into a Data Sharing Agreement (DSA) to share the data from all surveys conducted for the national cross-site evaluation. The DSAs have been reviewed by all the institutions and they all committed to sign them. In addition, awardee institutions have consulted and discussed the evaluation study with their Institutional Review Board. An evaluation ID will be assigned to each graduate student and postdoctoral scientist. The NIH Secure Email/File Transfer Service (SEFT) will be used to send and receive data securely on a secure socket layer (SSL)/encrypted connection.

In instances where the awardee institution administers the surveys, the awardee institution will provide the NIH with the evaluation IDs and survey responses within 30 days of the close of the survey. The awardee institutions will keep a key file containing evaluation IDs, participant names, and emails in a secure location according to the protocols approved by their Institutional Review Board. The email invitation for graduate students and postdoctoral scientists will inform respondents that their survey data will be shared with the NIH and the NIH contractor.

In instances where the NIH administers the surveys, the awardee will provide the evaluation IDs and the email addresses of the graduate students and postdoctoral scientists, and the NIH will provide the survey responses with evaluation IDs to the awardees. The NIH will provide to the awardee institutions a codebook for each survey to standardize the data collection.

The following procedures will be used for all online surveys:
- The online surveys will be designed to be clear and easy to navigate. As appropriate, the online surveys will use a skip-pattern so that each respondent is only presented with questions that are relevant to his or her specific situation.

- An e-mail invitation will be sent to graduate students and postdoctoral scientists from the participating graduate programs/departments. The email will explain the purpose of the online survey and provide a hyperlink to the survey website.
- One week after the e-mail invitation, a reminder e-mail will be sent to all non-respondents. The e-mail will encourage those who have not yet followed the link to participate in the survey.
- One week after the first reminder e-mail, a second e-mail reminder will be sent to all non-respondents. The e-mail will reinforce the purpose and relevance of the survey.
- One week after the second reminder e-mail, a final e-mail reminder will be sent to all non-respondents. The e-mail will reinforce the purpose and relevance of the survey.

Procedures for Phone Interviews with Program Staff – Individual interviews with the PI, co-PIs, Program Director, and local evaluator from each awardee institution will be conducted annually. The interviews will be conducted within a three-month period at the end of each calendar year throughout the duration of the BEST grant. Interviews will be scheduled at the convenience of the interviewees. The participation in phone interviews is voluntary. Program staff will be asked if they agree to be interviewed and if they agree to audio recording of their interview. Attachment B.2.1.5 contains the invitation for the phone interviews.

Procedures for Data Form for PIs – The Data Form consists of four sections. Four excel files, one for each section, will be provided to PIs. Section 1 and 2 will be submitted yearly. Section 3 will be submitted only in FY 2015, and Section 4 will be submitted once in year four of the BEST award. PIs will be asked to submit each section along with the required NIH annual Research Performance Progress Report (RPPR).

B.2.2. Analysis Procedures

Qualitative Data – Content analysis will be used to analyze the data collected from phone interviews and open-ended questions from the surveys.
Quantitative Data – The first desired outcome will be assessed using a two-level repeated measures design with multiple measurement points nested within individual. This approach allows the main and interaction effects in a two-group repeated measures design to be captured. All models will be specified using the Mplus statistical software package. Mplus provides significant flexibility in modeling and the management of a variety of data structures (Muthén & Muthén, 1998-2011)[4]. Mplus also offers a robust set of routines for handling missing data.

The second desired impact, improved employment outcomes, will be assessed using survival analysis techniques, specifically discrete time survival analysis. For graduate students, the outcome of interest is possible differences in time from graduation to research-related employment. For postdoctoral scientists, the outcome of interest is the duration of the postdoctoral appointment. Shorter terms are assumed to imply gainful research-related employment beyond the postdoctoral appointment. The discrete time survival analysis models allow the analyst to predict the occurrence and timing of non-repeatable events (Allison, 1995; Teachman, 1983)[5] such as initial non-terminal employment. Discrete (as opposed to continuous)

---

[4] Muthén, L. K., & Muthén, B. O. (1998-2011). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

[5] Allison, P. D. (1995). Survival analysis using SAS:A practical guide. Cary, NC: SAS Institute; Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards models. Social Science Research, 12, 263–301.

form of the model will be used because the measure of interest is the status of employment in units of months, where there are likely many ties at each discrete time point (i.e., many respondents are likely to experience the employment event of interest within any month). The discrete time model will provide adjusted probabilities of experiencing the non-terminal employment for each time period in the series. The initial observation period will span from the time of the exit survey administration (origin time) in phase one of the study to the time period in which the two-year follow-up survey is administered, or 24 months.

