FINAL INTEGRATED REPORT

Evaluation of the National Institutes of Health (NIH) Broadening Experiences in Scientific Training (BEST) Program

Submitted To:

The National Institutes of Health (NIH)
Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI),
Office of Strategic Coordination (OSC)

By:

Julieta Scalo, PhD, PharmD
Larry Freauff, PhD

June 30, 2020

Preface

This report presents findings from evaluation of the Broadening Experiences in Scientific Training (BEST) awards granted by the National Institutes of Health (NIH) Common Fund as part of the "Strengthening Biomedical Research Workforce" program.

This evaluation was undertaken by zCore Business Solutions, Inc. (zCore) (Round Rock, Texas), a Small Disadvantaged Business, Economically Disadvantaged Woman Owned Small Business, Texas Historically Underutilized Business.

Windrose Vision, LLC (Fairfax, VA) collected the data for this evaluation between 2014 and 2019; zCore conducted the data analyses between November 2018 and June 2020.
zCore and its team members have no financial, personal, business, or volunteer affiliations that constitute a conflict of interest for this project, and are committed to upholding the Evaluators' Ethical Guiding Principles ratified by the American Evaluation Association (www.eval.org).

The National Institutes of Health (NIH), Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI), Office of Strategic Coordination (OSC)

zCore Business Solutions, Inc.

810 Hesters Crossing Road, Suite 206
Round Rock, TX 78681
DUNS number: 034439328
CAGE Code: 1YTD3

Table of Contents

List of Tables and Figures

1 EXECUTIVE SUMMARY

The National Institutes of Health (NIH) **Broadening Experiences in Scientific Training (BEST)** program is part of a transformational effort to ensure that the future biomedical, behavioral, social, and clinical research workforce has the training they need to succeed in any biomedical research careers of their choice, whether it is within or outside academia.[1,2] Seventeen awardees received research grants to support the development and evaluation of novel training programs aimed at better preparing graduate student (GS) and postdoctoral scientist (PD) trainees for the varied career paths that make up the contemporary biomedical research enterprise. The goals for this evaluation of the BEST program were to:

1. Assess changes in understanding of career opportunities, confidence to make career decisions, and attitudes towards career opportunities;
2. Determine reduced time or no increase in time to desired, non-training, non-terminal career opportunities, and reduced time in postdoctoral positions; and
3. Identify creation and/or further development of institutional infrastructure to continue BEST-like activities.

Each program year, awardee sites submitted an exhaustive Data Form to report on BEST activities and institutional characteristics. Participating and non-participating trainees at all seventeen BEST program sites took one or more of four surveys (Entrance, Interim, Exit, and Post-Exit), administered at different time points to track knowledge and attitudes about career choices, knowledge of and participation in BEST activities, and their trajectory from training to subsequent employment (and/or further training).

Formal participation in the BEST program was defined as participation in at least one of the activities for which individual attendance could be tracked. About 1,800 trainees participated during the first year, and over 3,000 participated every year thereafter. The highest participation totals were for workshops/seminars/symposia, which were also the most frequently offered activities. Certificate programs had the highest average participation (29 trainees per program), but it should be noted that these were not singular events and may have spanned an entire academic term or year. Professional and peer mentoring programs also had high levels of participation, averaging 15 and 20 trainees per activity, respectively. Both graduate students and postdoctoral scientists participated in a median of four activities each year.

One goal of the BEST awards was to broaden trainees' exposure to career options outside the academic research path. On Entrance Surveys, most graduate students (70.1%) reported they were familiar with "all" or "most" of 20 science-related career paths, while slightly fewer postdoctoral scientists (64.8%) reported familiarity with "all" or "most" career paths. Among trainees who completed both Entrance and Exit Surveys, familiarity with career paths increased significantly; 82% of graduate students and 73% of postdoctoral scientists reported familiarity with "all" or "most" career paths on Exit Surveys. For both graduate students and postdoctoral scientists, there was a significant relationship between BEST participation and being familiar with an increased number of career paths between entrance and exit (for graduate students: t = -1.9919, p-value = 0.0466, for postdoctoral scientists: t = -2.0060, p-value = 0.0451).

Trainee agency for making career decisions was also measured by asking respondents to rate the extent to which they were currently considering each of the twenty Individual Development Plan (IDP) career paths. The most strongly considered career paths were research in industry, combined research and teaching in academia, and Principal Investigator in a research-intensive institution. For graduate students, and especially those in BEST programs, consideration tended to decrease between the entrance and exit surveys for nearly all of the career paths. This likely represents not a decreased interest overall, but rather a narrowing of career focus by ruling out certain paths. For postdoctoral scientists, interest decreased primarily among the most highly considered career paths, especially among BEST program participants. This likely reflects a shift in focus from traditional careers in academia to multiple other options (where slight increases in interest can be observed).

Another measure of trainee career-development agency was trainee confidence. On Entrance Surveys, 86% - 90% of all graduate students and 84% - 92% of all postdoctoral scientists reported being moderately, highly, or completely confident across six measures, including identifying potential employers and discussing career goals with mentors. For trainees who completed both Entrance and Exit Surveys, confidence scores increased across all groups, but only for graduate students was BEST participation associated with a significantly greater increase in confidence (t-test = -4.1777, $p > 0.001$).

One goal of the BEST program is that participation would either reduce or not increase time to complete graduate studies or postdoctoral training. Based on Entrance and Exit Survey responses, graduate students' mean time to PhD completion was not significantly different for BEST participants (mean = 5.8 years; SD = 1.0 years) versus the comparison group (mean = 5.6 years; SD = 1.2 years; t-test = -1.8716; p-value = 0.0218). Among postdoctoral scientists, BEST participants did spend more time in training (mean = 3.5 years; SD = 1.9 years) versus the comparison group (mean = 3.1 years; SD = 1.6 years) but the difference was only about four and a half months (difference = 0.37 years; SD = 0.26 years; t-test = -1.4312; p-value = 0.0131). Hierarchical regression modeling was conducted to control for demographic and other factors and to account for awardee site-related clustering effects. No correlations were found in these models between BEST participation and time to degree or time in postdoctoral training.

The third outcome assessed in this program evaluation is institutional infrastructure to continue BEST-like activities. Factors that impact sustainability include program leadership, faculty attitudes toward the program, and external partnerships that can provide instructional and/or financial support. Most sites assembled both an advisory board (usually for hands-off guidance) and a steering committee (for guidance, review, and—in some cases—direct administration). All sites measured faculty attitudes toward the BEST program at least once during the five-year period, and many conducted surveys each year. All reported that faculty generally supported trainees' pursuit of non-academic careers and participation in career development activities, but opinions were mixed on the amount of time considered appropriate, and whether participation reduced research productivity. Between thirteen and sixteen sites reported non-BEST funding for their programmatic activities each year. The number of funding sources decreased in later years, but the median funding amount from each source increased, along with a shift from short-term to long-term funding. More than half of the individual funding sources and more than seventy-five percent of total funding originated from sites' institutional departments and programs.

2 PROGRAM EVALUATION OVERVIEW

Starting with the 2013–2014 academic year, seventeen awardee sites received five-year research grants to support the development and evaluation of novel training programs aimed at better preparing graduate student (GS) and postdoctoral scientist (PD) trainees for the varied career paths that make up the contemporary biomedical research enterprise.[3,4]

Although common program elements were employed (e.g., mentoring, workshops), specific methods of implementation vary, and each awardee represented a unique experiment. Taken together, however, these experiments provide a larger view of potential educational enhancements, including their strengths and limitations.

This evaluation takes a formative approach, with an overall objective of providing academic institutions with an evidence base for further development and improvement of training activities to prepare biomedical trainees for a wide range of careers beyond academia. The NIH has outlined three program goals to be assessed with this cross-site evaluation:[5]

1. Improvements to understanding of career opportunities, confidence to make career decisions, and attitudes towards career opportunities.
2. Reduction, or no increase, in time to degree for graduate students and time to desired, non-training, non-terminal career opportunities for postdoctoral positions.
3. Creation and/or further development of institutional infrastructure to continue BEST-like activities.

In support of these goals, the evaluation seeks to answer the following questions:

1. Trainee career-development agency.
   a. Does knowledge of a broad range of careers change with BEST participation?
   b. Does confidence to make career decisions change with BEST participation?
   c. Does consideration of various career paths change with BEST participation?
2. Trainee time expenditures
   a. Does time to degree change with BEST participation?
   b. Does to time in postdoctoral position change with BEST participation?
3. BEST program implementation and sustainability
   a. How are sites disseminating and expanding BEST activities?
   b. How are sites implementing BEST activities and addressing challenges?
   c. How sites engaging with external partners to support BEST activities?

2.1 Methods

2.1.1 Roles in evaluation

Data collection instruments were developed through collaborations between NIH staff, awardee institution personnel, and Windrose Vision, LLC (Fairfax, VA). Survey administration and data collection were conducted by Windrose Vision and awardee personnel. NIH contracted zCore Business Solutions, Inc. (Round Rock, TX) to perform the independent evaluation contained in this report.

2.1.2 Evaluation period

Awardee sites started their BEST programs in two waves, corresponding with the timing of the 5-year funding periods: ten sites began in 2013, and seven in 2014. It should be noted that Dr. Lenzi et. al (2020)[5] have earlier summarized BEST program implementation and activities using data collected between 2013 and 2017. Those data were divided into year of Common Fund program support, meaning Year 1 (2013-2014) included only the first cohort of sites. To allow cohorts to be evaluated together, analyses for this evaluation are based on program year. Thus, Year 1 comprises data from 2013-2014 for the first cohort as well as data from 2014-2015 for the second cohort.

2.1.3 Data collection

Three formal methods were used to collect program data from the awardee sites:

1. Interviews and site visits with BEST site coordinators to gather contextual information
2. Annual reporting by awardee sites using a custom Data Form
3. Surveys of individual trainees at awardee sites

2.1.3.1 Interviews and site visits

Annual phone conversations were conducted with awardees. These conversations included NIH staff, Windrose Vision staff, BEST principal investigators, local BEST site evaluators, and/or BEST program managers. Uniform questions guided the conversation to assess infrastructure development and sustainability of BEST activities and provide contextual information. In the first year of each award, NIH staff also performed site visits with institutional administrative leaders, faculty involved in BEST activities, participating trainees, and the faculty and staff responsible for program implementation.

2.1.3.2 Data Forms collected from awardee sites

Each program year, awardee sites submitted a Data Form to report on BEST activities and institutional characteristics. The annual Data Form Section 1 described specific activities and logged trainee participation. Section 2 described characteristics of the awardee sites. In the first reporting year, awardee sites completed Section 3, detailing baseline data for the awardee institution. Data definitions were agreed upon by the sites to facilitate global comparisons. A complete copy of the Data Form and the Data Definitions is available online in the Data Collections tool and Data Definitions files respectively.

In Data Form Section 1, BEST activities labeled as "aggregate level" are those for which it was not practical or feasible to track individual attendance. For these events, sites estimated total attendance. Multiple aggregate activities of a particular type (e.g., several Mixer/Networking Events), were all listed on a single table (e.g., 7g), with attendance estimates for each event. For activities where was tracked individually ("individual level"), each unique activity was tracked on a separate worksheet, along with a corresponding table of attendees. Individual trainees (graduate students and postdoctoral scientists) were assigned a unique evaluation number to allow tracking across activities and program years.

2.1.3.3 Surveys administered to trainees

Starting in academic year 2014-2015, participating and non-participating trainees at all seventeen BEST program sites took one or more of four surveys, administered at different times to track

changes over time. Entrance Surveys were administered upon entrance into eligibility to participate in BEST programming and Exit Surveys were administered near graduation or completion of postdoctoral training. Interim Surveys were administered once (academic year 2016-2017) and only to graduate students. Two years after the Exit Survey, Post-Exit surveys were administered to follow up on trainees' career trajectories and provide information on the long-term influence of the BEST program. Trainees were identified with the same unique evaluation numbers used to track attendance in the Data Forms. **Table 1** provides the data collection schedule for trainee surveys.

Table 1. Data collection schedule for trainee surveys

Academic Year	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
Entrance Survey*	✓	✓	✓	✓	
Interim Survey (Graduate Students only)			✓		
Exit Survey		✓	✓	✓	✓
Post-Exit Survey				✓	✓

* "Entrance" does not correspond to training stage, but rather to time of entrance eligibility for BEST activities.

Surveys were designed to capture data about trainees, including their demographic information, knowledge and attitudes about career choices, knowledge of and participation in BEST activities, and their trajectory from training to subsequent employment (and/or further training). Exit Surveys also included questions related to trainee career-search activities and post-training employment. Survey instruments can be found in the Data Collections tool file.

Some BEST awardee sites self-administered the surveys and provided data to Windrose Vision. For the remaining sites, Windrose Vision was responsible for survey administration and data collection. Data collection ended with academic year 2018-2019. Thus, a large number of trainees who completed Entrance Surveys, but did not complete their training in that timeframe are not represented in Exit and/or Post-Exit datasets.

2.2 Data analysis

Analyses for this report were performed using Excel (Microsoft, 2018) and Stata 15 (StataCorp, 2017). A criterion of $p < 0.05$ was used to test for statistical significance, and 95% confidence intervals were calculated to characterize precision for parameter estimates. To account for the possibility of increased Type I error from repeated statistical testing, the significance criterion was lowered to $p < 0.01$ for some analyses.

2.2.1 Study group designations

Each BEST awardee institution determined from which institutional departments and/or programs they recruited BEST participants. Trainees were assigned to the "participation" group if they participated in at least one "tracked" BEST program activity (for which attendance was individually tracked and recorded in attendance tables on Data Forms). Trainees who were eligible to participate but did not appear in official attendance tables made up the "comparison" group. Note that for some activities, attendance was not tracked and members of the comparison group may have participated in one of more untracked activities.

2.2.2 Summary statistics and bivariate comparisons

Data were summarized using counts and percentages for categorical variables and means and standard deviations (SD) for continuous variables. Continuous variables that were not normally distributed were described with the median, minimum and maximum values, and the interquartile range. Categorical variables were characterized by their counts and percentages. Comparison across groups were made using the statistical tests shown in **Table 2**.

Table 2. Statistical tests used for group comparisons

Variable Type	Two groups	Three or more groups
Continuous, normally distributed	Student's *t*-test	ANOVA
Continuous, non-normally distributed	Wilcoxon rank-sum test	Kruskall-Wallis ANOVA

2.2.3 Regression modeling

Multiple regression modeling was used to measure associations between BEST participation and outcomes while controlling for potential confounders such as demographic characteristics. Diagnostic tests included evaluations of multicollinearity, normality in the residuals, linearity between predictor variables and outcomes, independence of errors, and homoscedasticity. Extensions of linear regression modeling were applied as needed and are described below.

Hierarchical regression modeling was used to account for non-independence of observations within sampling units. For cross-sectional evaluations, individual observations modeled as Level 1 were nested within awardee sites modeled as Level 2.

Level 2 Awardee sites
Level 1 Survey or Data Form responses for one point in time

For surveys repeated over time, an additional level was added account for dependence in repeated measures. The three levels were modeled as follows:

Level 3 Awardee sites
Level 2 Individual graduate students or postdoctoral scientists
Level 1 Survey responses at different times

To confirm the need for multi-level modeling, intraclass correlation coefficients were calculated to evaluate whether a substantial proportion of total variation was attributable to the higher-level unit(s). Fixed and random predictors were evaluated individually for inclusion, and model fit was evaluated by χ^2 comparisons of deviance statistics and information criteria statistics.

Generalized linear modeling was used for outcomes with non-normally distributed errors. For example, scaled outcomes (e.g., responses ordered from most negative to most positive) were modeled using ordinal logistic regression. Model refinement included testing of alternative link functions, linear model specifications, and variance structures. Goodness-of-fit was evaluated using global measures (e.g., deviance, generalized Pearson's X^2) and examination of residuals. Competing models were compared using likelihood ratio tests and information criteria statistics.

2.3 Challenges

As with any evaluation, limitations related to study design and data collection must be considered.

2.3.1 Sample selection

Selection bias is a potential concern for program evaluation because trainee participation in BEST programs was not based on random sampling. At the awardee level, selection was based on grant proposals and award criteria. For individual trainees, determinants of program participation varied across individuals themselves (e.g., personal preference, peer pressure) as well as at department and site levels (e.g., advisor expectations, program eligibility criteria). Therefore, the results of this evaluation could be biased by unknown factors influencing individual or institutional participation in the BEST program.

2.3.2 Sample sizes

A large portion of data for this BEST program evaluation was collected primarily via surveys of trainees. Uneven and/or low response rates can introduce response bias and may reduce statistical power. Loss to follow-up is also a potential source of bias for analyses that involve comparisons of change over time. Entrance Surveys were administered in the first four program years, but because the entire data-collection period was only five years, many trainees who took Entrance Surveys had not completed their training in time to take Exit or Post-Exit Surveys.

2.3.3 Data collection

Entrance Surveys were used for baseline cross-sectional analyses and Exit Survey data were added for longitudinal comparisons, but variance in data collection across sites and over time presented limitations. Interim surveys were excluded from longitudinal analyses because they were administered in one year only and highly disproportionally from site to site (ranging from 2.7% to 20.4% of all surveys administered at a site). In addition, some Exit and Entrance survey questions were tailored to the administration timeframe and therefore were not identical. For example, questions on the Entrance Survey about completing an Individual Development Plan did not appear on the Exit Survey and, therefore, could not be evaluated longitudinally.

2.3.4 Sample heterogeneity

The BEST program comprised seventeen awards granted to academic institutions across the US. Awardees represented a diversity of enrollment sizes, academic infrastructures, trainee demographics, geographic regions, and other characteristics. Most important is the considerable heterogeneity among the interventions (i.e., BEST programs) themselves. Although many awardees shared common BEST programmatic elements (e.g., mentoring, internships, certificate programs, etc.), the content and implementation of these elements vary widely across institutions. As a result of these differences and the limited sample, pooling data across institutions may not be suitable for all analyses, and findings may have limited generalizability outside the awardee sites.

2.3.5 Group assignment

A limitation specific to this evaluation is that a true control group could not be defined. Because one purpose of the BEST program was to change institutional culture, trainees who did not actively participate in BEST activities were expected to be passively influenced by the program and—in many cases—had access to BEST events, such as large symposia, workshops, or networking events.

