CUR Quarterly
COUNCIL ON UNDERGRADUATE RESEARCH

Spring 2012 | Volume 32, Number 3

The Challenge of the Count

www.cur.org

Cover: Rachel Wright, Fine Arts and Philosophy, experimented with plein air painting style and techniques pioneered by some of the Hudson River school painters and other European schools. Rachel participated in Buffalo State's summer research program in 2010 and was mentored by Lin Xia Jiang, Professor of Fine Arts. photo credit: Used with permission from The Buffalo News (photographer: Derek Gee)

Jill Singer, Bridget Zimmerman, *SUNY-Buffalo State*

Evaluating a Summer Undergraduate Research Program: Measuring Student Outcomes and Program Impact

At SUNY-Buffalo State we have undertaken a multi-year effort aimed at developing and field-testing an evaluation methodology for measuring student learning and related outcomes in our summer undergraduate research program. Our goal was to extend the findings of the many valuable studies that had already been done on the impact of undergraduate research on participating students (Gregerman *et al*. 1998; Alexander *et al*. 2000; Merkel 2001; Bauer and Bennett 2003; Seymour *et al*. 2004; Lopatto 2004; and Hunter *et al*. 2007). As described in Singer and Weiler (2009), our aim was to obtain reliable independent assessments of program impact without creating a measurement burden, and at the same time provide information to participating students that could help them gain new insights into their academic strengths and weaknesses. We also were interested in obtaining information from faculty mentors on how the summer undergraduate research program influenced their teaching.

The difficulty in achieving our goal originates mainly from the fact that the outcomes of undergraduate research generally are not measurable by objective tests. Rather, they encompass independent student projects whose results often are a written paper, a new work of art, a laboratory or field experience, or a range of other experiences. As a result, outcome assessments largely rely on perceptions and judgments by faculty mentors. These assessments are subject to challenge on the grounds that faculty may have had inadequate exposure to students' work, that assessment parameters rarely are made explicit, and/or that assessments are not comparable across different disciplines or across different faculty and student participants. We engaged in a multi-year effort to design an evaluation that could overcome these concerns. It is worth noting that it is not ordinarily practical to conduct evaluations of undergraduate research that rely on classic comparison-group designs, given the difficulties of matching students who have not applied or not been accepted into a program with "comparison" students, as well as the difficulty of asking faculty mentors to assess "comparison" students without being able to meet with and observe them.

Our strategy for developing the evaluation began with a two-day retreat during the summer of 2006. Faculty representing a range of disciplines (including arts, humanities, social sciences, and sciences) identified 11 student outcome categories of interest. Each outcome category included lists of specific outcome components (see Table 1). In addition to the 11 categories, it was decided that mentors and students would be able to add more outcomes as appropriate to their specific research interests.

Next we discussed and designed an assessment procedure that uses a five-point scale linked to an explanatory rubric to denote that a student always (5), usually (4), often (3), seldom (2), or never (1) displays a given outcome for each component in the 11 outcome categories. Faculty mentors rate students on each component, and students evaluate their own progress using an identical instrument. As a result of these and other decisions about the design of the evaluation, described more fully below, the evaluation now has six essential features:

- Repeated assessments (pre-research, mid-research, and end-of-research);
- Assessments in which faculty mentors and students all use the same outcome categories and components;
- A scoring rubric, used by all mentors and students, that defines the meaning of each assessment score on the five-point scale;
- A "confidence" judgment in which mentors are asked to indicate their level of confidence in the accuracy of each of their assessment scores, using a five-point scale ranging from "very confident" to "not confident at all";
- Student self-assessments and mentors' assessments of students, performed independently; and
- Student-mentor discussions to compare their independent assessments following each of the three assessment periods.

