Section **Page**

GENERIC CLEARANCE FOR THE NASA OFFICE OF STEM ENGAGEMENT/ PERFORMANCE MEASUREMENT AND EVALUATION (TESTING) SUPPORTING STATEMENT

A. JUSTIFICATION

1. NECESSITY FOR INFORMATION COLLECTION:

The National Aeronautics and Space Administration inspires the world with our exploration of new frontiers, our discovery of new knowledge, and our development of new technology in support of the vision to discover and expand knowledge for the benefit of humanity.

The NASA Office of STEM Engagement (OSTEM) supports that mission by deploying programs to advance the next generation's educational endeavors and expand partnerships with academic communities (see Appendix A).

NASA has a long history of engaging the public and students in its mission through educational and outreach activities and programs. NASA's endeavors in education and public outreach began early on, driven by the language in Section 203 (a) (3) of the Space Act, "to provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof, and to enhance public understanding of, and participation in, the Nation's space program in accordance with the NASA Strategic Plan." NASA's education and outreach functions aim to inspire and engage the public and students, each playing a critical role in increasing public knowledge of NASA's work and fostering an understanding and appreciation of the value of STEM, and enhancing opportunities to teach and learn. By augmenting NASA's public engagement and communicating NASA's work and value, the Agency contributes to our Nation's science literacy. NASA is committed to inspiring an informed society; enabling the public to embrace and understand NASA's work and value, today and tomorrow; engaging the public in science, technology, discovery, and exploration; equipping our employees to serve as ambassadors to the public, and providing unique STEM opportunities for diverse stakeholders.

The OSTEM Performance and Evaluation (P&E) Team supports the performance assessment and evaluation of NASA's STEM Engagement investments executed through headquarters and across the ten Center STEM Engagement Offices. The P&E Team became lead for performance measurement and program evaluation activities within OSTEM on October 1, 2017. Responsibilities include recommending and implementing agency-wide strategy for performance measurement and evaluation; ensuring the collection of high-quality data; process documentation of NASA Education projects; formative and outcome evaluations; training and technical assistance on performance measurement and evaluation. The P&E Team's goal is to provide support that improves education policy and decision-making, provides better education services, increase evaluation rigor and accountability, and ensures more effective administration of investments. The Educational Platform and Tools Team supports the NASA STEM Engagement community in the areas of information technology, dissemination and Web services, and communications and operations support. These two teams in collaboration support the overall performance assessment of NASA STEM Engagement investments across the agency.

The purpose of this request is to renew the clearance for methodological testing in order to

continue to enhance the quality of OSTEM's data collection instruments and overall data management through interdisciplinary scientific research, utilizing best practices in educational, psychological, and statistical measurement. OSTEM is committed to producing the most accurate and complete data within the highest quality assurance guidelines for reporting purposes by OSTEM leadership and by authority of the Government Performance and Results Modernization Act (GPRMA) of 2010 that requires quarterly performance assessment of Government programs for purposes of assessing agency performance and improvement. It is with this mission in mind, then, that this clearance package is submitted.[1]

Under the current clearance (2700-0159 OMB Control Number) for the NASA OSTEM Performance Measurement and Evaluation (Methodological Testing) the following information collections were approved for pilot testing.

- NASA Intern Survey
- NASA Internship Applicants and Awardees Survey
- NASA Office of STEM Engagement Engineering Design Challenge Impact Surveys: Parent Survey
- NASA Office of STEM Engagement Engineering Design Challenge Impact Surveys: Educator Feedback Survey
- NASA Office of STEM Engagement Engineering Design Challenge Impact Surveys: Student Retrospective Survey

The P&E Team conducted an internal assessment of the OSTEM information collections above to determine the outcome and results of the methodological testing. Available documentation and testing technical reports provided the example summary of methodological testing results for the NASA Intern Survey and the NASA Office of STEM Engagement Engineering Design Challenge Surveys.

[1] The entire GPRMA of 2010 can be accessed at http://www.gpo.gov/fdsys/pkg/BILLS-111hr2142enr/pdf/BILLS- 111hr2142enr.pdf.

NASA Intern Survey Methodological Testing

Methodological testing was conducted with a sample of 50 interns from the summer 2021 internship session. The findings are summarized below.

Construct Survey Items Analysis

Rasch (1960/1980) measurement is employed to assess the construct sections of the NASA Intern Survey using the Winsteps 4.4.0.6 software (Linacre, 2019). More specifically, Andrich's (1978) rating scale model is implemented because we are using polytomous survey response data. Through an iterative evaluative approach, multiple facets of each construct are investigated: rating scale function, item fit indices, item measure, as well as item and person reliability and separation. ***Rating scale function*** is studied with four criteria established by Linacre (2002). Linacre's guidelines indicate that rating scales function best when there are at least 10 observations in each category; average measures

advance monotonically with categories; outfit mean squares are less than 2.0; and step calibrations advance by at least 1.4 for a 5-pooint scale or 1.0 for a 4-point scale. ***Item fit indices*** investigated are infit, outfit, and point-biserial. An item with a negative point-biserial should be removed immediately as it does not fit with the measure. Item infit and outfit is productive for measurement between 0.5 and 1.5; is less productive but not degrading when below 0.5 or between 1.5 and 2.0; and should be considered for removal when over 2.0 as the item may distort or degrade the measurement system (Linacre, 2002). ***Item measures*** are used to determine content redundancy in a construct. When items possess a similar measure, one can usually be removed to streamline the survey provided other indices remain stable or are improved. ***Item/Person reliability and separation*** measure consistency and clarity of the construct and are considered excellent (0.90+ reliability; 3.00+ separation), good (0.80-0.89 reliability; 2.00-2.99 separation), or acceptable (0.70-0.79 reliability; 1.50-1.99 separation) according to Duncan and colleagues' (2003) criteria.

Descriptive Survey Items Analysis
Descriptive statistics are used to describe participating intern responses to these individual items as they do not form a construct or scale. Items are investigated for category use.

Integrated Findings and Recommendations
Findings are presented by item type. Recommendations related to each survey component are presented at the end of their respective sections in orange text.

Evaluative Survey Items Findings
Survey participants were asked three evaluative questions at the end of the survey. On average, survey completers said it took them 18.28 minutes (*SD*=8.63 minutes) to finish the survey. Large proportions of survey completers reported agreement or strong agreement related to survey instructions being clear (89%) and survey questions being understandable (87%). ***Reducing survey length by removing redundant of misfitting items should be considered to help lessen survey participation burden and bring the survey down to approximately 10-15 minutes for completion.***

Construct Survey Items Findings
Multiple construct analysis trials were conducted to determine the best survey data fit to the Rasch model. However, all iterations are not presented in the findings. Instead, only the initial run and other substantive iterations that assist with understanding final conclusions are documented.

STEM Engagement in Internship Construct
Scale Analysis. This section is comprised of 11 items on a 5-point scale (Not at all, At least once, Monthly, Weekly, Every day). Scale analysis showed the 5-point scale worked acceptably in all areas except Step Calibrations. Disagree and Neutral categories were overlapping as well as Agree and Strongly Agree overlapped suggesting the use of a 3-point scale which resolved these issues.

	Scale Analysis			
	10+ Observation	**Measures Advance**	**Outfit MNSQ<**	**Step Calibrations**

Scale	s	Monotonically	2.0	Acceptable
Original (5-point scale; 11 items)	Yes	Yes	Yes	No
Revised 1 (5-point scale; 10 items)	Yes	Yes	Yes	No
Revised 2 (3-point scale; 10 items)	Yes	Yes	Yes	Yes

Item Fit. All trials resulted in no negative point-biserial items and all items were acceptable in terms of infit and outfit. However, both trials on 5-point scales (11 items and 10 items) showed fewer borderline problematic items in terms of fit.

Scale	Item Fit Indices		
	-PtBis	**Problematic Infit**	**Problematic Outfit**
Original (5-point scale; 11 items)	None	8, 10	2, 8
Revised 1 (5-point scale; 10 items)	None	8, 10	2, 8
Revised 2 (3-point scale; 10 items)	None	10	1, 2, 6, 8

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. Items most difficult for interns to endorse are at the top and items easier for interns to endorse are at the bottom.

- Item 7 and below are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant in terms of measure and should be considered for removal.
- Note: Q2 is highlighted due to a grammatical issue in the item. It should read – *Work with a STEM researcher on a project of your own choosing.*

Item		Recommendation
Q4	Present my STEM research to a panel of judges from a relevant industry	Keep
Q6	Use laboratory procedures and tools	Keep
Q3	Design my own research or investigation based on my own question(s)	Keep
Q10	Build or make a computer model	Keep
Q2	Work with a STEM researcher project of your own choosing	Keep
Q7	Identify questions or problems to investigate	Keep
Q8	Analyze data or information and draw conclusions	Keep
Q9	Work collaboratively as part of a team	Keep
Q5	Interact with STEM researchers	Keep
Q1	Work with a STEM researcher or company on a real-world STEM research project	Keep

Q11	Solve real world problems	Remove – similar content and redundant measure to Q1

Person/Item Reliability & Separation Indices. While item and person reliability and separation are acceptable or better in all trials, both 5-point scale runs produced higher levels (Good and Excellent).

Scale & Items	Person		Items	
	Separation	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 11 items)	2.20	0.83	4.28	0.95
Revised 1 (5-point scale; 10 items)	2.09	0.81	4.25	0.95
Revised 2 (3-point scale; 10 items)	1.71	0.74	3.47	0.92

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. While scale categories may be "too narrow" or present "too many category options" for participants (Linacre, 1999), will not degrade interpretation of findings. Remove item 11 as it is redundant in terms of content and measure, but high levels of reliability/separation are maintained with its removal. Revise Q2 to "Work with a STEM researcher on a project of your own choosing."***

STEM Engagement in School Construct

Scale Analysis. This section is comprised of 11 items on a 5-point scale (Not at all, At least once, Monthly, Weekly, Every day). Scale analysis showed the 5-point scale worked acceptably in all areas except Step Calibrations. Disagree and Neutral categories were overlapping as well as Agree and Strongly Agree overlapped suggesting the use of a 3-point scale which resolved these issues.

Scale	Scale Analysis			
	10+ Observations	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 11 items)	Yes	Yes	Yes	No
Revised 1 (5-point scale; 10 items)	Yes	Yes	Yes	No
Revised 2 (3-point scale; 10 items)	Yes	Yes	Yes	Yes

Item Fit. All trials resulted in no negative point-biserial items and all items were acceptable in terms of infit and outfit. However, both trials on 5-point scales (11 items and 10 items) showed slightly more borderline problematic items in terms of fit.

