REU Program Evaluation Instrument

Phase II: Instrument Recommendations

Chris Aberson

Jenna Barry

Humboldt State University

September 25, 2009

Table of Contents

Summary

This report focuses on the development of measures of Self-Efficacy, Applications to Graduate School/Retention in the Major, Attitudes toward Computer Science, Academic Help-Seeking and Coping, and Satisfaction with the REU experience. We briefly review theory and research for relevant measures, and then provide recommendations for the measurement of each construct. For two of the domain areas, self-efficacy and graduate school intentions, we developed measures based on state-of-the-art psychological theory. For attitudes toward computing science and academic coping/help-seeking, we evaluated existing measures and selected the measure that best addressed the construct. Finally, for satisfaction, we developed a common (general) measure and provide a template for addition of site-specific items.

Self-Efficacy

As discussed in the Phase I report, there was considerable agreement that improved research self-efficacy is an important outcome of REU programs. Self-efficacy reflects beliefs in one's capabilities to execute actions required to manage prospective situations. Self-efficacy is an important predictor of choices to pursue various actions, the amount of effort put into those actions, and how long individuals persist in those actions (Bandura, 1997). Many sites made efforts to address self-efficacy, however few of the measures addressed the construct adequately.

Our primary recommendation is to include a) a set of items for use by all sites and b) additional items reflecting the site-specific abilities.

Issues with Measures Presently In Use

One issue with some measures currently used by REU sites is the focus on broad capabilities such "ability to work independently on a project." Greater specificity predicts future behaviors far better than a general questions (see Bandura, 2006). For example, a good self-efficacy measure addressing weight loss would address whether the respondent can regulate the types of foods purchased, can exercise control over eating habits, and can maintain a regular schedule of physical activity. A bad self-efficacy measure would simply ask whether they respondent feels they can lose weight.

Several scales in current use present response options ranging from Strongly Disagree-Somewhat Disagree-No Opinion-Somewhat Agree-Strongly Agree. The inclusion of the "no opinion" option as the center point of the scale makes the scale invalid for most parametric analyses (e.g., t-tests, ANOVA) and damages measurement properties of the scale. Students may perceive "no opinion" as a middle point between disagree and agree, or may take it to mean that they legitimately have no opinion about the issue. Any instrument using this format needs restructuring. The center should read "neither agree nor disagree" or be eliminated.

A final issue is the language used in questions. All items should include "can do" language rather than "will do" statements. Can reflects capability. Will reflects intentions. Self-efficacy involves perceived abilities so "can" correctly taps the construct of interest.

General Self-Efficacy Measure

In developing general items, we borrowed content from Oklahoma State's instrument. Some of their items provide the specificity and range of content items required for an adequate measure. Students respond to 15-20 questions regarding whether students can "locate the primary scientific research literature in your field" and "statistically analyze data." The Oklahoma items informed selection of items (e.g., what specific skills to include) but required revision and expansion to allow for adequate measurement.

The first two questions included in the scale require sites to specify the area of research (e.g., computer science, engineering). Best practice recommendations suggest inclusion of practice items on an unrelated topic. The recommended practice item appears in Appendix A.

Scale Scoring

The overall scale score is a simple sum of the scores on the set of items.

Scale 1: General Self-Efficacy

How certain are you that you could perform each of the following activities **right now**?