The first four of these features are designed to overcome the concerns about faculty assessments summarized above by ensuring that (1) faculty mentors have multiple opportunities to familiarize themselves with student work; (2) assessments are conducted according to standards that are explicit and uniform across disciplines and across different student-faculty pairs; and (3) assessments are weighted to reflect the amount and quality of information underlying the scores. The last two evaluation features outlined above are designed to provide opportunities for students to improve their understanding of their academic strengths and weaknesses.

During the summer of 2007 we conducted a pilot implementation of the evaluation methodology that included both first-time and experienced mentors familiar with the summer research program. At the end of the summer, focus groups composed of student researchers and faculty mentors provided feedback on the evaluation methodology and on the clarity of the overall process. Based on this feedback, modifications were made to the evaluation instruments and the overall process was simplified to help mentors and students better understand the sequence of steps involved

Table 1: Evaluation Outcome Categories and Components

Communication
Uses and understands professional and discipline-specific language.
Expresses ideas in an organized, clear, concise, and accurate manner.
Writes clearly and effectively in discipline-specific formats.
Creativity
Brings new insights to the problem at hand.
Shows ability to approach problems from different perspectives.
Combines information in new ways and/or demonstrates intellectual resourcefulness.
Effectively connects multiple ideas/approaches.
Autonomy
Demonstrates the ability to work independently and identify when input, guidance, and feedback are needed.
Accepts constructive criticism and applies feedback effectively.
Displays high level of confidence in ability to meet challenges.
Uses time well to ensure work gets accomplished and meets deadlines.
Ability To Deal With Obstacles
Learns from and is not discouraged by setbacks and unforeseen events.
Shows flexibility and a willingness to take risks and try again.
Practice And Process Of Inquiry
Demonstrates ability to formulate questions and hypotheses within the discipline.
Demonstrates ability to properly identify and/or generate reliable data.
Shows understanding of how knowledge is generated, validated, and communicated within the discipline.
Nature Of Disciplinary Knowledge
Shows understanding of the way practitioners think within the discipline (e.g., as an earth scientist, sociologist, or artist) and view the world around them.
Shows understanding of the criteria for determining what is valued as a contribution to the discipline.
Shows understanding of important current individuals within the discipline.
Critical Thinking And Problem Solving
Trouble-shoots problems, searches for ways to do things more effectively, and generates, evaluates, and selects between alternatives.
Recognizes discipline-specific problems and challenges established thinking when appropriate.
Recognizes flaws, assumptions, and missing elements in arguments.
Understanding Ethical Conduct
Shows understanding and respect for intellectual property rights.
Predicts, recognizes, and weighs the risks and benefits of the project for others.
Recognizes the severity of creating, modifying, misrepresenting, or misreporting data, including omission or elimination of data/findings or authorship.
Intellectual Development
Demonstrates growth from basic to more complex thinking in the discipline.
Recognizes that problems are often more complicated than they first appear to be and the most economical solution is usually preferred over convoluted explanations.
Approaches problems from a perspective that there can be more than one right explanation or model or even none at all.
Displays accurate insight into the extent of his/her own knowledge and understanding and an appreciation for what isn't known.
Culture Of Scholarship
Is involved in the scholarly community of the discipline and/or professional societies.
Behaves with a high level of collegiality and ethical responsibility.
Content Knowledge Skills/Methodology
Displays detailed and accurate knowledge of key facts and concepts.
Displays a thorough grasp of relevant research methods and is clear about how these methods apply to the research project being undertaken.
Demonstrates an advanced level of requisite skills.

in completing the various instruments. For a more complete description of the evaluation-development process and methodology, see Singer and Weiler (2009).

A full-scale implementation of the evaluation has now been conducted with three groups of student researchers and their faculty mentors. This article reports on evaluation findings for the period 2008 to 2010. Static versions of all instruments referred to in this paper and data tables supporting our findings can be found at: http://www.buffalostate.edu/undergraduateresearch/x561.xml.

