Evaluation of Research Integrity Training Program (ERITP) Final Report

November 18, 2015

Produced by:

GENERAL DYNAMICS
Information Technology

In partnership with:

NETE®

Contents

Introduction

The Office of Extramural Research – Research Integrity (OER-RI) is responsible for providing training on research integrity to staff throughout the NIH. Historically, this training has been offered using seven different delivery approaches including:

1. In-person training offered through the ICs
2. Video training
3. Webinar training (hosted by OER-Research Integrity)
4. Online information (IC or OER websites)
5. The annual NIH extramural research integrity training (hosted by OER-Research Integrity)
6. ESA training
7. NIH Core Curriculum

Background

In 2014, OER-RI contracted with NETE to evaluate the effectiveness of the research integrity training. GDIT was tasked with supporting the contract with Instructional Design and Expert Evaluation support. During the initial three phases of the program, GDIT assisted NETE by providing consultation and advice to support:

- Designing and validating the questions for use in the online research integrity knowledge assessment tool.
- Assisting NETE Developers with tool specific knowledge and subject matter expert advice.
- Offering expert guidance to establish the reliability of both the assessment tool and the assessment questions.

Note that changes in NIH project leadership during the project may have created some confusion and not all recommendations were implemented. However, the project did move forward with subsequent phases.

Current Project Phases

The remainder of GDIT's involvement covers phases four and five of the program, which are detailed in this report. GDIT's responsibilities during these last phases include:

Phase IV: Assessments, Adjustments & Preliminary Data Evaluation

- Conducting a preliminary data evaluation on the baseline assessment
- Calculating descriptive statistics to answer the following questions:
    - What are the characteristics of staff who participated in the baseline assessment?

- What is the level of staff knowledge in the two categories of research integrity: a) federal regulations and policies on research misconduct; and b) NIH policies and recommendations on the handling of misconduct allegations that have come to the attention of the staff?
- How many staff members previously have had training in these matters here at NIH, outside NIH, or not at all? How long has it been since this training occurred?
- If staff had participated in such training at NIH, how many participated in each of the training delivery types (in-person or online)?

Phase V: Final Data Evaluation and Reporting

- Performing an analysis of the data to answer the following questions:
  - What is the impact of staff participation in the NIH extramural research integrity training program?
  - Of the two areas of staff knowledge of federal regulations and policies on research misconduct as well as NIH policies and recommendations on the handling of misconduct allegations, which are the area(s) that require more coverage in the future? In order to optimize resources, which of the educational delivery methods is most effective?
- Providing a detailed report to the NIH.

This report provides summary and detailed data on the preliminary and final evaluation assessments and recommendations for future training.

Executive Summary

Preliminary Assessment

The preliminary assessment looked at a variety of descriptive data. This included:

- **The characteristics of staff who participated in the baseline assessment**

    - A total of 355 individuals participated in the study. This included:
        - A total of 230 NIH personnel the Baseline Assessment.
        - A total of 191 participated in Final Assessment, including 66 who participated in Baseline Assessment and 125 new participants.
    - Members of 27 NIH institutes participated, with the largest share coming from NCI (10%) and NIAID (9%) in Baseline and CSR (14%) and NIAID (10%) in Final Assessment (See Figure 1 based on Table 1).
    - Participants were most commonly from Program Staff (Baseline - 48%; Final Assessment - 55%) and Scientific Review (Baseline - 22%; Final Assessment - 28%). (See Figure 2 based on Table 2).

Figure 1: Rate of Participation by NIH Affiliation

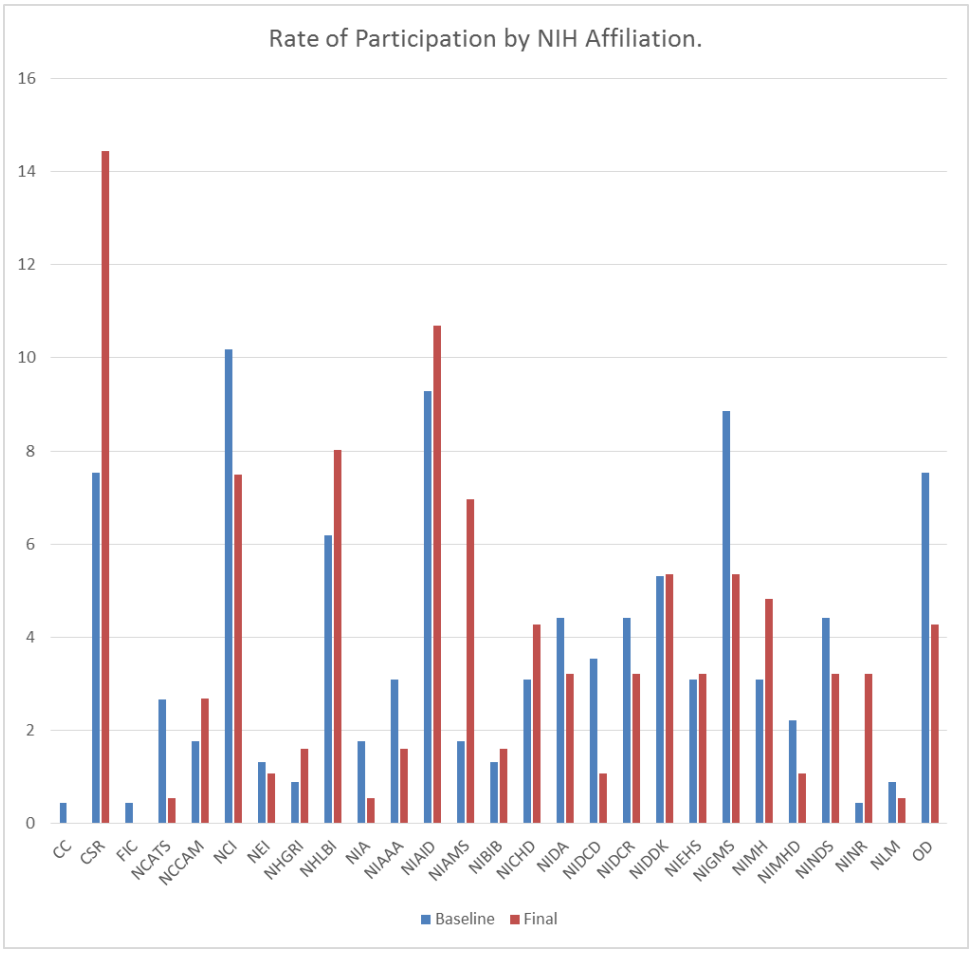
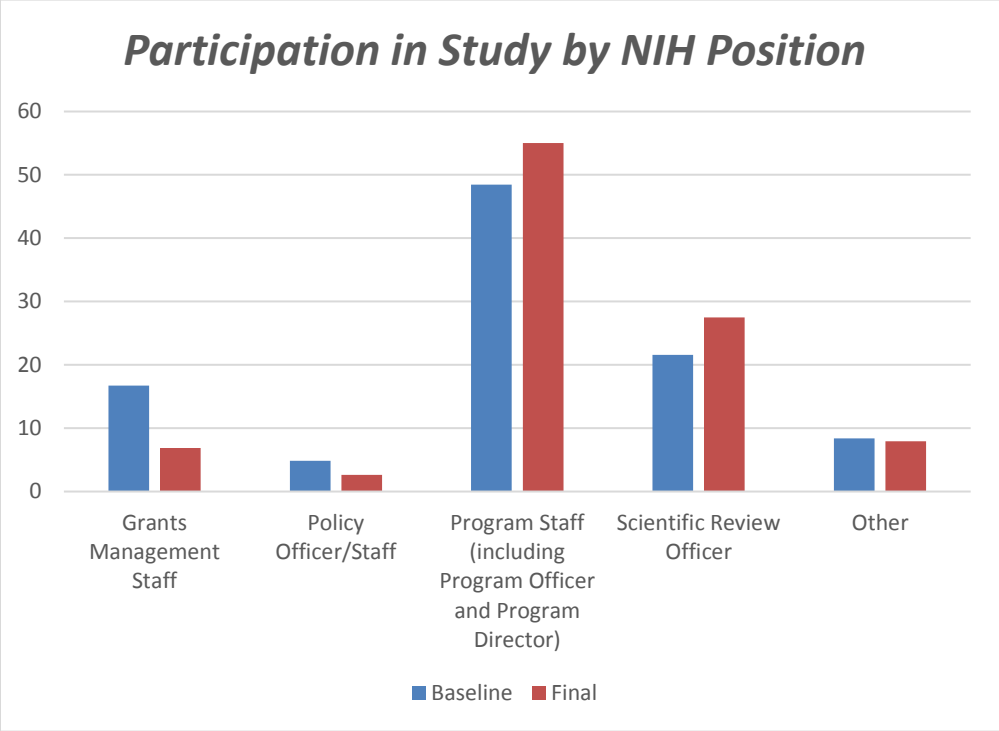