The discrete time survival model allows for the specification of both fixed and time-varying covariates. Of specific interest in this analytic component will be the possible effect of the BEST programming intervention on the shape of the survival function. The hypothesis is that those participating in the BEST programming intervention will experience a shorter time to employment windows. Therefore, the focus is on specific "treatment," or BEST program effect.

Missing data – It is anticipated that missing data will be encountered at level-2 for both the graduate student and postdoctoral scientists. The use of listwise deletion of cases with missing data will be avoided, as it is widely known that listwise deletion can result in a tremendous loss of data and biased parameter estimation. The traditional solutions provided in most software programs are listwise, or pairwise deletion or mean substitution of missing data. In most situations, none of these would be considered as optimal (or acceptable) approaches (Enders & Bandalos, 2001)[6]. For example, listwise deletion leads to inflated standard errors when the data are Missing Completely at Random (MCAR), and biased parameter estimates when the data are Missing at Random (MAR) (Allison, 2002; Larsen, 2011)[7]. Mean substitution treats individuals with missing data as if they were on the "grand mean" (MCAR), which is also likely to introduce bias in most situations (e.g., by reducing variance). Therefore, Mplus software package's set of multiple imputation (MI) data procedures will be used, allowing identification of patterns of missing data that can then be substituted with plausible values imputed using the Expectation-Maximization (EM) algorithm. EM is a common method for obtaining maximum likelihood estimates with incomplete data, and has been shown to reduce bias due to missing data (Peugh & Enders, 2004)[8]. Obtaining estimates involves an iterative, two-step process where missing values are first imputed, and then a covariance matrix and mean vector are estimated. This repeats until the difference between covariance matrices from adjacent iterations differs by a trivial amount. The imputed data sets can be saved as separate data sets and then analyzed. It is often the case, for example, that even with 25-35% missing data, the analyst can impute a number of "random" plausible values for individuals in order to generate a number of new data sets that can be saved for further analysis.

B.3. Methods to Maximize Response Rates

<u>Online surveys.</u> Consistent with sound survey methodology, the design of all online surveys will

---

[6]Enders, C. K., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(430–457).

[7]Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage. Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling, 18*(4), 649–662.

[8]Peugh, J. A. & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556.

include approaches to maximize response rates, while retaining the voluntary nature of the effort. Below is an overview of the approach to maximize response rate.

- All 17 awardee institutions have agreed to participate in the national cross-site evaluation and their leadership is supportive of the evaluation. All institutions will include information about the evaluation on their BEST program websites and will encourage their graduate students and postdoctoral scientists to participate in the surveys. The NIH and the awardee institutions will disseminate information about the evaluation study in conferences, journals, and on the NIH website for the BEST program.
- The online surveys will be designed to be clear and easy to navigate. As appropriate, the online surveys will use a skip-pattern so that each respondent is only presented with questions that are relevant to his or her specific situation. Also, the online survey contains multiple choice and closed-ended questions.
- The introductory e-mail invitations for graduate students and postdoctoral scientists will inform participants that their participation will help to improve biomedical research training in their institutions and nationwide. The invitation will also explain the different surveys they will be asked to complete and the type of information that will be requested. The email for graduate students and postdoctoral scientists will contain enough information to generate interest in the online surveys. The emails will provide a point of contact for additional information.
- Reminder emails to complete the online surveys will be sent to participants who have not accessed the online survey. The reminder emails will reinforce the purpose and encourage participation.
- Graduate students will be followed up every 6 months via email after they graduate from the awardee institutions to ensure that the email address they provided in the surveys is correct. Also, awardee institutions will provide to the NIH a list of graduates. Within the follow-up email, the respondent will be asked to click on a link and either confirm that the email address is correct or provide up to two additional email addresses. Participants of the pilot recommended the 6 month follow-up.
- Postdoctoral scientists will be followed up every 6 months via email after they complete the Exit survey to ensure that the email address they provided is correct. Within the follow-up email, the respondent will be asked to click on a link and then either confirm that the email address is correct or provide up to two additional email addresses.

Phone Interviews. To maximize the response rates for the phone interviews with program staff (PIs, co-PIs, Program Director, local evaluator) from each awardee institution, sufficient time for data collection will be provided. Phone interviews will be carried out over the course of three months to make sure that the busy schedules of respondents can be accommodated. Also, email reminders will be sent to respondents to encourage participation in the phone interviews.