Implementation of the Evaluation Methodology

Evaluation Stages

Table 2 shows the summer research program and its evaluation divided into three stages: Pre-to-Early Research, Mid-Research, and End-to-Post Research.

Pre- to Early Research

The summer research program starts with the student-faculty mentor teams attending a two-hour group orientation session. The orientation includes a thorough explanation of the evaluation that emphasizes its dual purpose of assessing program outcomes and providing a means for students to learn more about their academic strengths and weaknesses. Following the orientation session, students complete an online survey that asks them about their motivation for participating in the undergraduate summer research program, their knowledge and expectations, and their current understanding of their academic strengths and weaknesses. The survey provides students with a structured opportunity to explore their goals, knowledge, and readiness for the summer research program, and it provides mentors with insights into student knowledge and thinking as an aid to completing their pre-research assessments.

After the mentors review students' survey responses, students and mentors meet in order to give mentors an opportunity to query students on their survey responses and to formulate preliminary student assessments. The mentors' version of the survey shows the relationship between each survey question and relevant student outcomes, as delineated in the assessment categories and components. Following this meeting, the students and mentors each independently complete the pre-research assessment. They then meet to discuss why their scores on outcome components of interest were the same or different. These procedures provide both students and mentors with an environment in which they can be unbiased in their scoring and open and forthcoming in discussing their rationale for assigning a particular score. Following completion of the pre-research assessment and student-mentor meetings, the students begin their research projects.

Mid-Research

To help students and mentors keep track of their experiences and progress, we encourage them to keep a journal. An electronic form (with access restricted to user only) is available. In the absence of a formal journal (electronic or paper), we recommend that students and mentors keep track of the experience by noting times when particular obstacles were encountered or when a particular accomplishment was achieved. About halfway through the summer, students and mentors complete a short report that responds to several questions about research progress, changes from original research plans, and plans for the second half of the summer. In addition to answering these questions, students and mentors each complete the mid-research assessment and meet to compare their scores on each outcome component and discuss scores that changed from their initial assessments. A feature of the online assessment allows students and mentors to review their scores from the pre-research assessment and asks them to reflect on the reason(s) for any score changes. As with the pre-research assessment, students and mentors cannot directly review one another's assessment scores before meeting to discuss their respective assessments. Student research then continues for the remainder of the summer.

End- to Post-Research

At the conclusion of the summer research program, the students and mentors complete a final report. This is longer than the mid-summer progress report and students provide a short (3-to-4 page) report on their project, including their methods/approach, findings, and suggestions for places where they might present their work. The report is uploaded as a document, and often includes figures, data tables, and illustrations. Mentors are asked about their project experiences and the extent to which the program has helped them reconsider their approach to classroom teaching. Both students and mentors complete the post-research assessment and meet one final time to discuss how they each scored the outcome components. As before, students and mentors can review their pre- and mid-research assessment scores but cannot see each other's scores prior to their final meeting.

Evaluation Modifications

Based on interim findings following our experiences in 2009, we made four significant modifications to the evaluation:

Orientation was improved. To ensure that program participants followed each step of the evaluation in the proper order and at the appropriate time in the research experience, clearer instructions were provided to better prepare students and mentors to follow the sequence of steps (summarized in Table 2).

Student confidence scores were dropped. Students' responses in 2008 and 2009 supported the elimination of the "confidence" score from all three stages of students' self-assessments. The confidence scores were replaced by a single question at the end of the assessment that asks students how certain they are about their skill levels and responses.