2.3.6 Measurement error

Data collection for BEST program evaluation was heavily reliant on surveys completed by individual trainees and data forms completed by sites. Survey data are subject to several biases, including response bias, recall bias, and others, that must be considered when reporting and interpreting results. For data forms completed by sites, sources of error include data entry mistakes and variable interpretation of instructions among sites. Despite quality-control measures employed during the data-collection period (e.g., establishment of specific definitions for key terms, yearly review of Data Forms for completeness) missing data and data entry errors could not be entirely avoided.

It is worth noting that because the sample of awardee sites is small, any missing data could create substantial bias. All sites completed the Data Forms each program year except for one that spent no award funds on trainee activities in the first program year, but instead used that time for planning. For the remaining Data Forms, reporting was remarkably thorough with a few exceptions. Tables 9 and 11 were made non-mandatory in September 2017, and the overall level of missing responses increased in Program Year 5.

3 RESULTS

3.1 BEST participant characteristics

Over the five-year period, a total of 23,860 trainees (13,759 graduate students; 10,101 postdoctoral scientists) were invited to complete the BEST Entrance Surveys (**Table 3**). The response rate for all invitees was 45%. The final sample of Entrance Survey respondents comprised 6,265 graduate students and 4,538 postdoctoral scientists. In the total sample, 54.2% were female, 9.1% were Hispanic/Latinx, and 61.5% were white. The median age was 26 for graduate students and 32 for postdoctoral scientists. **Figure 1**, **Figure 2**, and **Figure 3** show these demographic characteristics stratified by trainee type (graduate students, postdoctoral scientists) and by evaluation group (BEST participants, comparison group).

The comparison and participant groups did not differ in terms of ethnicity, but there were statistically significant differences for both gender and race. There was a higher proportion of females in the participant group versus the comparison group for both graduate students (Chi^2 = 75.8723, $p < 0.001$) and postdoctoral scientists (Chi^2 = 24.5965, $p < 0.001$). The distribution of race was similar between the comparison and participant groups of postdoctoral scientists, but among graduate students the participant group had a larger proportion of Asian trainees and a smaller proportion of White trainees versus the comparison group (Chi^2 = 15.6289, $p = 0.008$).

Table 3. Trainees invited to complete BEST Entrance Surveys

a. Graduate Students

Responded	BEST Participants N	%	Comparison group N	%	Total N	%
No	5867	64.1%	1627	35.3%	7494	54.5%
Yes	3284	35.9%	2981	64.7%	6265	45.5%
Total	**9151**	**100.0%**	**4608**	**100.0%**	**13759**	**100.0%**

b. Postdoctoral Scientists

Responded	BEST Participants N	%	Comparison group N	%	Total N	%
No	4619	60.2%	942	38.9%	5561	55.1%
Yes	3059	39.8%	1479	61.1%	4538	44.9%
Total	**7678**	**100.0%**	**2421**	**100.0%**	**10099**	**100.0%**

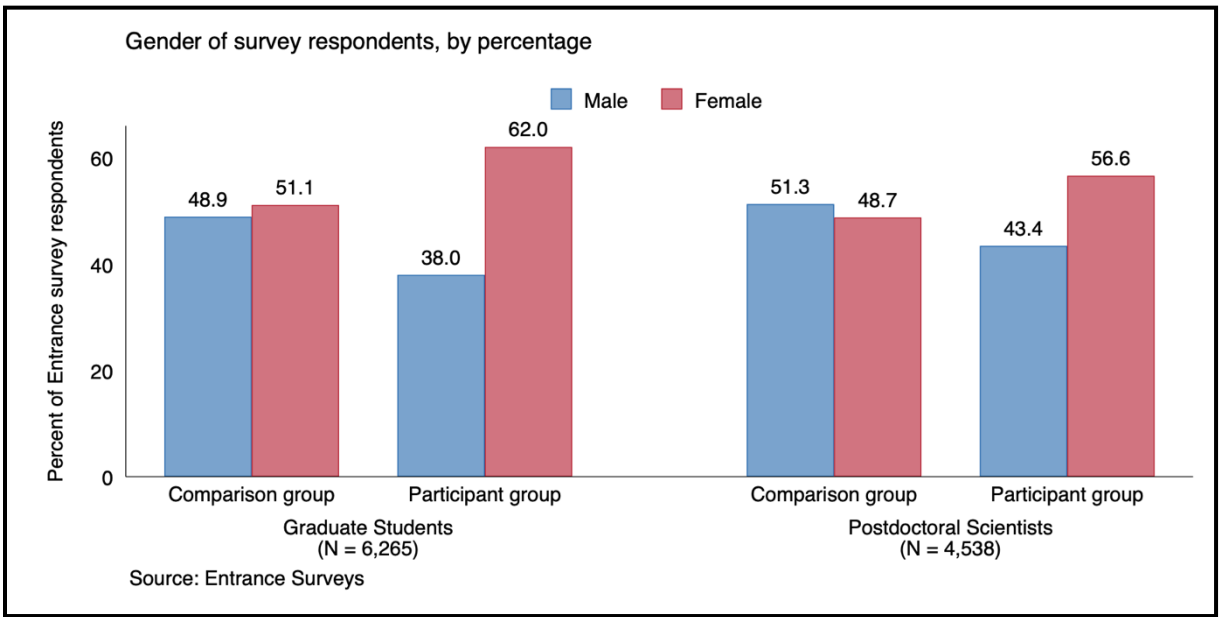

Figure 1. Gender distribution (by percentage) of Entrance Survey respondents, stratified by trainee type and BEST participation.

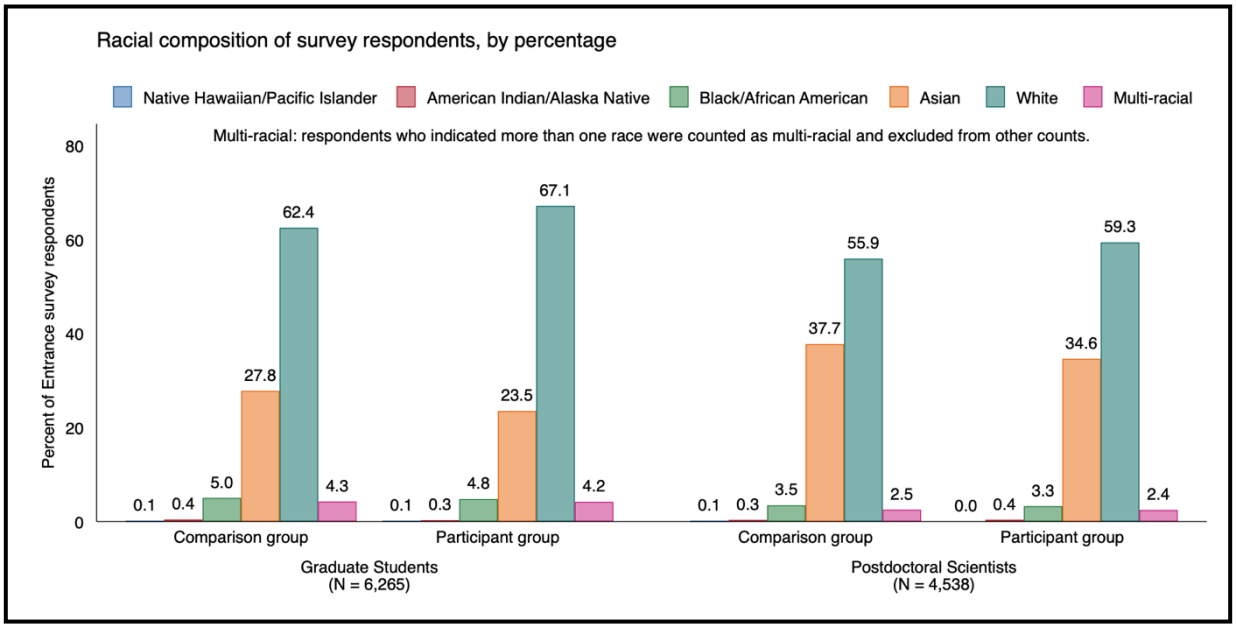

Figure 2. Racial distribution (by percentage) of Entrance Survey respondents, stratified by trainee type and BEST participation.

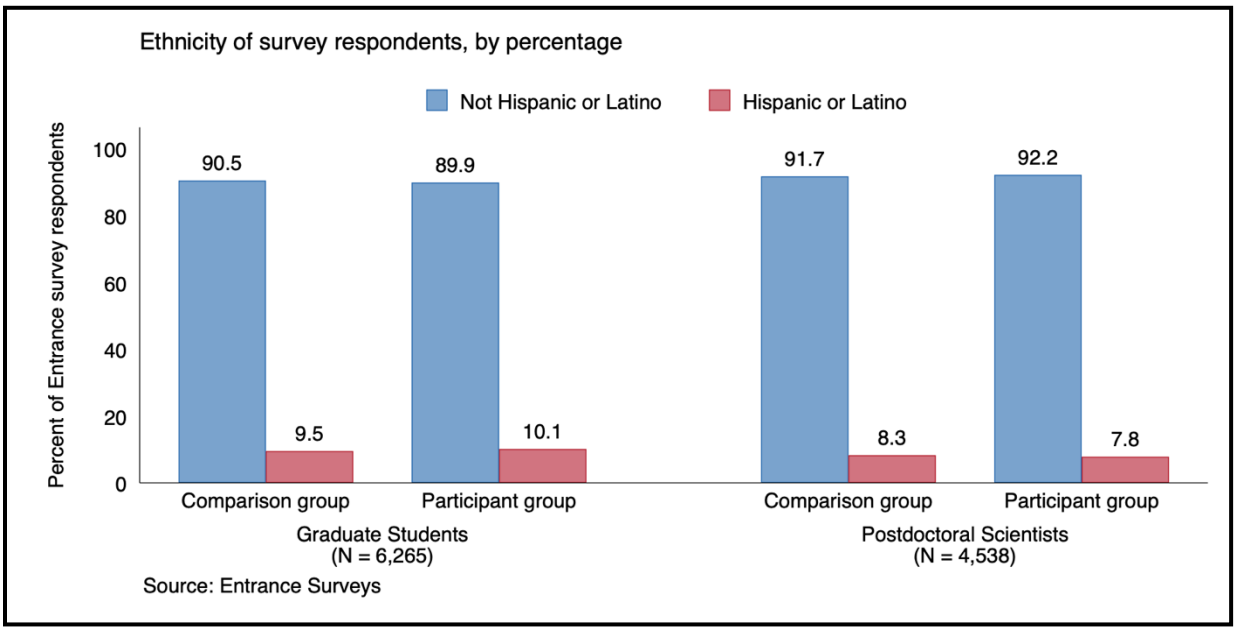

Figure 3. Ethnic distribution (by percentage) of Entrance Survey respondents, stratified by trainee type and BEST participation.

3.2 BEST program characteristics

3.2.1 Program eligibility

Eligibility for participation in BEST programming varied. Some awardees developed an open access model allowing any trainee from target departments—and, in some cases, non-affiliated departments—to participate. The majority of awardees, however (ten), required trainees to submit applications to participate, often requiring a new application each year. Several awardees had a combination of cohort and open access; some started small and gradually opened program activities to more participants due to demand, whereas others provided both open-access and application-only BEST programming by design. One awardee organized their program into phases, with the first phase open to all trainees and applications required for subsequent phases. Other awardees allowed free access to most of the program but required applications for certain activities either because of demand or to ensure that the trainee was at the appropriate stage of development to participate. Five awardees specified PI or advisor approval as a requirement, four required graduate students to have completed their qualifying exams and/or thesis proposals, and three awardees specified time-in-training requirements (at least two years for graduate students and six months for postdoctoral scientists). Other examples of criteria for participation included: academic good standing, submission of a *my*IDP completion certificate, and a demonstrated interest in exploring careers outside of academia.

3.2.2 Program certificates

Several awardee sites provided certificates of completion of participation to document trainees' participation in BEST activities (**Figure 4**). Certificates of completion were generally awarded to trainees who completed a pre-defined series or number of activities. For example:

- Completion of a 40-hour workshop, a 3-credit course, and at least 72 hours of professional shadowing; preparation of a resume; and participation in mentorship as both a mentee and mentor.
- Earning a minimum number of points through participation in various workshops and activities.

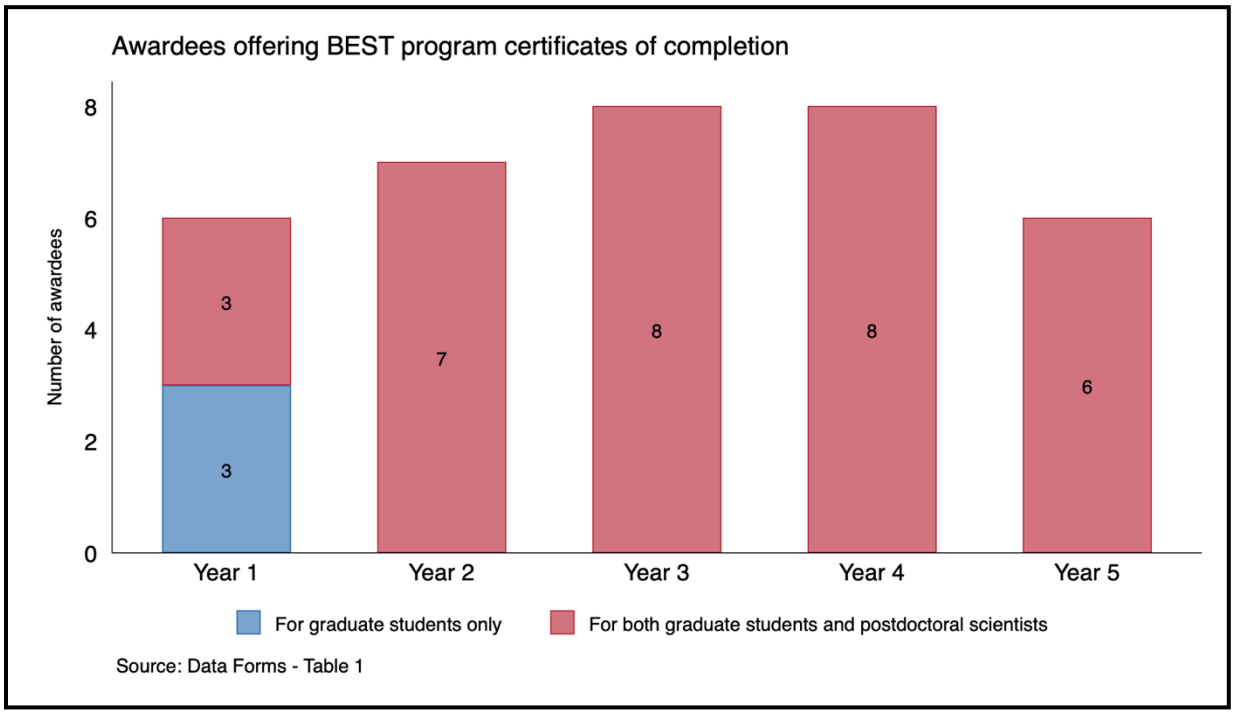

Figure 4. Number of awardee sites offering certificates of completion (non-accredited) for BEST program participation, by program year.

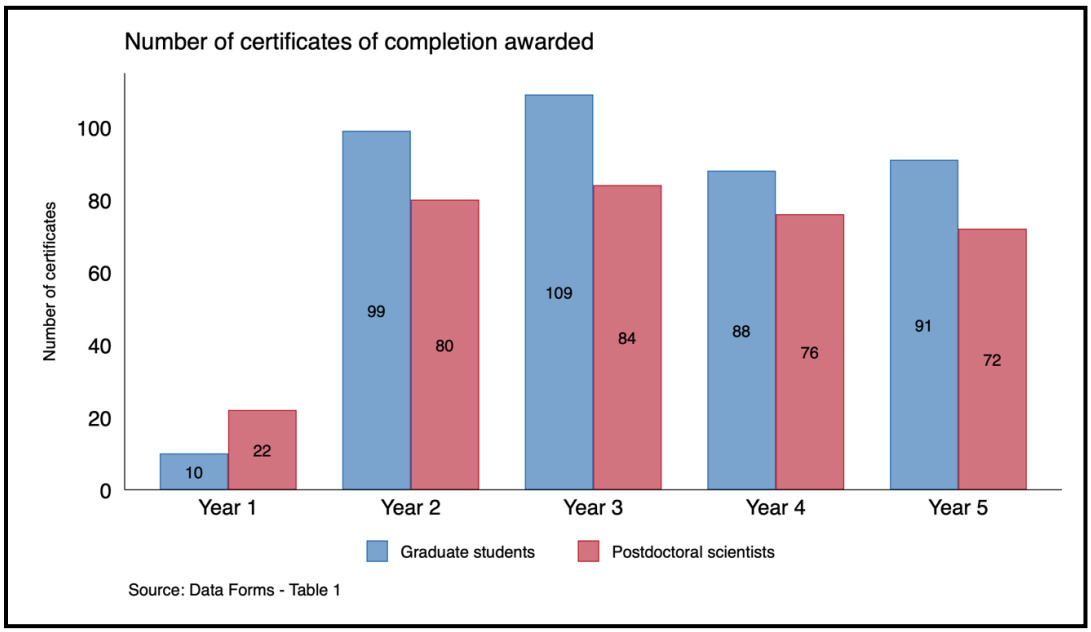

Figure 5. Number of BEST Program certificates of completion (non-accredited) awarded by all awardee sites, by program year.

Two awardees did not require a minimum level of activity, but instead issued certificates to a designated number of trainees who were most active in their programs. Certificates of completion for BEST program participation were not recordable in official transcripts, but some awardees encouraged their trainees to include the certificates on resumes, and one awardee specifically suggested that trainees post their certificates on LinkedIn. A total of 381 graduate students and 334 postdoctoral scientists received certificates of completion over the five-year period (**Figure 5**).