Figure 2: Participation in Study by NIH Position

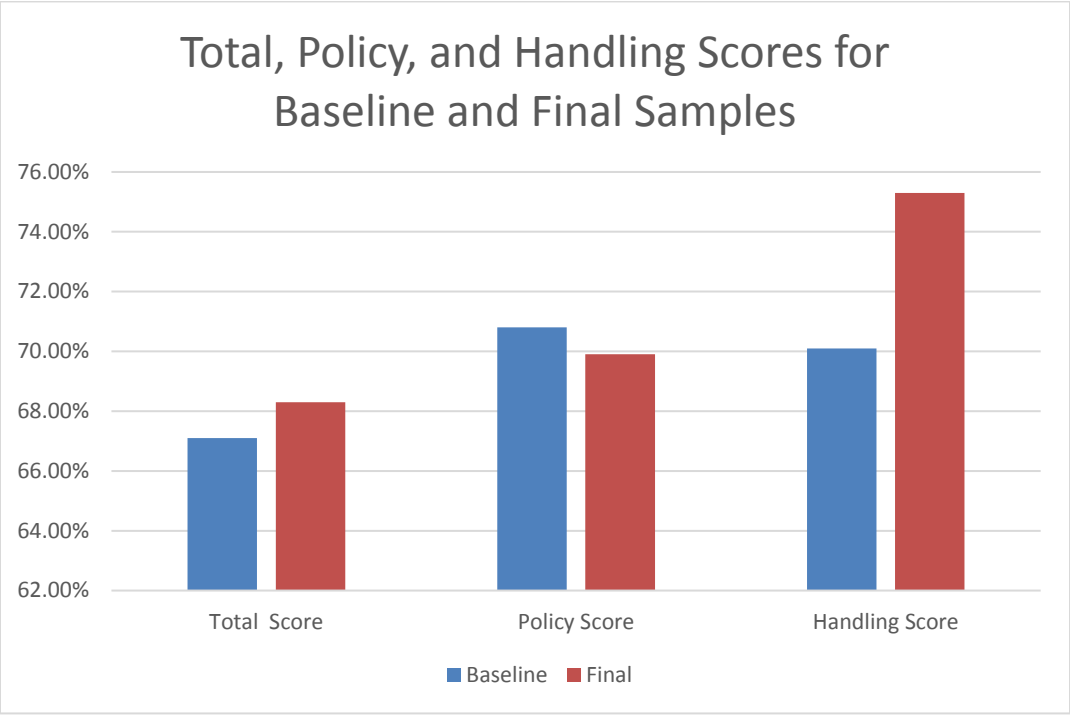


- Baseline participants averaged 14.7 years as an NIH employee or contractor, including 9.7 years tenure in the current position. Final Assessment participants averaged 13.3 years as a Federal employee or contractor, including 9.2 years in the current position. (See Table 3).
- In the past 5 years, Baseline participants encountered an average of 1.7 instances of research misconduct, of which .53 occurred during the past year, compared to 1.8 instances of misconduct encountered by Final Assessment participants, of which .69 occurred during the last year (See Table 4).

- **The level of staff knowledge in the two categories of research integrity: a) federal regulations and policies on research misconduct; and b) NIH policies and recommendations on the handling of misconduct allegations that have come to the attention of the staff**

  - Baseline scores indicated that staff had considerable knowledge within these two categories. Participants responded correctly to 71% of the Policy questions and 70% of the Handling questions. (See Figure 3 based on Table 6a).

Figure 3: All Baseline and Final Assessment Participants' Performance on Total, Policy and Handling Tests



- **The numbers of staff members who previously had training in these matters here at NIH, outside NIH, or not at all and the length of time since this training occurred**

    - The numbers of staff with previous training within the past 5 and the past 1 year is shown in Figure 4 (based on Table 11).

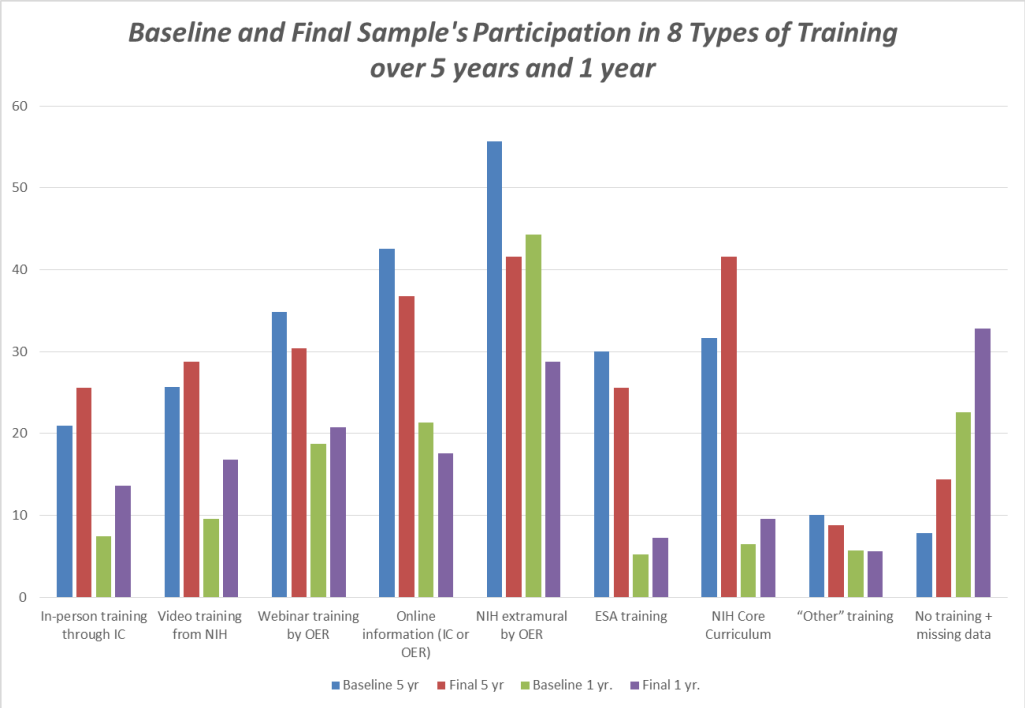
- **If staff had participated in such training at NIH, the number who had participated in each of the training delivery types (in-person or online)**

    - Baseline participants participated in 8 types of training in the 5 years prior to the assessment. The percentages of participants participating in each type of training are listed below.
        - In-person training offered through your IC: 21%
        - Video training (on http://grants.nih.gov/grants/research_integrity): 26%
        - Webinar training (hosted by OER-Research Integrity): 35%
        - Online information (IC or OER websites): 43%
        - The annual NIH extramural research integrity training (hosted by OER-Research Integrity): 56%

- ESA training: 30%
- NIH Core Curriculum: 32%
- "Other" training: 10%

- The numbers of staff participating in the various types of training is shown in Figure 4.

Figure 4: Baseline and Final Assessment Participants in 8 Types of Training over 5 Years and 1 Year



Final Assessment

The goal of the final assessment was to answer several questions regarding the OER-RI research integrity training program. These questions and the results of the final assessment appear below.

1. What is the impact of staff participation in the NIH extramural research integrity training program?

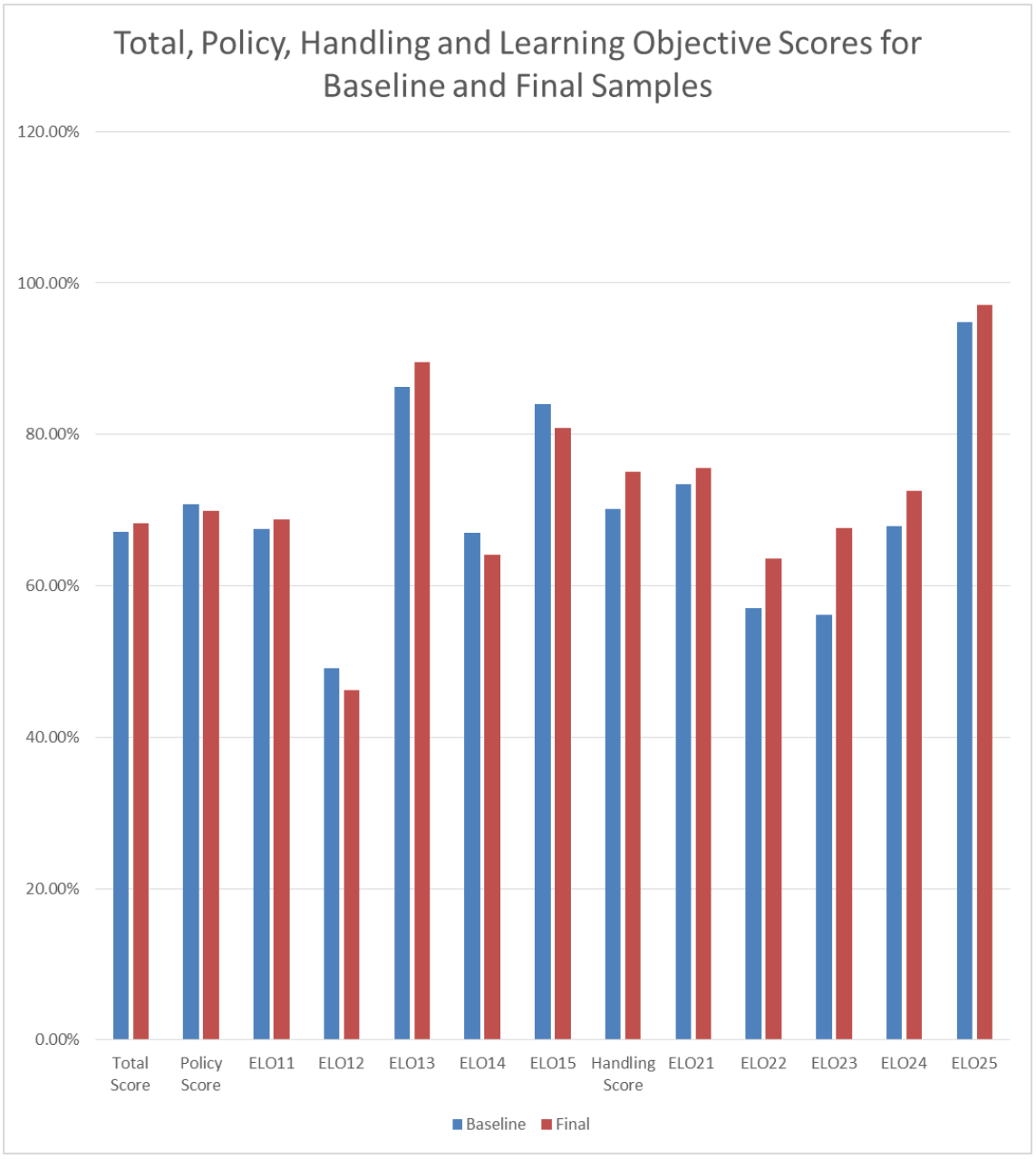
a. Does it result in improved judgment, knowledge, or decision-making in the total of the 2 areas of knowledge (a. federal regulations and policies on research misconduct, or, b. NIH policies and recommendations on the handling of misconduct allegations) from pre-training to post-training?