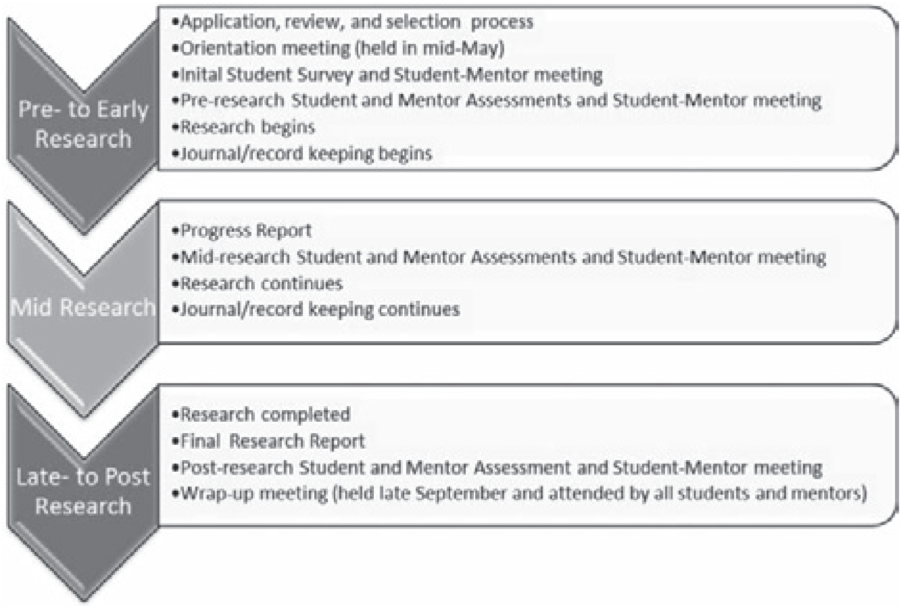
One mentor confidence score was dropped. Mentors' responses in 2008 and 2009 supported the elimination of "confidence" scores from the mid-research assessments. The measure is retained in both the pre- and post-research assessments.

The format of the instrument was improved. The assessment instrument was simplified so that additional optional outcomes now are identified at the end of the instrument rather than after each outcome category.

Analysis of Assessment Instruments

In order to ensure that we had obtained a realistic measure of program outcomes, we began by making certain that our assessment instruments provided reliable and valid impact measures. First, Cronbach's alpha was used to determine

Table 2. Stages of Summer Research Program and Evaluation



the strength of association of the individual assessment components. This measure was calculated for the 34 component questions for 2008 (n=17), 2009 (n=20), and 2010 (n=24) data, to determine how well the items hold together as the scales currently are defined. The mentor assessment (N= 61, three years combined) had an overall coefficient alpha of .964. The student self-assessment (N=61, three years combined) had an overall coefficient of .951. These alpha coefficients are evidence that the measured items have high internal consistency. The high coefficient scores suggest that the categories and components represent the multiple facets of one holistic construct—student intellectual and professional growth.

We also were sensitive to the overall length of the assessments and wanted to eliminate possibly redundant and non-discriminating items in order to reduce response fatigue for mentors and students. To this end, we completed a factor analysis/principal components analysis for all three years combined (n=61) for both the mentors and the students. This was done to ascertain whether all the questions added value to defining the summer research program's impact on the participants. The 34 items were highly correlated with each other. The analysis suggested a minor reduction in the number of questions, but there were differences between the mentor and student assessments regarding the particular questions that could be eliminated.

Given the lack of persuasive statistical evidence to support a reduction in the length of the assessments and our desire to keep the mentor and student versions of the assessments the same, only minor modifications in format were made, as noted above. This decision was further supported by two central characteristics of the evaluation: (1) The assessment instrument's contents pertained to issues that faculty had identified during the program's initial retreat as important potential outcomes of students' education and research experience, and (2) the assessment categories and components were being used strategically (not just as an exercise in psychometrics) in the student-mentor discussions to help students understand their own strengths and weaknesses and thus help them grow academically and think professionally.

Findings

In order to understand the impact of the summer research program, we first applied a repeated measures analysis of variance (ANOVA) test to student and mentor pre-, mid-, and post-research assessments (N =61, 2008–2010). Mentors' confidence scores were analyzed with a paired samples T-test, as there were only pre- and post-research assessment confidence measures.