Rate your degree of confidence by recording a number from 0 to 100 using the scale below:

0 10 20 30 40 50 60 70 80 90 100

Cannot do at all — Moderately certain can do — Highly certain can do

I can ….	**Confidence (0 to 100)**
1. Locate primary research literature in [specific field] (e.g. journal articles)	
2. Understand primary research literature in [specific field]	
3. Formulate a research hypothesis	
4. Design an experimental test of a hypothesis	
5. Collect data	
6. Statistically analyze data	
7. Interpret data analyses	
8. Reformulate a research hypothesis	
9. Orally communicate the results of research projects	
10. Write a research paper for publication	
11. Work with others to investigate a research problem	
12. Discuss research with other students	
13. Discuss research with graduate students	
14. Discuss research with professors	
15. Discuss research at a professional meeting or conference	

Site-specific Self-Efficacy Measures

As each site provides training in specialized areas, we cannot provide items that address site-specific abilities. However, we provide a template for doing so and examples of good measures. Iowa State's REU includes questions that address site-specific areas of self-efficacy. They ask students about confidence in their abilities to perform activities such as "create a 3D model from 2D side and front views of an object" and "Compile a C++ program." These items are an excellent example of the level of specificity required for a measure of this construct. The sample scale below uses some of the Iowa State items. Each site should aim to adapt this scale to specific skills.

Scale 2: Site-Specific Self-Efficacy Example

How certain are you that you could perform each of the following activities **right now**?

Rate your degree of confidence by recording a number from 0 to 100 using the scale below:

0 10 20 30 40 50 60 70 80 90 100

Cannot do at all — Moderately certain can do — Highly certain can do

I can ….	**Confidence (0 to 100)**
1. Create a 3D model from 2D side and front views of an object	
2. Compile a C++ program that I have written using existing classes	
3. Compile a program that uses OpenGL	
4. Compile a program that uses VR Juggler	

Intentions to Attend Graduate School

The most common approach regarding intention to attend graduate schools utilized in instruments involved addressing student attitudes toward graduate programs. For example, several sites use the Attitudes toward Graduate Studies measure developed by John Carpinelli. Whereas attitudes are an important predictor of future behaviors, there are several other areas of equal importance for measurement. Also, there are more comprehensive ways to address both intentions and factors that influence intentions.

The approach we suggest reflects research on the Theory of Planned Behavior/Reasoned Action (see Ajzen, 2001; Fishbein & Ajzen, 2010). This is widely recognized as one of the most enduring and widely applied behavior prediction theories in psychology. This approach involves measurement of intentions and factors that influence intentions such as attitudes, subjective norms, and perceived behavioral control.

Attitude questions focus on how students evaluate applying to graduate school in terms of items such as whether it is good or bad. Subjective norms refer to issues such as whether people who are important to the student think they should or should not apply to graduate school. Perceived behavioral control items address how strongly students believe they have control over whether or not they apply. Finally, intention measures address the how strongly students believed they will apply to a graduate program.

Although intention to attend graduate school is the construct of primary interest, inclusion of the other variables allows sites to determine how they can improve intentions. For example, a site with low subjective norm scores might increase their focus on enhancing perceptions that others are applying, which in turn should improve intentions to apply to graduate school.

Scale wording could be changed to address different target behaviors such as taking the GRE, asking for letters of recommendation, and so on, if there was a desire to get at behaviors that are more specific. Of course, this would increase questionnaire length considerably.

Scale Scoring

Score items based on each subscale. Subscales are noted with separate letters. A = Attitudes, S = Subjective Norms, P = Perceived Behavioral Control, I = Intentions. An asterisk (*) notes that the item should be reverse coded (e.g., 7 becomes 1) prior to scoring.

Scale 3: Intentions to Apply to Graduate School

1. For me to apply to graduate school is …(A)

Extremely Bad 1 2 3 4 5 6 7 Extremely Good

2. Most people who are important to me think that _____ apply to graduate school (S*)

I Should 1 2 3 4 5 6 7 I Should Not

3. For me to apply to graduate school is (P)

Extremely Difficult 1 2 3 4 5 6 7 Extremely Easy

4. I plan to apply to graduate school (I*)

Extremely Likely 1 2 3 4 5 6 7 Extremely Unlikely

5. For me to apply to graduate school is …(A*)

Extremely Valuable 1 2 3 4 5 6 7 Extremely Worthless

6. It is expected of me that I will apply to graduate schoo1(S)

Definitely False 1 2 3 4 5 6 7 Definitely True

7. Whether or not I apply to graduate school is completely up to me (P*)

Strongly Agree 1 2 3 4 5 6 7 Strongly Disagree

8. I will make an effort to apply to graduate school (I)

I Definitely Will Not 1 2 3 4 5 6 7 I Definitely Will

9. For me to apply to graduate school is …(A)

Extremely Harmful 1 2 3 4 5 6 7 Extremely Beneficial

10. Most of the students in the [*fill in name of program*] with whom I am acquainted plan to apply to graduate schoo1 (S*)