- There was no significant difference between the Baseline participant scores and the Final Assessment participant scores on the Total Knowledge Score, or

Policy Scores. However, there was an improvement in the Handling score. (See Figure 3, above)

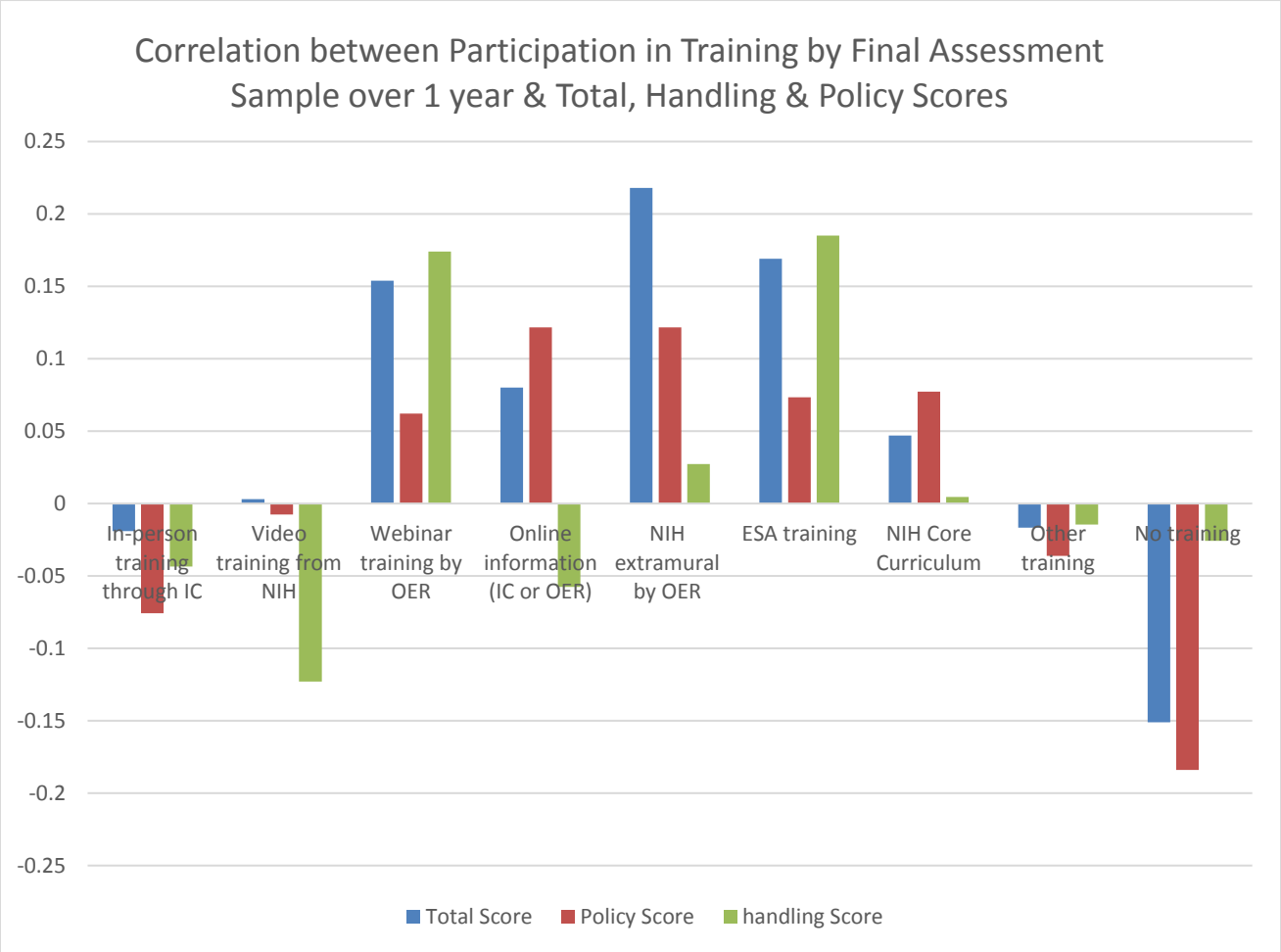
- Final Assessment participants performed significantly better than in Baseline on Learning Objective 2.3 “Identify appropriate federal and NIH resources associated with research integrity and research misconduct” (See Table 6b). However, because the same people were not necessarily in the Baseline and Final Assessment samples, interpretation of the difference must be tentative. (See Figure 5, below).

Figure 5: Total, Policy, Handling and Learning Objective Scores for Baseline and Final Samples



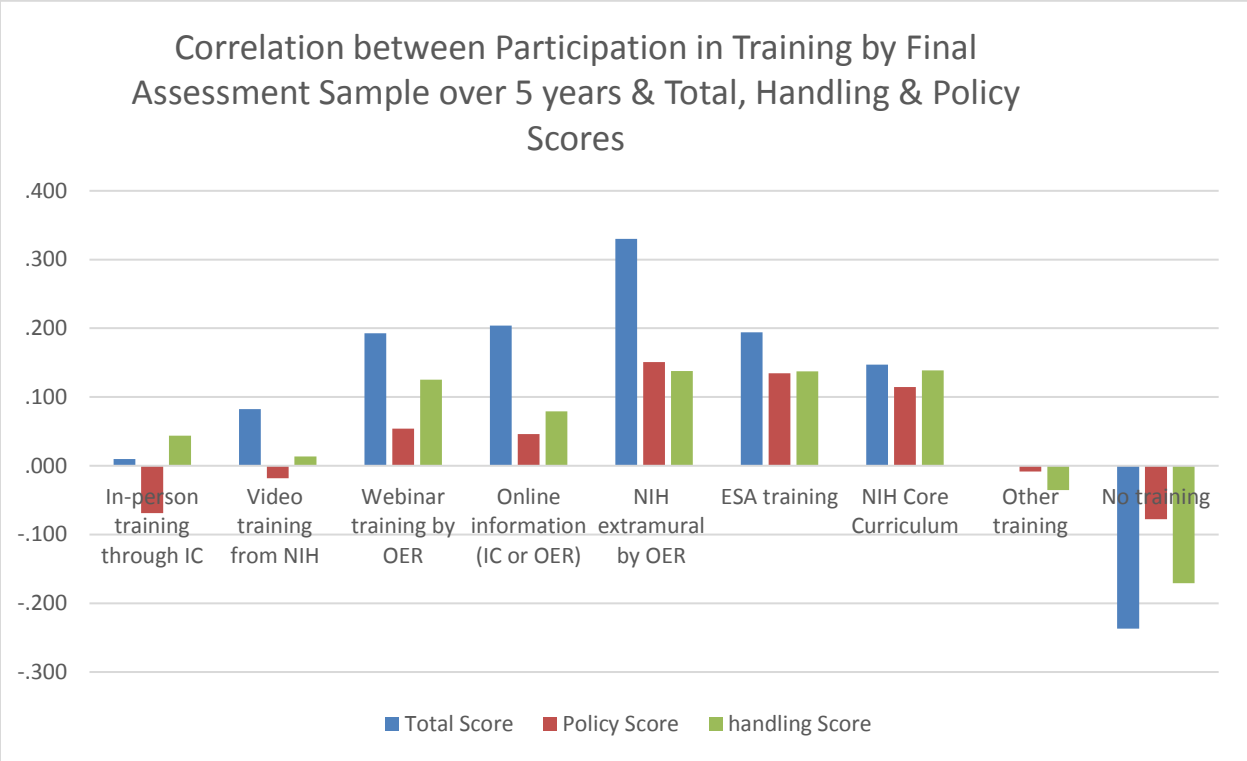
- Overall performance on the Final Assessment test was significantly improved by experience over the past year in (See Figure 6 partially based on Table 13):
    - The annual NIH extramural research integrity training
    - ESA Training
    - Webinar training
- Participation in Webinar training and ESA training over the past year was related to higher performance in the Handling portion of the test.
- Taking *no* training over the past year was *negatively* related to performance on both the Total test and Policy portion of the test.

Figure 6: Correlation between Participation in Training in the Past Year and Performance on Final Total, Policy and Handling Tests



- Performance on the Final Assessment test was also significantly related to experience over the past 5 years in (See Figure 7 partially based on Table 13):
    - The annual NIH extramural research integrity training
    - Online information
    - ESA Training
    - Webinar training
    - NIH Core Curriculum
    - Performance was *negatively* related to taking *no* training

Figure 7: Correlation between Participation in Training in the Past Year and Performance on Final Total, Policy and Handling Tests by all who took the Final Assessment

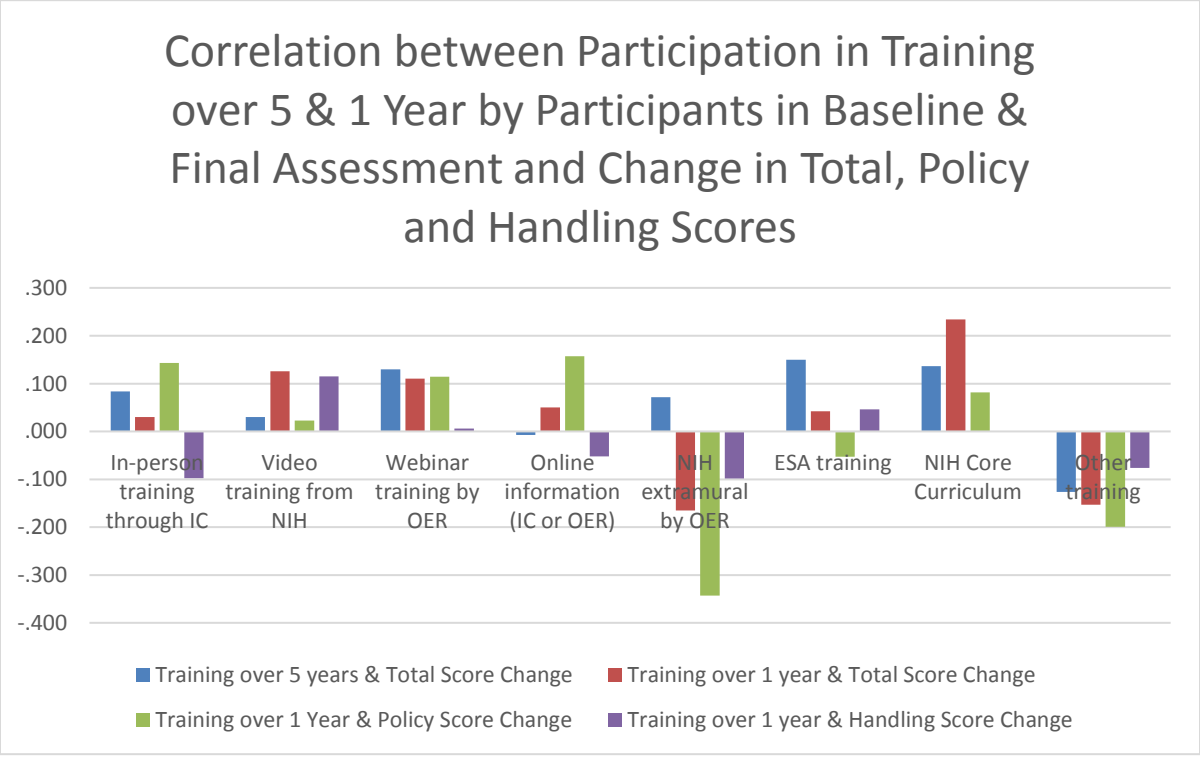


b. Does it result in improved judgment from pre-to-post training in individuals who previously had training here at NIH, outside NIH, or not at all?