We found that the mentors' largest adjustments in appraisals of their students' abilities—with the strongest confidence levels—took place between the pre- and mid-research assessments, with few additional assessment differences and little additional growth in confidence between the mid- and post-research assessments. This finding is consistent with the expectation that the greatest gain in mentors' understanding of students' abilities would likely take place within the first half of the summer research program as the mentors began to work with the students, and that additional gains in understanding would be minimal because most of their knowledge about the students had already been gained by mid-summer.

In their pre-research assessments, student self-ratings were on average somewhat higher than mentor ratings of their students on the same outcome components. This finding suggests that many students at first over-estimated their academic strengths. This is confirmed by many students' comments on the post-research assessment form, where they wrote that they thought they knew "a lot" at the outset of their research experience, but that by the time they concluded their research, they realized how much they didn't know. One student wrote:

> *"I was glad to see that there was more than one assessment survey given over time. In this way, people can see the change (for better or worse) they underwent. We (mentor and student) found that I had improved in some areas but declined in others because I over-scored myself in the beginning."*

Table 3. Pre-Post Mean Scores for Students and Mentors

Scale: Always (5); Usually (4); Often (3); Seldom (2); and Never (1)

Outcome Components	Student			Mentor		
	Pre	Post	Significance	Pre	Post	Significance
Uses and understands professional and discipline-specific language	3.87	4.19	***	3.92	4.28	***
Expresses ideas in an organized, clear, concise, and accurate manner	3.80	4.14	**	3.93	4.23	**
Writes clearly and effectively in discipline-specific formats	3.77	4.13	**	3.77	3.70	
Brings new insights to the problem at hand	3.70	4.13	***	3.83	4.26	***
Shows ability to approach problems from different perspectives	3.82	4.22	**	3.92	4.25	***
Combines information in new ways and/or demonstrates intellectual resourcefulness	3.89	4.18	*	3.95	4.28	**
Effectively connects multiple ideas/approaches	3.87	4.03		3.88	4.30	***
Demonstrates the ability to work independently and identify when input, guidance and feedback are needed	4.31	4.53	*	4.32	4.54	**
Accepts constructive criticism and applies feedback effectively	4.36	4.60	*	4.35	4.62	**
Displays high level of confidence in ability to meet challenges	4.10	4.22		4.23	4.34	
Uses time well to ensure work gets accomplished and meets deadlines	4.08	4.05		4.27	4.23	
Learns from and is not discouraged by set-backs and unforeseen events	4.13	4.25		3.92	4.41	***
Shows flexibility and a willingness to take risks and try again	4.28	4.35		4.14	4.61	***
Demonstrates ability to formulate questions and hypotheses within discipline	3.77	3.95	***	3.71	4.07	*
Demonstrates ability to properly identify and/or generate reliable data	3.43	3.97	**	3.10	4.05	***
Shows understanding of how knowledge is generated, validated and communicated within the discipline	3.72	4.10	*	3.77	4.07	
Shows understanding of the way practitioners think within the discipline and view the world around them	3.90	4.15		3.85	4.21	**
Shows understanding of the criteria for determining what is valued as a contribution in the discipline	3.95	4.07		3.78	4.10	**
Shows understanding of important current individuals within the discipline	3.38	3.63		3.30	3.82	***
Trouble-shoots problems, searches for ways to do things more effectively, and generates, evaluates and selects between alternatives	3.85	4.20	*	4.02	4.30	**
Recognizes discipline-specific problems and challenges established thinking when appropriate	3.69	3.92		3.45	3.84	**
Recognizes flaws, assumptions, and missing elements in arguments	3.79	3.73		3.52	3.82	**
Shows understanding and respect for intellectual property rights	4.62	4.67		4.12	4.44	**
Predicts, recognizes, and weighs the risks and benefits of the project for others	4.02	4.27		3.44	3.74	
Recognizes the severity of creating, modifying, misrepresenting, or misreporting data including omission or elimination of data/findings or authorship	4.48	4.62		3.97	4.48	
Demonstrates growth from basic to more complex thinking in the discipline	3.85	4.13		4.17	4.43	*
Recognizes problems are often more complicated than they first appear to be and the most economical solution is usually preferred over convoluted explanations	3.82	3.93		3.75	4.10	*
Approaches problems from a perspective that there can be more than one right explanation or model or even none at all	3.98	4.22		4.02	4.15	
Displays accurate insight into the extent of his/her own knowledge and understanding and an appreciation for what isn't known	4.03	4.17		4.08	4.35	*
Is involved in the scholarly community of the discipline and/or professional societies	3.39	3.67		3.23	3.35	
Behaves with a high level of collegiality and ethical responsibility	4.41	4.50		4.65	4.74	
Displays detailed and accurate knowledge of key facts and concepts	3.80	4.02		3.90	4.34	***
Displays a thorough grasp of relevant research methods and is clear about how these methods apply to the research project being undertaken	3.38	3.98	***	3.67	4.16	**
Demonstrates an advanced level of requisite skills	3.54	3.78		4.02	4.23	