Scale	Item Fit Indices		
	-PtBis	**Problematic Infit**	**Problematic Outfit**
Original (5-point scale; 11 items)	None	6	6
Revised 1 (5-point scale; 10 items)	None	6	6
Revised 2 (3-point scale; 10	None	None	None

items)			

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. These do not need to be evaluated for this section as this scale must have parallel items to STEM Engagement in Internship.

Person/Item Reliability & Separation Indices. While item and person reliability and separation are Good to Excellent in all trials, both 5-point scale runs produced higher levels.

	Reliability		**Separation**	
Scale & Items	**Person**	**Item**	**Person**	**Item**
Original (5-point scale; 11 items)	2.57	0.87	4.49	0.95
Revised 1 (5-point scale; 10 items)	2.50	0.86	4.57	0.95
Revised 2 (3-point scale; 10 items)	2.23	0.83	3.81	0.94

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. While scale categories may be "too narrow" or present "too many category options" for participants (Linacre, 1999), will not degrade interpretation of findings. Remove item 11 as it is redundant in terms of content and measure, but high levels of reliability/separation are maintained with its removal. Revise Q2 to "Work with a STEM researcher on a project of your own choosing."***

STEM Knowledge Construct

Scale Analysis. This section is comprised of 5 items on a 5-point scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree). Scale analysis showed the 5-point scale did not function properly in two areas: 10+ Observations per category for SD and D, as well as Step Calibrations for the same categories. The main issue is that participants all reported feeling very positively towards all items in this set and removing or collapsing lower end categories would not leave a scale. Thus, no additional varying scale trials were reported on as they do not conceptually work.

	Scale Analysis			
Scale	**10+ Observations**	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 5 items)	No (SD & D)	Yes	Yes	No
Revised 1 (5-point scale; 4 items)	No (SD & D)	Yes	Yes	No

Item Fit. No negative point-biserial items were found and all items were productive for measurement.

	Item Fit Indices		
Scale	**-PtBis**	**Problematic Infit**	**Problematic Outfit**

Original (5-point scale; 5 items)	None	None	None
Revised 1 (5-point scale; 4 items)	None	None	None

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. More difficult items for interns to agree with are at the top and items easier to agree with are at the bottom.

- All items are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant in terms of measure and should be considered for removal.

Item		**Recommendation**
Q3	Knowledge of research processes, ethics, and rules for conduct in STEM	Keep
Q5	Knowledge of what everyday research work is like in STEM	Keep
Q2	Knowledge of research conducted in a STEM topic or field	Possibly Remove – similar content to other items and redundant in measure
Q1	In depth knowledge of a STEM topic(s)	Keep
Q4	Knowledge of how scientists and engineers work on real problems in STEM	Keep

Person/Item Reliability & Separation Indices. Reliability and separation are Good in both trials.

Scale & Items	**Person**		**Items**	
	Separation	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 5 items)	2.24	0.83	1.65	0.73
Revised 1 (5-point scale; 4 items)	2.15	0.82	1.66	0.73

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Collapsing categories will not improve scale function due to so few negative responses. Remove item 2 as it is redundant in terms of content and measure, but good levels of reliability/separation are maintained with its removal.***

STEM Competencies Construct

Scale Analysis. This section is comprised of 13 items on a 5-point scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree). Scale analysis showed the 5-point scale worked acceptably in all areas except Step Calibrations. Disagree and Neutral categories were overlapping suggesting the use of a 4-point scale which resolved this issue.

	Scale Analysis

Scale	10+ Observations	Measures Advance Monotonically	Outfit MNSQ< 2.0	Step Calibrations Acceptable
Original (5-point scale; 13 items)	Yes	Yes	Yes	No
Revised 1 (5-point scale; 8 items)	Yes	Yes	Yes	No
Revised 2 (4-point scale; 8 items)	Yes	Yes	Yes	Yes

Item Fit. All trials resulted in no negative point-biserial items and all items were acceptable in terms of infit and outfit. Item fit indices functioned similarly regardless of trial.

Scale	Item Fit Indices		
	-PtBis	Problematic Infit	Problematic Outfit
Original (5-point scale; 13 items)	None	9	None
Revised 1 (5-point scale; 8 items)	None	3	None
Revised 2 (4-point scale; 8 items)	None	3	None

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. Items most difficult for interns to agree with are at the top and easier to agree with items are at the bottom.

- All items are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant in terms of measure and should be considered for removal.

Item		Recommendation
Q7	Carrying out an experiment and recording data accurately	Keep
Q2	Creating a hypothesis or explanation that can be tested in an experiment/problem	Keep
Q5	Designing procedures or steps of an experiment or designing a solution that works	Remove – similar in content to Q3
Q11	Identifying the strengths and limitations of data or arguments presented in technical or STEM texts	Keep
Q13	Defending an argument based upon findings from an experiment or other data	Remove – similar in content to Q12
Q9	Considering multiple interpretations of data to decide if something works as intended	Remove – slight infit problem
Q12	Presenting an argument that used data and/or findings from an experiment or investigation	Keep
Q1	Defining a problem that can be solved by developing a new or improved product or process	Remove – construct functions better with Q1 removed instead of Q6

Q6	Identifying the limitations of the methods and tools used for collecting data	Keep
Q8	Creating charts or graphs to display data and find patterns	Remove – redundant content with multiple other items related to data use
Q4	Making a model to show how something works	Keep
Q10	Supporting an explanation with STEM knowledge	Keep
Q3	Using my knowledge and creativity to suggest a solution to a problem	Keep

Person/Item Reliability & Separation Indices. While item and person reliability and separation are Good in all trials, both 5-point scale runs produced higher levels for persons and higher levels for items with only 8 items on the 5-point scale.

	Person		Items	
Scale & Items	**Separation**	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 13 items)	2.60	0.87	2.11	0.82
Revised 1 (5-point scale; 8 items)	2.14	0.82	2.71	0.88
Revised 2 (4-point scale; 8 items)	2.02	0.80	2.64	0.87

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Collapsing D/N categories does not improve scale function and will not degrade interpretation of findings. Remove items 1, 5, 8, 9, and 13 as they are redundant in terms of measure, and in removing these items reliability/separation improve suggesting a stronger construct.***

21st Century Skills Construct

Scale Analysis. This section is comprised of 15 items on a 5-point scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree). Scale analysis showed the original 5-point scale with 15 items worked acceptably in areas other than Step Calibrations and Outfit MNSQ<2.0. When 4 items were removed, the scale functioned better with only Step Calibrations being too narrow as Disagree and Neutral categories were overlapping suggesting the use of a 4-point scale which resolved this issue.

	Scale Analysis			
Scale	**10+ Observations**	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 15 items)	Yes	Yes	No (SD)	No
Revised 1 (5-point scale; 12 items)	Yes	Yes	Yes	No
Revised 2 (4-point scale; 12 items)	Yes	Yes	Yes	Yes

Item Fit. All trials resulted in no negative point-biserial items and all items were acceptable in terms of infit and outfit. Item fit indices functioned best in the 5-point scale with 11 items trial.

Scale	Item Fit Indices		
	-PtBis	Problematic Infit	Problematic Outfit
Original (5-point scale; 15 items)	None	None	9, 13
Revised 1 (5-point scale; 12 items)	None	None	None
Revised 2 (4-point scale; 12 items)	None	None	6

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. Items most difficult for interns to agree with are at the top and easier to agree with items are at the bottom.

- Item 5 and below are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant in terms of measure and should be considered for removal.

Item		Recommendation
Q13	Creating media products like videos, blogs, social media	Keep
Q12	Analyzing media (news) – understanding points of view in the media	Keep
Q5	Evaluating others' evidence, arguments, and beliefs	Keep
Q14	Using technology as a tool to research, organize, evaluate, and communicate information	Keep
Q2	Working creatively with others	Remove – construct functions better when Q2 removed instead of Q3
Q3	Using my creative ideas to make a product	Keep
Q8	Collaborating with others effectively and respectfully in diverse teams	Keep
Q10	Accessing and evaluating information efficiently (time) and critically (evaluates sources)	Keep
Q11	Using and managing data accurately, creatively, and ethically	Remove – construct functions better when Q11 removed instead of Q10
Q1	Thinking creatively	Keep
Q15	Adapting to change when things do not go as planned	Keep
Q6	Solving problems	Keep
Q7	Communicating clearly (written and oral) with others	Keep – content is not covered in other items, so it is necessary to align with the 21st Century Framework
Q9	Interacting effectively in a respectful and professional manner	Remove – slight misfit and similar in content to Q8

Q4	Thinking about how systems work and how parts interact with each other	Keep

Person/Item Reliability & Separation Indices. Item and person reliability and separation are Good or Excellent across trials.

Scale & Items	Person		Items	
	Separation	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 15 items)	2.60	0.87	3.94	0.94
Revised 1 (5-point scale; 11 items)	2.47	0.86	4.28	0.95
Revised 2 (4-point scale; 11 items)	2.35	0.85	5.11	0.96

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Collapsing D/N categories does not meaningfully improve scale function and will not degrade interpretation of findings. Remove items 2, 9, and 11 as they are redundant in terms of measure, and in removing these items reliability/separation improve from original trial suggesting a stronger construct.***

STEM Identity Construct

Scale Analysis. This section consists of 8 items on a 5-point scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree). Scale analysis showed the 5-point scale did not function properly in 10+ Observations per category (SD & D) and Step Calibrations for the same categories. Participants report feeling positively towards all items in this set. Collapsing all lower end categories would not leave a scale. Thus, no other scale variations work conceptionally and are not reported on.

Scale	Scale Analysis			
	10+ Observations	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 8 items)	No (SD, D)	Yes	Yes	No
Revised 1 (5-point scale; 6 items)	No (SD, D)	Yes	Yes	No

Item Fit. All items had acceptable fit indices. However, the 6-item survey trial functioned slightly better.

Scale	Item Fit Indices		
	-PtBis	**Problematic Infit**	**Problematic Outfit**
Original (5-point scale; 8 items)	None	2	2
Revised 1 (5-point scale; 6 items)	None	None	None

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to

difficulty. Items most difficult for interns to agree with are at the top and easier to agree with items are at the bottom.

- All items are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant and should be considered for removal.

Item		Recommendation
Q6	Patience for the slow pace of STEM research	Keep
Q2	Deciding on a path to pursue a STEM career	Remove – slightly problematic infit & outfit
Q1	Interest in a new STEM topic	Keep
Q5	Confidence to try out new ideas or procedures on my own in a STEM project	Remove – construct functions better when Q5 removed instead of Q8
Q8	Connecting a STEM topic or field to my personal values	Keep
Q3	Sense of accomplishing something in STEM	Keep
Q4	Feeling prepared for more challenging STEM activities	Keep
Q7	Desire to build relationships with mentors who work in STEM	Keep

Person/Item Reliability & Separation Indices. Reliability and separation are Acceptable/Good in both trials with the 6-item scale producing stronger results.