- There were 66 participants who completed the tests at both Baseline and Final Assessment. There was no significant overall change in Total Score from Baseline to Final Assessment (See Table 10). This may be attributed to the wide variety of training experiences that individual participants pursued during that year.

- Individual improvement in the Final Assessment Total score was marginally associated with participation in the NIH Core Curriculum during the past 1 year. The correlation between such participation and improvement in the Final score was modest, r = .23, p = .07 (See Figure 7 which was partially based on Table 15).
- Analyses also were conducted on changes in the two forms of knowledge: Policy and Handling. No form of training was significantly associated with improvement in Policy scores from Baseline to Final Assessment, nor were there any significant effects of tenure on such changes. There was an apparent negative impact of annual NIH extramural research integrity training (hosted by OER-Research Integrity) on Policy scores (r = -.34, p <.01), but in light of the small sample it would be unwise to draw strong conclusions
- For those who took both tests, there was a significant overall change in Handling scores from Baseline to Final Assessment 71.8% to 77. 4%. (See Table 10). The impact was particularly pronounced on LO 2.3, which saw a change of 50.1% to 72.1%.
- There was a significant impact on Handling score improvement from participating in the NIH Core Curriculum r = .27, p <.05.

Figure 8 Correlation between Participation in Training over 5 & 1 Year by Participants Who Took Baseline and Final Assessments and Changes in Total, Policy and Handling Scores



c. Does improvement in total test score from pre-to-post training vary by type of training delivery (in-person versus online versus both)?

- Because of the small sample, improvement scores are less informative than the performance of the Final Assessment sample as a whole. As reported in Figure 7 and Table 13, Performance on the Final Assessment test was significantly related to a variety of training experiences over the past 5 years:
  - In-person training
    - Participation in the annual NIH extramural research integrity training (hosted by OER-Research Integrity)
    - ESA Training,
    - NIH Core Curriculum
  - Online training
    - Online information (IC or OER websites)
  - Webinar training
    - Webinar training (hosted by OER-Research Integrity),
- Using a Multiple Regression approach, overall performance on the Final Assessment test was significantly predicted by experience over the past year in the annual NIH extramural research integrity training; Online information , and ESA training, which each independently predicted Final Assessment tests scores. (See Table 14a).
- Participation in the past year in both "In-person training offered through your IC." and "Video training (on http://grants.nih.gov/grants/research_integrity)" were negatively correlated with performance on LO 1.1 in the Final Assessment (see Table 16). It is not clear if less knowledgeable or less motivated employees chose those forms of training, or if there was something confusing about the content in those trainings.

d. Does length of time since training impact on learning retention?

Participants were asked if they had taken a given type of training within the past 5 years and within the past year. Information on the specific date of training and the time between training and testing were not collected. As a consequence, fine grained analysis is precluded.

e. Does improvement (pre-to-post training) on total test score occur differentially based on group and training delivery, regardless of amount of experience?

- The impact of group and training delivery occurred independently of experience, which was assessed as tenure at NIH.

2. Of the two areas of staff knowledge of federal regulations and policies on research misconduct as well as NIH policies and recommendations on the handling of misconduct allegations, which are the area(s) that require more coverage in the future?

Participants performed better on the test items for Handling Misconduct in the Final Assessment than in Baseline. This suggests that more coverage of Policy topics and issues may be helpful in the future.

In order to optimize resources, which of the educational delivery methods is most effective?

NIH Extramural Research training, Online training, and ESA training over the past year were significant at both the univariate and multivariate levels on Final Assessment scores. This indicates that these were particularly effective forms of training.

There was, however, a trend for Final Assessment participants to take both NIH Extramural Research training and ESA training over the past five years, so it is more difficult to separate out the most effective forms of training. As a consequence, it might be useful to perform an experimental investigation in which participants are randomly assigned to training.

Analysis and Results

Methods

Sample

A total of 230 NIH personnel from a wide variety of institutes and positions participated in Baseline. A total of 191 participated in Final Assessment, including 125 new participants and 66 who participated in Baseline, for a total of 355 individual participants.

Participation was obtained from 27 NIH institutes, with the largest share coming from NCI (10.18%) and NIAID (9.29%) in Baseline and CSR (14.44%) and NIAID (10.70%) in Final Assessment (See Table 1). Participants were most commonly from Program Staff (48.46%; 55.03%) and Scientific Review (21.59%; 27.52%) in both Baseline and Final Assessment phases, respectively (See Table 2). Baseline participants averaged 14.69 years as an NIH employee or contractor, including 9.72 years tenure in the current position. Final Assessment participants averaged 13.259 years as a Federal employee or contractor, including 9.18 years in the current position. (See Table 3). In the past 5 years, Baseline participants encountered an average of 1.96 instances of research misconduct, of which .53 occurred during the past year, compared to 1.81 instances of misconduct encountered by Final Assessment participants, of which .69 occurred during the last year (see Table 4).

Instrument design

A total of 96 items were written to assess 10 Learning Objectives (LO) in two categories of 5 Learning Objectives each dealing with Policy and Handling Misconduct, respectively. Each Learning Objective had 8 to 14 items assigned to it. The same fifty-six items were used in the Baseline and Final Assessment testing, ranging from 4 to 8 items per LO (See Table 5).

Random Assignment of Assessment Items

For the assessment of individual participants, one item was randomly chosen from each of the 10 LO pools, with the exception of LO 2.1 and 2.4, from which two items were chosen. Thus, each participant answered a total of 12 items.

Randomization was generally effective, with each item receiving responses from a reasonable number of participants. There was some variability, however, in the use of individual items. In Baseline, some items were randomly assigned and answered by as few as 31 of the 212 participants, whereas others were answered by as many as 62. Similarly in Final Assessment, some items were randomly assigned and answered by as few as 26 of the 173 participants, whereas others were answered by as many as 55. (See Tables 6a, 6b).

The ideal would be to have all items answered by the same number of participants. If the items were not equal in difficulty level, differential use of items will affect the scores both of individuals and of the Learning Outcomes.

Results

Consistency of Item Difficulty Level

The difficulty level of the items was examined in terms of the participants' rate of providing a correct answers. Across all 56 items, the average item pass rate in Baseline was 71.64% and in Final Assessment it was 74.19%. Across the total score calculated based on the randomly selected items that were used, the average pass rate at Baseline was 67.1% and in Final Assessment it was 68.3%. This suggests that more difficult items tended to be used more often than the less difficult items. (See Table 6a, 6b).

There was quite a bit of variability in item difficulty level. The most difficult items were passed by just 23.08% of participants who responded to those items in Baseline and 28.13% who responded in Final Assessment. The least difficult items were passed by 100% in both Baseline and Final Assessment Phases (see Table 7). This indicates substantial diversity in item difficulty level.

There also was variability in pass rates within Learning Objectives, so additional analyses were conducted on the individual items within each of the 10 Learning Objectives (see Table 8). Across items within Learning Objectives in Baseline, the average difference in pass rates between the most difficult item and the least difficult item was 38%. This ranged from a 6.25% difference in item difficulty level between the hardest and easiest item in LO 1.3 to a difference of 63.36% in LO 2.1. Items with major differences in pass rates are not equivalent and cannot effectively serve as substitutes for one another in estimating participants' knowledge, if that estimate is calculated by the unweighted sum of correct answers.

Consistency of Difficulty Level across Learning Objectives

The difficulty level of the Learning Objectives was estimated by examining the average pass rates of each of them. The pass rates for the LOs in Baseline varied from 49.06% for LO1.2 to 94.81% for LO 2.5, with an average of 70.31% (based on the average of the 2 items in LO 2.2 and 2.4; see Table 6a). The pass rates for the LOs in Final Assessment varied from 46.24% for LO1.2 to 97.11% for LO 2.5, with an average of 72.63%.

Many of the differences between in difficulty level of the Learning Objectives were statistically significant (see Table 8). Such results are challenging to interpret, because each Learning Outcome score represents the average pass rate of different items that were randomly selected within the Learning Outcome. There was no reason to expect that the Learning Objectives

would be equal it difficulty level, so this outcome is not a problem, except for calculating reliability, as described below.

Reliability

In classical test construction, a sample of participants takes all items and then the consistency of responses to items within a category, such as a Learning Objective, is calculated using a formula such as Cronbach's Alpha. Such analysis indicate the homogeneity of the items, which indicates the extent to which one item can be substituted for another. In this study, each participant responded to just one item in eight of the ten Learning Objectives, so that was impossible.

As a substitute, the five Learning Objections within the two Learning Objective categories were treated as a 5-item scale for two analyses and a third analysis was conducted across all ten Learning Objectives. As Table 9 indicates, the relations among Learning Objective categories were modest. Cronbach Alphas for the 10 Learning Objectives were $\alpha = .39$ for Baseline, $\alpha = .10$ for Final Assessment and $\alpha = .29$ for both Baseline and Final Assessment phases combined. Alphas were smaller within the combinations of 5 Learning Objectives. This indicates that the Learning Objectives sample a diverse array of knowledge, and performance on one Learning Objective has only a small relation to performance on a different Learning Objective.

Two Areas of Knowledge

Participants' knowledge of "Policy" pertaining to research integrity and procedures for "Handling" research misconduct were assessed by summing the 5 Learning Objective scores within each domain. Because LO 2.1 and 2.4 were assessed using two items each, scores were calculated both using the simple sums, and the average of the two items. The latter approach made the LO scores more comparable, and allowed more appropriate comparison of the Policy and Handling scores. (See Table 6a, 6b). These procedures were followed in both Baseline and Final Assessment. Total scores were calculated as the sum of all 12 items completed by the individual participant.