** p <.05, ** p <.01, *** p <.001*

Another student reported:

> *"I thought I was good at research until I actually started doing it and then I realized how little I knew."*

One mentor captured this realization with the comment:

> *"Every student is different and one thing I believe I learned this summer was that very good students assess themselves more poorly than I do and moderately good students with a more naive approach assess themselves better than I do. I didn't really anticipate this and found it interesting and something I need to take into account while mentoring."*

Students reported growth on all 34 outcome components from pre- to mid-research assessment and again from mid- to post-research assessment. However, actual differences between student pre- and post-research assessments were statistically significant less frequently than were the comparable mentors' assessments. Student self-assessment scores showing pre- to post-research academic growth on 13 of 34 assessment components were statistically significant at $p < .05$ or better, which is strong evidence of knowledge growth on the 13 items. Student's open-ended comments also focused on the impact of the program on gains in their knowledge, their contribution to the discipline, value for future endeavors (e.g., applying to graduate school, and listing the research program on their resume), and knowledge gained above and beyond the classroom setting.

Table 3 shows average student self-assessment and mentor pre-research and post-research scores for each of the 34 outcome components. Pre- to post-scores shown in asterisks were statistically significant at $p < .05$ or better.

As noted above, mentors tended initially to rate the students lower than the students rated themselves, but by the end of the program, the mentors' assessment scores of students were, on average, higher than the students' self-assessment scores. Twenty-four of 34 items showing these higher ratings were statistically significant at $p < .05$ or better, providing strong evidence that the mentors saw growth in student knowledge on the 24 items. Two items with decreased scores ("writes well" and "is involved in the community") suggest that the mentors may have initially over-rated their students on these outcomes and, after more experience with the students, adjusted for this by lowering the scores. On the other hand, the decreased scores on these items could have masked any improvement the students actually may have made in these areas. Mentors' confidence in assigning scores from pre-research to post-research assessments were statistically significant at $p < .05$ or better for 33 of 34 items.

In their responses to open-ended questions on the assessment forms, a number of students wrote about the value of the program's emphasis on blending assessment and educational goals (as described at length above). Students said, for example:

About the orientation:

> *"It helped me to realize the nature of research and that things don't always turn out as you planned. This made me more open to learning new things and making more connections between ideas."*

> *"It prepared me for the amount of work this really is, as well as getting me even more motivated seeing how prestigious this program is and the standards you are held to."*

About the assessment:

> *"It was useful in that it allowed me to compare my own perception of my strengths and weaknesses to my mentor's interpretation of the same."*

> *"I was able to see if I was improving /declining in any areas and just gave me the ability to assess myself and set a new goal of where I wanted to be by the end of the summer."*