Scale & Items	Person		Items	
	Separation	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 8 items)	2.11	0.82	1.77	0.76
Revised 1 (5-point scale; 6 items)	2.16	0.82	2.10	0.82

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Remove items 2 and 5 as removing these items improves overall variable functioning and suggests a stronger construct.***

Future STEM Engagement Construct

Scale Analysis. This section consists of 10 items on a 5-point scale (Much less likely, Less likely, About the same before and after, More likely, Much more likely). Scale analysis showed the 5-point scale did not function properly in 10+ Observations per category (SD & D) and Step Calibrations. Participants report feeling positively towards all items in this set and removing or collapsing lower end categories would not leave a scale. No other scale variations work conceptionally and are thus not reported on.

Scale	Scale Analysis			
	10+ Observations	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 10 items)	No (SD, D)	Yes	Yes	No

Revised 1 (5-point scale; 7 items)	No (SD, D)	Yes	Yes	No

Item Fit. All items had acceptable fit indices. However, the 7-item survey trial functioned slightly better.

Scale	Item Fit Indices		
	-PtBis	**Problematic Infit**	**Problematic Outfit**
Original (5-point scale; 10 items)	None	None	5, 4
Revised 1 (5-point scale; 7 items)	None	None	5

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. Items most difficult for interns to agree with are at the top and easier to agree with items are at the bottom.

- All items are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant and should be considered for removal.

Item		**Recommendation**
Q1	Watch or read non-fiction STEM	Keep
Q3	Worn on solving mathematical of scientific puzzles	Remove – construct functions better when Q3 removed instead of Q1
Q8	Participate in a STEM camp, club, or competition	Keep
Q4	Use a computer to design or program something	Remove – slight overfit and redundant content
Q2	Tinker (play) with a mechanical or electrical device	Keep
Q9	Take an elective (not required) STEM class	Remove – construct functions better when Q9 removed instead of Q8
Q7	Help with a community service project related to STEM	Keep
Q6	Mentor or teach other interns about STEM	Keep
Q10	Work on a STEM project or experiment in a university or professional setting	Keep
Q5	Talk with friends or family about STEM	Keep

Person/Item Reliability & Separation Indices. Reliability and separation are Good for Persons in both trials. However, item reliability/separation are unacceptable with 10-items and Acceptable with 7-items.

Scale & Items	Person		Items	
	Separation	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 10 items)	2.79	0.89	1.41	0.67
Revised 1 (5-point scale; 7 items)	2.59	0.85	1.83	0.77

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99);

Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Remove items 3, 4, and 5 as removing these items improves overall variable functioning and suggests a stronger construct.***

Overall Internship Impacts Construct

Scale Analysis. This section is comprised of 9 items on a 5-point scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree). Scale analysis showed the 5-point scale only worked acceptably in terms of the Measures Advancing Monotonically. Again, participants reported very favorable responses leaving the lower end of the scale with very few responses. Additionally, the Disagree and Neutral categories were overlapping suggesting the use of a 4-point scale which resolved this issue but could not correct the 10+ Observations concern in the lower end of the scale.

Scale	Scale Analysis			
	10+ Observations	**Measures Advance Monotonically**	**Outfit MNSQ< 2.0**	**Step Calibrations Acceptable**
Original (5-point scale; 9 items)	No (SD, D)	Yes	No (SD)	No
Revised 1 (4-point scale; 9 items)	No (SD)	Yes	Yes	Yes

Item Fit. All trials resulted in no negative point-biserial items and all items were acceptable in terms of infit and outfit. The 5-point scale functioned slightly better in terms of item fit.

Scale	Item Fit Indices		
	-PtBis	**Problematic Infit**	**Problematic Outfit**
Original (5-point scale; 9 items)	None	8	None
Revised 1 (4-point scale; 9 items)	None	4	4

Infit & Outfit MNSQ Criteria = Productive for Measurement (0.5-1.5) – Keep; Less Productive, but not degrading (<0.5 or 1.5-2.0) – Keep; Distorts or Degrades Measurement System (>2.0) – Consider Removal (Linacre, 2002)

Item Difficulty & Redundancy. Items are sorted in the table below according to difficulty. Items most difficult for interns to agree with are at the top and easier to agree with items are at the bottom.

- Item 9 and below are "easy" for interns to agree with (below intern response mean).
- Items highlighted in the same color are redundant in terms of measure and should be considered for removal.

Item		Recommendation
Q5	I am more interested in taking STEM classes at my college	Keep
Q1	I am more confident in my STEM knowledge, skills, abilities	Keep – construct function does not improve if removed
Q6	I am more interested in pursuing a career in STEM	Keep – construct function does not

		improve if removed
Q2	I am more interested in participating in STEM activities outside of college requirements	Keep
Q3	I am more aware of other internship opportunities	Keep
Q4	I am more interested in participating in other internships	Keep
Q9	I am more interested in pursuing a STEM career with NASA	Keep
Q8	I have a greater appreciation of NASA	Keep
Q7	I am more aware of NASA research and careers	Keep

Person/Item Reliability & Separation Indices. Item and person separation and reliability are considered Good across both trials reported on.

	Person		Items	
Scale & Items	**Separation**	**Reliability**	**Separation**	**Reliability**
Original (5-point scale; 9 items)	2.46	0.86	2.04	0.81
Revised 1 (4-point scale; 9 items)	2.53	0.87	2.35	0.85

Excellent (Reliability=0.90+; Separation=3.00+); Good (Reliability=0.80-0.89; Separation=2.00-2.99); Acceptable (Reliability= 0.70-0.79; Separation=1.50-1.99); Unacceptable (Reliability=below 0.70; Separation=below 1.50) (Duncan et al., 2003)

Scale Recommendations: ***Keep original 5-point scale. Do not remove any items as removing these items does not improve variable functioning and suggests a stronger construct keeping all 9 items.***

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 (α> 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.*

Towards monitoring performance of its STEM Engagement activities, NASA Office of STEM Engagement will use rigorously developed and tested instruments administered and accessed through the approved survey management tools and/or NASA STEM Gateway system.[2] Each data collection form type possesses unique challenges which can be related to respondent characteristics, survey content, or form of administration. In the absence of meticulous methods, such issues impede the effectiveness of instruments and would decrease the value of the data gathered through these instruments for both NASA Office of STEM Engagement and the Agency.

The central purpose of measurement is to provide a rational and consistent way to summarize

the responses people make to express achievement, attitudes, or opinions through instruments such as achievement tests or questionnaires (Wilson, 2005, p. 5). In this particular instance, our interest lies in attitude and behavior scales, surveys, and psychological scales related to the goals of NASA STEM engagement activities. Yet, since NASA Education captures participant administrative data from activity application forms and program managers submit administrative data, P&E Team extends the definition of instruments to include electronic data collection screens, project activity survey instruments, and program application forms, as well.[3] Research-based, quality control methods and techniques are integral to obtaining accurate and robust data, data of high quality to assist leaders in policy decisions.

The following research techniques and methods may be used in these studies:

- Usability testing: Pertinent are the aspects of the web user interface (UI) that impact the User's experience and the accuracy and reliability of the information Users submit (Kota, n.d.; Jääskeläinen, 2010).

[2] The NASA STEM Gateway (Universal Registration and Data Management System) is a comprehensive tool designed to allow learners (i.e., students, educators, and awardee principal investigators) to register for and apply to NASA STEM engagement opportunities (e.g., internships, fellowships, challenges, educator professional development, experiential learning activities, etc.) in a single location. This web-based application enables the NASA Office of STEM Engagement to manage its participant application and data collection and reporting capabilities agency-wide. Major goals achieved through the system include 1) an enterprise solution to Registration, Application, and Data Management reducing the burden, cost and time of STEM Engagement community; 2) a structure for linking applicant information with participant information; 3) elimination of duplication and reduction in burden of student profile data (i.e., demographics and geographic distribution); 4) improvement in the overall data quality, integrity and analysis/reporting capabilities; and 5) providing a means to monitor project performance data for the purposes of determining and assessing the outputs and outcomes of STEM Engagement Investments.

[3] If constituted as a form and once approved by OMB, forms will be submitted to NASA Forms Management according to NASA Policy Directive (NPD) 1420. Thus, forms used under this clearance, will have both an OMB control number and an NPD 1420 control number that also restricts access to NASA internal users only. Instruments not constituted as forms will display an OMB control number only.

Think-aloud protocols: This data elicitation method is also called 'concurrent verbalization', meaning subjects are asked to perform a task and to verbalize whatever comes to mind during task performance. The written transcripts of the verbalizations are referred to as think-aloud protocols (TAPs) (Jääskeläinen, 2010, p 371) and constitute the data on the cognitive processes involved in a task (Ericsson & Simon, 1984/1993).

Focus group discussion: With groups of nine or less per instrument, this qualitative approach to data collection comprises the basis for brainstorming to creatively solve remaining problems identified after early usability testing of data collection screen and program application form instruments (Colton & Covert (2007), p. 37).

- Comprehensibility testing: Comprehensibility testing of program activity survey instrumentation will determine if items and instructions make sense, are ambiguous, and are understandable by those who will complete them (Colton & Covert, 2007, p. 129).

- Pilot testing: Testing with a random sample of at least 200 respondents to yield preliminary validity and reliability data (Haladyna, 2004; Komrey and Bacon, 1992; Reckase, 2000; Wilson, 2005).

- Large-scale statistical testing: Instrument testing conducted with a statistically

representative sample of responses from a population of interest. In the case of developing scales, large-scale statistical testing provides sufficient data points for exploratory factor analysis, a "large-sample" procedure (Costello & Osborne, 2005, p. 5).

- Item response approach to constructing measures: Foundations for multiple-choice testing that address the importance of item development for validity purposes, address item content to align with cognitive processes of instrument respondents, and that acknowledge guidelines for proper instrument development will be utilized in a systematic and rigorous process (DeMars, 2010).

- Split-half method: This method is an efficient solution to parallel-forms or test/retest methods because it does not require developing alternate forms of a survey and it reduces burden on respondents, requiring only participation via a single test rather than completing two tests to acquire sufficient data for reliability coefficients.

The P&E Team's goal and purpose for data collection through methodological testing is to provide support that improves education policy and decision-making, provides better STEM Engagement services, increases accountability, and ensures more effective administration within OSTEM. More in depth descriptions of techniques and methods can be found in Appendix D.