Performance in Final Assessment versus Baseline

The performance of all participants who took the Final Assessment test was compared to performance of all who took part in Baseline, at the level of the total score, the 5 item Policy and Handling scores, the ten Learning Objectives and the 56 items.

There was no significant difference between the Baseline sample and the Final Assessment sample on the Total Knowledge Score, or Policy Scores, but there was a difference on the Handling scores, such that participants performed better on the test items for Handling misconduct in Final Assessment than in Baseline, 3.77 vs. 3.47, $F(1,384) = 8.47$, $p = .004$ (see Table 6b).

One of the 10 Learning Objectives also showed a significant difference between Baseline and Baseline. In the Final Assessment, participants performed better than in Baseline on Learning Objective 2.3 "Identify appropriate federal and NIH resources associated with research integrity and research misconduct" (.56. vs. .68, F (1,383) = 5.36, p <.02 (See Table 6b). Because the same people were not necessarily in the Baseline and Final Assessment samples, interpretation of the difference must be tentative.

Final Assessment differences, compared to Baseline scores, were found on 2 of the 56 individual items (Item 1.1001 "According to the Public Health Service (PHS) Policies on Research Misconduct 42 C.F.R. 93, which of the following may be considered research misconduct?" .42 vs. .67, F(1,68) = 4.43, p < .04); Item 1.3041. "According to the PHS Policies for research misconduct 42 C.F.R. 93, institutions involved in PHS supported research activities are required to develop and implement policies and procedures for:" .83 vs. 1.00, F(1,84) = 7.42, p < .008). Because there is a 4% chance that 2 in 56 items will be statistically significantly different, it would be a mistake to over-interpret these findings. (See Table 7).

Individual Improvements

Comparisons were made between the Baseline vs. the Final Assessment performance of the 66 participants who took the test twice, of whom 61 provided complete data. Repeated measures analysis found no significant improvement in Total Scores from Baseline to Final Assessment (8.94 vs. 9.18, t (61) = .84, p = .41; or in Policy scores (3.84 vs. 3.75, t (60) = -.59, p = .56. There was, however, a significant improvement in Handling scores (3.59 vs. 3.86, t (60) = 1.97, p=.05, similar to what was found in the larger sample as reported above There also was a significant difference on Learning Objective 2.3, (.56 vs. .72m F (1, 60) = 4.05, p = .05). (See Table 10).

Participation in 7 forms of NIH training

Test participants reported whether or not they participated in each of 7 forms of NIH training over the past five years: (a) In-person training offered through your IC; (b) Video training (on http://grants.nih.gov/grants/research_integrity); (c) Webinar training (hosted by OER-Research Integrity); (d) Online information (IC or OER websites); (e) The annual NIH extramural research integrity training (hosted by OER-Research Integrity); (f) ESA training, (g) NIH Core Curriculum. The survey instrument did not ask precisely when the participant took each form of training. Participants also were asked if they were involved in "other" training, besides the 7 mentioned above. That response, and a report of having taken no training, were included in selected analyses.

Some participants may have taken the training 4 or more years previously, and some may have taken the training during the past month within the 5 year period. The latter response, however, will be captured by a separate question asking about training during the past year.

The most frequent form of training participation in the past 5 years was in the annual NIH extramural research integrity training for both Baseline and Final Assessment participants (55.7%; 50.8%) , which was also the most frequent form of participation over the past year (44.3%; 37.7%) (See Table 11, 12).

Effect of training on Baseline test scores

Correlations were calculated in the Baseline data between binary reports of experience in the seven forms of training over the past 5 years and performance on the sum of the 12 assessment items. One form of training was significantly predictive of performance (see Table 13). Participation in the annual NIH extramural research integrity training (hosted by OER-Research Integrity) correlated r =.20, p = .004 with Total test scores. There also was trend for experience in that form of training within the past year to predict Total test performance r = .12, p = .09. (See Table 13).

Effect of training on Final Assessment test scores

Performance on the Final Assessment test was significantly related to experience over the past 5 years in the annual NIH extramural research integrity training (hosted by OER-Research Integrity) r= .33, p = .0001; Online information (IC or OER websites) r = .20, p = .005.; ESA Training, r = .19, p = .007; Webinar training (hosted by OER-Research Integrity), r = .19, p = .007; and NIH Core Curriculum, r =.15, p =.04 (see Table 13).

Overall performance on the Final Assessment test was significantly predicted by experience over the past 1 year in the annual NIH extramural research integrity training (hosted by OER-Research Integrity) r= .20, p = .0002, ESA Training, r = .17, p = .02; and Webinar training (hosted by OER-Research Integrity), r = .15, p = .03. (See Table 13).

Because participants could participate in more than one form of training, a series of multiple regression analyses were conducted to determine if each of the forms of training taken during the past year that were significant when evaluated in combination. Experience during the past year in the annual NIH extramural research integrity training (Beta = .37); Online information (Beta = .18) and ESA training (Beta = .14) each independently predicted Final Assessment tests scores. (See Table 14a & 14b).

Effect of training on improvement from Baseline to Final Assessment in test scores

As noted above, there were 66 cases who completed the tests at both Baseline and Final Assessment. There was no significant effect of training experiences over the past 5 years, as reported at Final Assessment, on improvement from Baseline to Final Assessment. There was a marginal trend for participation in the NIH Core Curriculum during the past year, however, to be correlated with improvement in the Total score, r = .23, p = .07 (See Table 15).

Analyses also were conducted on changes in the two forms of knowledge: Policy and Handling. There was no significant overall change in Policy scores from Baseline to Final Assessment ($F(1,61) = .39$, $p = .54$, nor was there any significant effect of training experiences or tenure on such changes. There also was no significant overall change on Handling scores from Baseline to Final Assessment ($F(1,61) = .88$ $p = .35$), nor was there any significant effect of tenure.

Paralleling the results for Total change scores, however, was a significant impact of participating in the NIH Core Curriculum on Policy score improvement, $r = .27$, $p = .04$. Stronger results might have been obtained with greater statistical power and less diversity in training opportunities. Nonetheless, the NIH Core Curriculum joins the NIH Extramural training, Online training, and ESA training as promising approaches.

Effects of types of training on specific Learning Objectives

Exploratory analyses were conducted on the relationship between training experiences in the past year and performance on specific Learning Objectives. There was no significant effect of any one form of training on Policy knowledge scores, but Handling was significantly affected by participation in the Webinar training (hosted by OER-Research Integrity), $r=.18$, $p =.02$ followed by participation in ESA, $r = .17$, $p =.03$. (See Table 13b). This was paralleled by comparable effects on Learning Objective 2.3, which was the LO that showed the most difference from Baseline to Final Assessment samples. It, too, was most influenced by participation in the Webinar training, $r=.14$, $p =.06$ followed by participation in ESA, $r = .14$, $p =.07$. (See Table 16).

Null or Negative Effects of Specific Forms of Training

Two forms of training, **"**In-person training offered through your IC." and "Video training (on http://grants.nih.gov/grants/research_integrity)" did not register significant positive effects on Baseline or Final Assessment scores. In fact, those forms of training were negatively correlated with performance on LO 1.1 in Final Assessment: In-person training, $r = -.15$, $p=.05$; Video training $r = -.18$, $p = .02$ (See Table 13a). It is not clear if less knowledgeable or less motivated employees chose those forms of training, or if there was something confusing about the content in those trainings. Further research is needed on those topics, and also on the inexplicable association between participation in the Annual NIH extramural research integrity training and more negative performance from Baseline to Final Assessment on Policy scores, $r = -.343$, $p = .007$ (See Table 15).

Conclusion

Knowledge of policy regarding research integrity and the ability to respond to perceived research integrity issues are both key elements of achieving the core mission of the NIH.

Data from this study revealed that staff across the NIH have a solid foundation of knowledge regarding both basic policy and appropriate responses.

The most well-attended types of training over both the past 5 years and 1 year were:

- NIH extramural training offered by OER
- Online information offered by an IC or OER
- Webinar training offered by OER

ESA training was also well attended over the past five years, but showed a decline in attendance over the past 1 year compared to other types of training.

Currently, OER offers a wide variety of options for learning about research integrity policy and appropriate responses to perceived violations. However, the diversity of learning options presents several challenges:

1. Learners may not know which options they should take
2. Content and messaging may vary across the options
3. Maintaining current and consistent content across the options challenging

The data also revealed that participating in current research integrity training opportunities was not consistently related to overall improvement. However, there are several reasons why this conclusion may not be definitive.

- Only 66 respondents participated in both Baseline and Final assessments. This is a small sample from which to draw definitive conclusions.
- The Baseline and Final assessment tests consisted of 12 questions. This may be too few test items for definitive conclusions.
- While the data reveals how many participants took training of various types, the study did not examine the content of each type of training. Variations in content may have had a significant impact on acquisition and retention of the knowledge addressed in the tests.

Recommendations

- To follow up on the results of the ERITP project:
  - Investigate why the NIH extramural research integrity training, ESA training and webinar training were more effective than other approaches.
  - Consider repeating this study using participants who are randomly assigned training conditions, to allow more effective assessment of impact of specific types of training.
  - Consider repeating the assessments using more test questions for each participant. It would be helpful if a group of participants could take all items so as to determine internal consistency.
  - Investigate the correlation between performance on the ERITP test questions and overall job performance and the employee's track record of addressing research misconduct.
  - Conduct additional research on the small number of negative correlations noted between training and test scores.
- A more integrated approach to ensuring staff knowledge in the two research integrity content areas would be to develop a comprehensive strategy for research integrity training. This strategy might include:
  - Defining indicators of acceptable performance, both for testing and job performance.
  - Finalizing the learning objectives used for the ERITP project for use in future research integrity training
  - Developing a comprehensive e-learning course on research integrity which would be available to all staff on an as-needed basis
  - Focusing the other delivery approaches on more narrowly focused topics and the application of policy to specific scenarios

Appendix A: Tabular Data

Table 1: Rate of Participation by NIH Affiliation.