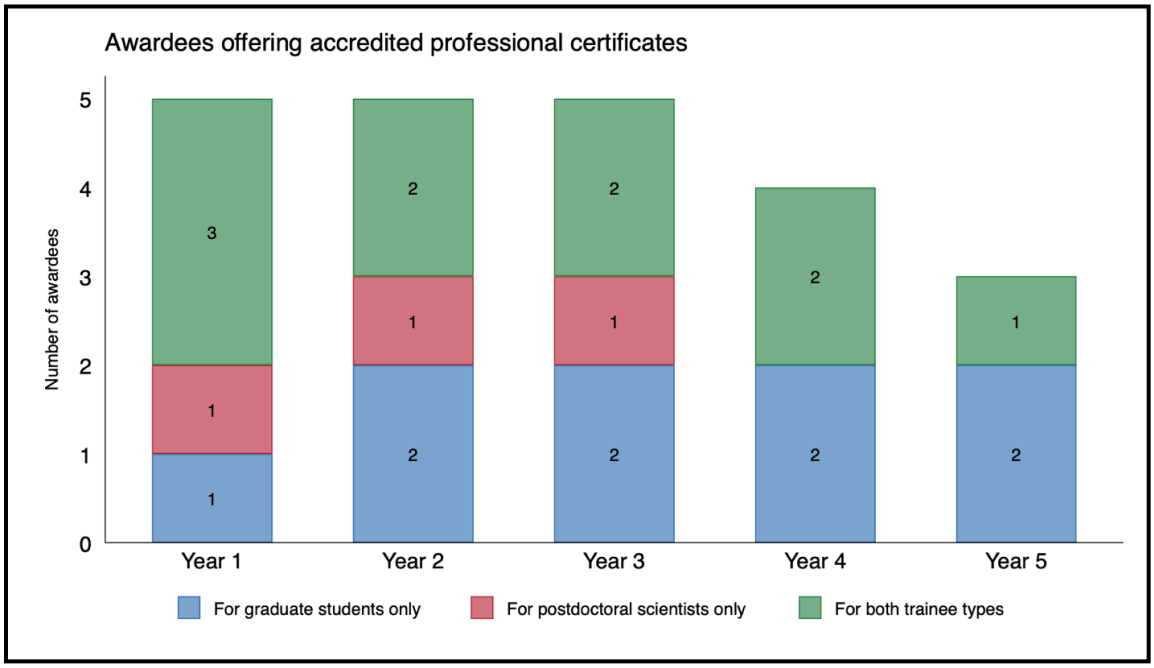

Figure 6. Awardees offering accredited professional certificates (these were not official components of BEST programs but were already available at awardee institutions).

Five awardees also offered accredited (transcript-recordable) professional certificates (**Figure 6**), including three of the awardees who offered certificates of completion. In all cases, however, the accredited certificate programs had been established at their academic institutions prior to the initiation of the BEST program (for example, a Teaching Excellence program offered through the Graduate Division). Not all BEST programs tracked the number of accredited professional certificates earned, but most gave examples of the available options. The most commonly cited subject areas for accredited certificate programs were: bioinformatics, translational science, entrepreneurship, and communication.

3.2.3 Individual Development Plans

The IDP is becoming almost standard in higher education and beyond as a tool to help trainees identify and work toward their career goals. All sites reported use of IDPs, but only about half made it a general requirement for trainees. Nearly all sites also incorporated the IDP formally into their BEST programs, with about half making the IDP a requirement for BEST program participation. The majority of awardees used both the online interactive tool *my*IDP[6] (myidp.sciencecareers.org) developed by the American Association for the Advancement of Science (AAAS) and an IDP modified for their institution (**Table 4**).

Table 4. Institutional use of AAAS Individual Development Plans in BEST programs

Type of IDP used	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5***
None	1	1	0	0	0
*my*IDP only	5	2	3	4	1
Institutional IDP only	1	1	1	1	1
*my*IDP & Institutional	10	13	13	12	14
Totals	**17**	**17**	**17**	**17**	**16**
Trainees required to complete IDP	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
None	10	7	6	7	6
Graduate students only	3	6	6	6	5
Postdoctoral scientists only	0	0	0	0	1
Both trainee types	4	4	5	4	4
Totals	**17**	**17**	**17**	**17**	**16**
BEST program requirement	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Not part of BEST program	3	1	1	1	1
Mandatory	9	9	8	9	8
Optional	5	7	7	7	7
Totals	**17**	**17**	**17**	**17**	**16**
Where used	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Not applicable	1	0	1	1	0
Institution-wide	2	4	4	3	3
Varies by department	5	4	5	6	8
Varies by graduate program	9	9	7	7	5
Totals	**17**	**17**	**17**	**17**	**16**

* Not all awardees submitted complete information.

As reported on Entrance Surveys by trainees themselves, however, only 37.7% of all graduate students and 34.4% of postdoctoral scientists competed an IDP in the prior twelve months (**Table 5**). One possible explanation for the disparity between awardee and trainee reports of IDP

use is that trainees may not have understood what the IDP is or may not have recognized it by name on the survey. It is also possible that sites requiring trainees to complete an IDP were not enforcing those policies. Notably, BEST participants—both graduate students and postdoctoral scientists—were significantly more likely to report completing an IDP than trainees in the comparison group (Chi^2 = 391.6077, $p < 0.001$ for graduate students, and Chi^2 = 114.2436, $p < 0.001$ for postdoctoral scientists), yet there were still about half that reported "No" or "I do not know/I do not remember" for this question (**Figure 7**).

Table 5. Use of Individual Development Plans reported by trainees on Entrance Survey

	Graduate Students		Postdoctoral Scientists		Total	
Completed IDP within previous 12 months					*Chi^2 p-value > 0.001*	
No	3,094	*51.92%*	2,407	*55.78%*	5,501	*53.54%*
Yes	2,248	*37.72%*	1,483	*34.37%*	3,731	*36.31%*
I do not know/remember	617	*10.35%*	425	*9.85%*	1,042	*10.14%*
Total	**5,959**	***100%***	**4,315**	***100%***	**10,274**	***100%***
Discussed IDP with mentor					*Chi^2 p-value > 0.001*	
No	1,004	*50.91%*	628	*43.22%*	1,632	*47.65%*
Yes	890	*45.13%*	790	*54.37%*	1,680	*49.05%*
I do not know/remember	78	*3.96%*	35	*2.41%*	113	*3.3%*
Total	**1,972**	***100%***	**1,453**	***100%***	**3,425**	***100%***

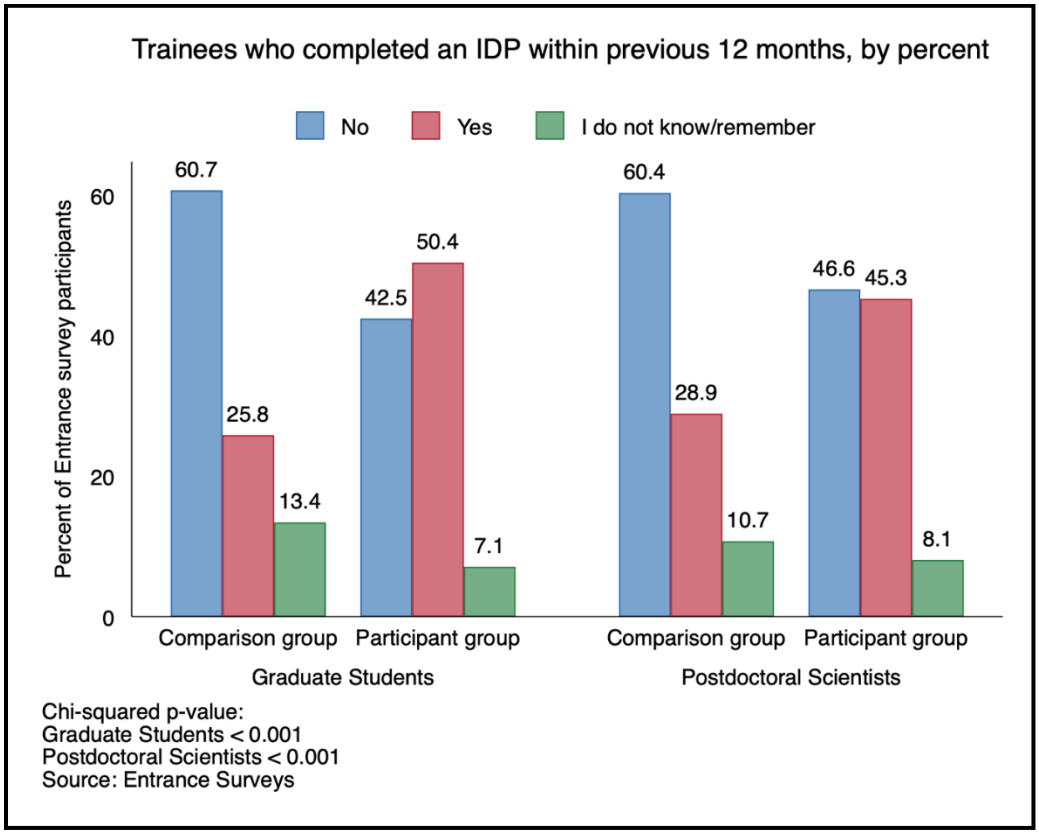

Figure 7. BEST participants were more likely than the comparison group to have completed an Individual Development Plan within the previous twelve months (as reported on Entrance Surveys)

A somewhat surprising finding from the Entrance Surveys (**Figure 8**) was that trainees from the comparison group were significantly more likely to discuss the IDP with their mentors than were BEST participants (for graduate students: $Chi^2 = 6.6122$, $p = 0.0367$, for postdoctoral scientists: $Chi^2 = 37.7255$, $p < 0.001$). Conversations between NIH staff and BEST trainees at site visits confirmed that many completed the IDP in isolation from their mentors and many revealed reluctances to discuss their career plans with their mentors.

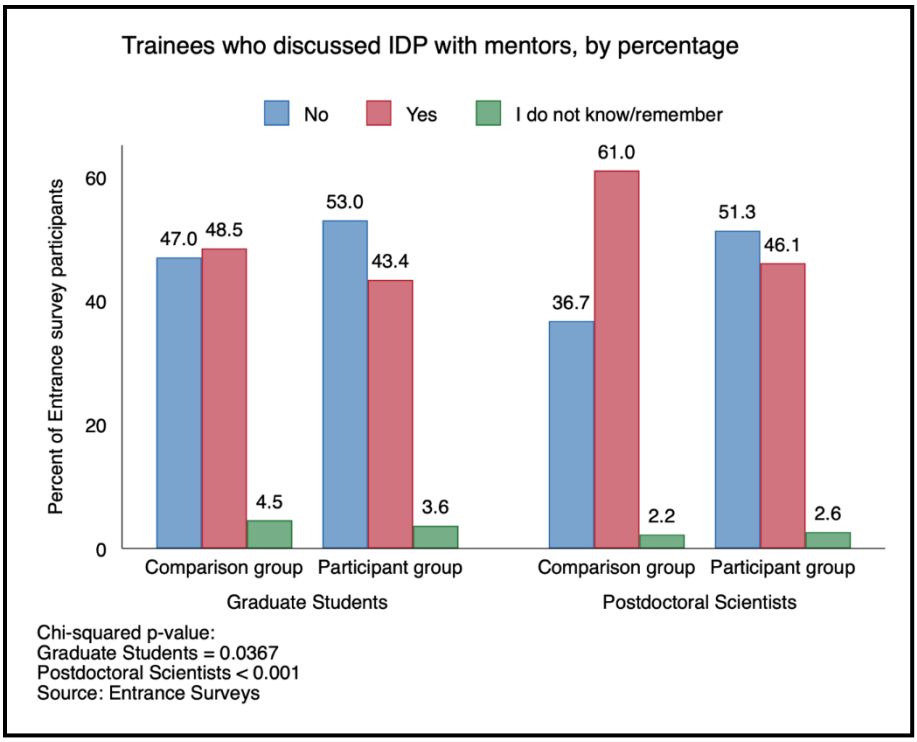

Figure 8. BEST participants were less likely than the comparison group to discuss their completed Individual Development Plan with their mentors.

Nonetheless, the majority of trainees who completed an IDP assessed it to be valuable (**Table 6**). On a scale ranging from 1 = "Not at all valuable" to 5 = "Extremely valuable", two-thirds of graduate students (66.34%) and an even larger proportion of postdoctoral scientists (75.48%) scored the IDP as moderately to extremely valuable ($Chi^2 = 91.5403$, $p < 0.001$). As seen in **Figure 9**, however, BEST participants and the comparison group did not differ in their assessments (graduate students: Chi^2 p-value = 0.386, postdoctoral scientists: Chi^2 p-value = 0.167).

Table 6. Entrance Survey responses on the value of completing an Individual Development Plan

Res	Graduate Students		Postdoctoral Scientists		Total	
Extremely valuable	112	5.01%	169	11.42%	**281**	**7.57%**
Very valuable	484	21.67%	416	28.11%	**900**	**24.23%**
Moderately valuable	886	39.66%	532	35.95%	**1,418**	**38.18%**
Slightly valuable	485	21.75%	235	15.88%	**721**	**19.41%**
Not at all valuable	237	10.61%	113	7.64%	**350**	**9.42%**
I do not know/remember	29	1.3%	15	1.01%	**44**	**1.18%**
Total	**2,233**	**100%**	**1,480**	**100%**	**3,714**	**100%**

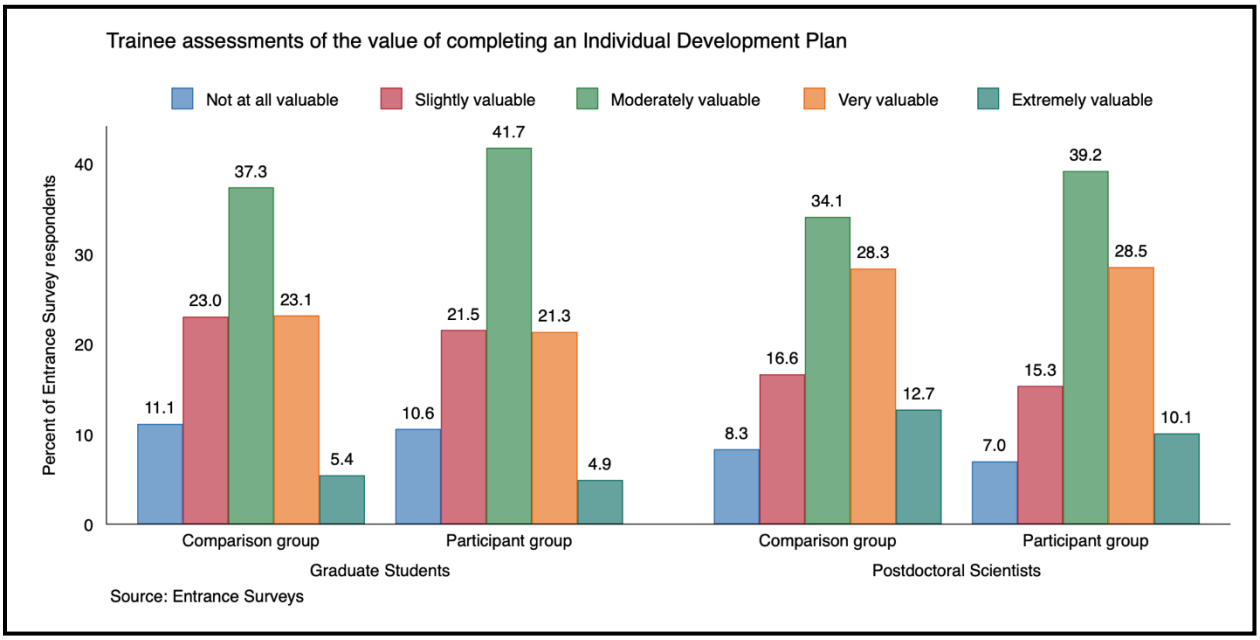

Figure 9. Trainee assessments (from Entrance Surveys) of the value of completing an Individual Development Plan. No differences were found between BEST participants and the comparison group.

These results are noteworthy. Low rates of IDP completion overall may indicate a disconnect between institutional requirements and mentors' expectations or may reflect a need for improved communication. Although BEST participants were more likely to complete an IDP, they were less likely to discuss their responses with a mentor, which ignores perhaps its most important purpose. Nonetheless, both BEST trainees and the comparison group overwhelmingly reported receiving some value from completing the IDP.

3.3 BEST program activities

Each BEST institution developed its own program to achieve its defined goals. To facilitate cross-site evaluation, program activities were classified and defined as shown in **Table 7**. **Figure 10** shows the number of sites offering each type of activity over the five program years, and **Figure 11** shows the total number of unique activities offered for each type. Seminars, workshops, and symposia—especially single-day events—were among the most widely implemented activities, both in terms of the number of sites offering them as well as the number of unique offerings of each activity. Experiential learning activities, including internships, courses, and visits to employer sites, were another major focus.

Table 7. Definitions for BEST program activities

Activities	Definition
Certificate Programs	Different from a Certificate of Participation for which a student/postdoc receives a physical and formal certificate in recognition of their participation in the BEST program from the BEST program itself. A formal Certificate Program that is more likely to be defined by a competitive application process, prerequisites, and graded coursework and may frequently be run under the authority of the university itself.
Clubs	An association or structured organization of two or more people united by a common professional interest or goal that meets with some regularity. Must have originated or been enhanced from BEST efforts even though they now are run independently. Report as 'new' until they are disbanded.
Courses	Courses (credit or non-credit) are institutionally recognized, possibly with a listing in the course catalogue or time-table and will likely have tuition consequences. Courses typically span an academic period (like a quarter or semester) and have a fixed roster of students covering a specific idea or topic.
Externships	Job shadowing a Job shadowing a professional at work for the purpose of observing and experiencing the work environment and learning about the expectations of a profession.
Internships	Working in a professional setting for the purpose of receiving hands-on training. Assumes the trainee is able to develop some skills during the experience and results in a deliverable.
Mixers/ Networking	Gathering of students/postdocs and/or professionals with the purpose of networking. This event may take place in person or in an online setting. Does not include small group discussions, such as a lunch with a speaker.
Peer mentoring	Peer Mentoring is when another trainee serves as a resource, provides assistance, or advice, or a sounding board, or referrals for training opportunities and career development. The role of peer mentors is to provide support, encouragement, and information to trainees. It is a formal program which facilitates this experience.
Professional mentoring	Professional Mentoring is a relationship between two or more people with the goal of professional and personal development. Professionals could include faculty members (other than PI/thesis advisor), institutional staff members, alumni, career coaches, or professionals in any industry. Mentoring can take place in a group setting or one-on-one. Would not include 'oneoff' counseling sessions with a career coach. If the interaction is a onetime session, then the interaction should be recorded in as 'other' and indicate as advising.
Self-assessment	Participant completes a career or personality self-assessment tool, such as the Meyers-Briggs Personality Indicator (MBTI) or a Career Assessment Worksheet. As self-assessment may be an independent activity, or it could be a component of another activity, such as a workshop or course. Additionally, it may require a follow-up with faculty or staff members to discuss the results.
Workshop	An event for the purpose of gaining knowledge or skills which involves hands-on activities and active participation by attendees. A workshop may take place in person or in an online setting. This activity may occur over one or more days.
Symposium	This is defined as an event with multiple sessions and speakers addressing more than one topic. This activity may occur over a half-day, one day, or more than one day.
Seminar	This is defined as an event at which one or more speakers give a presentation or lecture on one topic for the purpose of education or training. A seminar may take place in person or in an online setting.
Site visit	One or more trainees visit a work setting to learn more about an organization and tour the facility. The visit is an observational experience that usually takes place over the course of one day or less.
Other	Examples of "Other" activities include presentations, luncheons, and advisory meetings that did not meet the formal definitions above.

