The Office of Education Performance Assessment and Evaluation Information Management (PAEIM) Team conducted an internal assessment of information collections approved under OMB Control Number 2700-0159 NASA Office of Education Performance Measurement and Evaluation (Testing) to determine the outcome and results of the methodological testing. Available documentation and testing technical reports provided the following example summaries of results.

NASA Education STEM Challenges Impact Surveys (Student Baseline and Follow-Up Instruments, and Educator Retrospective)

NASA Education STEM Challenges Impact Surveys Methodological Testing

Methodological testing was conducted with educator and student respondents in the 21st Century Learning Community Centers (21stCCLC)/NASA Phase 3 Collaboration. In conducting the methodological testing analysis of our instruments, we included several survey items to address: the amount of time to complete the surveys, if survey questions were understandable, clarity of the survey instructions and if respondents had any survey feedback.

Type of Validity and Reliability Assessment

We measured validity and reliability of the instruments. Instrument validity occurs when the answers correspond to what they are intended to measure. There are four types of validity:

1. *Content – domain covered in its entirety;*
2. *Face – general appearance, design or layout;*
3. *Criterion – how effective are the questions in measuring what is purports to measure;*
4. *Construct – how the questions are structured to form a relationship or association (Bell, 2007).*

Reliable instruments are assessments that produce consistent results in comparable settings. For example, reliability is increased when there are consistent scores across more than one organization that serves populations in a rural setting (Bell, 2007)

We examined the instrument items and its subscales. As such, we calculated conventional measures of reliability for each scale. Cronbach's α, which can be interpreted as the average correlation (or loading usually denoted by λ) between the latent dimension and the items measuring the latent dimension. The squared multiple correlation (SMC), sometimes referred to as Guttman's λ_6, represents the proportion of the variance in the true score explained by the items. For each item, we also calculated the SMC and an examination of each item's contribution to α by examining α if we deleted the item.

Construct validity was used to identify questions that assessed students' skills, attitudes and behaviors toward STEM. The multi-scale measures described below are from the PEAR Institute Common Instrument Suite Survey 3.0 (PEAR Institute, 2016).

The common instrument suite survey has been administered over 30,000 times to students enrolled in informal science programs across the U.S., and it has shown strong reliability in previous work ($\alpha > 0.85$) (https://www.thepearinstitute.org/common-instrument-suite, Allen et al, 2016).

Respondent Characteristics

Our sample consisted of 70 EDC sites chosen at random and all 12 GLOBE SRC pilot sites. Together these 82 evaluation sites provided all the data (e.g., implementation information collected from participation logs, educator feedback forms, and in-depth interviews) for this evaluation.

From these sites we collected a total of 992 surveys from EDC students and 151 surveys from GLOBE SRC students at pre-test. During the post-test, 671 EDC students and 81 GLOBE SRC students provided responses. This represents a retention rate of 68 percent for EDC and 54 percent for GLOBE SRC. High attrition rates are common in OST programs; previous research has found that between 31 and 41 percent who start such programs go on to finish them (Apsler, 2009; Weisman and Gootfredson 2001).

All 992 EDC participants contributed to our analysis, but we retained only 151 of the 159 participants from GLOBE SRC due to one school dropping out of the study prior to post-test. Of the 992 EDC pre-test participants, 671 (or 68%) participated at post-test, where 321 were lost to attrition. An additional 183 participants provided data only at post-test; however, these participants likely only had partial exposure to the EDC program. As a result, we excluded this from our analysis. Considering comparable numbers for GLOBE SRC, of the 151 pre-test participants, 81 (or 54%) participated at post-test and 70 were lost to attrition.

Findings

Key findings from the performance assessment of the student and educator surveys and analysis are as follows:

1. *EDC and GLOBE SRC students required more than the projected average 10 minutes to complete the pre- or post-test surveys;*
2. *EDC and GLOBE SRC educators required more than the projected average 15 minutes to complete the post-test (retrospective) surveys;*
3. *Students responded that the pre- and post-test survey items were understandable and that the instructions were clear;*
4. *Of those students who provided suggestions for improvement of the EDC and GLOBE SRC pre- and post-test surveys, the most common suggestion was to add more response options, followed by provide additional/more interesting questions;*
5. *Among educators, four responses/suggestions for improving the EDC and GLOBE SRC educator surveys were to provide greater clarity to the questions, reduce the use of reverse coding, that the retrospective reporting may have proved challenging for some respondents, and more time was spent on open-ended responses;*

6. *Survey items and scales for each of the EDC and GLOBE SRC (pre- and post-test) surveys, as well as the EDC and GLOBE SRC educator surveys (retrospective) performed as expected and yielded acceptable reliability readings.*

Recommendations

Based on the findings from the survey item and subscale analysis, and the methodological testing survey item analysis, the contract evaluator made the following recommendations:

1. ***Create a shorter (fewer questions) and simpler (language) version of the student surveys*** *to achieve a 10-minute survey experience for students, especially if the plan in the future is to survey younger elementary school aged children (e.g., 4th grade);*
2. ***Create a shorter (fewer questions) version of the educator surveys*** *to achieve a 15-minute survey experience for educators;*
3. ***Consider modifying the student and educator instruments to be applicable for older student populations*** *(e.g., 9th and 10th grades) and include 9th and 10th grade students in future evaluations to examine effects of 21st CCLC on older students;*
4. ***Maintain separate EDC and GLOBE SRC student instrume****nts (do not combine the two instruments);*
5. ***Conduct a comparative analysis with other available data on STEM attitudes and beliefs****;*
6. ***Continue scaling the EDC and GLOBE SRC programs and use revised survey instruments*** *to collect student pre- and post-test data and educator post-test data;*
7. ***Continue to collect and analyze student and educator data and contribute to the research literature*** *regarding successes and challenges of 21st CCLC programs teaching engineering and science skills.*