Definitely True 1 2 3 4 5 6 7 Definitely False

11. I am confident that if I wanted to I could apply to graduate schoo1 (P)

Definitely False 1 2 3 4 5 6 7 Definitely True

12. I intend to attend to apply to graduate school (I*)

Strongly Agree 1 2 3 4 5 6 7 Strongly Disagree

13. For me to apply to graduate school is …(A*)

Extremely Pleasant 1 2 3 4 5 6 7 Extremely Unpleasant

14. Most people whose opinions I value would approve of applying to graduate school (S)

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

15. For me to apply to graduate school is (P*)

Possible 1 2 3 4 5 6 7 Impossible

Computer Science Attitudes

In examining computer science attitudes, we examined both general measures of science attitudes and some measures that addressed particular science domains. We evaluated scales addressing general attitudes toward science with an eye toward adapting measures to CS and reviewed scales specifically addressing computer science attitudes. Our recommendation reflects the instrument that best combined strong psychometric properties and length (i.e., we selected the shortest instrument that demonstrated good measurement properties).

The measure we recommend focuses on students' attitudes and perceptions toward computing science (Hoegh & Moskal, 2009). The scale includes several subscales addressing confidence in CS abilities, interest in CS, gender equality in CS, usefulness of CS to career goals, and perceptions of computer scientists' personal lives. The subscales demonstrate strong psychometric properties (Cronbach's alpha ranging from .77 to .93). This scale seems relevant to the REU undergraduate experience as many of the items included in the scale overlap questions found in existing instruments. This measure contains only 38 items, compared to the 50-plus items included in the other scales reviewed.

Scale Scoring

Score items based on each subscale. Subscales are noted with separate letters. C = confidence; I = interest; G = gender; U = usefulness; P = personal. An asterisk (*) notes that the item should be reverse coded (e.g., 4 becomes 1) prior to scoring.

Scale 4: Attitude toward Computer Science

For each item, please respond using the following scale:

4 = Strongly Agree 3 = Somewhat Agree 2 = Somewhat Disagree 1 = Strongly Disagree

	Agreement (0 to 100)
1. I am comfortable with learning computing concepts. (C)	
2. I would not take additional computer science courses if I were given the opportunity. (I*)	
3. Computing is an appropriate subject for both men and women to study. (G)	
4. Developing computing skills will not play a role in helping me achieve my career goals. (U*)	
5. I have little self-confidence when it comes to computing courses. (C*)	
6. I hope that my future career will require the use of computer science concepts. (I)	
7. I doubt that a woman could excel in computing courses. (G*)	
8. Students who are skilled at computer science are just as popular as other students. (P)	
9. I can learn to understand computing concepts. (C)	
10. I think computer science is boring. (I*)	
11. Women and men can both excel in careers that involve computing. (G)	
12. My career goals do not require that I learn computing skills. (U*)	
13. I do not think that I can learn to understand computing concepts. (C*)	
14. I like to use computer science to solve problems. (I)	
15. It is not appropriate for women to study computing. (G*)	
16. A student who performs well in computer science is likely to have a life outside of computers. (P)	
17. I can achieve good grades (C or better) in computing courses. (C)	
18. The challenge of solving problems using computer science does not appeal to me. (I*)	