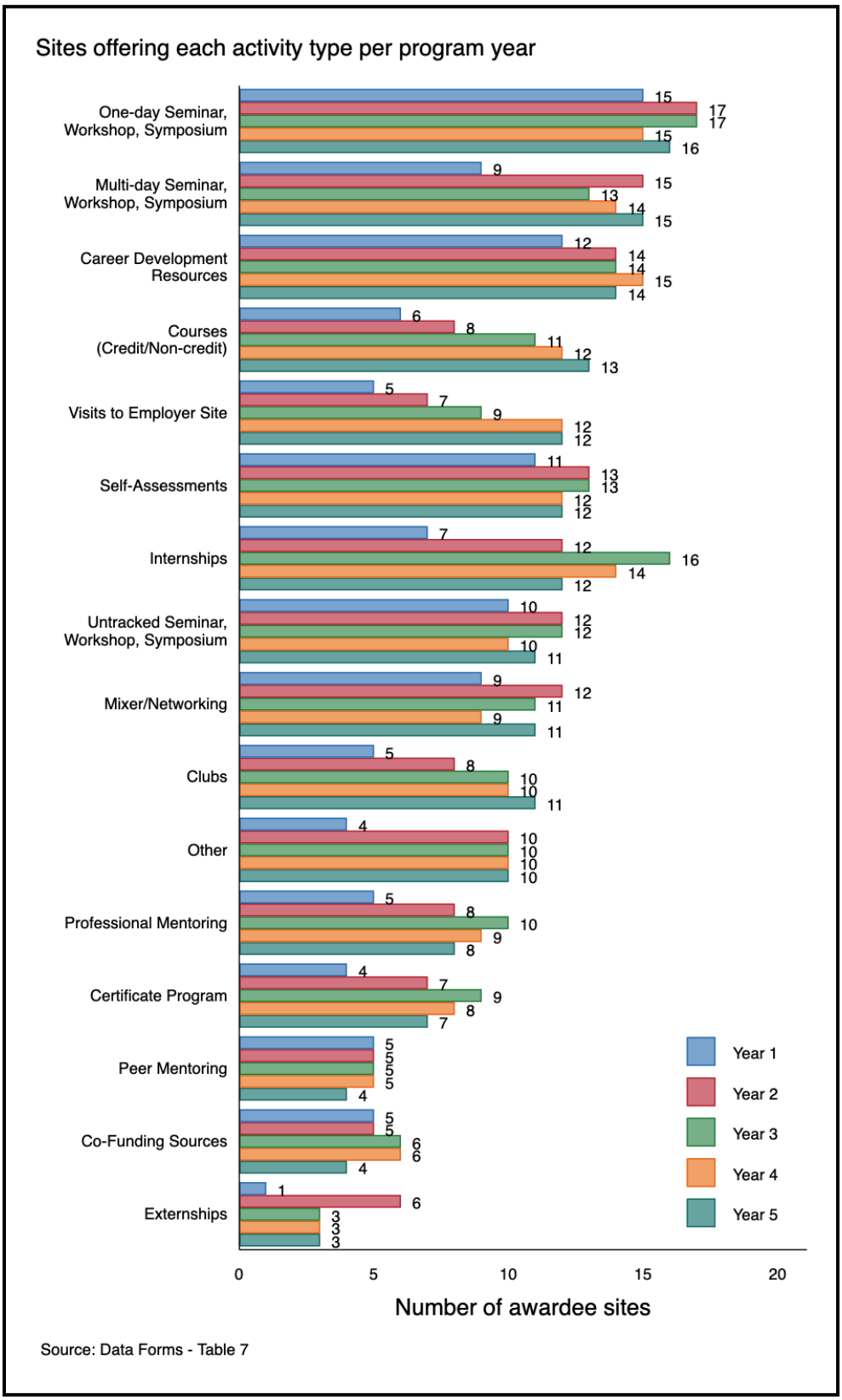

Figure 10. Number of awardee sites offering each type of career-development activity over the five-year award period, ordered by number of activities in Year 5.

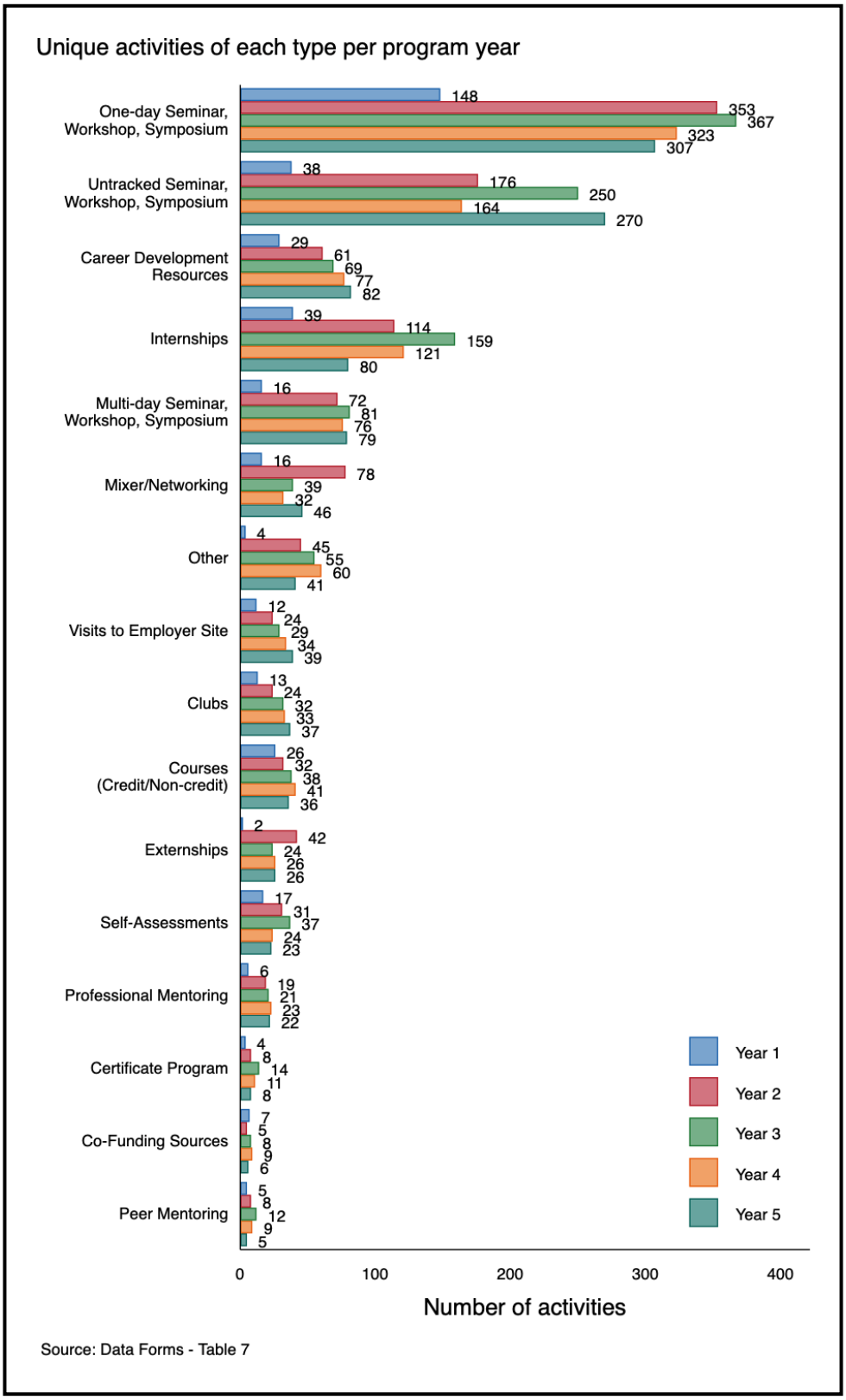

Figure 11. Number of unique program activities of each type offered over the five-year award period, ordered by number of activities in Year 5.

For many types of activity, offerings increase then decrease over the BEST award period. This trend likely reflects a learning process in which new activities were piloted over time but not all were deemed to have ongoing value or sustainability. Overall, however, there was a substantial broadening over the award period of the number of sites offering each type of activity as well as the number of activities offered (344 activities in Year 1 to 5,124 activities in Year 5).

As shown in **Figure 12**, BEST programs included a combination of activities that awardees reported as either a) already existing prior to the BEST award, b) enhancements to existing activities, or c) new activities initiated during the funding period. Proportionally, nearly three-quarters (72.6%) of activities were newly initiated during the award period, and more than one-third of those (37.4%) were multi-day seminars/workshops/symposia (**Table 8**). Twenty-one percent of activities were in place prior to the BEST award and the remaining six percent were enhancements or expansions of existing activities. Determination of whether an activity was existing/enhanced/new was made by the awardees.

Table 8. Development and expansion of BEST activities, by activity type, for all five program years

Activity	Existing Activities		Enhanced Activities		New Activities		Total Activities	
Clubs	18	1.7%	29	9.9%	91	2.6%	138	2.9%
Co-Funding Sources	10	1.0%	0	0.0%	25	0.7%	35	0.7%
Externships	42	4.1%	42	14.3%	89	2.5%	173	3.6%
Internships	0	0.0%	0	0.0%	120	3.4%	120	2.5%
Mixer/Networking	36	3.5%	28	9.5%	447	12.8%	511	10.6%
Peer Mentoring	62	6.0%	8	2.7%	141	4.0%	211	4.4%
Professional Mentoring	3	0.3%	4	1.4%	31	0.9%	38	0.8%
Career Development Resources	5	0.5%	13	4.4%	73	2.1%	91	1.9%
Self-Assessments	25	2.4%	8	2.7%	285	8.1%	318	6.6%
One-day Seminar, Workshop, Symposium	46	4.5%	5	1.7%	81	2.3%	132	2.7%
Multi-day Seminar, Workshop, Symposium	141	13.7%	46	15.6%	1311	37.4%	1498	31.0%
Untracked Seminar, Workshop, Symposium	50	4.8%	31	10.5%	243	6.9%	324	6.7%
Visit to Employer Site	556	53.9%	33	11.2%	309	8.8%	898	18.6%
Other	20	1.9%	14	4.8%	104	3.0%	138	2.9%
Totals	**1,031**	***100 %***	**294**	***100 %***	**3,505**	***100 %***	**4,830**	***100 %***
Proportion of all activities	***21.3%***		***6.1%***		***72.6%***		***100%***	

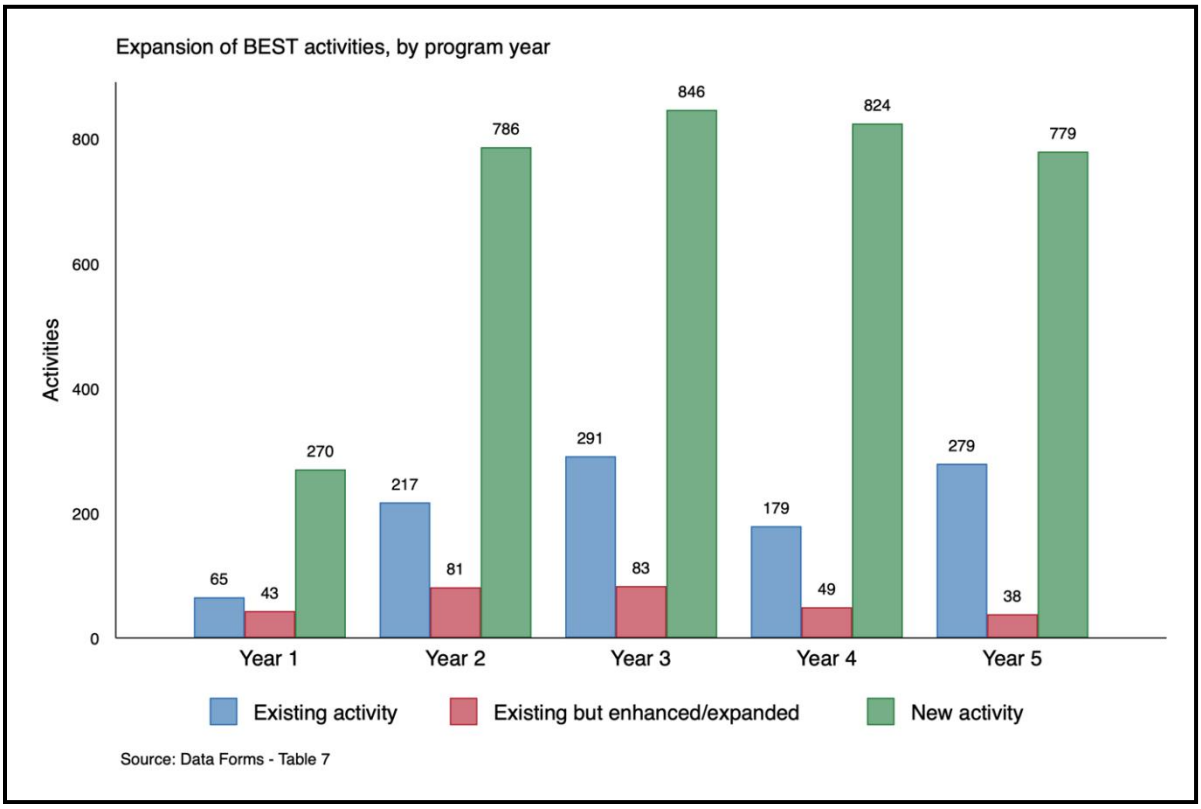

Figure 12. Yearly development and expansion of BEST program activities, for all activity categories combined.

3.4 Trainee participation

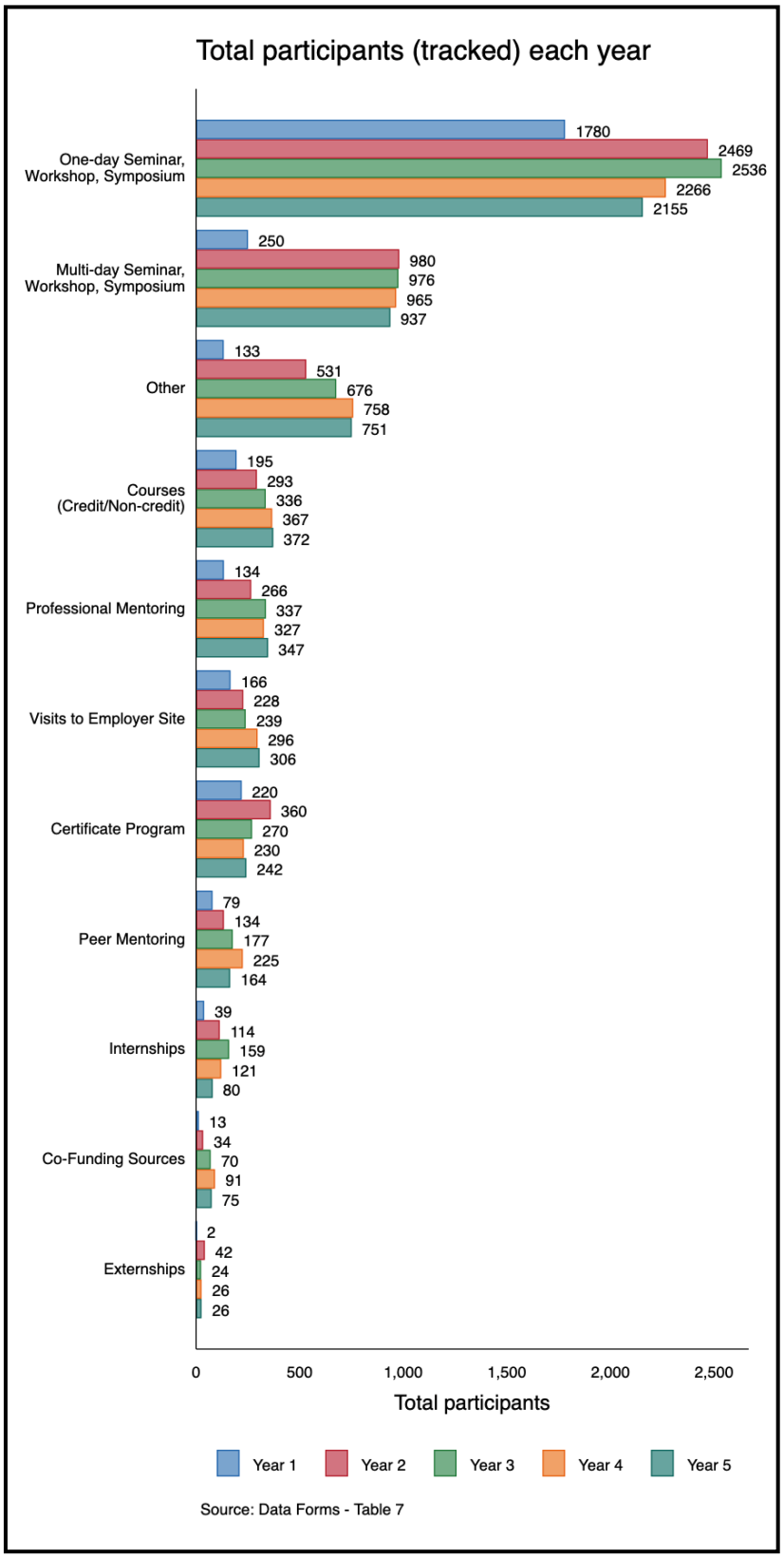

Figure 13. Yearly participation in BEST activities for which attendance was tracked.