2. USES OF INFORMATION

The purpose of this data collection by the P&E Team is to ultimately improve our Federal data collection processes through scientific research. Theories and methods of cognitive science, in combination with qualitative and statistical analyses, provide essential tools for the development of effective, valid, and reliable data collection instrumentation.

The P&E Team's methodological testing is expected to 1) improve the data collection instruments employed by OSTEM, 2) increase the accuracy of the data produced by execution of OSTEM project activities upon which policy decisions are based, 3) increase the ease of administering data collection instruments for both respondents and those responsible for administering or providing access to respondents, 4) increase response rates as a result of reduced respondent burden, 5) increase the ease of use of the data collection screens within the STEM Gateway system, and 6) enhance OSTEM's confidence in and respect for the data collection instrumentation utilized by the OSTEM community.

The application of cognitive science, psychological theories, and statistical methods to data collection is widespread and well established. Neglecting accepted research practices and relying on trial and error negatively impact data quality and unfairly burden respondents and administrators of data collection instruments. For example, without knowledge of what respondents can be expected to remember about a past activity and how to ask questions that effectively aid in the retrieval of the appropriate information, researchers cannot ensure that

respondents will not take shortcuts to avoid careful thought in answering the questions, or be subject to undue burden. Similarly, without investigating potential respondents' roles and abilities in navigating electronic data collection screens, researchers cannot ensure that respondents will read questions correctly with ease and fluency, navigate electronic data screens properly or efficiently, or record requested information correctly and consistently. Hence, consequences of failing to scientifically investigate the data collection process should and can be avoided.

In light of the Administration's call for increased sharing of federal STEM education resources through interagency collaborations, OSTEM may make available results of methodological testing to other federal STEM agencies in the form of peer-reviewed methods reports or white papers describing best practices and lessons learned. For instance, from inception NASA has supported the Federal Coordination in STEM (FC-STEM) Graduate and Undergraduate STEM Education interagency working groups' efforts determine cross-agency, common metrics and share effective program evaluations. *Coordination Objective 2: Build and use evidence based approaches* calls for agencies to:

> Conduct rigorous STEM education research and evaluation to build evidence about promising practices and program effectiveness, use across agencies, and share with the public to improve the impact of the Federal STEM education investment. (National Science and Technology Council, 2013, p. 45)

The methods to be employed in developing and testing data collection instruments will be methodologically sound, rigorously obtained, and will thus constitute evidence worthy of dissemination through appropriate vehicles. Data collection instruments appropriate for a participant in a postsecondary OSTEM research experience and are specific to the category of participant: undergraduate student, graduate student, mentor participant. One survey instrument explores a participant's preparation for a research experience while its complement explores a participant's attitudes and behaviors pre- and post-experience (undergraduate or graduate student) (Crede & Borrego, 2013.) Two non-cognitive competency scales explore a participant's developmental levels of affect (grit and mathematics self-identity & self-efficacy) as related to participation in a NASA Education research experience (Duckworth, Peterson, Matthews, & Kelly, 2007; National Center for Education Statistics, 2009.) Lastly, the mentor survey explores a mentor's attitudes and behaviors associated with participation as a mentor of an OSTEMresearch experience (Crede & Borrego, 2013.) Additional information collections will be submitted separately under this clearance with justification information and evidence-based methodology for methodological testing. Appendix G shows the explanatory content that will accompany each information collection for methodological testing purposes.

3. CONSIDERATIONS OF USING IMPROVED TECHNOLOGY

The P&E Team in collaboration with the Educational Platform and Tools Team will plan, conduct, and interpret field and laboratory research that contributes to the design of electronic data collection screens, project activity survey instruments, and program application forms

used within the context of the OSTEM community.. These efforts are supported in two ways, by use of information technology applications and strategic efforts to improve the overall information technology data collection systems used by OSTEM.

Use of Information Technology (IT) Application

IT applications will be used to bridge the distance between the P&E Team of researchers mostly based at NASA Glenn Research Center in Cleveland, OH, and the Educational Platform and Tools Team at NASA Headquarters in Washington, DC. Multiple modes of technology may be used to bring the laboratory environment to study participants at various Center locales. In addition, data management and analyses applications have been made available to study leads to optimize data collection and analyses.

Different laboratory methods may be used in different studies depending on the aspects of the data collection process being studied. Computer technology will be used when appropriate to aid the respondents and interviewers, and to minimize burden. For instance, the P&E Team and/or contractor support may use Adobe Connect, Microsoft TEAMS, or Webex to conduct focus groups and cognitive interviews if indeed there is inadequate representation of participant populations at area NASA research centers.[4,5] All of these platforms are used throughout the NASA research centers and have the potential to facilitate instrument development by providing access to appropriate study participants. The P&E Team has direct access and is also training in using other IT applications to facilitate this work as described below.

- Adobe Connect: Adobe Systems Incorporated describes Adobe Connect as "a web conferencing platform for web meetings, eLearning, and webinars [that] powers mission critical web conferencing solutions end-to-end, on virtually any device, and enables organizations […] to fundamentally improve productivity."
- SurveyMonkey: This application may be used to collect non-sensitive, non-confidential qualitative responses to determine preliminary validity. This online survey software provides an electronic environment for distributing survey questionnaires.[6] For the purpose of NASA Office of STEM Engagement, SurveyMonkey is a means by which feedback can be collected from a variety of participants such as from subject matter experts when in the early stages of instrument development when operationalizing a construct is vital to the process of instrument development.

[4] More information on Adobe applications is available at http://www.adobe.com/products/adobeconnect.html
[5] More information on WebEx applications is available at https://www.webex.com/video-conferencing
[6] More information on SurveyMonkey can be found at https://www.surveymonkey.com/mp/take-a- tour/?ut_source=header.

A process referred to as operationalization is another tangible means to measure a construct since a construct cannot be observed directly (Colton & Covert, 2007, p. 66). The qualitative feedback of subject matter experts, in addition to the research literature, provides the factors or variables associated with constructs of interest. SurveyMonkey will facilitate the gathering of such information and interface with NVivo 10 for Windows qualitative software for analyses and consensus towards developing valid items and instruments.

- SurveyMonkey: This application may be used to collect non-sensitive, non-confidential qualitative responses to determine preliminary validity. This online survey software provides an electronic environment for distributing survey questionnaires.[6] For the purpose of NASA Education, SurveyMonkey is a means by which feedback can be collected from a variety of participants such as from subject matter experts when in the early stages of instrument development when operationalizing a construct is vital to the process of instrument development. A process referred to as operationalization is another tangible means to measure a construct since a construct cannot be observed directly (Colton & Covert, 2007, p. 66). The qualitative feedback of subject matter experts, in addition to the research literature, provides the factors or variables associated with constructs of interest. SurveyMonkey will facilitate the gathering of such information and interface with NVivo 10 for Windows qualitative software for analyses and consensus towards developing valid items and instruments.
- NASA Google G-Suite (Google Form): This application may be used to collect non-sensitive, non-confidential qualitative responses to determine preliminary validity. This online survey application provides an electronic environment for distributing survey questionnaires. For the purpose of NASA Education, Google Form is a means by which feedback can be collected from a variety of participants such as from subject matter experts when in the early stages of instrument development when operationalizing a construct is vital to the process of instrument development. The NASA Google G-Suite also provides a file storage and synchronization service that allows users to store files on their servers, synchronize files across devices, and share files with NASA/non-NASA credentialed.
- NVivo 10 for Windows: This software is a platform for analyzing multiple forms of unstructured data. The software provides powerful search, query, and visualization tools. A few features pertinent to instrument development include pattern based auto-coding to code large volumes of text quickly, functionality to create and code transcripts from imported audio files, and convenience of importing survey responses directly from SurveyMonkey. [7]

[7] More information is available at http://www.qsrinternational.com/products_nvivo.aspx

- STATA SE v14: This data analysis and statistical software features advanced statistical

functionality with programming that accommodates analysis, testing, and modeling from large data sets with the following characteristics: Maximum number of variables-32,767; Maximum number of right-hand variables- 10,998; and unlimited observations. These software technical specifications allow for the statistical calculations to determine and monitor over time item functioning and psychometric properties of NASA Office of Education data collection instrumentation. [8]

Strategic Planning and Designing Improved Information Technology Data Collection Systems

The P&E Team has invested much time and effort in developing secure information technology applications that will be leveraged on behalf of instrument piloting and for the purposes of routine deployment that will enable large-scale statistical testing of data collection instruments. New information technology applications, the Composite Survey Builder and Survey Launcher, are in development with the new NASA STEM Gateway System. The Survey Launcher application will allow the P&E Team to reach several hundred OSTEM project activity participants via email whereas the Composite Survey Builder will allow the P&E Team to administer data collection instruments approved by the Office of Management and Budget (OMB) Office of Information and Regulatory Affairs via emailed web survey links. This same technology will be leveraged to maximize response rates for piloting and routine data collection instrument deployment.

Most recently, OSTEM has acquired a full-time SME specifically tasked with strategizing approaches to enhance the Office's IT systems and applications to be more responsive to Federal mandates as well as to the needs of the OSTEM community. This person's work is intended to lay the foundation for fiscally responsible IT development now and in the future.

Recall, participants in focus groups and cognitive interviews must mirror in as many characteristics as possible the sample of participants upon which the instrument will eventually be tested and then administered. Using technology to employ qualitative and quantitative methods is a means to establish validity from the onset prior to field testing and quantitative measures to determine instrument reliability and validity while monitoring and minimizing burden on study participants. Having the proper IT foundations in place for this work is a NASA OSTEM priority.

[8] More information is available at http://www.stata.com/products/which-stata-is-right-for-me/#SE

4. EFFORTS TO IDENTIFY DUPLICATION

Because developing new valid and reliable data collection instrumentation is still a relatively new procedure for NASA OSTEM, many participants within our community have yet to

participate in this kind of procedure. Participation in instrument development or testing is not mandatory.
Further, to reduce burden, any participant within our community recruited to participate in instrument development will only be solicited to contribute effort towards a single instrument, unless he or she volunteers for other opportunities. The P&E Team will attempt to reduce some of the testing burden by identifying appropriate valid and reliable instruments/scales through Federal resources or the educational measurement research literature.

5. Efforts to Minimize Burden on Small Business

Not applicable. NASA OSTEM does not collect information from any small business or other small entities.

6. Consequences of Less Frequent Data Collection

This planned collection of data will allow the P&E Team the opportunity to design appropriate valid and reliable data collection instrumentation, and the prerogative to modify and alter instruments in an on-going manner in response to changes in respondent demographics and the NASA OSTEM portfolio of activities. Because this collection is expected to be an on-going effort, it has the potential to have immediate impact on all data collection instrumentation within OSTEM. Any delay would sacrifice potential gains in development of and modification to data collection instrumentation as a whole.