Data Form. To maximize the response rates, four excel files have been created, one for each section of the Data Form. The excel files are user-friendly. The four sections were reviewed by PIs from the awardee institutions, so they are already familiar with the data requested and the estimated time of completion for each section. PIs will be able to retrieve the information requested prior to completing a section. Also, per the suggestion of the PIs, the schedule of the submission of the sections coincides with the submission of the NIH Research Performance Progress Report (RPPR). Email reminders will be sent to PIs and the NIH Program Officer and the NIH contractor will be available to answer questions.

B.4. Test of Procedures or Methods to be Undertaken

All surveys for graduate students and postdoctoral scientists were pilot tested in November and December of 2014 under OMB # 0925-0046-07. Participation in the pilot testing was voluntary and the responses were kept private. The results were only used to improve the surveys. Participants were provided with an URL link to access the Entrance, the Interim, or the Exit survey so each participant only commented on one survey. The 2-year Post-Exit survey was tested via phone interviews. The graduate students and postdoctoral scientists who participated in the pilot test were asked to provide feedback on the flow of the survey questions, appropriateness of the skip patterns, and the length of time to complete the online survey. They were also asked to comment on the wording of specific survey questions and provide their overall impression of the online survey. Their feedback was incorporated into the final version of each online survey.

Fifty graduate students from awardee institutions were invited to participate in the pilot and 37 responded, which yielded a response rate of 74 percent. Fifty postdoctoral scientists from awardee institutions were invited to participate and 34 responded, which yielded a response rate of 68 percent. The overall response rate was 71 percent.

The pilot test of the 2-year Post-Exit survey included 10 interviews with five graduate students and five postdoctoral scientists. During the interview, they were asked to provide their feedback on the instructions and wording of questions, and comment on the overall impression of the survey. Revisions were made to the survey based on their comments during the interview.

The NIH BEST Evaluation Subcommittee and some members of the NIH Strengthening the Biomedical Research Workforce Working Group for the BEST program (see Attachment A.8.3 for a list of members) provided feedback on the data collection instruments for the graduate students, postdoctoral scientists, and program staff. To ensure a successful implementation of the national cross-site evaluation, the BEST awardee institutions were also consulted and provided feedback. The design and implementation plans for the national cross-site evaluation were discussed at the awardees annual meetings held in Bethesda in October 2013 and 2014.

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Staff from Windrose Vision, a company that specializes in program evaluation, worked with the NIH staff to develop the design for the national cross-site evaluation study. Windrose Vision staff has extensive experience evaluating NIH programs and developing surveys for a variety of audiences such as grant applicants, reviewers, and awardees. In addition, an expert on multi-level model analysis was consulted on the design and statistical analysis of this study. Scott L. Thomas, Ph.D., professor and dean of the School of Educational Studies at Claremont Graduate University (CGU), conducted the power analyses and will conduct the statistical analyses for this study. He is a co-director of CGU's Howard R. Bowen Institute for Policy Studies in Higher Education. His methodological research focuses multilevel models and sample design.

Dr. Thomas' work on methodological topics work can be found in a series of books addressing applied issues in multilevel modeling. These books are published by Taylor & Francis and

include <u>An Introduction to Multilevel Modeling Techniques</u>[9], <u>Multilevel and Longitudinal Analysis Using IBM SPSS</u>[10], and <u>Multilevel Modeling of Categorical Outcomes with IBM SPSS</u>[11]. His most recent book (with Ron Heck), <u>An Introduction to Multilevel Modeling: MLM and SEM approaches using Mplus,</u> will be released in 2015.

Dr. Thomas is the editor-in-chief at the *Journal of Higher Education*, the premier journal in the field of higher education. He co-edits (with David Palfreyman and Ted Tapper of Oxford University) the book series, <u>International Studies in Higher Education</u> (published by Taylor & Francis), that currently boasts more than 20 volumes on a variety of topics related to international higher education. He is president-elect of the Association for the Study of Higher Education.

Please email windose@windrosevision.com for any questions related to the power and statistical analyses.

---

[9] Heck, R. H. & Thomas, S. L. (2009). An introduction to multilevel modeling techniques, 2nd edition. New York: Routledge/Taylor & Francis.

[10] Heck, R. H. & Thomas, S. L., and Tabata, L. (2010). Multilevel and longitudinal analysis using SPSS. New York: Routledge/Taylor & Francis.

[11] Heck, R. H. & Thomas, S. L., and Tabata, L. (2012). Multilevel modeling of categorical outcomes with IBM SPSS. New York: Routledge/Taylor & Francis.