NASA Office of Education Undergraduate Internship Impact Surveys (Retrospective and Traditional Development Surveys, Student Baseline Instruments No. 1 and Follow-up Instruments #1)

NASA Internship Expectations Post-Survey and Development Retrospective Methodological Testing NASA International Internships (I^2)

- *Deployed Spring 2016*
- *N=20*
- *STEM-related Outcomes Constructs of interest:*
  - *Internship Expectations*
  - *Development Outcomes: a dependent variable for student learning as well as an additional construct to understand students' intention to complete their degrees and satisfaction with their programs. (Retrospective)*

Success Story

NASA International Internships Post-Experience Survey

Summary (N=16) Response rate: 76%

Inspire Engage, Educate, Employ.

✓ **Who are I^2 Students?**

When asked whether the amount of work in this internship was more than expected they offered the following:

Aerospace Engineering

Mechanical Engineering

Engineering Physics

Biomedical Engineering

Biotechnology

Software Engineering

Biological Sciences

Physiology

"Although the amount of work in this internship was more than I expected, I did enjoy the challenge and opportunity to further develop my skills."

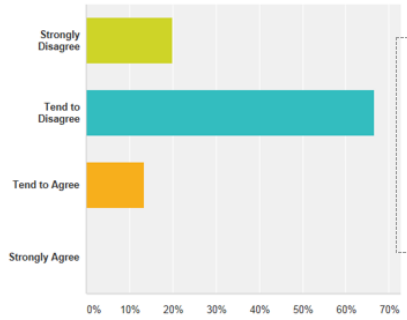
"Compared to other internships I've had, I did a lot more at NASA and really enjoyed it."

"Working with NASA as a research associate is incredible; this is not "work"."

"I came to NASA expecting to work. That's exactly what I did. In the case that I felt I wasn't given enough work I went out and found more as to not put the opportunity to waste."

25% of I^2 students are repeat participants!

✓ **I^2 students are industrious, hardworking, and eager to learn... And when asked if the internship experience was *easier* than expected:**



"I expected it to be tough and it was, but in a good way."

"It was challenging but in a positive way."

"I was challenged day in and day out with my projects."

The take away? Every student had a positive experience in I^2!

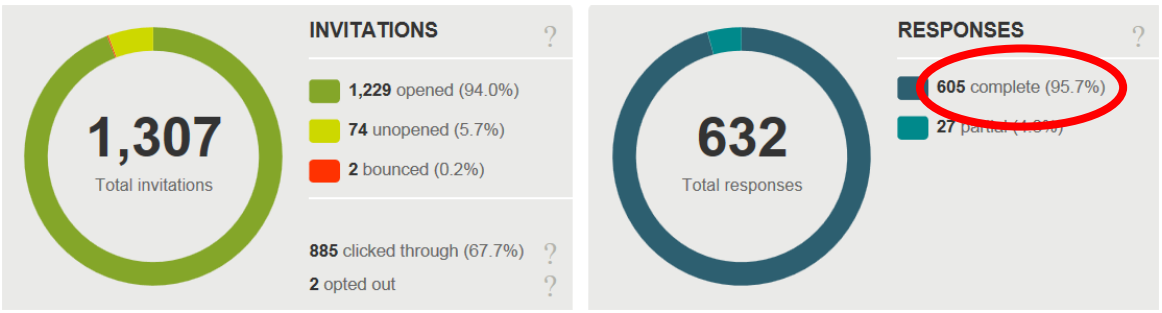
- I walked away from this internship with multidisciplinary knowledge and the insight of what it is like to conduct research.
- Learned a lot while being able to contribute to research was very satisfying.
- I have loved NASA and space since childhood. Being a NASA research associate is satisfying beyond words.
- It was a great opportunity! I am thrilled to have opportunity to contribute to an amazing project.
- The Aeromechanics department at Ames provides the best internship experience for anyone looking to network, learn and overall grow as a person.
- Working at NASA was a childhood dream, and now I know that the people and the facilities are exactly as amazing as I imagined.
- The opportunity was great. I feel that it is one of the best overall internship experiences I have ever heard of. (In regards to the particular branch I was in, Rotorcraft Aeromechanics)

Comprehensibility & Response Rate Results:

Table 1. Calculation chart to determine statistically relevant number of respondents

Data Collection Source	(N) Population Estimate for FY 2015 Q4	(A) Sampling Error +/- 5% (.05)	(Z) Confidence level 95%/ Alpha 0.05	(P) *Variability (based on consistency of intervention administration) 50%	Base sample size	Response Rate	(n) Number of Respondents
OSSI Internship Placements	1,379.00	0.0025	3.84	0.50	300	0.60	500

Table 2. Response rates for the NASA Internship Expectations Pre-Survey Summer 2016



Grit Results: Principal Component Analysis & Deployment June 2017

Rotation of the factor structure has realized three factors and variables loading highly on two factors primarily, with two variables distinct and apart in the rotated solution and presenting as a third factor.

What we learned and how it can be used:

- *Statistical means to ensure population representation in testing and routine administration*
- *Certain of the auto-reminder/auto-send frequency pattern to ensure high, representative response rates*
- *Retrospective survey format for attitude & behavior scales yields statistically relevant STEM-related outcomes data*
- *Comprehensibility questions (2) will aid in OMB clearance reporting*