7. Special Circumstances

Not applicable. This data collection does not require any one of the reporting requirements listed.

8. Federal Register Announcement and Consultation Outside the Agency

- The 60-day Federal Register Notice, Volume 86, Number 069 was published on 4/13/2021. No comments were received from the public.
- The 30-day Federal Register Notice, Volume 86, Number 145 was published on 8/2/2021.

OSTEM will continue to leverage its civil servant and contractor workforce to develop strategies, design programs, sustain operations, implement new application and capabilities, develop business processes and training guidance, and provided support to stakeholders and end users. Key to an effective portfolio of programs is having a more rigorous approach to planning and implementation of activities through the use of evidence-based effective practices for STEM education and evaluation. An important component of these performance assessment and evaluation activities, is the review and input by a panel of nationally recognized experts in STEM. For this reason, OSTEM will also consult with relevant expertise from individuals outside of the agency through a Performance Assessment and Evaluation Expert Review Panel

(ERP) to obtain views and feedback on performance measurement activities including, but not limited to: internal and external performance measures and recommended data collection sources, process and tools, as well as NASA evidence-based decision making. The ERP will act as a technical review working group providing expertise and feedback in the following areas: program structure and evaluation, K12/higher education and diversity, building technical research capacity at higher education institutions, information technology systems/social media and emerging technologies, science literacy and large scale public engagement campaigns.

9. PAYMENT OR GIFTS TO RESPONDENTS

Not applicable. NASA OSTEM does not offer payment or gifts to respondents.

10. ASSURANCE OF CONFIDENTIALITY

OSTEM is committed to protecting the confidentiality of all individual respondents that participant in data collection instrumentation testing. Any information collected under the purview of this clearance will be maintained in accordance with the Privacy Act of 1974, the e-Government act of 2002, the Federal Records Act, and as applicable, the Freedom of Information Act in order to protect respondents' privacy and the confidentiality of the data collected (See Appendix E.)

The data collected from respondents will be tabulated and analyzed only for the purpose of evaluating the research in question. Laboratory respondents will be asked to read and sign a Consent form, a personal copy of which they are provided to retain. The Consent form explains the voluntary nature of the studies and the use of the information, describes the parameters of the interview (taped or observed), and provides assurance of confidentiality as described in NASA Procedural Requirements (NPR) 7100.1.[9]

The consent form administered will be edited as appropriate to reflect the specific testing situation for which the participant is being recruited (See Appendix C). The confidentiality statement, edited per data collection source, will be posted on all data collection screens and instruments, and will be provided to participants in methodological testing activities per NPR 7100.1 (See Appendix E.)

[9] The entire NPR 7100.1 Protection of Human Research Subjects (Revalidated 6/26/14) may be found at: http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_7100_0001_&page_name=main

11. JUSTIFICATION FOR SENSITIVE QUESTIONS

Assuring that students participating in OSTEM projects are representative of the diversity of the Nation requires OSTEM to capture the race, ethnicity, and disability statuses of its

participants. Therefore, to assure the reliability and validity of its data collection instruments, the P&E Team will need to ascertain that study participants are representative of students participating in NASA STEM Engagement projects. Race and ethnicity information is collected according to Office of Management and Budget (1997) guidelines in "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity."[10] Although disclosure of race and ethnicity are not required to be considered for opportunities at NASA, respondents are strongly encouraged to submit this information. The explanation given to respondents for acquiring this information is as follows:

> In order to determine the degree to which members of each ethnic and racial group are reached by this internship/fellowship program, NASA requests that the student select the appropriate responses below. While providing this information is optional, you must select decline to answer if you do not want to provide it. Mentors will not be able to view this information when considering students for opportunities. For more information, please visit http://www.nasa.gov/about/highlights/HP_Privacy.html.

Information regarding disabilities is collected according to guidelines reflected in the "Self- Identification of Disability" form SF-256 published by the Office of Personnel Management (Revised July 2010) and is preceded by the following statement:

> An individual with a disability: A person who (1) has a physical impairment or mental impairment (psychiatric disability) that substantially limits one or more of such person's major life activities;
> (2) has a record of such impairment; or (3) is regarded as having such an impairment. This definition is provided by the Rehabilitation Act of 1973, as amended (29 U.S.C 701 et. seq.)[11]

Regulations safeguarding this information is provided to study participants on the informed consent form as governed by NPR 7100.1.

[10] http://www.whitehouse.gov/omb/fedreg_1997standards
[11] http://www.opm.gov/forms/pdf_fill/sf256.pdf

12. ESTIMATE OF RESPONDENT BURDEN

The estimate of respondent burden for methodological testing is as follows (See Table 1):

Table 1: Estimate of Respondent Burden for Methodological Testing

Data Collection Sources	Respondent Category	Statistically Adjusted Respondents	Frequency of Response	Total minutes per Response	Total Response Burden in Hours
NASA STEM Gateway System	Students (15 and younger)	9,200	1	15	2,300 hours
	Students (16 and older)	9,200	1	15	2,300 hours
	Educators and Parents	4,000	1	15	1,000 hours
	Total Burden for Methodological Testing	**22,400**			**5,600**

The estimate of annualized cost to respondents for methodological testing is as follows (See Table 2). Annualized Cost to Respondents is calculated by multiplying Total Response Burden in Hours by Wage specific to Respondent Category (Bureau of Labor Statistics, 2014).

Table 2: Estimate of Annualized Cost to Number of Respondents Required for Methodological Testing

Data Collection Sources	Respondent Category	Total Response Burden in Hours	Wage	Annualized Cost to Respondents
NASA STEM Gateway	Students (15 years of age and younger)	2,300	$5.36/hr	$12,328
	Students (16 years of age and older)	2,300	$7.25/hr	$16,675
	Educators and/or Parents	1,000	$25.09/hr	$25,090
	Total Burden for Methodological Testing	**5,600**		**$54,093**

13. COST BURDEN TO RESPONDENTS

Not applicable. Participation in testing does not require respondents to purchase equipment, software, or contract out services. The instruments used will be available in electronic format only. OSTEM's expectation is all targeted respondents can access the NASA STEM Gateway System forms/instruments electronically for the purposes of testing as they have in the past when applying to NASA opportunities.

14. COST BURDEN TO FEDERAL GOVERNMENT

The total annualized cost estimate for this information collection is $0.7 million based on existing contract expenses that include contract staffing, staff training for data collection, data cleaning, validation, and management, and reporting relating to contract staffing for online systems including but not limited to the NASA STEM Gateway data collection suite.

15. REASON FOR CHANGE IN BURDEN

This is a renewal application for methodological testing of data collection instrumentation within OSTEM by the P&E Team. Adjustments to burden in Items 13 and 14 reflect new projected respondent population universe and minutes per response for testing in alignment with the OSTEM evaluation strategy including collection through the NASA STEM Gateway System and/or other survey management tools.

16. SCHEDULE FOR INFORMATION COLLECTION AND PUBLICATION

OSTEM may make available results of methodological testing to other federal STEM agencies in the form of peer-reviewed methods reports or white papers describing best practices and lessons

learned on an as-appropriate basis determined by OSTEM leadership. Although there is no intent to publish in academic journals, standards for drafting will reflect peer-reviewed, publication-level standards of quality.

17. DISPLAY OF OMB EXPIRATION DATE

The OMB Expiration Date will be displayed on every data collection instrument, once approval is obtained.

18. EXCEPTION TO THE CERTIFICATE STATEMENT

NASA does not take exception to the certification statements below:

> The proposed collection of information –
>
> (a) is necessary for the proper performance of the functions of NASA, including that the information to be collected will have practical utility;
>
> (b) is not unnecessarily duplicative of information that is reasonably accessible to the agency;
>
> *(c)* reduces to the extent practicable and appropriate the burden on persons who shall provide information to or for the agency, including with respect to small entities, as defined in the Regulatory Flexibility Act *(5 U.S.C. 601*(6)), the use of such techniques as:
>
> > (1) establishing differing compliance or reporting requirements or timelines that take into account the resources available to those who are to respond;
> >
> > (2) the clarification, consolidation, or simplification of compliance and reporting requirements; or
> >
> > (3) an exemption from coverage of the collection of information, or any part thereof;
>
> (d) is written using plain, coherent, and unambiguous terminology and is understandable to those who are targeted to respond;
>
> (e) indicates for each recordkeeping requirement the length of time persons are required to maintain the records specified;
>
> (f) has been developed by an office that has planned and allocated resources for the efficient and effective management and use of the information to be collected, including the processing of the information in a manner which shall enhance, where appropriate, the utility of the information to agencies and the public;
>
> (g) when applicable, uses effective and efficient statistical survey methodology appropriate to the purpose for which the information is to be collected; and
>
> (h) to the maximum extent practicable, uses appropriate information technology to reduce burden and improve data quality, agency efficiency and responsiveness to the public; and
>
> (i) will display the required PRA statement with the active OMB control number, as validated on www.reginfo.gov

Name, title, and organization of NASA Information Collection Sponsor certifying statements

above:

NAME: Richard L. Gilmore Jr., M.Ed.

TITLE: Performance Assessment and Evaluation Program Manager

ORG: Office of STEM Engagement

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psycometrika, 43,* 561-573.

Bureau of Labor Statistics. (2014). Retrieved from http://www.bls.gov/home.htm.

Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social reserch and evaluation.* San Francisco: John Wiley and Sons, Inc.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Crede, E., & Borrego, M. (2013). From ethnography to items: A mixed methods approach to developing a survey to examine graduate engineering student retention. *Journal of Mixed Methods Research, 7*(1), 62-80.

Davidshofer, K. R., & Murphy, C. O. (2005). *Psychological testing: Principles and applications.* (6th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

DeMars, C. (2010). *Item response theory.* New York: Oxford University Press.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perserverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087- 1101.

Duncan, P., Bode, R., Lai, S., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation, 84,* 950-963.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis.* New York, NY: Oxford University Press.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jaaskelainen, R. (2010). Think-aloud protocol. In Y. Gambier, & L. Van Doorslaer (Eds.), *Handbook of translation studies* (pp. 371-373). Philadelphia, PA: John Benjamins.

Komrey, J. D., & Bacon, T. P. (1992). Item analysis of acheivement tests based on small numbers of examinees. *Paper presented at the annual meeting of the American Educational Research Association.* San Francisco.