Formal participation in the BEST program was defined as participation in at least one of the activities for which individual attendance could be tracked. By far the highest participation totals were for workshops/seminars/symposia (**Figure 13**) which is unsurprising, as these were also the most frequently offered activities. Per uniquely-offered activity, however, certificate programs had the highest average participation (29 trainees) among activities for which participation was tracked, but it should be noted that these were not singular events and may have spanned an entire academic term or year (**Table 9**). Professional and peer mentoring programs also had high levels of participation, averaging 15 and 20 trainees per activity, respectively.

Table 9. Number of BEST activities held each program year, total trainee participation, and average participation per event

	Program Year 1			Program Year 2			Program Year 3			Program Year 4			Program Year 5			All Years		
Tracked attendance	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**
Certificate	220	4	55.0	360	8	45.0	270	14	19.3	230	11	20.9	242	8	30.3	1322	45	29.4
Co-funding Sources	13	7	1.9	34	5	6.8	70	8	8.8	91	9	10.1	75	6	12.5	283	35	8.1
Courses	195	26	7.5	293	32	9.2	336	38	8.8	367	41	9.0	372	36	10.3	1563	173	9.0
Externships	2	2	1.0	42	42	1.0	24	24	1.0	26	26	1.0	26	26	1.0	120	120	1.0
Internships	39	39	1.0	114	114	1.0	159	159	1.0	121	121	1.0	80	80	1.0	513	513	1.0
Peer Mentoring	79	5	15.8	134	8	16.8	177	12	14.8	225	9	25.0	164	5	32.8	779	39	20.0
Professional Mentoring	134	6	22.3	266	19	14.0	337	21	16.0	327	23	14.2	347	22	15.8	1411	91	15.5
Single-day Workshop, Seminar, Symposium	1780	148	12.0	2469	353	7.0	2536	367	6.9	2266	323	7.0	2155	307	7.0	11206	1498	7.5
Multi-day Workshop, Seminar, Symposium	250	16	15.6	980	72	13.6	976	81	12.0	965	76	12.7	937	79	11.9	4108	324	12.7
Visit to Employer	166	12	13.8	228	24	9.5	239	29	8.2	296	34	8.7	306	39	7.8	1235	138	8.9
Other	133	4	33.3	531	45	11.8	676	55	12.3	758	60	12.6	751	41	18.3	2849	205	13.9
Estimated attendance*	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**	**T**	**A**	**T/A**
Clubs	608	13	46.8	907	24	37.8	1253	32	39.2	2029	33	61.5	2179	37	58.9	6976	139	50.2
Mixers/ Networking	603	16	37.7	1739	78	22.3	1947	39	49.9	1981	32	61.9	1970	46	42.8	8240	211	39.1
Self-assessments	1125	17	66.2	1318	31	42.5	1746	37	47.2	1443	24	60.1	1326	23	57.7	6958	132	52.7
Untracked Workshop, Seminar, Symposium	3395	38	89.3	7693	176	43.7	10327	250	41.3	6090	164	37.1	9042	270	33.5	36547	898	40.7
Totals	**8742**	**353**	**24.76**	**17108**	**1031**	**16.59**	**21073**	**1166**	**18.06**	**17215**	**986**	**17.46**	**19972**	**1025**	**19.48**	**84110**	**4561**	**18.44**

T: Trainees; A: Activities; T/A: Average number of trainees per activity

* Attendance for these activities was estimated by institution staff and may have included trainees who were not identified as part of the "treatment" group.

For BEST activities that were open events where individual participation was not tracked, attendance was estimated by BEST program staff. Estimated total attendance at these events was nearly six thousand in Year 1 and well over ten thousand each subsequent program year (**Figure 14**). Note that these estimates likely included many trainees from the comparison group, illustrating a potential influence of the program on trainees who were not defined as part of the participant group and that these programs were in a position to influence the overall culture at the BEST institutions.

Table 9 Error! Reference source not found. and **Figure 13** and **Figure 14** present totals for the n umber of unique participants per activity type, thus a single trainee may be represented in multiple or even all activities. **Figure 15** shows the number of unique trainees who participated in any BEST activities each year (i.e., each trainee is counted only once per year) out of all trainees in each participating department (i.e. those trainees who were potentially eligible to participate in BEST activities). About 1,800 trainees participated during the first year, and over 3,000 participated every year thereafter. Proportionally, BEST participants comprised 12.4% to 20.8% of the total number of trainees in participating departments. Both graduate students and postdoctoral scientists participated in a median of four activities each year (**Figure 16**). The large ranges for participation counts (1 to 42 for graduate students, 1 to 43 for postdoctoral scientists) can be attributed to variability in how awardees measured certain activities. For peer and professional mentoring, in particular, some awardees defined a series of meetings over an academic term as a single mentorship activity, whereas others counted each meeting as a separate event.

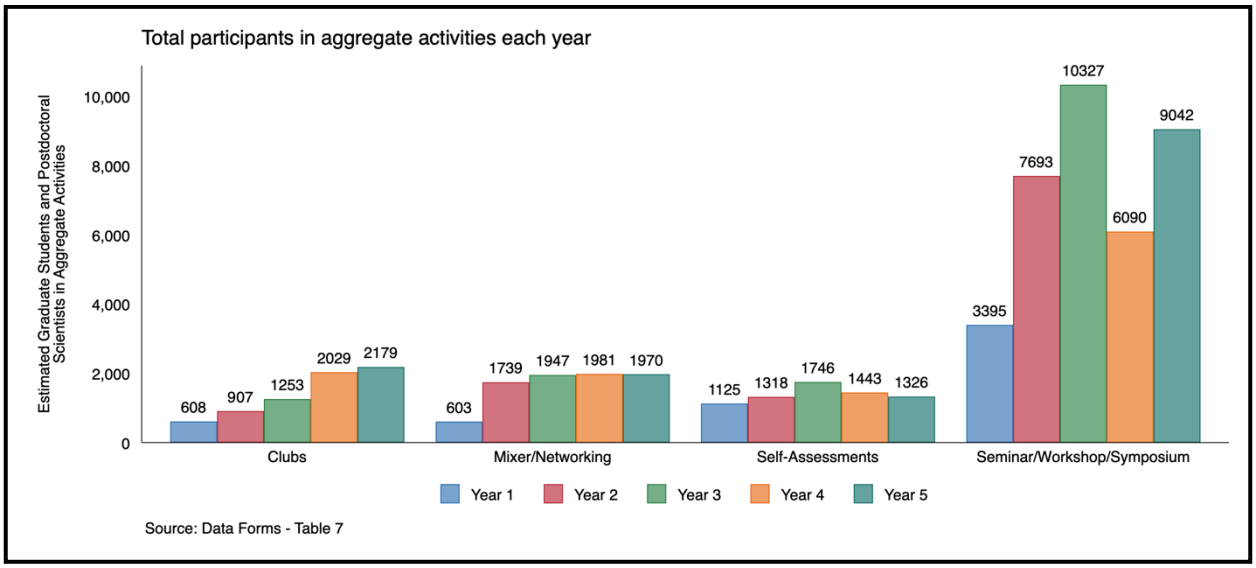

Figure 14. Estimated yearly participation in BEST activities for which attendance could not be tracked.

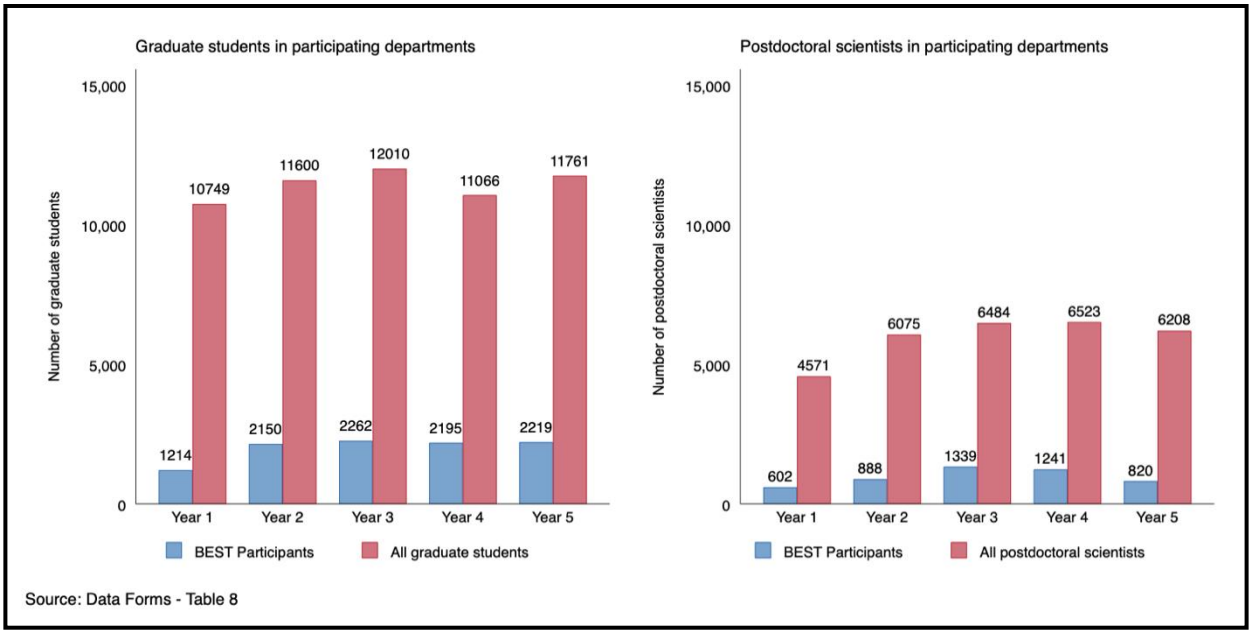

Figure 15. Total number trainees participating in BEST programs from among all trainees in participating departments each year.

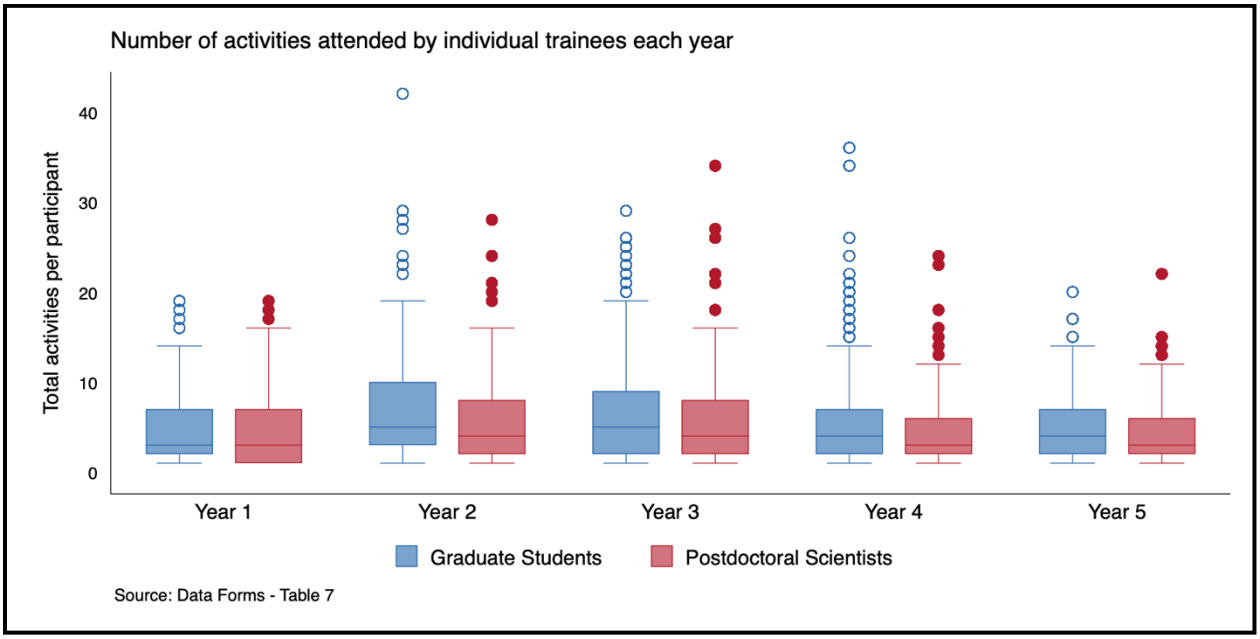

Figure 16. Total number of BEST program activities each individual trainee participated in each year (excluding activities for which attendance was not individually tracked).

3.5 Trainee career-development agency

3.5.1 Exposure to career paths

One goal of the BEST awards was to broaden trainees' exposure to career options outside the academic research path. From the list of twenty science-related careers designated in the *my*IDP (**Table 10**), awardees reported on the career paths to which they provided exposure in their

program. Three sites also included data science/data analysis, and three others specified that their BEST programs would support trainees exploring any career path interest. Other career paths mentioned were contract research organizations, science illustration and animation, and community or non-profit organizations in health- or biomedical-related domains.

Table 10. The twenty science-related career paths listed in the Individual Development Plan (myIDP), and the number of sites that incorporated exposure to those career paths into their BEST program each year

Career Paths	Year 1 Sites	Year 2 Sites	Year 3 Sites	Year 4 Sites	Year 5 Sites
Business of science	16	17	17	17	16
Entrepreneurship	16	17	17	17	16
Intellectual property	16	17	17	17	16
Research in industry	16	17	17	17	16
Science policy	16	17	17	17	16
Science writing	15	17	17	17	16
Drug/device approval and production	16	16	17	17	15
PI, research-intensive institution	15	16	16	16	15
Research administration	15	16	16	16	15
Combined research and teaching careers	14	16	16	16	15
Research staff, research-intensive institution	14	16	16	16	15
Teaching-intensive careers in academia	14	16	16	16	15
Science education for non-scientists	13	15	15	15	15
Sales/marketing of science-related products	13	15	14	14	15
Public health related careers	14	15	14	14	13
Support of science-related products	13	14	14	14	14
Clinical research management	14	14	13	13	12
Science education for K-12 schools	12	14	13	13	11
Scientific/medical testing	12	13	12	12	11
Clinical practice	8	10	10	10	8

3.5.2 Familiarity with career pathways

On Entrance Surveys, most graduate students (70.1%) reported they were familiar with “all” or “most” of the IDP career paths, while slightly fewer postdoctoral scientists (64.8%) reported familiarity with “all” or “most” career paths. Familiarity with career paths was significantly greater on Exit Surveys (**Table 11**), with the majority of both graduate students (82.0%) and postdoctoral scientists (73.2%) reporting familiarity with “all” or “most” career paths.

Table 11. Trainee-reported familiarity with the twenty IDP career paths at Entrance and Exit

Graduate Students	Entrance Survey		Exit Survey		Total		Chi^2
	N	%	N	%	N	%	*p*-value
I am familiar with all	274	24.5	387	34.9	661	29.7	50.31
I am familiar with most (between 13 and 19)	520	46.4	521	47.0	1041	46.7	<0.001
I am familiar with some (between 7 and 12)	240	21.4	145	13.1	385	17.3	
I am familiar with a few (between 1 and 6)	84	7.5	56	5.0	140	6.3	
I am not familiar with any	2	0.2	0	0.0	2	0.1	
Total	**1120**	**100.0**	**1109**	**100.0**	**2229**	**100.0**	
Postdoctoral Scientists	**Entrance Survey**		**Exit Survey**		**Total**		**Chi^2**
	N	**%**	**N**	**%**	**N**	**%**	***p*-value**
I am familiar with all	224	23.7	286	30.7	510	27.2	22.40
I am familiar with most (between 13 and 19)	387	41.0	399	42.9	786	41.9	<0.001
I am familiar with some (between 7 and 12)	209	22.1	142	15.3	351	18.7	
I am familiar with a few (between 1 and 6)	118	12.5	99	10.6	217	11.6	
I am not familiar with any	7	0.7	5	0.5	12	0.6	
Total	**945**	**100.0**	**931**	**100.0**	**1876**	**100.0**	

The mean change in familiarity scores was evaluated using Student's *t*-tests. For both graduate students (**Figure 17**) and postdoctoral scientists (**Figure 18**), there was a significant relationship between BEST participation and being familiar with an increased number of career paths between entrance and exit (for graduate students: t = -1.9919, p-value = 0.0466, for postdoctoral scientists: t = -2.0060, p-value = 0.0451).

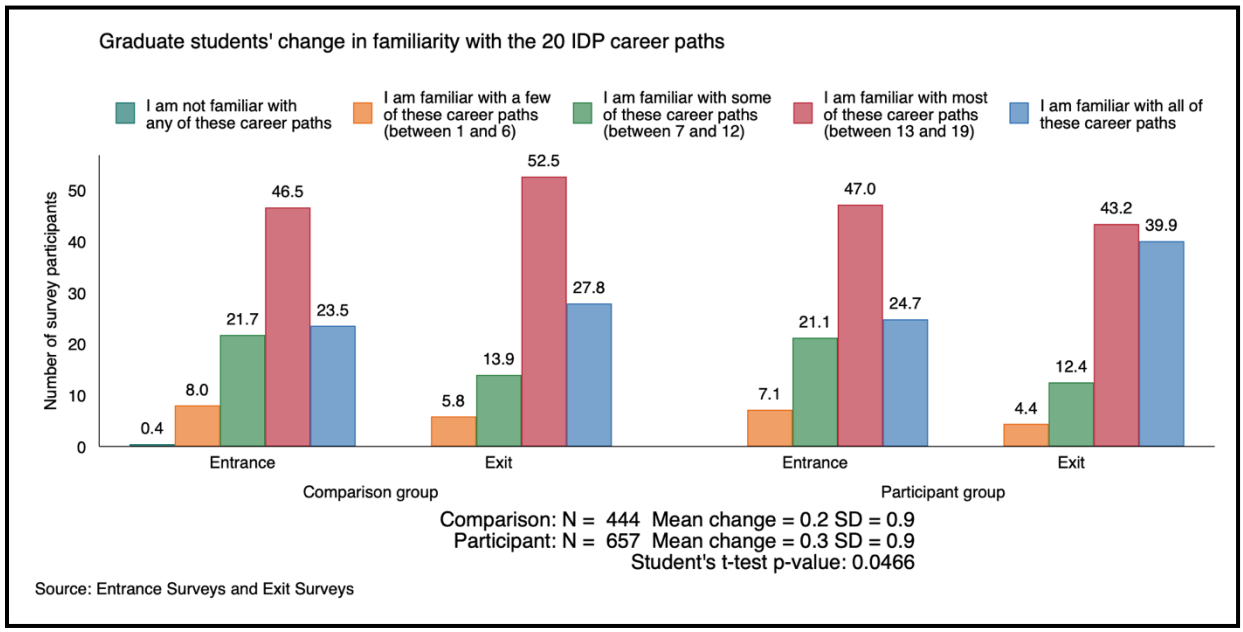

Figure 17. Change in the number of IDP career paths graduate students were familiar with on Exit Surveys as compared to Entrance Surveys.