Affiliation	Baseline Frequency	Baseline Percent	Final Assessment Frequency	Final Assessment Percent
CC	1	0.442	0	0
CSR	17	7.522	27	14.439
FIC	1	0.442	0	0
NCATS	6	2.655	1	0.535
NCCAM	4	1.77	5	2.674
NCI	23	10.177	14	7.487
NEI	3	1.327	2	1.07
NHGRI	2	0.885	3	1.604
NHLBI	14	6.195	15	8.021
NIA	4	1.77	1	0.535
NIAAA	7	3.097	3	1.604
NIAID	21	9.292	20	10.695
NIAMS	4	1.77	13	6.952
NIBIB	3	1.327	3	1.604
NICHD	7	3.097	8	4.278
NIDA	10	4.425	6	3.209
NIDCD	8	3.54	2	1.07
NIDCR	10	4.425	6	3.209
NIDDK	12	5.31	10	5.348
NIEHS	7	3.097	6	3.209
NIGMS	20	8.85	10	5.348
NIMH	7	3.097	9	4.813
NIMHD	5	2.212	2	1.07
NINDS	10	4.425	6	3.209
NINR	1	0.442	6	3.209
NLM	2	0.885	1	0.535
OD	17	7.522	8	4.278

Table 2: Participation in Study by NIH Position

NIH position	Baseline Frequency	Baseline Percent	Final Assessment Frequency	Final Assessment Percent
Grants Management Staff	38	16.74	13	6.878
Policy Officer/Staff	11	4.846	5	2.646
Program Staff (including Program Officer and Program Director)	110	48.458	104	55.026
Scientific Review Officer	49	21.586	52	27.513
Other	19	8.37	15	7.937

Table 3: Participants' Tenure in Months

	Baseline How long have you been in your current position or role?	Baseline In total, how long have you worked at NIH as a Federal employee and/or contractor? (	Final Assessment How long have you been in your current position or role?	Final Assessment In total, how long have you worked at NIH as a Federal employee and/or contractor?
N of Cases	230	230	191	191
Minimum	0	0	0	0
Maximum	519	550	378	480
Mean	116.609	176.274	110.199	159.021
Standard Deviation	85.064	109.132	77.96	105.462

Table 4: Participants' Experience with Research Misconduct

	Baseline Based on what you currently know about research misconduct, how many times in the past 5 years have you come across issues or allegations that may have been associated with research misconduct?	Baseline How many of these have occurred in the past year?	Final Assessment Based on what you currently know about research misconduct, how many times in the past 5 years have you come across issues or allegations that may have been associated with research misconduct?	Final Assessment How many of these have occurred in the past year?
N of Cases	223	221	180	179
Minimum	0	0	0	0
Maximum	50	11	30	30
Mean	1.96	0.534	1.806	0.687
Standard Deviation	4.785	1.326	4.356	2.967

Table 5: Learning Objectives and Items

Policy and Regulations	# of Questions for this LO	# of T1/T2 Control Questions	Question ID	Difficulty Range
TLO 1.0 Explain critical regulations and policies on research misconduct	46			
ELO 1.1 Define the components and application of research misconduct under 42 CFR 93	14	5	11-001,002,009,011,013	1.8-2.0
ELO 1.2 Explain the requirements and conditions for a finding of research misconduct	7	5	12-016,017,018,020,021	2.6-3.0
ELO 1.3 Explain key regulations and policies associated with research misconduct	9	4	13-025,028,031,041	2.2-2.8
ELO 1.4 Explain advanced concepts of regulations and policies associated with research misconduct	9	6	14-029,030,033,036,037,047	3.0-3.5
ELO 1.5 Define common terms associated with research integrity	7	5	15-043,046,049,050,101	2.6-3.2
Handling Research Misconduct				
TLO 2.0 Respond appropriately to situations associated with research misconduct	50			
ELO 2.1 Recognize research misconduct as defined by 42 CFR 93	8	8	21-051,052,053,054,055, 056,057,058	2.2-3.2
ELO 2.2 Explain the roles of federal agencies, NIH and other institutions play in matters associated with research misconduct	10	5	22-061,063,064,065,066	2.6-3.0
ELO 2.3 Identify appropriate federal and NIH resources associated with research integrity and research misconduct	8	5	23-069,070,072,074,076	1.8-2.8
ELO 2.4 Appropriately apply the concept of protecting confidentiality for allegations	11	7	24-077,079,082,083,085,086,087	2.4-3.0

Policy and Regulations	# of Questions for this LO	# of T1/T2 Control Questions	Question ID	Difficulty Range
ELO 2.5 Apply key concepts of federal regulation and NIH policies on research misconduct to scenarios that NIH staff could expect to encounter	13	6	25-089,090,091,092,093,097	2.4-2.6
	96	56		

Table 6a: Baseline Total scores and Learning Objective scores

	Num. of items	Range of N per item	Range of item means in LO	Baseline Mean	Baseline Std. Deviation	Baseline Percentage Correct Score
Total Score	12			8.048	2.812	67.1%
Policy Score	5			3.538	1.064	70.8%
ELO11	5	31-54	.42-.93	.675	.470	67.5%
ELO12	5	32-53	.32-.51	.491	.501	49.1%
ELO13	4	48-60	.82-.90	.863	.344	86.3%
ELO14	6	31-47	.44-.87	.670	.471	67.0%
ELO15	5	35-50	.61-.91	.840	.368	84.0%
Handling Score	7			4.906	1.216	70.1%
Handling Score Avg.	5			3.493	.919	69.9%
ELO21	8	42-62	.28-.92	1.467	.611	73.4%
ELO21a (Avg. of 2 items)				.733	.305	73.3%
ELO22	5	37-46	.23-.84	.571	.496	57.1%
ELO23	5	38-49	.26-.86	.561	.497	56.1%
ELO24	7	50-67	.38-.92	1.358	.670	67.9%
ELO24a (Avg. of 2 items)				.679	.335	67.9%
ELO25	6	37-41	.86-1.00	.948	.222	94.8%

N = 212. Items were scored as 0=incorrect, 1 = correct.

Table 6b: Final Assessment Total scores and Learning Objective scores and Baseline vs. Final Assessment ANOVAs

	Num. of Poss. Items	Range of N per item	Range of item means in LO	Final Assessment Mean	Final Assessment Std. Deviation	Final Assessment Percentage Correct	Baseline v. Final Assessment ANOVA F	Sig.
Total Score	12			8.199	2.761	68.3%	.306	.580
Policy Score	5			3.497	.944	69.9%	.154	.695
ELO11	5	31-40	.58-93	.688	.465	68.8%	.078	.781
ELO12	5	26-39	.32-.58	.462	.500	46.2%	.301	.584
ELO13	4	38-46	.77-1.00	.896	.306	89.6%	.951	.330
ELO14	6	29-33	.42-.79	.641	.481	64.1%	.335	.563
ELO15	5	27-43	.59-.97	.809	.394	80.9%	.609	.436
Handling Score	7			5.249	1.235	75.0%	7.468	.007
Handling Score Avg.	5			3.766	.911	75.3%	8.469	.004
ELO21	8	30-48	.35-.97	1.51	.606	75.5%	.579	.447
ELO21a (Avg. of 2 items)				.757	.303	75.7%	.579	.447
ELO22	5	28-47	.28-.87	.636	.483	63.6%	1.680	.196
ELO23	5	29-41	.38-93	.676	.469	67.6%	5.355	.021
ELO24	7	41-55	.52-.91	1.451	.623	72.6%	1.927	.166
ELO24a (Avg. of 2 items)				.725	.312	72.5%	1.927	.166
ELO25	6	26-35	.96-1.00	.971	.168	97.1%	1.261	.262

N = 173

Table 7: Item Pass Rates within Baseline & Final Assessment and Differences in Item Pass Rate across Two Assessments

	Baseline N	Baseline Mean	Baseline Std. Deviation	Final Assessment N	Final Assessment Mean	Final Assessment Std. Deviation	Baseline v Final Assessment F	Sig.
LO11001	31	.41935	.501610	39	.66667	.477567	4.430	.039
LO11002	43	.93023	.257770	33	.93939	.242306	.025	.875
LO11009	54	.66667	.475831	40	.65000	.483046	.028	.868
LO11011	41	.65854	.480091	30	.60000	.498273	.249	.619
LO11013	43	.62791	.489083	31	.58065	.501610	.165	.686
LO12016	32	.68750	.470929	38	.57895	.500355	.863	.356
LO12017	53	.41509	.497454	26	.46154	.508391	.150	.700
LO12018	36	.50000	.507093	32	.40625	.498991	.588	.446
LO12020	41	.63415	.487652	39	.53846	.505035	.743	.391
LO12021	50	.32000	.471212	38	.31579	.471069	.002	.967
LO13025	60	.88333	.323732	43	.86047	.350605	.117	.733
LO13028	48	.89583	.308709	46	.97826	.147442	2.690	.104
LO13031	56	.83929	.370591	46	.76087	.431266	.975	.326
LOQ13041	48	.83333	.376622	38	1.00000	0.000000	7.423	.008
LO14029	47	.72340	.452151	33	.69697	.466694	.065	.800
LO14030	36	.44444	.503953	26	.42308	.503831	.027	.870
LOQ14033	35	.54286	.505433	25	.64000	.489898	.553	.460
LOQ14036	31	.67742	.475191	31	.54839	.505879	1.071	.305
LOQ14037	30	.86667	.345746	29	.72414	.454859	1.844	.180
LOQ14047	33	.78788	.415149	29	.79310	.412251	.002	.961
LOQ15043	35	.91429	.284029	43	.76744	.427463	3.035	.086
LOQ15046	50	.90000	.303046	26	.96154	.196116	.878	.352
LOQ15049	33	.60606	.496198	42	.59524	.496796	.009	.926
LOQ15050	46	.95652	.206185	35	.97143	.169031	.121	.729
LOQ15101	48	.77083	.424744	27	.85185	.362014	.697	.407