Kota, K. (n.d.). Testing your web application: A quick 10-step guide. *Retrieved from http://www.adminstrack.com/articles/testing_web_apps.pdf*.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Linacre, J. M. (2019). *Winsteps ® (Version 4.4.0.6)* [Computer Software]. Winsteps.com

Linacre, J. M. (1999). Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Measurement Transactions, 13*(1), p. 675.

National Center for Education Statistics, U. (2009). High School Longitudinal Study of 2009, First Follow-up. *OMB No: 1850-0852*.

National Science and Technology Council. (2013). *Federal science, technology, engineering, and mathematics (STEM) education 5 year strategic plan.* Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/stem_stratplan_2013.pdf.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), with foreword and afterword by B.D. Wright. The University of Chicago Press.

Reckase, M. D. (2000). The minimum sample size needed to calibrate items using the three-parameter logistic model. *Paper presented at the annual meeting of the American Educational Research Association.* New Orleans.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* New York: Psychology Press.

APPENDIX A: NASA Office of STEM Engagement Goals

Section 203 (a) (3) of the Space Act, directs NASA "to provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof, and to enhance public understanding of, and participation in, the Nation's space program in accordance with the NASA Strategic Plan." In support of this directive, NASA engages the public and students in its mission through a portfolio of STEM programs and activities. The 2020 NASA Strategic Plan reinforces the Agency's commitment to inspiring an informed society; engaging the public in science, technology, discovery and exploration; and providing unique STEM opportunities for diverse stakeholders. NASA's investments in these areas are guided by Strategic Goal 3: *Address national challenges and catalyze economic growth*, and Strategic Objective 3.3: *Inspire and engage the public in aeronautics, space, and science.* NASA's support of U.S. industry and academia seeks to foster economic development and growth, embody American ingenuity, and serve as a magnet for the STEM workforce.

NASA is committed to defining and implementing a portfolio of STEM Engagement programs, projects, activities and products directed toward achieving the objectives and strategies above, driving a coherent and coordinated set of activities across the Agency. Ultimately, the work dedicated to this strategy will contribute to achieving NASA's STEM Engagement vision to immerse students in NASA's work, attract students to STEM, and inspire the next generation to explore.

Central to this effort is an architecture designed to enable relevant student contributions to NASA's mission and work, driven by requirements from NASA's Mission Directorates. This structure aligns appropriated STEM engagement programs, existing and emerging relevant projects, and activities and products from across the Agency, into an overarching framework and strategy. The result is an effective and coherent approach and outcomes, see Figure 1.

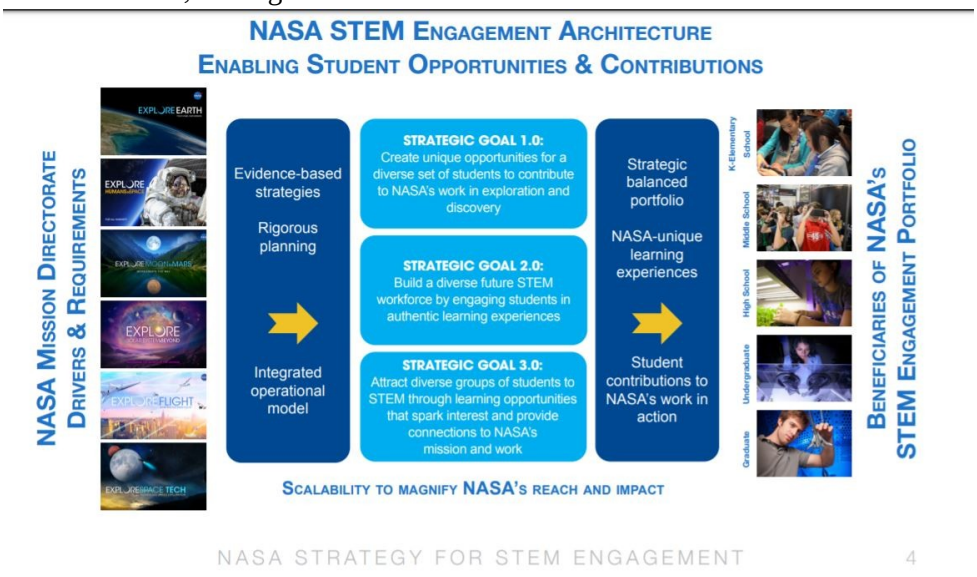

Figure 1. Office of STEM Engagement (OSTEM) architecture.

At the core of NASA's efforts in STEM Engagement are the following cross-cutting design and operational principles. These principles guide the STEM engagement community in the planning and execution of work in direct support of achieving the objectives.

- Mission-driven authentic STEM experiences
- Evidence-based practices
- Diversity and inclusion
- Scalability through partnerships and networks
- Outcome-driven

NASA's STEM engagement function plays a critical role in achieving the Agency's Strategic Objective 3.3 by implementing activities within three focus areas: 1) Create unique opportunities for a diverse set of students to contribute to NASA's work in exploration and discovery; 2) Build a diverse future STEM workforce by engaging students in authentic learning experiences with NASA's people, content and facilities; and 3) Attract diverse groups of students to STEM through learning opportunities that spark interest and provide connections to NASA's mission and work. The goals and objectives for NASA STEM Engagement are:

Goal 1.0: Create unique opportunities for a diverse set of students to contribute to NASA's work in exploration and discovery.

Objective 1.1: Provide student work experiences that enable students to contribute to NASA's missions and programs, embedded with NASA's STEM practitioners.

Objective 1.2: Create structured and widely-accessible, experiential learning opportunities for students to engage with NASA's experts and help solve problems that are critical to NASA's mission.

Goal 2.0: Build a diverse future STEM workforce by engaging students in authentic learning experiences with NASA's people, content and facilities.

Objective 2.1: Develop and deploy a continuum of STEM experiences through authentic learning and research opportunities with NASA's people and work to cultivate student interest, including students from underrepresented and underserved communities, in pursuing STEM careers and foster interest in aerospace fields.

Objective 2.2: Design the portfolio of NASA STEM engagement opportunities to contribute toward meeting Agency workforce requirements and serving the nation's aerospace and relevant STEM needs.

Goal 3.0: Attract diverse groups of students to STEM through learning opportunities that spark interest and provide connections to NASA's mission and work.

Objective 3.1: Develop and deploy targeted opportunities and readily available NASA STEM engagement resources and content, to attract students to STEM.

Objective 3.2: Foster student exposure to STEM careers through direct and virtual experiences with NASA's people and work.

APPENDIX B: NASA Center STEM Engagement Offices

Strategic management of the NASA STEM Engagement portfolio requires the participation of the Office of STEM Engagement (headquarters), the four Mission Directorates and all ten NASA Centers. This extensive participation provides STEM Engagement opportunities with NASA content, people and facilities. Close and effective consultation, coordination and cognizance among all entities are critical to the optimal fulfillment of NASA's objectives relative to its education investment.

OSTEM provides integration and evaluation support to the community. As such, the Educational Platform and Tools Team maintains a centralized database of all NASA OSTEM activities and investments, and supports coordination of evaluation and assessment of the Agency education portfolio. The Performance and Evaluation Team (P&E) works closely with the Office of the Chief Information Officer (OCIO) to develop Paperwork Reduction Act (PRA) guidance and training resources for Center STEM Engagement Offices. Upon improved compliance of the Center STEM Engagement Offices, all Centers will submit data collection instruments for development and clearance through the P&E Team first and then approval by the NASA OMB liaison prior to submission to OMB. This process will reduce burden on the STEM Engagement community while optimizing data collection.

Center STEM Engagement Offices are responsible for implementing NASA OSTEM programs, projects and activities for the Mission Directorates and OSTEM, as well as planning and implementing STEM Engagement projects that are unique to and funded by their Centers. Centers are responsible for execution of programs and projects and for institutional assets. The Center STEM Engagement Offices provide expertise in state standards and requirements in their area of geographic responsibility for K-12 education, and provide valuable field-based input into STEM Engagement program planning.

Locations of NASA Center STEM Engagement Offices

Ames Research Center

Ames specializes in research geared towards creating new knowledge and new technologies that span the spectrum of NASA interests.

Armstrong Flight Research Center

As the lead for flight research, Armstrong continues to innovate in aeronautics and space technology. The newest, fastest, the highest -- all have made their debut in the vast, clear desert skies over Armstrong.

Glenn Research Center

Glenn Research Center develops and transfers critical technologies that address national priorities through research, technology development, and systems development for safe and reliable aeronautics, aerospace, and space applications.

Goddard Space Flight Center

The mission of the Goddard Space Flight Center is to expand knowledge on the Earth and its environment, the solar system, and the universe through observations from space.

Jet Propulsion Laboratory

The Jet Propulsion Laboratory, managed by the California Institute of Technology is NASA's lead center for robotic exploration of the Solar System.

Johnson Space Center

From the early Gemini, Apollo, and Sky Lab projects to today's Space Shuttle and International Space Station programs, Johnson Space Center continues to lead NASA's effort in Human Space Exploration.

Kennedy Space Center

Kennedy Space Center is America's Gateway to the Universe -- leading the world in preparing and launching missions around the Earth and beyond.

Langley Research Center

Langley continues to forge new frontiers in aviation and space research for aerospace, atmospheric sciences, and technology commercialization to improve the way the world lives.

Marshall Space Flight Center

Bringing people to space; bringing space to people. Marshall Space Flight Center is world leader in the access to space and use of space for research and development to benefit humanity.

Stennis Space Center

Stennis is responsible for NASA's rocket propulsion testing and for partnering with industry to develop and implement remote sensing technology.

APPENDIX C: Data Instrument Collection Testing Participation Generic Consent Form[12]

In accordance with the Privacy Act of 1974, as amended (5 U.S.C. 552a), you are hereby notified that this study is sponsored by the National Aeronautics and Space Administration (NASA) Office of Education Performance Assessment and Evaluation Information Management (PAEIM) Team, under authority of the Government Performance and Results Modernization Act (GPRMA) of 2010 that requires quarterly performance assessment of Government programs for purposes of assessing agency performance and improvement. Your participation is important to the success of this study. The information we collect will help us improve the nature of NASA education project activities and the accuracy with which NASA Office of Education can report to the stakeholders about the project activities offered. The NASA PAEIM Team will use the information provided for statistical purposes related to data collection instrument development only and will hold the information in confidence to the full extent permitted by law. Information will be secured and removed from this server and location upon guidelines set out by the NASA Records Retention Schedule 1392, 68-69. Although the following efforts will be taken to ensure confidentiality, there remains a remote risk of personal data becoming identifiable. A non-identifying code number will be assigned to participants' data records, which will be stored in accordance with federal regulatory procedures and accessible only to the investigator. Any use of individual data to illustrate specific assessment results will be labeled in a manner to preserve the participants' anonymity. In no way does refusing participation in this instrument development study preclude you from eligibility for NASA education project activities now or in the future.