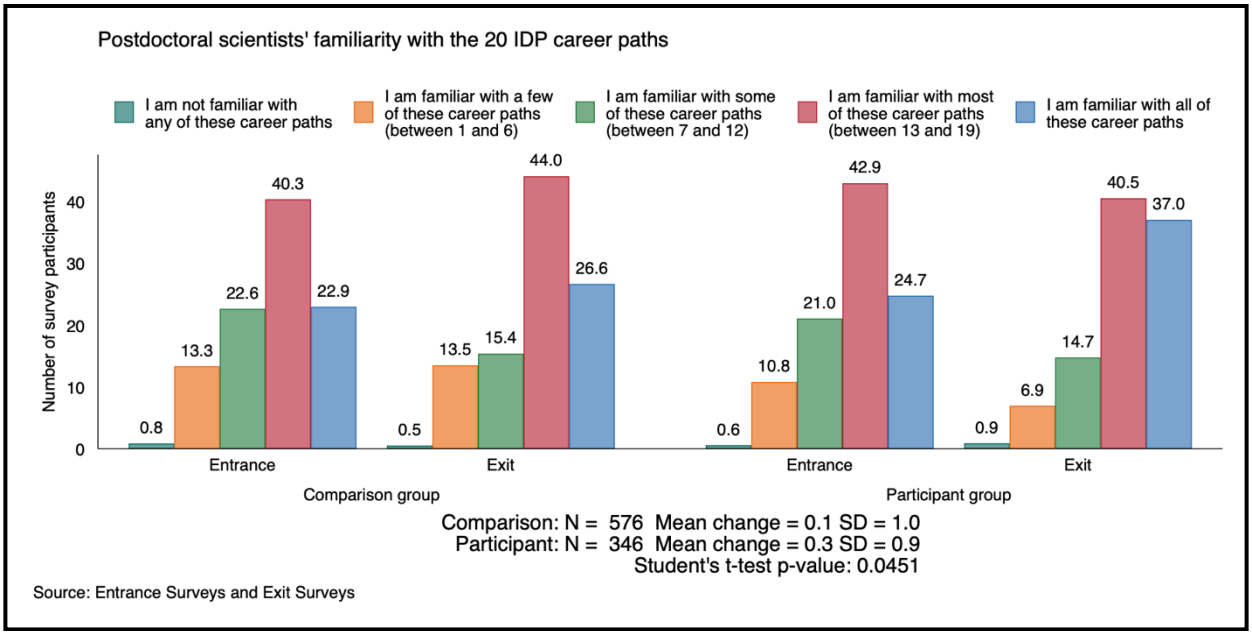

Figure 18. Change in the number of IDP career paths postdoctoral scientists were familiar with on Exit Surveys as compared to Entrance Surveys.

3.5.3 Consideration of career pathways

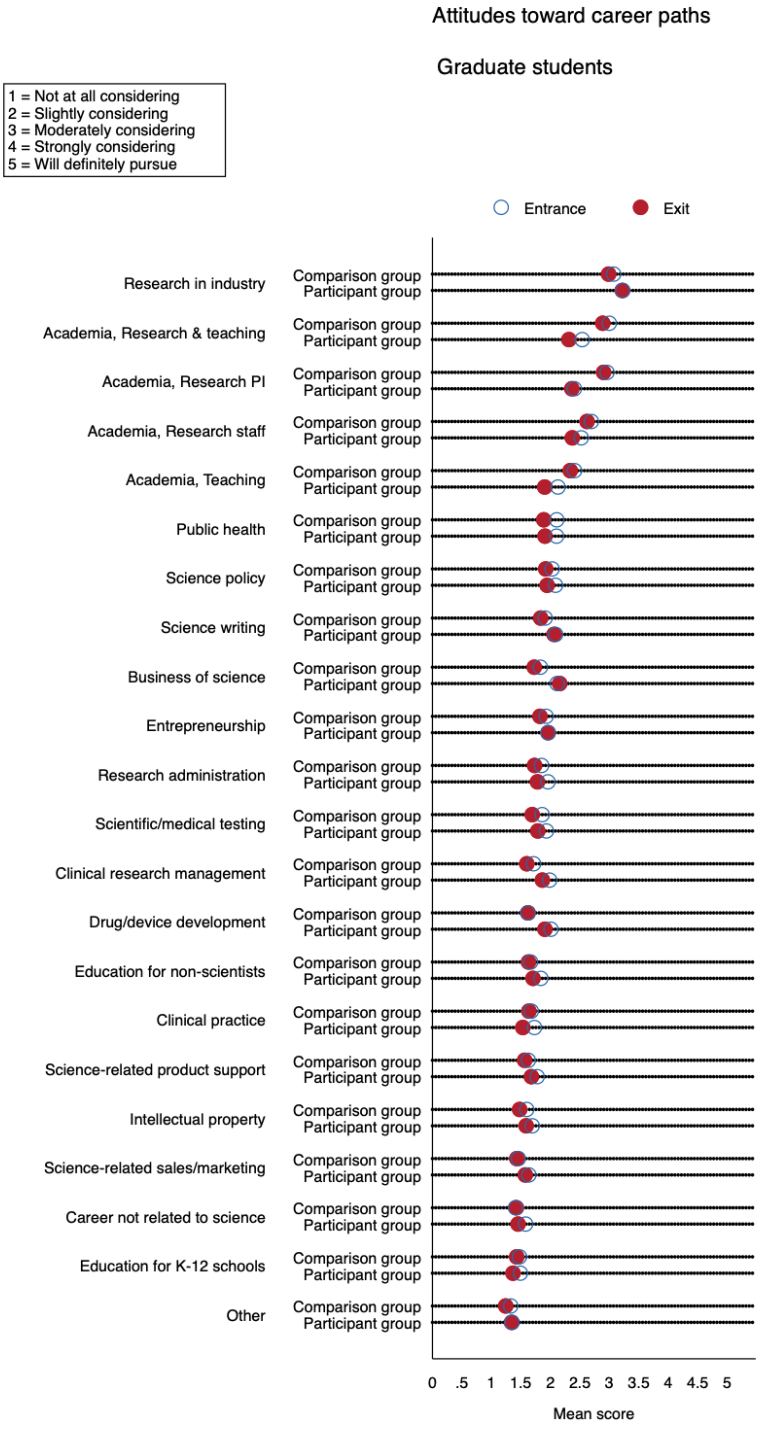

Figure 19. Graduate students' interest in pursuing the twenty IDP science-related career paths, measured on Entrance Surveys and Exit Surveys.

Trainee agency for making career decisions was also measured by asking respondents to rate the extent to which they were currently considering each of the twenty IDP career paths. Note that these were twenty separate questions scored from 1 = "Not at all considering" to 5 = "Will definitely pursue", rather than a sequential ranking of preference. Therefore, trainees could "strongly consider" multiple career paths. The most strongly considered career paths overall were research in industry, combined research and teaching in academia, and Principal Investigator in a research-intensive institution.

Figure 19 and show the mean score for each career path considered at entrance and exit, stratified by evaluation group. (Because of the possibility of increased Type I error with multiple comparisons, individual entrance-to-exit *t*-test comparisons for each career path were not part of the planned analyses, and changes are instead described as overall trends. It may be of interest to note, however, that non-overlapping circles generally correspond with *p*-values < 0.01).

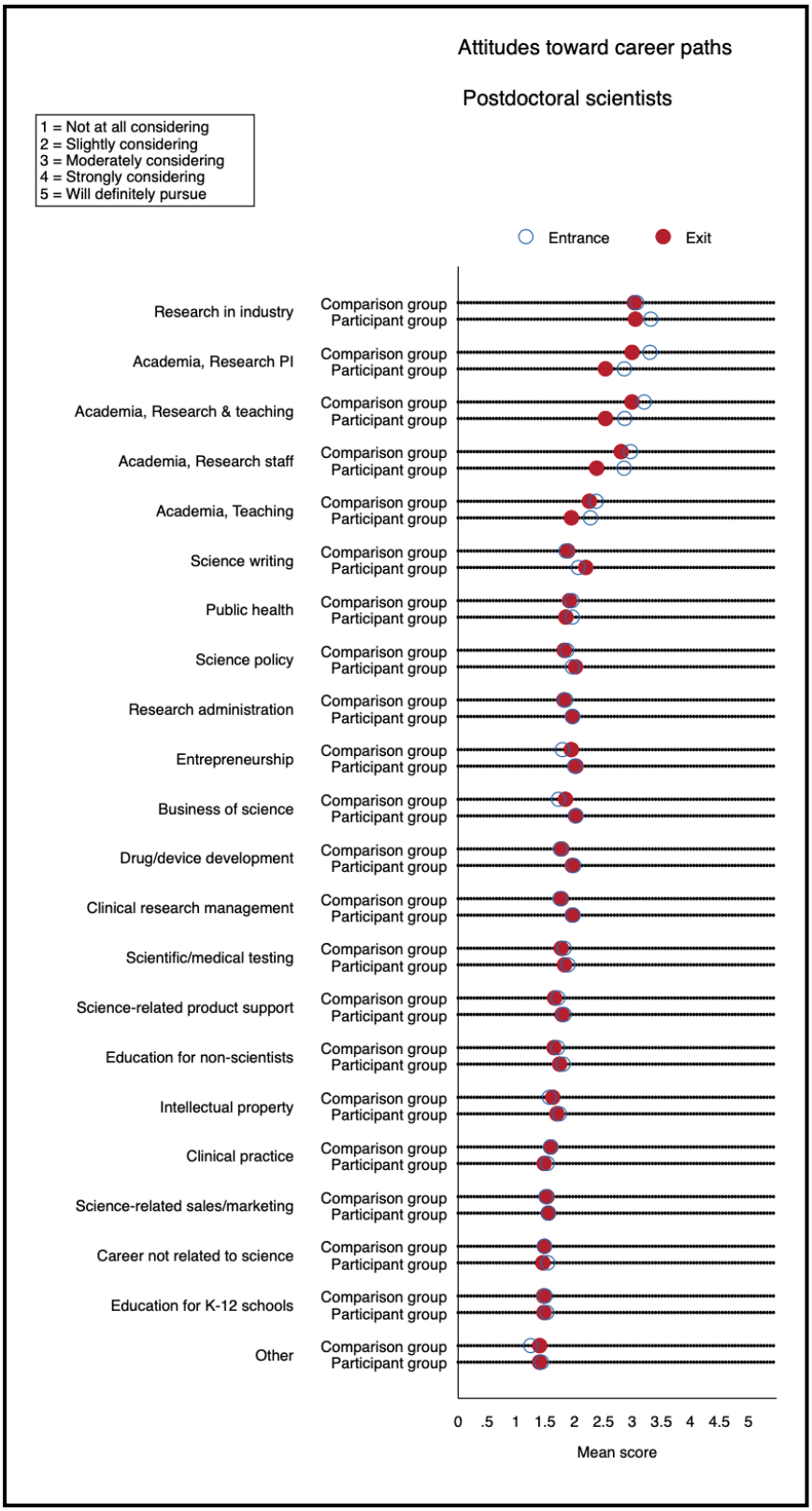

Figure 20. Postdoctoral scientists' interest in pursuing the twenty IDP science-related career paths, measured on Entrance Surveys and Exit Surveys.

Among graduate students, and especially those in BEST programs, consideration tended to decrease between the Entrance and Exit surveys for nearly all of the career paths. This global trend might represent not a decreased interest overall, but rather a narrowing of career focus by ruling out certain paths. Notably, neither BEST participants nor the comparison group changed how strongly they were considering Principal Investigator in a research-intensive institution ("Academia, Research PI") as a career path.

For postdoctoral scientists, a trend of decreased interest can be observed mostly among the most highly considered career paths, especially among BEST program participants. This likely reflects a shift in focus from traditional career paths—such as research in industry, research as a PI in academia, or research and teaching in academia—to multiple other options (where slight increases in interest can be observed).

3.5.4 Trainee confidence

Another measure of trainee career-development agency was trainee confidence. Using a scale from 1 = "Not at all confident" to 5 = "Completely confident" trainees were asked to self-rate their confidence in being able to:

1. assess their ability pursue their desired career path,
2. determine steps needed to pursue their desired career path,
3. seek advice from professionals in that career path,
4. identify potential employers relevant to their desired career path,
5. achieve their career goals, and
6. discuss their career goals with their PI/thesis advisor.

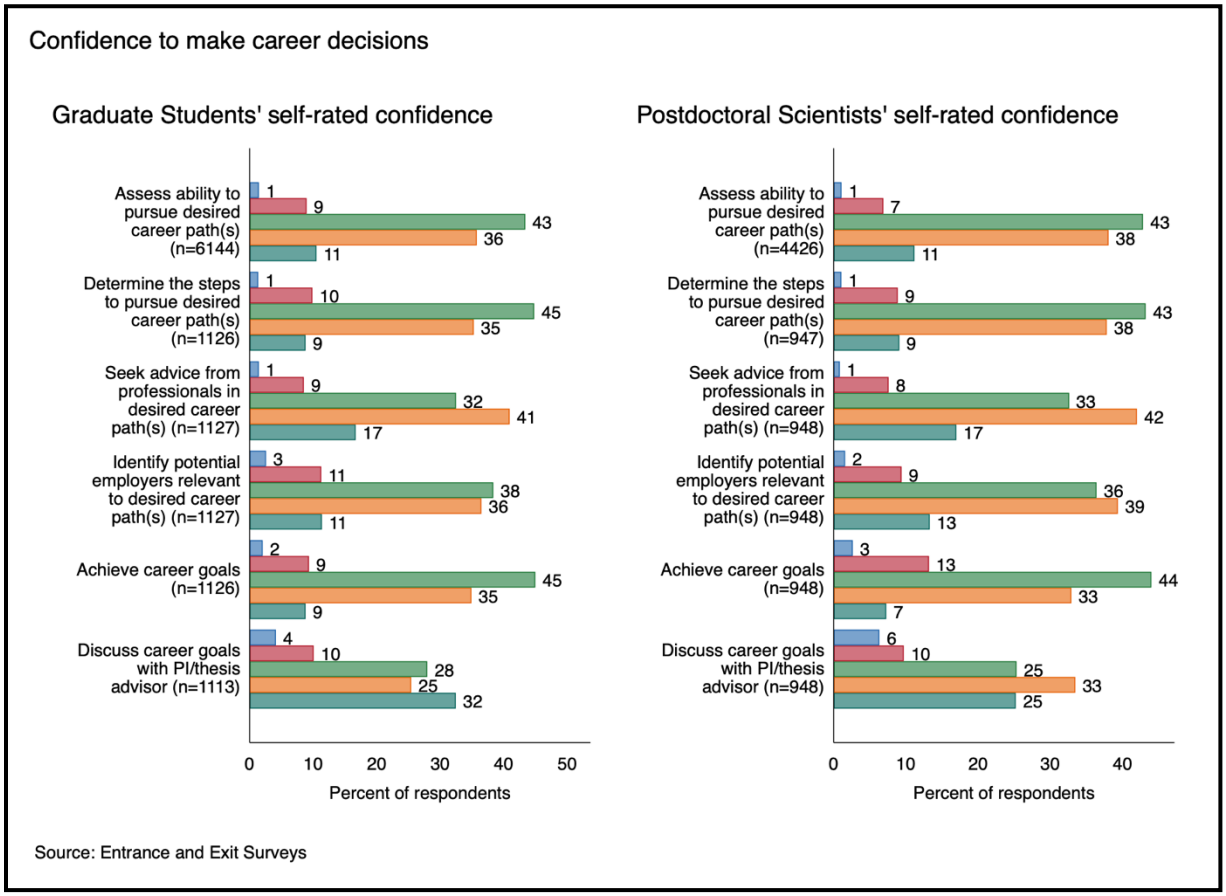

Figure 21. Trainees' self-rated confidence for six measures of career-development abilities, measured on the Entrance Survey.

In Entrance Surveys, 86% - 90% of all graduate students and 84% - 92% of all postdoctoral scientists reported being moderately, highly, or completely confident across all six measures (**Figure 21**).

Figure 22 shows confidence measures at entrance and exit, broken down by trainee type and evaluation group. For each subgroup, the mean score for each measure is represented. A general trend of increased confidence can be seen for all groups on almost all measures, but statistical

comparisons were not made for individual measures due to the risk of family-wise error from multiple comparisons. For statistical testing, the six survey items were averaged for each trainee to create a single composite confidence score.