19. Women produce the same quality work in computing as men. (G)	
20. Knowledge of computing skills will not help me secure a good job. (U*)	
21. I am not comfortable with learning computing concepts. (C*)	
22. I think computer science is interesting. (I)	
23. Men are more likely to excel in careers that involve computing than women are. (G*)	
24. I expect that learning to use computing skills will help me achieve my career goals. (U)	
25. I am confident that I can solve problems by using computer applications. (C)	
26. I hope that I can find a career that does not require the use of computer science concepts. (I*)	
27. Men and women are equally capable of solving computing problems. (G)	
28. A student who performs well in computer science will probably not have a life outside of computers. (P*)	
29. I doubt that I can solve problems by using computer applications. (C*)	
30. The challenge of solving problems using computer science appeals to me. (I)	
31. Men are more capable than women at solving computing problems. (G*)	
32. Knowledge of computing will allow me to secure a good job. (U)	
33. I do not like using computer science to solve problems. (I*)	
34. Men and women can both excel in computing courses. (G)	
35. Students who are skilled at computer science are less popular than other students. (P*)	
36. I would voluntarily take additional computer science courses if I were given the opportunity. (I)	
37. Men produce higher quality work in computing than women. (G*)	
38. Developing computing skills will be important to my career goals. (U)	

Note. C = confidence; I = interest; G = gender; U = usefulness; P = personal.

Academic Help Seeking and Coping Behavior

In examining academic help seeking and coping, we reviewed measures covering a range of coping and help-seeking responses. There are many psychometrically sound measures of coping and help seeking, however most are very long (50-plus items). Our recommendation is the 30-item Student Coping Scale (SCOPE; Struthers et al., 2000). The SCOPE assess students' thoughts, behaviors, and action strategies (e.g., coping styles) that associate with poor academic achievements. The scales include two subscales. One addressing problem-based coping and the other emotion focused coping. The subscales demonstrate adequate reliability (Cronbach's alpha of .70 to .80) and good validity as scores associate with relevant outcome such optimism (positive relationship and stress (negative relationship).There are stronger measures, but all are considerably longer.

The recommended scale addresses how students would respond to a poor exam grade. We believe it would be possible to modify the scale to address a different area such as a problem with a mentor. However, as the scale authors argue, response to a poor exam performance is a good indicator of general academic coping.

Scale Scoring

Score items based on each subscale. Subscales are noted with separate letters. P = Problem-Focused Coping, E = Emotion-Focused Coping. Scores are a simple sum of each subscale.

Scale 5: Help-seeking and Coping

The following scale addresses how you would react if you performed poorly on an exam. For each item, please respond using the following scale:

1 2 3 4 5 6 7 8 9 10

Extremely Uncharacteristic of Me — Extremely Characteristic of Me

When I do poorly on an important exam, typically ….	**Rating (1 to 10)**
1. I think about how I might best handle the problem (P)	
2. I act as though it hasn't happened (E)	
3. I feel competent (P)	
4. I skip class (E)	
5. I try to get emotional support from friends and family (E)	
6. I do what has to be done one step at a time (P)	
7. I let my feelings out (E)	
8. I buy a study guide (P)	
9. I make a plan of action (P)	
10. I refuse to believe that it happened (E)	
11. I feel confident (P)	
12. I reduce the amount of effort I put in to solving the problem (E)	
13. I discuss my feelings with someone (E)	
14. I think about the reason(s) why the situation occurred (P)	
15. I feel a lot of emotional distress and I find myself expressing those feelings (E)	
16. I use my study guide (P)	

17. I try to come up with a strategy about what to do (P)	
18. I say to myself ‘this isn’t real’ (E)	
19. I feel hopeful (P)	
20. I drop out of the class I’m doing poorly in (E)	
21. I talk to someone about how I feel (E)	
22. I concentrate my efforts on doing something about it (P)	
23. I get upset and let my emotions out (E)	
24. I try a different study technique (P)	
25. I think hard about what steps to take (P)	
26. I pretend that it hasn’t really happened (E)	
27. I feel motivated (P)	
28. I give up trying to reach my goal (E)	
29. I take additional action to try to get rid of the problem (P)	
30. I get upset and am really aware of it (E)	

Note. Subscales are P = Problem-Focused Coping, E = Emotion-Focused Coping

Satisfaction with Experience

Several existing instruments included strong measures of satisfaction with the REU. We provide a general measure that includes space for sites to add specific items. The materials below borrow from the UMASS student questionnaire.