Introduction

This research seeks to support the mission of the NASA Office of STEM Engagement by asking you to take part in a (focus group/cognitive interview/ instrument development testing) pertaining to our interest in the ways in which NASA project activities impact outcomes for participants.[13] The information we collect will help us to improve the nature of the project activity and the accuracy with which NASA Office of STEM Engagement can report to the community about the project activities it offers.

Purpose of the Study

Determine the degree to which this instrument accurately captures the ways participant outcomes are measured by this data collection instrument.

Description of Study Procedures

Participants will be asked to complete XXX.
There are no foreseeable risks to participants electing to participate in this study.

Estimation of Time Required

We estimate it will take you an average of [enter #] minutes to participate in this research (ranging from [enter #] minutes to [enter #] minutes).

---

[12] Once approved by OMB, this form will be submitted to NASA Forms Management according to NASA Policy Directive (NPD) 1420. Thus, this form, and all others used under this clearance, will have both an OMB control number and an NPD 1420 control number that also restricts access to NASA internal users only.

[13] This clearance package is to obtain permission to develop instruments to be used in testing that will be approved by OMB first for inclusion under this clearance prior to testing.

Securing Your Responses

Under no circumstances will the results of your surveys be shared with anyone without your explicit permission. The results of this research may be presented at meetings or in publications, however, your identity will not be disclosed. Presentations and manuscripts typically contain participants' quotes, but participants are never identified by name. Your involvement in the development of this instrument is entirely voluntary and you have the right to discontinue participation at any time.

Contact Persons

If you have any additional questions concerning the research, this informed consent, or confidentiality of responses, please contact Richard L. Gilmore Jr., Evaluation Manager, at richard.l.gilmore@nasa.gov or call (216)433-5493.

~~~~~~

I have read and understand the contents of this study information and informed consent form and have been encouraged to ask questions. I have received answers to the questions I have asked. I give my consent to participate freely in this research. I have signed and retained a copy of the information and consent form for my records and future reference. **I have signed and submitted this information and consent form for the researcher's records.**

_____ _____

Participant's signature Date

_____

Participant's printed name

_____

Researcher's signature

Privacy Notice: This is not a government application, the application is controlled and operated by a third party. NASA's Web Privacy Policy does not apply to this application. NASA will not maintain, use, or share Personally Identifiable Information (PII) that becomes available through the use of this third-party application unless expressly stated and consent is obtained from the user. For additional information on NASA's Third-Party Privacy Notice please go to http://www.nasa.gov/about/highlights/HP_Privacy.html.

Paperwork Reduction Act Statement: This information collection meets the requirements of 44 U.S.C. §3507, as amended by section 2 of the Paperwork Reduction Act of 1995. You do not need to answer these questions unless we display a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is 2700-0159 and expires mm/dd/yyyy. We estimate that it will take 15 minutes to read the instructions and answer the questions. Send only comments relating to this time estimate to: richard.l.gilmore@nasa.gov.
~~~~~~

APPENDIX D: Descriptions of Methodological Testing Techniques

- Usability testing: Pertinent are the aspects of the web user interface (UI) that impact the User's experience and the accuracy and reliability of the information Users submit. The ease with which Users navigate the data collection screens and the ease at which the User accesses the actions and functionality available during the data input process are equally important. User experience is also impacted by the look and feel of the web UI and the consistency of aesthetics from page to page, including font type, size, color scheme utilized and the ways in which screen real estate is used (Kota, n.d.). The foundation for Usability testing will be a think-aloud protocol analysis as described by Jääskeläinen (2010) that exposes distractions to accurate input of data whereas a short Likert Scale survey with qualitative questions will determine the extent of distraction and nature of the distractions that impede accurate data input.
- Think-aloud protocols (commonly referred to as cognitive interviewing): This data elicitation method is also called 'concurrent verbalization', meaning subjects are asked to perform a task and to verbalize whatever comes to mind during task performance. The written transcripts of the verbalizations are referred to as think-aloud protocols (TAPs) (Jääskeläinen, 2010, p 371) and constitute the data on the cognitive processes involved in a task (Ericsson & Simon, 1984/1993). When elicited with proper care and instruction, think-aloud does not alter the course or structure of thought processes, except with a slight slowing down of the process. Although high cognitive load can hinder verbalization by occupying all available cognitive resources, that property is of no concern regarding the tasks under analysis that are restricted to information actively processed in working memory (Jääskeläinen, 2010, p. 371). For the purposes of NASA Education, think-aloud protocols will be especially useful towards the improvement of existing and developing of new data collection screens, which are different in purpose from online applications. Whereas an online application is an electronic collection of fields that one either scrolls through or submits, completed page by completed page, data collection screens represent hierarchical layers of interconnected information for which user training is required. Since user training is required for proper navigation, think-aloud protocols capture the user experience to incorporate it into a more user-friendly design and implementation of this kind of technology. Lastly, data from think-aloud protocols is used to ensure that user experiences are reliable and consistent towards collecting robust data.
- Focus group interviews: With groups of nine or less per instrument, this qualitative approach to data collection is a matter of brainstorming to creatively solve remaining problems identified after early usability testing of data collection screen and program application form instruments (Colton & Covert, 2007, p. 37). Data from this type of research will include audiotapes obtained with participant consent, meeting minutes taken

by a subject matter expert in administration assistance, and reflective comments submitted by participants after conclusion of the focus group. Focus group interviews may be used to refine items that failed initial reliability testing for the purposes of retesting. Lastly, focus group interviews may be used with participants as a basis for a grounded theory approach to instrument development or for refining an already existing instrument to be appropriate to a specific audience.

- Comprehensibility testing: Comprehensibility testing of program activity survey instrumentation will determine if items and instructions make sense, are ambiguous, and are understandable by those who will complete them. For example, comprehensibility testing will determine if items are complex, wordy, or incorporate discipline- or culturally-inappropriate language (Colton & Covert, 2007, p. 129).
- Pilot testing: After program activity survey instruments have performed satisfactorily in readability and comprehensibility testing, the next phase is pilot testing with a sample of the target population that will yield statistically significant data, a random sample of at least 200 respondents (Komrey and Bacon, 1992; Reckase, 2000). The goal of pilot testing is to yield preliminary validity and reliability data to determine if items and the instrument are functioning properly (Haladyna, 2004; Wilson, 2005). Data gleaned from pilot testing will be used to fine-tune items and the instrument in preparation for more complex statistical analysis upon large-scale statistical testing.
- Large-scale statistical testing: Instrument testing conducted with a statistically representative sample of responses from a population of interest. In the case of developing scales, large-scale statistical testing provides sufficient data points for exploratory factor analysis (EFA), a multivariate statistical method used to uncover the underlying structure of a relatively large set of variables and is commonly used when developing a scale, a collection of questions used to measure a particular research topic (Fabrigar & Wegener, 2011). EFA is a “large-sample” procedure where generalizable and/or replicable results is a desired outcome (Costello & Osborne, 2005, p.5). This technique is particularly relevant to examining relationships between participant traits and the desired outcomes of NASA Education project activities.
- Item response approach to constructing measures: Foundations for testing that address the importance of item development for validity purposes, address item content to align with cognitive processes of instrument respondents, and that acknowledge guidelines for proper instrument development will be utilized in a systematic and rigorous process. Validity will be determined as arising from item development, from statistical study of item responses, and from exploring item response patterns via methods prescribed by Haladyna (2004) and Wilson (2005.)
- Split-half method: This method for determining test reliability is an efficient

solution to parallel-forms or test/retest methods. Split-half method does not require developing alternate forms of a survey and it places a reduced burden on respondents in comparison to other methods, requiring participation in a single test scenario rather than requiring retesting at a later date. This method involves administering a test to a group of individuals, dividing the test in half along odd and even item numbers, and then correlating scores on one half of the test with scores on the other half of the test (Davidshofer & Murphy, 2005).

APPENDIX E: Privacy Policies and Procedures

- Information collected under the purview of this clearance will be maintained in accordance with the Privacy Act of 1974, the e-Government act of 2002, the Federal Records Act, NPR 7100.1, and as applicable, the Freedom of Information Act in order to protect respondents' privacy and the confidentiality of the data collected.[14]
- Data is maintained on secure NASA servers and protected in accordance with NASA regulations at 14 CFR 1212.605.
- Approved security plans are in place for the Office of Education Performance Measurement (OEPM) system in accordance with the Federal Information Security Management Act of 2002 and Office of Management and Budget, Circular A-130, *Management of Federal Information Resources*.
- Only authorized personnel requiring information in the official discharge of their duties are authorized access to records from workstations within the NASA Intranet or via a secure Virtual Private Network (VPN) connection that requires two-factor hardware token authentication.
- NASA STEM Gateway resides in a certified NASA data center and has met strict requirements relating to application security, network security, and backup/recovery of the NASA Office of the Chief Information Officer's security plan.
- Data will be secured and removed from this server and location upon guidelines set out by the NRRS/1392, 68-69. Specific guidelines relevant to the STEM Gateway system include the following:
    - Project management records documenting basic information about projects and/or opportunities, including basic project descriptions, funding amounts and sources, project managers, and NASA Centers, will be destroyed when 10 years old or when no longer needed, whichever is longer.
    - Records of participants (in any format), maintained either as individual files identified by individual name or number, or in aggregated files of multiple participants identified by name or number, including but not limited to application forms, personal information supplied by the individuals, will be destroyed 5 years after the last activity with the file.
    - Survey responses and other feedback (in any format) from project participants and the general public concerning NASA educational programs, including interest area preferences, participant feedback, and reports of experiences in projects, will be destroyed when 10 years old or when no longer needed, whichever is longer.

[14] http://www.nasa.gov/privacy/nasa_sorn_10EDUA.html

The following Confidentiality Statement and Paperwork Reduction Act (PRA) statement, edited per data collection source, will be posted on all data collection screens and instruments, and will be provided to participants in methodological testing activities per NPR 7100.1:

Privacy Act Statement: In accordance with the Privacy Act of 1974, as amended (5 U.S.C. 552a), you are hereby notified that this study is sponsored by the National Aeronautics and Space Administration (NASA) Office of Education, under authority of the Government Performance and Results Modernization Act (GPRMA) of 2010 that requires quarterly performance assessment of Government programs for purposes of assessing agency performance and improvement. Your participation is important to the success of this study. The information we collect will help us improve the nature of NASA education project activities and the accuracy with which NASA Office of Education can report to the stakeholders about the project activities offered.