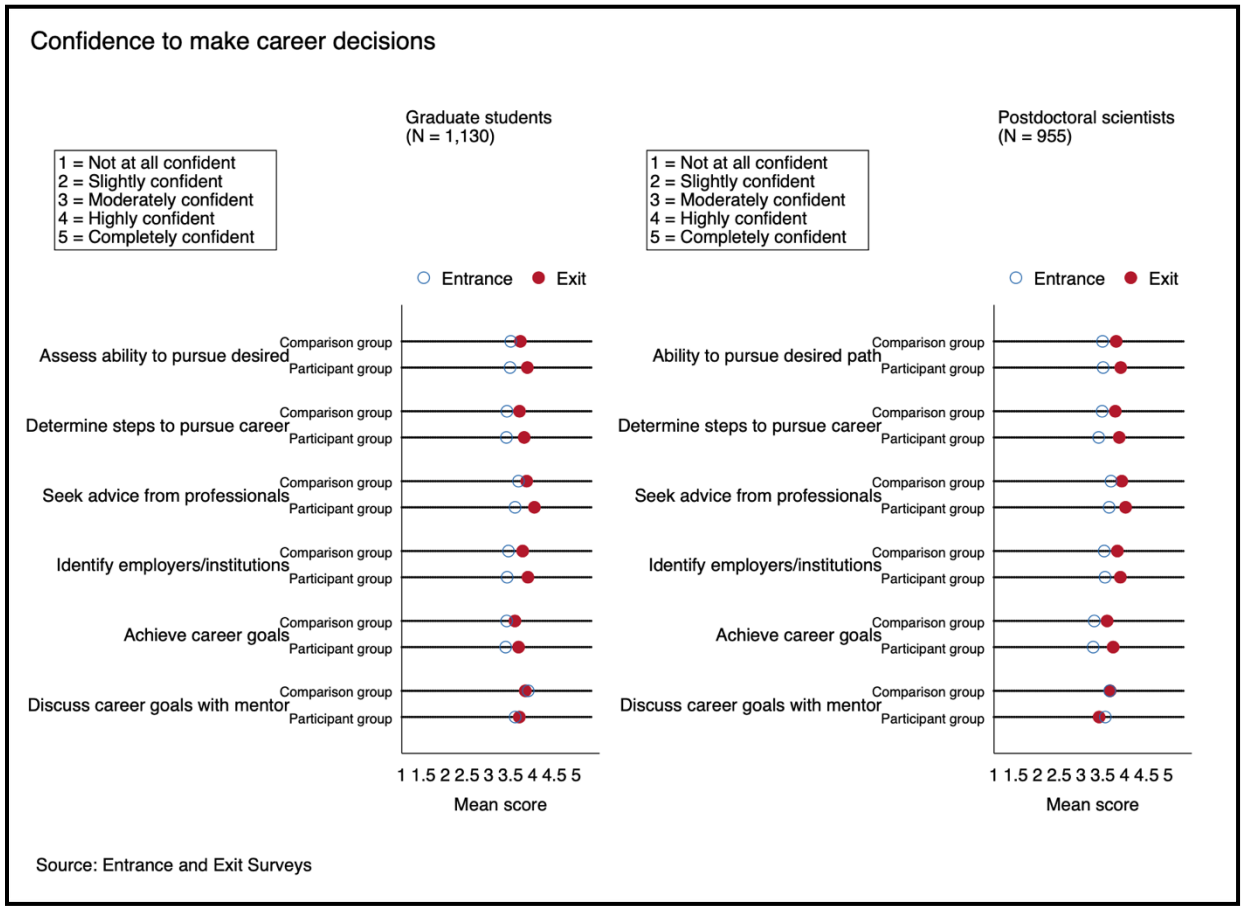

Figure 22. Changes from Entrance Survey to Exit Survey in six items measuring trainee confidence.

Table 12. Increase in trainee composite confidence scores from Entrance Survey to Exit Survey

Graduate Students	Mean	SE	95% Confidence Interval		*t*-test	*p*-value
Comparison group (n = 447)	0.19	0.03	0.13	0.25	-4.1777	> 0.001
BEST trainees (n = 663)	0.36	0.03	0.31	0.41		
Total	**0.29**	**0.02**	**0.25**	**0.33**		

Postdoctoral Scientists	Mean	SE	95% Confidence Interval		*t*-test	*p*-value
Comparison group (n = 587)	0.24	0.03	0.19	0.29	-1.8089	0.0708
BEST trainees (n = 349)	0.32	0.04	0.25	0.39		
Total	**0.27**	**0.02**	**0.23**	**0.31**		

To evaluate whether participation in the BEST program was associated with change in overall confidence, the change in composite confidence score from entrance to exit was calculated for each trainee who completed both Entrance and Exit Surveys. As shown in **Table 12**, confidence

scores increased across all groups, but only for graduate students was BEST participation associated with a significantly greater increase in confidence (*t*-test = -4.1777, $p > 0.001$).

3.6 Trainee time expenditures

3.6.1 Self-reported time to complete training

One of the goals of the BEST program is that participation would result in either reduced time or no increase in time to complete graduate studies or transition from postdoctoral positions to desired, non-training, non-terminal career opportunities. For graduate students (**Figure 23**), mean time to PhD completion was not significantly different for BEST participants (mean = 5.8 years; SD = 1.0 years) versus the comparison group (mean = 5.6 years; SD = 1.2 years; *t*-test = -1.8716; *p*-value = 0.0218).

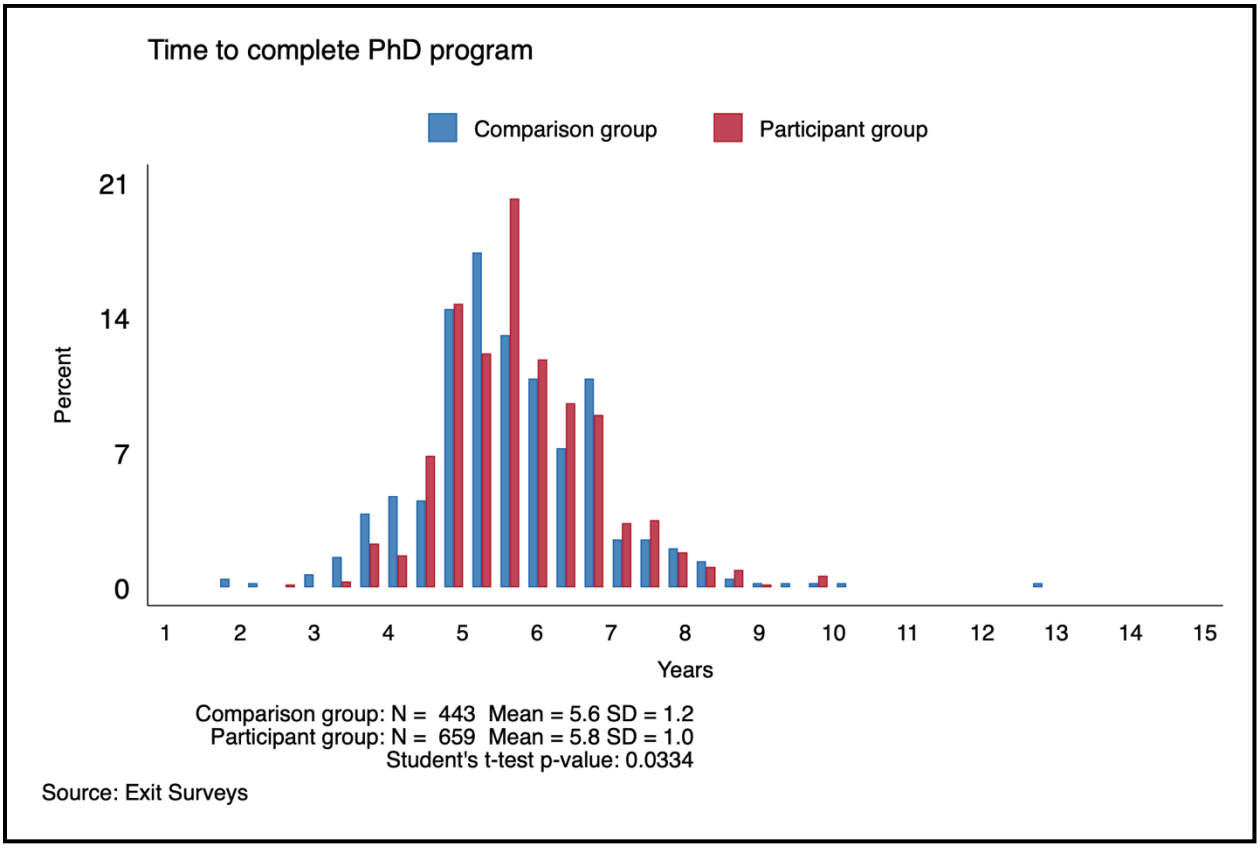

Figure 23. Comparison of graduate students' self-reported time to complete PhD for BEST participants versus non-participants.

Among postdoctoral scientists (**Figure 24**), BEST participants did spend statistically significantly more time in training (mean = 3.5 years; SD = 1.9 years) versus the comparison group (mean = 3.1 years; SD = 1.6 years) but the difference was only about four and a half months (difference in means = 0.37 years; SD = 0.26 years; *t*-test = -1.4312; *p*-value = 0.0131).

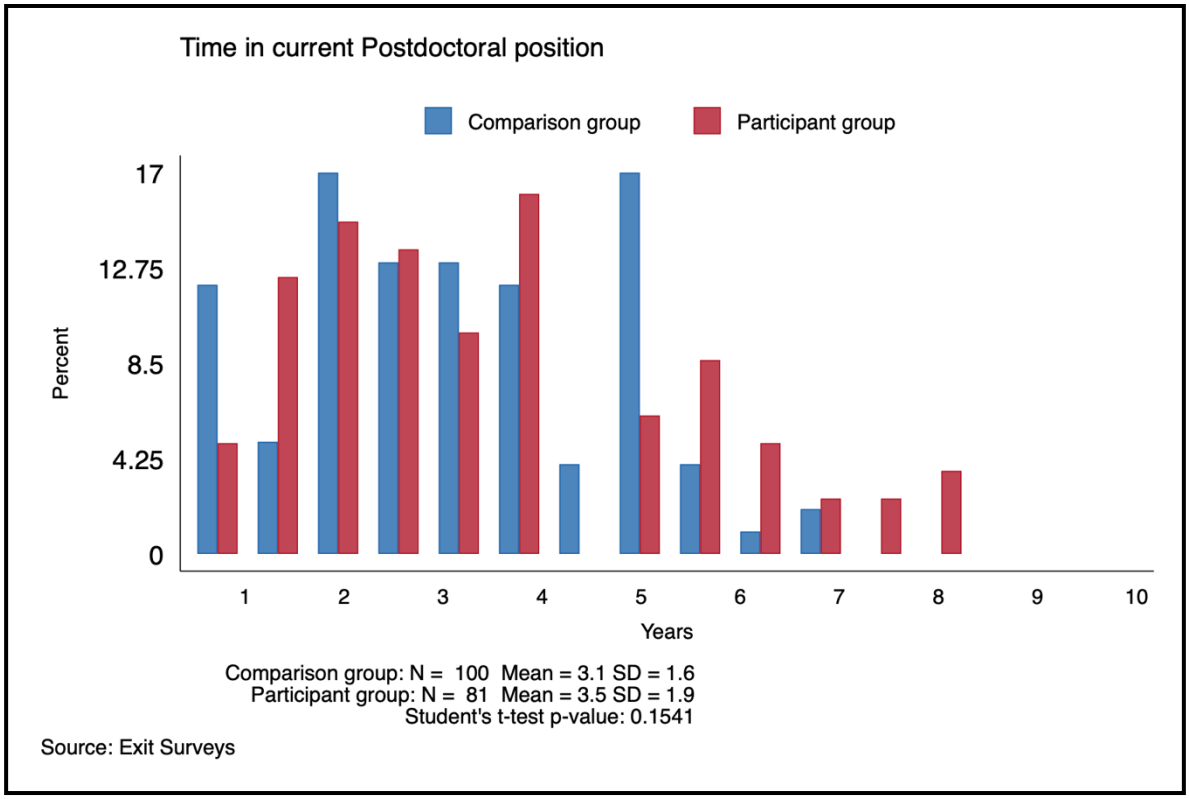

Figure 24. Comparison of postdoctoral scientists' self-reported time in training for BEST participants versus non-participants.

Hierarchical regression modeling was conducted to evaluate whether a relationship could be detected between BEST participation and time in training, while controlling for demographic and other factors and accounting for sources of correlation within the different awardee sites. No relationships were found in these models between BEST participation and time to degree or time in postdoctoral training.

It is worth mentioning that some significant relationships appeared between awardee sites and time in training. Comparisons across sites are not within the scope of this evaluation but, as the BEST awards were considered research grants, individual awardees were expected to propose and test hypotheses about how well their program and interventions informed and prepared trainees for broad career options. The BEST awardees have begun to report on their individual experiences elsewhere.[7]

3.6.2 Awardee-reported time to complete training

A separate regression model was developed using data from the Data Forms, in which awardees were asked to estimate the median time to PhD for each academic department participating in BEST programming. These estimates were aggregated by taking the mean of the medians, weighted by the number of graduate students in each department. Awardees were also asked to estimate the median time to degree for the five years prior to initiating their BEST programs. For all awardees combined, the mean of the median historical time to PhD was 5.62 years (standard error = 0.38). The mean-median times to PhD during the BEST program years are shown in

Table 13. Hierarchical modeling was used to account for correlations within awardee institutions and over time. Model parameters for the fixed effects are presented in **Table 14**.

Table 13. Means (with standard errors) of median years to complete PhD as estimated by awardee sites for each participating department, weighted by number of trainees per department

	1st Year Mean (SE)	2nd Year Mean (SE)	3rd Year Mean (SE)	4th Year Mean (SE)	5th Year Mean (SE)	All Years Mean (SE)
Comparison group	5.66 (0.49) n = 1,344	5.75 (0.25) n = 1,338	5.67 (0.53) n = 1,154	5.58 (0.43) n = 390	5.48 (0.54) n = 163	5.66 (0.44) n = 4,389
Participant group	5.86 (0.34) n = 77	5.41 (0.59) n = 182	5.46 (0.68) n = 309	5.86 (0.67) n = 204	5.87 (0.83) n = 95	5.60 (0.64) n = 867
All trainees	5.67 (0.44) n = 1,618	5.68 (0.24) n = 1,683	5.64 (0.43) n = 1,523	5.78 (0.48) n = 633	5.94 (0.74) n = 258	5.64 (0.41) n = 5,715

Table 14. Fixed effects parameters for the model of awardee-reported time to PhD

Fixed Effects Parameters	Regression Coefficient	Standard Error	*p*-value	95% Confidence Interval	
BEST participation	-0.53	0.24	0.026	-1.00	-0.06
Program year	-0.01	0.06	0.104	-0.22	0.02
Interaction (participation * year)	0.14	0.10	0.138	-0.05	0.33
Mean of median time to PhD in five years preceding BEST	0.88	0.09	<0.001	0.70	1.05
Participating departments/programs	< 0.01	< 0.01	0.0129		
Doctorate degrees awarded to all students	< 0.01	< 0.01	0.4445		
Doctorate degrees awarded to BEST participants	< 0.01	< 0.01	0.5202	0.70	1.05

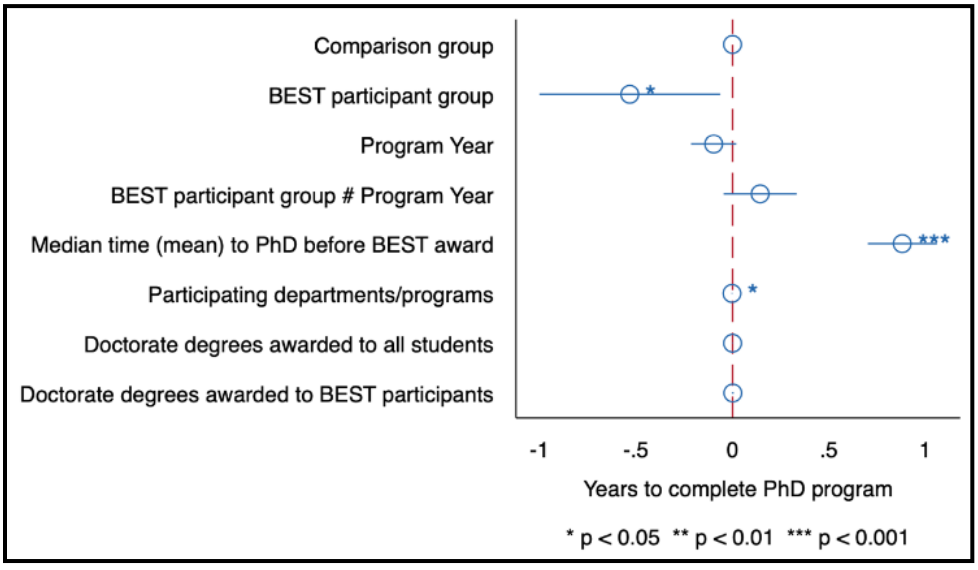

Figure 25. Plot of (fixed effects) coefficients for regression model of awardee-reported time to PhD. Circles represent point estimates for regression coefficients, bars represent 95% confidence intervals.

The median time to completion of a PhD was, on average, about 6 months less (regression coefficient = 0.53 years) for BEST participants as compared to the comparison group. The historical time to PhD completion (for the five years preceding the BEST award) was also a strong predictor; the average median time to complete a PhD during BEST program years was approximately 11 months longer (regression coefficient = 0.88 years) for every one-year increase in historical time to PhD. This finding is unsurprising but worth noting, as it may help to explain the strong associations between awardee site and time in training found in models based on trainee Survey data.

3.7 BEST Program Infrastructure and Sustainability

The third outcome assessed in this program evaluation is establishment or further development of institutional infrastructure to continue BEST-like activities. Factors that impact sustainability include program leadership, faculty attitudes toward the program, and partnerships with external individuals and organizations that can provide instructional and/or financial support.

3.7.1 Program oversight

The majority of awardee sites assembled both an advisory board and steering committee to guide BEST program implementation and development. **Table 15** summarizes the composition and sizes of these governing bodies. Generally, advisory boards provided advice, feedback, and suggestions, but took a more active role at some sites by liaising with external partners. Functions of steering committees ranged from providing relatively hands-off program review and implementation advice to directly engaging in program administration and activities.

Table 15. BEST program leadership structures: type, composition, and size

Leadership type	Year 1	Year 2	Year 3	Year 4	Year 5
Advisory board only	6	7	7	7	7
Steering committee only	2	1	0	0	0
Both	8	9	10	10	9
Composition	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Advisory Boards					
Internal members only	1	1	1	1	1
External members only	2	2	3	3	2
Both	11	13	13	13	13
Steering Committees					
Internal members only	9	8	7	8	7
External members only	0	0	0	0	0
Both	1	2	3	2	2
Size (summarized for all years)	**Minimum**	**Median**	**Maximum**		
Advisory Boards	4	12	34		
Steering Committees	4	8	25		

* Not all awardees submitted complete information.