	Baseline N	Baseline Mean	Baseline Std. Deviation	Final Assessment N	Final Assessment Mean	Final Assessment Std. Deviation	Baseline v Final Assessment F	Sig.
LOQ21051	42	.28571	.457230	48	.35417	.483321	.472	.494
LOQ21052	47	.91489	.282057	30	.83333	.379049	1.167	.283
LOQ21053	53	.88679	.319878	35	.97143	.169031	2.064	.154
LOQ21054	62	.91935	.274512	44	.95455	.210707	.509	.477
LOQ21055	45	.88889	.317821	37	.94595	.229243	.835	.364
LOQ21056	40	.82500	.384808	38	.84211	.369537	.040	.842
LOQ21057	53	.92453	.266679	45	.95556	.208409	.401	.528
LOQ21058	50	.60000	.494872	48	.70833	.459340	1.259	.265
LOQ22061	46	.50000	.505525	28	.57143	.503953	.348	.557
LOQ22063	37	.51351	.506712	36	.61111	.494413	.693	.408
LOQ22064	45	.84444	.366529	30	.86667	.345746	.069	.793
LOQ22065	45	.71111	.458368	47	.78723	.413688	.700	.405
LOQ22066	39	.23077	.426833	32	.28125	.456803	.231	.632
LOQ23069	41	.63415	.487652	36	.77778	.421637	1.885	.174
LOQ23070	49	.85714	.353553	41	.92683	.263652	1.086	.300
LOQ23072	42	.66667	.477119	33	.75758	.435194	.724	.398
LOQ23074	42	.30952	.467901	29	.37931	.493804	.365	.548
LOQ23076	38	.26316	.446258	34	.44118	.503995	2.527	.116
LOQ24077	50	.92000	.274048	46	.91304	.284885	.015	.903
LOQ24079	54	.74074	.442343	50	.84000	.370328	1.527	.219
LOQ24082	52	.80769	.397959	41	.78049	.419058	.102	.750
LOQ24083	65	.92308	.268543	43	.90698	.293903	.086	.770
LOQ24085	67	.55224	.501017	55	.65455	.479899	1.308	.255
LOQ24086	53	.81132	.394998	41	.82927	.380949	.049	.825
LOQ24087	53	.37736	.489364	50	.52000	.504672	2.121	.148
LOQ25089	37	.97297	.164399	26	.96154	.196116	.063	.803
LOQ25090	37	.86486	.346583	23	.95652	.208514	1.309	.257

	Baseline N	Baseline Mean	Baseline Std. Deviation	Final Assessment N	Final Assessment Mean	Final Assessment Std. Deviation	Baseline v Final Assessment F	Sig.
LOQ25091	33	.96970	.174078	35	.97143	.169031	.002	.967
LOQ25092	36	.97222	.166667	29	.96552	.185695	.023	.879
LOQ25093	41	.92683	.263652	33	.96970	.174078	.645	.425
LOQ25097	28	1.00000	0.000000	27	1.00000	0.000000	.000	.000

Table 8: Contrasts of differences in difficulty level between Learning Objectives (LSD Post-Hoc Contrasts)

		Baseline Mean Difference	Std. Error	Sig.[b]	Final Assessment Mean Difference	Std. Error	Sig.[b]
LO11	LO12	.184*	.044	.000	.225*	.053	.000
	LO13	-.189*	.039	.000	-.208*	.040	.000
	LO14	.005	.048	.923	.046	.050	.360
	LO15	-.165*	.038	.000	-.121*	.046	.009
	LO21	-.059	.037	.112	-.069	.042	.103
	LO22	.104*	.044	.020	.052	.049	.287
	LO23	.113*	.044	.010	.012	.052	.824
	LO24	-.005	.037	.899	-.038	.040	.343
	LO25	-.274*	.035	.000	-.283*	.036	.000
LO12	LO11	-.184*	.044	.000	-.225*	.053	.000
	LO13	-.373*	.041	.000	-.434*	.045	.000
	LO14	-.179*	.046	.000	-.179*	.052	.001
	LO15	-.349*	.040	.000	-.347*	.051	.000
	LO21	-.243*	.038	.000	-.295*	.044	.000
	LO22	-.080	.050	.110	-.173*	.051	.001
	LO23	-.071	.047	.136	-.214*	.053	.000
	LO24	-.189*	.038	.000	-.263*	.047	.000
	LO25	-.458*	.036	.000	-.509*	.041	.000
LO13	LO11	.189*	.039	.000	.208*	.040	.000
	LO12	.373*	.041	.000	.434*	.045	.000
	LO14	.193*	.039	.000	.254*	.044	.000
	LO15	.024	.034	.485	.087*	.038	.025
	LO21	.130*	.029	.000	.139*	.032	.000
	LO22	.292*	.040	.000	.260*	.046	.000

		Baseline Mean Difference	Std. Error	Sig.[b]	Final Assessment Mean Difference	Std. Error	Sig.[b]
	LO23	.302*	.040	.000	.220*	.039	.000
	LO24	.184*	.032	.000	.171*	.031	.000
	LO25	-.085*	.027	.002	-.075*	.027	.006
LO14	LO11	-.005	.048	.923	-.046	.050	.360
	LO12	.179*	.046	.000	.179*	.052	.001
	LO13	-.193*	.039	.000	-.254*	.044	.000
	LO15	-.170*	.042	.000	-.168*	.049	.001
	LO21	-.064	.038	.096	-.116*	.045	.011
	LO22	.099*	.047	.036	.006	.055	.917
	LO23	.108*	.045	.017	-.035	.052	.504
	LO24	-.009	.039	.807	-.084	.044	.058
	LO25	-.278*	.034	.000	-.329*	.039	.000
LO15	LO11	.165*	.038	.000	.121*	.046	.009
	LO12	.349*	.040	.000	.347*	.051	.000
	LO13	-.024	.034	.485	-.087*	.038	.025
	LO14	.170*	.042	.000	.168*	.049	.001
	LO21	.106*	.030	.001	.052	.037	.160
	LO22	.269*	.042	.000	.173*	.049	.001
	LO23	.278*	.040	.000	.133*	.046	.004
	LO24	.160*	.032	.000	.084*	.037	.026
	LO25	-.108*	.028	.000	-.162*	.034	.000
LO21a	LO11	.059	.037	.112	.069	.042	.103
	LO12	.243*	.038	.000	.295*	.044	.000
	LO13	-.130*	.029	.000	-.139*	.032	.000
	LO14	.064	.038	.096	.116*	.045	.011
	LO15	-.106*	.030	.001	-.052	.037	.160

		Baseline Mean Difference	Std. Error	Sig.[b]	Final Assessment Mean Difference	Std. Error	Sig.[b]
	LO22	.163*	.041	.000	.121*	.039	.002
	LO23	.172*	.040	.000	.081	.041	.052
	LO24	.054	.031	.079	.032	.032	.327
	LO25	-.215*	.025	.000	-.214*	.027	.000
LO22	LO11	-.104*	.044	.020	-.052	.049	.287
	LO12	.080	.050	.110	.173*	.051	.001
	LO13	-.292*	.040	.000	-.260*	.046	.000
	LO14	-.099*	.047	.036	-.006	.055	.917
	LO15	-.269*	.042	.000	-.173*	.049	.001
	LO16	-.163*	.041	.000	-.121*	.039	.002
	LO23	.009	.044	.832	-.040	.050	.421
	LO24	-.108*	.042	.011	-.090*	.043	.038
	LO25	-.377*	.037	.000	-.335*	.038	.000
LO23	LO11	-.113*	.044	.010	-.012	.052	.824
	LO12	.071	.047	.136	.214*	.053	.000
	LO13	-.302*	.040	.000	-.220*	.039	.000
	LO14	-.108*	.045	.017	.035	.052	.504
	LO15	-.278*	.040	.000	-.133*	.046	.004
	LO21	-.172*	.040	.000	-.081	.041	.052
	LO22	-.009	.044	.832	.040	.050	.421
	LO24	-.118*	.041	.004	-.049	.042	.238
	LO25	-.387*	.037	.000	-.295*	.037	.000
LO24a	LO11	.005	.037	.899	.038	.040	.343
	LO12	.189*	.038	.000	.263*	.047	.000
	LO13	-.184*	.032	.000	-.171*	.031	.000
	LO14	.009	.039	.807	.084	.044	.058

		Baseline Mean Difference	Std. Error	Sig.[b]	Final Assessment Mean Difference	Std. Error	Sig.[b]
	LO15	-.160*	.032	.000	-.084*	.037	.026
	LO21	-.054	.031	.079	-.032	.032	.327
	LO22	.108*	.042	.011	.090*	.043	.038
	LO23	.118*	.041	.004	.049	.042	.238
	LO25	-.269*	.027	.000	-.246*	.027	.000
LO25	LO11	.274*	.035	.000	.283*	.036	.000
	LO12	.458*	.036	.000	.509*	.041	.000
	LO13	.085*	.027	.002	.075*	.027	.006
	LO14	.278*	.034	.000	.329*	.039	.000
	LO15	.108*	.028	.000	.162*	.034	.000
	LO21	.215*	.025	.000	.214*	.027	.000
	LO22	.377*	.037	.000	.335*	.038	.000
	LO23	.387*	.037	.000	.295*	.037	.000
	LO24	.269*	.027	.000	.246*	.027	.000

Based on estimated marginal means, and the average of the two items used in LO

*. The mean difference is significant at the .05 level. b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table 9: Internal Consistency Analysis of the Learning objectives

Learning Objective	Baseline α	Final Assessment α	Combined Baseline and Final Assessment α
1.0 **POLICY** Explain critical regulations and policies on research misconduct, 5 indicators	0.203	-0.078	0.092
2.0 **HANDLING** Respond appropriately to situations associated with research misconduct, 5 indicators	0.096	0.236	0.175
Combined 1.0 and 2.0, 10 indicators	0.386	0.095	0.286

Table 10: Change from Baseline to Final Assessment in Total Score and Learning Objectives Scores for Participants involved in Both Phases

	Baseline Mean	Std. Deviation	%	Final Assessment Mean	Std. Deviation	%	F	Sig.	Observed Power
Total Score	8.935	1.638	74.5	9.177	1.921	76.5	0.702	.405	.131
Policy	3.836	.934	76.7	3.754	.888	75.1	0.348	.557	.089
LO 1.1	.770	.424	77.0	.754	.434	75.4	0.043	.837	.055
LO 1.2	.607	.493	60.7	.607	.493	60.7	0.000	1.000	.050
LO 1.3	.901	.300	90.1	.934	.250	93.4	0.396	.532	.095
LO 1.4	.672	.473	67.2	.623	.489	62.3	0.387	.536	.094
LO 1.5	.885	.321	88.5	.836	.373	83.6	1.000	.321	.166
Handling	3.590	.788	71.8	3.869	.856	77.4	3.880	.053	.491
LO 2.1	.762	.311	76.2	.770	.268	77.0	0.024	.877	.053
LO 2.2	.574	.499	57.4	.639	.484	63.9	0.567	.454	.115
LO 2.3	.557	.501	50.1	.721	.452	72.1	4.038	.049	.507
LO 2.4	.746	.311	74.6	.770	.297	77.0	0.207	.651	.073
LO 2.5	.950	.218	95.0	.967	.180	96.7	0.197	.658	.072