Paperwork Reduction Act Statement: This information collection meets the requirements of 44 U.S.C. §3507, as amended by section 2 of the Paperwork Reduction Act of 1995. You do not need to answer these questions unless we display a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is 2700-0159 and expires mm/dd/yyyy. Send comments to: richard.l.gilmore@nasa.gov.

APPENDIX F: Overview: NASA Office of STEM Engagement Data Collection Instrument Development Process

FROM OUTPUTS TO SCIENCE, TECHNOLOGY, ENGINEERING, AND MATHEMATICS (STEM) EDUCATION OUTCOMES MEASUREMENT: DATA COLLECTION INSTRUMENT DEVELOPMENT PROCESS

WORKING WITH THE PROJECT MANAGERS AND PROGRAM DIRECTORS

I. Develop a logic model

a. Information & training sessions to provide guidance
b. Facilitation of logic modeling upon request
c. Review and recommendations to ensure incorporation of evidence-based practice

II. Identify outputs and short-term outcomes from logic models for performance indicators

a. Identify outputs and outcomes across lines of business and projects aligned with CAP goals and FC-STEM investment priority areas
b. Convert outputs and outcomes into performance indicators and outcome measures, identifying required data elements and data collection methods

UNDERSTANDING THE IMPACT OF STEM EDUCATION PROJECT ACTIVITIES ON PARTICIPANTS

III. Develop survey instruments based on NASA STEM Engagement project performance indicators and outcome measures

a. Conduct a scholarly STEM education and measurement literature review (assures that the evidence base is rigorous and current)
b. Connect outcomes from literature review with identified outcome measures, given constraints of inputs and within the context of activities
c. Search the STEM education research and measurement literature for instrument candidates for adaptation (previous literature review augments this step)[15]
d. Create a draft instrument targeting a specific project activity to explore specific outcomes impacted by the quality of outputs (e.g., non-cognitive competencies associated with STEM degree attainment in the NASA Internships and Fellowships)[16]

[15] Provides opportunity to add to the research literature while using an instrument already determined to be reliable and valid for a particular respondent population.
[16] For example, reporting on STEM undergraduate attainment is much less meaningful without understanding what kinds of experiences contributed to degree attainment and the quality of their NASA experience.

i. Draft should be lengthy and exhaustive to allow editing down in the testing process
ii. Draft should reflect many questions that ask the same question to allow editing down
iii. Draft should demonstrate multiple items per construct as convergence is important

e. Obtain stakeholder feedback & edit instrument draft
   i. Editing question type
   ii. Adding new constructs and items
f. Conduct cognitive interviews with a small number (less than 10) of appropriate respondents & edit accordingly[17]
   i. Editing question language
   ii. Editing question type

DEVELOPING VALID AND RELIABLE DATA COLLECTION INSTRUMENTS

IV. Conduct field test of an instrument draft

a. Provide draft to OMB to approve for testing under the NASA OSTEM methodological testing generic clearance (no official timeline associated with this informal process)
b. Small scale field testing[18]
   i. Statistical analysis of responses
   ii. Remove items with low p-values
c. Large scale field testing
   i. Determine population/universe size for respondent audience
   ii. Implement steps to enhance response rate
   iii. Remove items with low p-values

OBTAINING AND MAINTAINING OMB-APPROVED DATA COLLECTION INSTRUMENTS

V. Obtain clearance from OMB for tested data collection instruments

a. Update OMB-approved drafts according to results obtained from large scale field testing
b. Submit tested data collection instrument for review by OMB, in accordance with the terms of clearance set upon approval of the plan as stipulated in the generic clearance.[19]

[17] Involves qualitative research skills and analysis using software NASA Ed has provided for this purpose.
[18] Involves statistical analysis skills and analysis using software NASA Ed has procured.
[19] PRA_Gen_ICRs_5-28-2010.pdf.

Accessed at https://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/PRA_Gen_ICRs_5-28-2010.pdf

VI. Reevaluate instrument function

a. Maintain first universe of collected responses as baseline data
b. On an annual basis, pool recently collected instrument responses with current data set and rerun statistical analyses
c. Take barely passing items back through process starting at III.f.
d. Integrate refreshed items into instrument and forward draft to OMB for approval under the NASA OE methodological testing generic clearance

VII. Reevaluate alignment of data collection instruments

a. Maintain alignment with portfolio as updated
b. Maintain alignment with line of business logic model as updated

APPENDIX G: Explanatory Content for Information Collections for Testing Purposes

Information collection for the purposes of methodological testing will be prefaced by a version of the information categories, ***edited to be appropriate for that particular instrument and audience***. Below are examples of information and a consent form describing the methods for testing that demonstrates the type of information and content that reflects the following (if applicable): 1.) Source of adaptation 2) Constructs of interest; 3) Bibliographic sources that support the particular adaption or instrument draft; 4) Instrument Introduction; 5) Purpose of the study; 6) Description of study procedures; 7) Estimate of time to complete the instrument; 8) Assurance of confidentiality; 9) Contact person's information; 10) Privacy statement; and 11) Paperwork Reduction Act Statement with Office of Management and Budget Control Number.

NASA Intern Survey Methodological Testing

Introduction

This research seeks to support the mission of the NASA Office of STEM Engagement by asking you to take part in a instrument development testing pertaining to our interest in the ways in which the NASA Internship Program impacts outcomes for participants.[20] The information we collect will help us improve the nature of the NASA Internship Program and the accuracy with which NASA Office of STEM Engagement can report on outcomes of the Internship Program.

Purpose of the Study

Test the survey for clarity and comprehensibility and determine the degree to which this instrument accurately captures the ways participant outcomes are measured by this data collection instrument.

Description of Study Procedures

NASA Intern Survey items will be placed into Survey Monkey online software, and a survey link will be distributed through email to 50 NASA Intern participants. Quantitative field tested will be conducted with an anticipated 30-50 NASA Intern participants. Rasch (1960/1980) modeling using the rating scale model (Andrich, 1978) for polytomous responses will be used for the psychometric analysis (described below). As such, this pilot sample size is minimally sufficient for providing approximately 99% confidence in measures within ±1 logit (Linacre, 1994).

Estimation of Time Required

We estimate it will take you an average of 20 minutes to complete the survey.

Securing Your Responses

Under no circumstances will the results of your surveys be shared with anyone without your

explicit permission. The results of this research may be presented at meetings or in publications, however, your identity will not be disclosed. Presentations and manuscripts typically contain participants' quotes, but participants are never identified by name. Your involvement in the development of this instrument is entirely voluntary and you have the right to discontinue participation at any time.

Contact Persons

If you have any additional questions concerning the research, this informed consent, or confidentiality of responses, please contact Richard L. Gilmore Jr., Evaluation Manager, at richard.l.gilmore@nasa.gov or call (216)433-5493.

~~~~~~

I have read and understand the contents of this study information and informed consent form and have been encouraged to ask questions. I have received answers to the questions I have asked. I give my consent to participate freely in this research. I have signed and retained a copy of the information and consent form for my records and future reference. **I have signed and submitted this information and consent form for the researcher's records.**

_____ _____

Participant's signature Date

_____

Participant's printed name

_____

Researcher's signature

Privacy Act Notification:

The information you provide via this form is protected from unauthorized disclosure in accordance with the Privacy Act of 1974. It will be used by NASA for the specific purpose of managing registrants, selecting applicants, implementing and evaluating STEM engagement investments. Collection of the information is authorized by the National Aeronautics and Space Act of 1958 § 403(a)(b), 42 U.S.C. § 2473 (c)(1). Provision of the requested information is strictly voluntary; however, failure to provide the information may result in NASA's inability to provide you with the information or STEM services you desire. NASA may disclose information to NASA administrators and managers, Office of Management and Budget officials, and members of Congress for the purposes of accountability and tracking of program and project efficiency and effectiveness. Elaboration and conditions of information disclosure may be found under "Routine Uses" of the full System of Records Notice at https://www.govinfo.gov/content/pkg/PAI-2013-NASA/xml/PAI-2013-NASA.xml#10euda and in Appendix B at https://www.govinfo.gov/content/pkg/PAI-2013-NASA/xml/PAI-2013-NASA.xml#appb.
~~~~~~

Paperwork Reduction Act Statement:

This information collection meets the requirements of 44 U.S.C. §3507, as amended by section 2 of the Paperwork Reduction Act of 1995. You do not need to answer these questions unless we display a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is 2700-0159 and expires mm/dd/yyyy.

NASA & U.S. Department of Education 21st Century Community Learning Centers Collaboration Phase 4: NASA Engineering Design Challenges

Passive Consent for Student Data Evaluation

Parental Passive Consent

Dear Parent:

Thanks for registering you child(ren) to participate in the NASA an U.S. Department of Education 21st Century Community Learning Collaboration Phase 4 out-of-school Science, Technology, Engineering, and Math (STEM) Engineering Design Challenge program. As part of this program, we will be collecting information in order to evaluate the program's overall goals and objectives: increasing students' interest and motivation in STEM while also working as a team to showcase their skills by participating in an exciting engineering science and space project. The information that we obtain through the evaluation will enable NASA and the U.S. Department of Education to understand how best to deliver the course content in out-of-school environments.

The survey is anonymous and voluntary there will be no identifying information on the surveys your child completes. Your child(ren)'s participation in the course does not depend on answering the questions. Your child does not have to fill out any part of the questionnaire that makes him or her feel uncomfortable or that or that you think your child should not answer. The program will benefit from this survey to the extent that we can identify best practices in the delivery of STEM-related programs that have the greatest chance of increasing students' interest and motivation in STEM classes and careers. If for any reason you do not wish your son or daughter to participate in the evaluation, please sign this form and return it within the first 72 hours of the program.

Student's Name (please print): _____

Parent Signature: _____

Date:_____Time: _____

Site Name: _____

State Name:_____

Privacy Notice: This is not a government application, the application is controlled and operated by a third party. NASA's Web Privacy Policy does not apply to this application. NASA will not maintain, use, or share Personally Identifiable Information (PII) that becomes available through the use of this third-party application unless expressly stated and consent is obtained from the user. For additional information on NASA's Third-Party Privacy Notice please go to http://www.nasa.gov/about/highlights/HP_Privacy.html.

Paperwork Reduction Act Statement: This information collection meets the requirements of 44 U.S.C. §3507, as amended by section 2 of the Paperwork Reduction Act of 1995. You do not need to answer these questions unless we display a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is 2700-0159 and expires 06/30/2021. We estimate that it will take 15 minutes to read the instructions and answer the questions. Send only comments relating to this time estimate to: richard.l.gilmore@nasa.gov.

List of Tables