3.7.2 Faculty engagement

Among all seventeen awardee sites, well over 400 departments participated in BEST programming each year (**Table 16**). On average, about four percent of faculty in graduate departments were active in delivering BEST programming, and another ten percent were not themselves active but supervised graduate students who were BEST participants. Participation rates were somewhat lower for faculty supervising postdoctoral scientists (one percent and three percent, respectively).

Table 16. Departmental and faculty participation in BEST programs

	Program Year	Participating Departments	All Faculty	BEST Faculty†	Non-BEST Faculty with BEST Trainees
Programs with graduate students*	1st	458	2,588	101	321
	2nd	465	3,849	195	535
	3rd	483	3,844	79	360
	4th	491	6,437	206	531
	5th	446	4,324	231	412
Departments with postdoctoral scientists*	1st	450	3,111	74	156
	2nd	522	4,925	71	128
	3rd	528	4,687	48	135
	4th	518	6,885	56	248
	5th	472	5,304	81	173

* Some awardees listed graduate programs and academic departments (for postdoctoral scientists) separately, but these often overlapped in terms of faculty composition. Totals are reported separately to avoid double-counting.
† In Year 2 and Year 3, one awardee did not distinguish between BEST-participating faculty and non-BEST-participating faculty. Those faculty (N=381 in Year 2, N=347 in Year 3) were removed from these calculations.

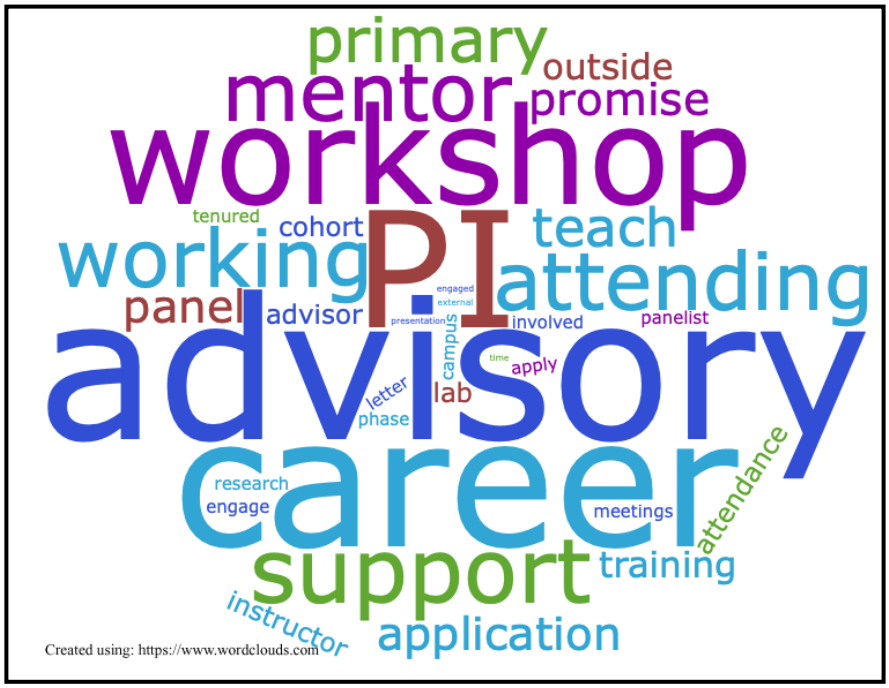

Figure 26. Word-cloud representation of the terms most frequently used to define faculty participation in BEST programs.

Definitions of faculty participation varied substantially among institutions, but common themes included participation in advisory committees, attending or contributing to program activities and curricula, and/or serving as a mentor to BEST trainees outside their own lab. **Figure 26** provides a representation of the terms most frequently used. Awardee sites also varied considerably in how they quantified faculty participation in BEST programs (summarized in **Table 17**). Reports of yearly activity frequency ranged from one to more than one thousand, and participation hours per year ranged from 0.5 to more than 150.

Table 17. Participation hours and frequency of faculty participation in BEST programming

Program Year	N*	Participation Hours		Activity Frequency
		Mean	(SD)	Median
1st	10	9.1	(6.6)	6
2nd	14	17.3	(24.5)	6
3rd	11	29.2	(31.2)	15
4th	10	34.1	(35.7)	13
Overall	**55**	**25.4**	**(33.2)**	**8**

* Not all seventeen awardees provided data each year

3.7.3 Faculty attitudes

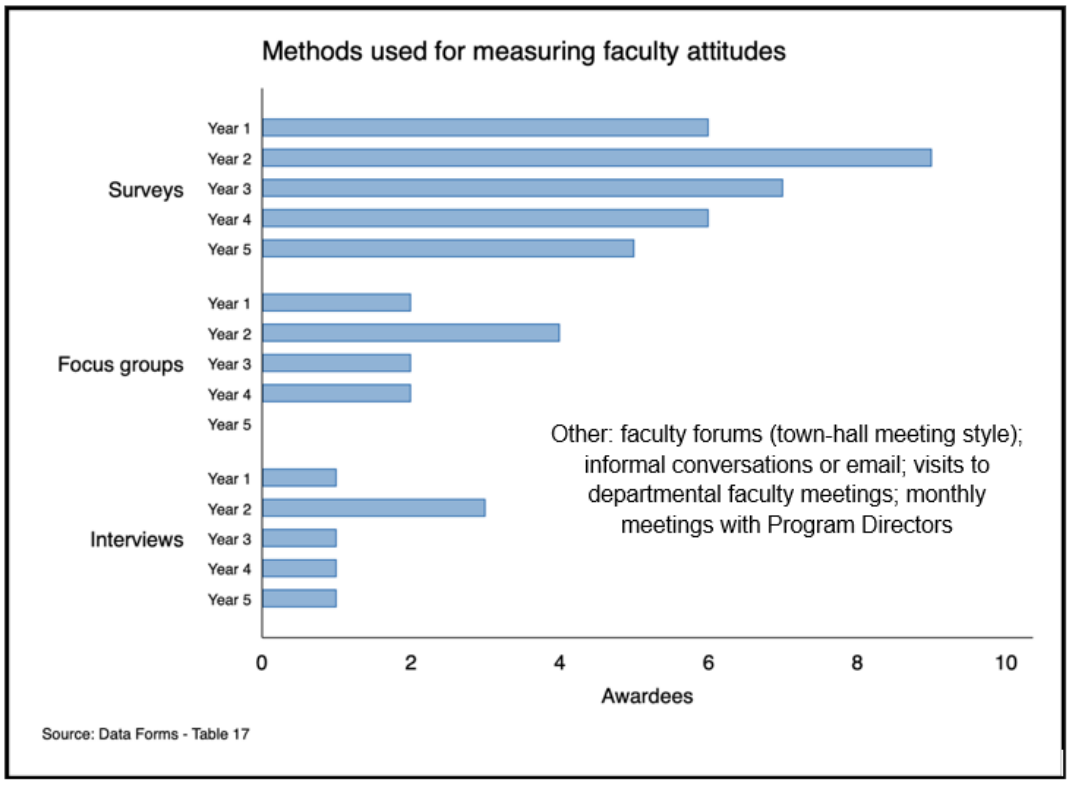

Figure 27. Methods used for measuring faculty attitudes toward BEST programs.

All awardee sites measured faculty attitudes toward the BEST program at least once during the five-year funding period, and many conducted surveys each year. Surveys were the most common method used for gathering faculty feedback, followed by focus groups (**Figure 27**). The BEST Data Form included suggested topics for faculty attitude assessments, such as attitudes toward trainee participation in career development activities and acceptance of trainees pursuing careers outside of academia. Many sites added topics including faculty awareness of BEST programming or other institutional resources for advising on non-academic careers, preferences for mentoring trainees interested in pursuing academic careers, and faculty familiarity/relationships with businesses or professionals outside of academia.

Because faculty surveys varied across institutions, it is not possible to provide summary statistics, but all awardees reported that faculty attitudes generally supported trainees' pursuit of non-academic careers and participation in career development activities. Opinions were often mixed, however, on the amount of time that was considered appropriate for such activities, and whether or not participation reduced research productivity. Many sites used survey results to identify areas in which faculty require more assistance, including lack of departmental support for extracurricular training, unfamiliarity with careers outside academia, and limited confidence for mentoring trainees pursuing non-science-related careers. Faculty also noted that, apart from benefits to trainees, BEST programs may benefit institutions and departments by enhancing recruitment efforts and adding value to grant applications. (See a 2019 publication by Watts SW, Chatterjee D, Rojewski JW, et. al. for a more extensive study of faculty knowledge and perceptions related to trainee career development at seven academic institutions offering BEST programming.[8])

3.7.4 External funding

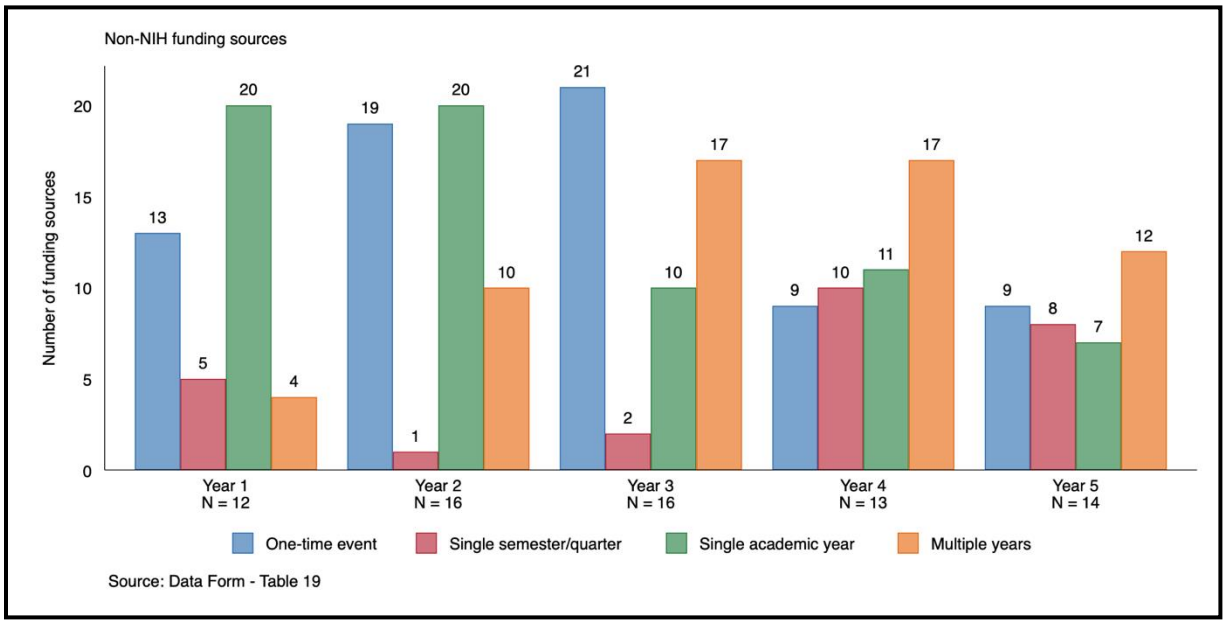

Figure 28. Number of funding sources external to the BEST awards reported each year.

Another measure of sustainability is the procurement of external funding that would allow programs to continue activities beyond the BEST award period. All 17 sites reported receiving some external funding each year. Between thirteen and sixteen sites reported non-BEST funding for their programmatic activities each year (**Figure 28**).

Although the number of funding sources decreased in later years, the median funding amount from each source increased (**Table 18**), in part due to a shift from short-term funding (single event, semester, or year) to long-term funding.

Table 18. Funding amounts external to the BEST awards reported by awardee sites each year

Year	Sites	Sources*	Median	Interquartile Range	Minimum	Maximum	Total Funding
Year 1	12	38	$12,500	$32,500	$35	$175,000	$1,042,388
Year 2	16	50	$5,000	$19,000	$100	$179,900	$962,737
Year 3	16	47	$4,577	$34,500	$200	$100,000	$1,034,258
Year 4	13	32	$15,000	$42,500	$500	$100,000	$864,517
Year 5	14	36	$3,700	$29,450	$250	$850,000	$1,431,350
Total	**71**	**203**	**$ 5,000**	**$ 33,532**	**$ 35**	**$ 850,000**	**$ 5,335,250**

* For some funding sources, monetary data were not reported.

Funding sources were not always explicitly documented (e.g., "various"), but from the available data it appears that institutional departments and programs provided more than half of the individual funds and more than seventy-five percent of total funding. Funding was also obtained from endowments, foundations (e.g., Burroughs Welcome Fund), associations (e.g., American Heart Association), private companies, private donors, and government agencies (National Science Foundation, National Institutes of Health, the World Health Organization). One awardee funded their program in part by charging a small yearly registration fee (< $20 per trainee) to enroll in the BEST program.

Figure 29. Terms most commonly used to describe how funds from sources external to the BEST awards were used in BEST programs.

Uses for the external funding varied widely across sites, but common themes included trainee support (e.g., stipends, internships), salaries for faculty and staff, development of curriculum and programs, and administrative costs (e.g., travel, food, marketing). **Figure 29** provides a representation of the terms most commonly used.

4 SUMMARY

The goals for this evaluation of the BEST program were to:

1. Assess changes in understanding of career opportunities, confidence to make career decisions, and attitudes towards career opportunities;
2. Determine reduced time or no increase in time to desired, non-training, non-terminal career opportunities, and reduced time in postdoctoral positions; and
3. Identify creation and/or further development of institutional infrastructure to continue BEST-like activities.

Although most graduate students and more than half of postdoctoral scientists reported familiarity with "all" or "most" career paths on Entrance Surveys, increased familiarity was still observed on the Exit Surveys for all trainees. BEST participation correlated with a greater increase in familiarity among both postdoctoral scientists and graduate students.

When trainees were asked about their interest in pursuing the twenty IDP career paths, the most highly considered careers were research in industry, combined research and teaching in academia, and Principal Investigator in a research-intensive institution. Between responding to an Entrance and Exit Survey, most graduate students, and especially those in BEST programs, indicated narrowed career interests as evidenced by decreased interest across most career paths. For postdoctoral scientists, interest decreased primarily among the most highly considered career paths, likely reflecting a shift in focus away from traditional careers in industry and academia.

For six measures of trainee confidence, 86% - 90% of all graduate students and 84% - 92% of all postdoctoral scientists reported on Entrance Surveys that they were moderately, highly, or completely confident. Confidence scores on Exit Surveys increased in all groups, but BEST participation was associated with a greater increase in confidence only among graduate students.

BEST participation was not associated with a difference in self-reported time to complete PhD, but postdoctoral scientists who were BEST participants spent about four and a half months longer in training than non-participants. In regression modeling to control for demographic and other factors, no correlations were found between BEST participation and time to degree or time in postdoctoral training.

Program leadership, faculty attitudes toward the program, and external partnerships for instructional and/or financial support were quantified to assess program sustainability. Most sites assembled both an advisory board (usually for hands-off guidance) and a steering committee (for guidance, review, and—in some cases—direct administration). In surveys, focus groups, and interviews, faculty were generally supportive of trainees' pursuit of non-academic careers and participation in career development activities, but opinions were mixed on the amount of time

considered appropriate for those activities, and whether participation negatively affected research productivity. Each program year, between twelve and sixteen sites procured non-BEST funding for programmatic activities each year. More than half of the individual funding sources and more than seventy-five percent of total funding originated from sites' institutional departments and programs.

5 LIST OF ACRONYMS AND ABBREVIATIONS

Acronym	Meaning
AAAS	American Association for the Advancement of Science
ANOVA	Analysis of Variance
BEST	Broadening Experiences in Scientific Training
DPCPSI	Division of Program Coordination, Planning, and Strategic Initiatives
GS	Graduate Student
IDP	Individual Development Plan
NIH	National Institutes of Health
OSC	Office of Strategic Coordination
PD	Postdoctoral Scientist
PI	Principal Investigator
SD	Standard deviation
SE	Standard error

6 REFERENCES

[1] Mathur A, Meyers FJ, Chalkley R, O'Brien TC, Fuhrmann CN. Transforming training to reflect the workforce. Sci Transl Med. 2015;7:285ed4.

[2] Meyers FJ, Mathur A, Fuhrmann CN, et al. The Origin and implemen- tation of the Broadening Experiences in Scientific Training programs: an NIH common fund initiative. FASEB J. 2016;30(2):507-514.

[3] National Institutes of Health. *NIH Director's Biomedical Workforce Innovation Award: Broadening Experiences in Scientific Training (BEST) Funding Opportunity Announcement RFA-RM-12-022*. Bethesda, MD: National Institutes of Health; 2013. https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-12-022.html. Accessed September 27, 2019.

[4] National Institutes of Health. *NIH Director's Biomedical Workforce Innovation Award: Broadening Experiences in Scientific Training (BEST) Funding Opportunity Announcement*

RFA-RM-13-019. Bethesda, MD: National Institutes of Health; 2014. https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-13-019.html. Accessed September 27, 2019.

[5] Lenzi RN, Korn SJ, Wallace M, Desmond NL, Labosky PA. The NIH "BEST" programs: Institutional programs, the program evaluation, and early data. *The FASEB Journal*. 2020;34:3570–3582.

[6] Fuhrmann CN, Hobin JA, Lindstaet B, Clifford PS. *my*IDP Individual Development Plan. IDP Version 1.1.11 (0914-PR) built on 9/14/2015 6:20:32 PM. http://myidp.sciencecareers.org/. Accessed September 27, 2019.

[7] St. Clair R, Hutto T, MacBeth C, Newstetter W, McCarty NA, Melkers J. The "new normal": adapting doctoral trainee career preparation for broad career paths in science. *PLoS One*. 2017;12(5):e0177035.

[8] Watts SW, Chatterjee D, Rojewski JW, et al. Faculty perceptions and knowledge of career development of trainees in biomedical science: what do we (think we) know? *PLoS One*. 2019;14(1):e0210189.