N=61

Table 11: Frequencies and Percent of Sample Participating in 8 Types of Training over the Last 5 Years and the Last 1 Year in Baseline and Final Assessment

Type of training	Baseline Last 5 years (n)	Baseline Last 5 years (%)	Baseline Last 1 year (n)	Baseline Last 1 year (%)	Final Assessment Last 5 years (n)	Final Assessment Last 5 years (%)	Final Assessment Last 1 year (n)	Final Assessment Last 1 year (%)
In-person training offered through your IC.	48	20.9	17	7.4	32	25.6	17	13.6
Video training (on http://grants.nih.gov/grants/ research_integrity).	59	25.7	22	9.6	36	28.8	21	16.8
Webinar training (hosted by OER-Research Integrity)	80	34.8	43	18.7	38	30.4	26	20.8
Online information (IC or OER websites)	98	42.6	49	21.3	46	36.8	22	17.6
The annual NIH extramural research integrity training (hosted by OER-Research Integrity)	128	55.7	102	44.3	52	41.6	36	28.8
ESA training	69	30.0	12	5.2	32	25.6	7	7.2
NIH Core Curriculum	73	31.7	15	6.5	52	41.6	12	9.6
"Other" training	23	10.0	13	5.7	11	8.8	7	5.6
Reported no training	11	4.8	42	18.3	11	8.8	31	24.8
Reported no training + missing data	18	7.8	52	22.6	18	14.4	41	32.8
Number of cases	230		230		125		125	

Note: participants could take more than one type of training, so percentages do not add up to 100.

Table 12: Number of Types of Training Taken over the Last 5 Years and the Last 1 Year in Baseline and Final Assessment

Types of Training Taken:	Baseline Last 5 years (n)	Baseline Last 5 years (%)	Baseline Last 1 year (n)	Baseline Last 1 year (%)	Final Assessment Last 5 years (n)	Final Assessment Last 5 years (%)	Final Assessment Last 1 year (n)	Final Assessment Last 1 year (%)
Only NIH training (any of 7 types)	188	81.7	164	71.3	96	76.8	77	61.6
Only "other" training	2	0.9	5	2.2	3	2.4	5	4.0
NIH training plus "other" training	21	9.1	8	3.5	8	6.4	2	1.6
No training + missing data	19	8.2	50	22.8	18	14.4	41	32.8
Number of Types of Training Taken:								
0	18	7.8	52	22.6	18	14.4	41	32.8
1	50	21.7	112	48.7	26	20.8	46	36.8
2	56	24.3	36	15.7	20	16.0	23	18.4
3	51	22.2	19	8.3	31	24.8	9	7.2
4	27	11.7	8	3.5	17	13.6	2	1.6
5	18	7.8	3	1.3	7	5.6	2	1.6
6	5	2.2	0	0.0	5	4.0	1	.8
7	5	2.2	0	0.0	1	.8	1	.8
8	0	0.0	0	0.0	0	0.0	0	0.0
Number of Cases	230	100.0	230	100.0	125	100.0	125	100.0

Table 13: Correlation between Training Experiences over the Past 5 Years and over 1 Year and Overall Test Performance in Baseline and Final Assessment.

		In-person training offered through your IC.	Video training (on http://grants.nih.gov/grants/research_integrity).	Webinar training (hosted by OER-Research Integrity)	Online information (IC or OER websites)	The annual NIH extramural research integrity training (hosted by OER-Research Integrity)	ESA training	NIH Core Curriculum
Over 5 years with Baseline Total Score	Correlation	.062	.012	-.018	.034	.197**	.023	.016
	Sig. (2-tailed)	.369	.864	.797	.625	.004	.742	.819
	N	212	212	212	212	212	212	212
Over 1 year with Baseline Total Score	Correlation	.009	-.039	-.115	.105	.115	.024	.013
	Sig. (2-tailed)	.901	.573	.095	.129	.093	.728	.848
	N	212	212	212	212	212	212	212
Over 5 years with Final Total Score	Corr.	0.01	0.083	.193**	.204**	.330**	.194**	.147*
	Sig. (2-tailed)	0.893	0.256	0.008	0.005	0	0.007	0.043
	N	191	191	191	191	191	191	191
Over 1 year with Final Score	Corr.	-0.019	0.003	.154*	0.08	.218**	.169*	0.047
	Sig. (2-tailed)	0.791	0.962	0.033	0.274	0.002	0.019	0.52
	N	191	191	191	191	191	191	191

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 14a: Multiple Regression Prediction of Test Performance in Final Assessment as a Function of Participation in 7 Training Experiences during the past year.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	6.703	.333		20.137	.000
In-person training offered through your IC.	-.402	.443	-.064	-.908	.365
Video training (on http://grants.nih.gov/grants/research_integrity).	-.067	.443	-.011	-.152	.879
Webinar training (hosted by OER-Research Integrity)	.322	.439	.054	.732	.465
Online information (IC or OER websites)	.924	.408	.165	2.267	.025
The annual NIH extramural research integrity training (hosted by OER-Research Integrity)	1.622	.393	.294	4.129	.000
ESA training	.736	.463	.122	1.590	.114
NIH Core Curriculum	.241	.446	.043	.541	.589

Table 14b: Multiple Regression Prediction of Test Performance in Final Assessment as a Function of Participation in 3 Training Experiences during the past year.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	6.701	.311		21.568	.000
The annual NIH extramural research integrity training (hosted by OER-Research Integrity)	1.673	.372	.304	4.495	.000
Online information (IC or OER websites)	1.006	.377	.179	2.666	.008
ESA training	.812	.408	.135	1.990	.048

Table 15: Correlation between Training Experiences over the Past 5 Years and over the Past 1 year with Improvement from Baseline to Final Assessment For Participants Who Took Both Tests .

		In-person training offered through your IC.	Video training (on http://grants.nih.gov/grants/research_integrity).	Webinar training (hosted by OER-Research Integrity)	Online information (IC or OER websites)	Annual NIH extramural research integrity training	ESA training	NIH Core Curriculum
Training over 5 years and Total Score Change Baseline to Final	Corr.	0.084	0.03	0.13	-0.007	0.071	0.15	0.137
	Sig. (2-tailed)	0.517	0.815	0.315	0.956	0.582	0.246	0.29
	N	61	61	61	61	61	61	61
Training over 1 Year and Total Score Change Baseline to Final	Corr.	0.03	0.126	0.11	0.05	-0.165	0.042	0.234
	Sig. (2-tailed)	0.817	0.33	0.393	0.699	0.199	0.743	0.067
	N	61	61	61	61	61	61	61
Training over 1 Year and Policy Score Change	Corr.	.143	.023	.114	.157	-.343**	-.053	.082
	Sig. (2-tailed)	.271	.862	.380	.226	.007	.684	.531
	N	61	61	61	61	61	61	61
Training over 1 Year and Handling Score Change	Corr.	-.098	.115	.006	-.052	-.098	.046	.265*
	Sig. (2-tailed)	.454	.378	.964	.691	.450	.725	.039
	N	61	61	61	61	61	61	61

**. Correlation is significant at the 0.01 level (2-tailed).

Table 16: Correlation between Training Experiences Over the Past Year and Performance on Specific Learning Objectives in Final Assessment

		In-person training offered through your IC.	Video training (on http://grants.nih.gov/grants/research_integrity).	Webinar training (hosted by OER-Research Integrity)	Online information (IC or OER websites)	The annual NIH extramural research integrity training (hosted by OER-Research Integrity)	ESA training	NIH Core Curriculum
Policy	Correlation	-.076	-.007	.062	.122	.122	.073	.077
	Sig. (2-tailed)	.322	.922	.417	.111	.111	.337	.312
	N	173	173	173	173	173	173	173
LO 1.1	Correlation	-.147	-.184*	.007	.022	.141	.086	-.058
	Sig. (2-tailed)	.054	.015	.924	.774	.064	.262	.445
	N	173	173	173	173	173	173	173
LO 1.2	Correlation	.064	.047	-.047	.111	.002	-.025	.003
	Sig. (2-tailed)	.407	.535	.539	.147	.977	.743	.973
	N	173	173	173	173	173	173	173
LO 1.3	Correlation	-.054	.086	.128	.011	.084	.093	.105
	Sig. (2-tailed)	.477	.259	.093	.884	.270	.223	.169
	N	173	173	173	173	173	173	173

LO 1.4	Correlation	.044	-.001	.146	.003	-.056	-.033	.059
	Sig. (2-tailed)	.565	.986	.056	.964	.465	.665	.441
	N	173	173	173	173	173	173	173
LO 1.5	Correlation	-.101	.074	-.077	.112	.125	.075	.097
	Sig. (2-tailed)	.188	.333	.312	.143	.101	.329	.203
	N	173	173	173	173	173	173	173
Handling	Correlation	-.026	-.102	.179*	-.046	.041	.171*	-.011
	Sig. (2-tailed)	.732	.181	.018	.552	.592	.025	.885
	N	173	173	173	173	173	173	173
LO 2.1	Correlation	.021	-.132	.129	.001	-.010	.144	.078
	Sig. (2-tailed)	.781	.083	.091	.986	.899	.059	.310
	N	173	173	173	173	173	173	173
LO 2.2	Correlation	.048	-.031	.122	.040	.028	.065	-.066
	Sig. (2-tailed)	.529	.689	.111	.598	.718	.397	.391
	N	173	173	173	173	173	173	173
LO 2.3	Correlation	-.059	.003	.142	-.063	.034	.140	.038
	Sig. (2-tailed)	.441	.966	.063	.408	.660	.066	.624
	N	173	173	173	173	173	173	173
LO 2.4	Correlation	-.117	-.060	.041	-.096	-.002	.095	-.025
	Sig. (2-tailed)	.126	.432	.597	.207	.978	.215	.742

	N	173	173	173	173	173	173	173
LO 2,5	Correlation	.062	-.125	-.082	-.009	.070	-.089	-.069
	Sig. (2-tailed)	.415	.100	.286	.903	.360	.246	.364
	N	173	173	173	173	173	173	173

*. Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).