Scale 6: Satisfaction with Experience

How satisfied are you with the following aspects of your REU experience?

Rate your satisfaction by recording a number from 0 to 100 using the scale below:

0 10 20 30 40 50 60 70 80 90 100

Extremely Dissatisfied — Extremely Satisfied

How satisfied were you with…	**Satisfaction (0 to 100)**
1. your faculty advisor?	
2. your housing arrangements?	
3. the program in general?	
4. your research experiences?	
5. your interaction with project staff?	
6. your interaction with other students?	
(Sites should add items here to address specific aspects of their program)	

General Comments on Measures

The measures included in this report are designed for "as is" use. Reducing the length of items through deletion of similar questions and other procedures destroys the instrument's psychometric properties. Additionally, it will not be possible to draw clear conclusions across sites if each site uses a different version of the instrument.

Analysis of Items

Data analyses should focus on overall scale scores or subscale scores rather than on individual items. For example, analysis of pre-post changes in the computer science attitudes might compare pre and post scores on the confidence, interest, gender, usefulness of CS, and personal life subscales. Analyses should not address changes on an item-by-item basis (e.g., 38 pre-post comparisons).

Shortening Scales

We recognize that the entire set of scales (particularly attitudes toward science and coping/help seeking) represent more items than desired. As noted above, simple deletion of items is not a sound strategy for reducing length. There are two empirically valid options for scale shortening. First, if there were agreement that only certain subscales are of value, then it would be appropriate to eliminate other subscales. For example, the attitudes toward computing science measure includes an 8-item confidence subscale (e.g. "I can learn to understand computing concepts") that overlaps the self-efficacy measures addressed earlier. It would be reasonable to remove these items. However, subscales removal makes overall scale scores meaningless, so data analyses should address the remaining subscales only.

Another option is empirical examination of scale data after administration. Several techniques exist to identify sets of variables from a scale that provide roughly the same quality of information. If we received data from sites in a common format, we could complete these analyses and recommendations with relatively little time commitment.

References

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*, 27-58.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Guide to the construction of self-efficacy scales. In F. Pajares & T. Urdan (Eds.). *Self-efficacy beliefs of adolescents*, (Vol. 5., pp. 307-337). Greenwich,CT: Information Age Publishing.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach.* New York: Psychology Press.

Hoegh, A., & Moskal, B. M. (October, 2009). Examining science and engineering students' attitudes toward computer science. Paper to be presented at 39th Frontiers in Education Conference, San Antonio, Texas.

Struthers, C. W., Perry, R. P., & Menec, V. H. (2000). An examination of the relationship among academic stress, coping, motivation, and performance in college. *Research in Higher Education, 41*, 581-593.

Appendix A: Practice Self-Efficacy Item

Practice Rating

To familiarize yourself with the rating form, please complete this practice item first. If you were asked to lift objects of different weights **right now**, how certain are you that you can lift each of the weights described below?

Rate your degree of confidence by recording a number from 0 to 100 using the scale below:

0 10 20 30 40 50 60 70 80 90 100

Cannot do at all — Moderately certain can do — Highly certain can do

I can …	**Confidence (0-100)**
Lift a **10** pound object	
Lift a **20** pound object	
Lift a **50** pound object	
Lift a **80** pound object	
Lift a **100** pound object	
Lift a **150** pound object	
Lift a **200** pound object	
Lift a **300** pound object	