> *"Meeting to review our responses allowed me to understand another perspective. It also allowed me to view weak areas that I needed to research and improve upon."*

About the journals:

> *"During the summer I kept a full journal of my notes, research and plans to develop a final installation piece incorporating historical paper cutting, and the integration of industrial processes. This journal was very important and still is very important to my future goals and plans for graduate school."*

> *"I did keep a very detailed journal about the process I went through this summer. ... It really allowed me to organize my thoughts and keep track of exactly what I had done already and what still needed to be completed."*

Comments made by mentors on their assessment forms often focused on the growth in their confidence about students' skills, abilities and limitations, and on the value of a collegial working relationship with their students. On the latter point, for example, one mentor noted:

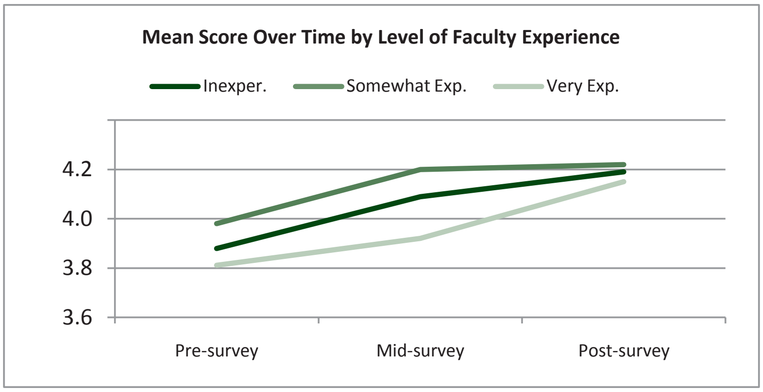
> *"We were able to get a better grasp on what we thought of each other and the project. The questions were not discipline-specific, so they were useful in getting to know the student's personality better and also the student's abilities, likes and dislikes, and aptitudes."*

Mentors also reflected on how the program influenced their teaching practices. For example:

> *"This experience reinforces the fact that each student is an individual and that one size does not fit all"*

> *"This year I had to mentor at a higher level because my student's knowledge base has grown. I had to learn to*

Table 4. Mean Score Over Time by Level of Faculty Experience



take less of a role and really stand by and monitor and advise during all phases of the research. This was difficult at times because I am used to being more hands on with students because they typically need more help."

"I think that I learned more about assessment/evaluation of student progress—especially in a student who is very desirous of doing well but doesn't necessarily have the tools yet to do so. That is, I think I am better able to pinpoint weaknesses and address them more quickly and effectively."

Students and mentors both commented, as well, on the value of the pre-research survey and follow-up conversation between student and mentor. Many reported that even when a student and mentor had worked together before and the mentor knew the student fairly well, the survey revealed information that was very helpful in establishing a "starting point" for the balance of their collaboration.

Other Findings

The evaluation data were analyzed to ascertain whether or not either mentor experience or academic discipline contributed to differential assessment of student outcomes. First, we coded the mentors as belonging to one of three experience categories: relatively inexperienced (less than three years mentoring undergraduates who are conducting research, n=25), somewhat experienced (between three and six years of mentoring, n=25) and very experienced (more than six years of mentoring, n=10). The data for the period 2008 to 2010 showed little difference either in the scores or patterns of scoring among inexperienced, somewhat experienced, and very experienced mentors. Very experienced mentors tended to be more conservative in their scoring on all three assessments than did inexperienced or somewhat experienced mentors. Taken as a whole, however, the mean scores from all mentor experience levels increased over time. Overall mean scores for all 34 items over time, coded by mentor experience, illustrate these trends, none of which were statistically significant. Mean differences were small. There also were no statistically meaningful differences in how the students of inexperienced, somewhat experienced, and very experienced mentors scored themselves on any of the 34 outcome components.

It is possible that differences in mentoring experience would ordinarily have shown up as marked differences in mentors' assessments of their students, although there is no way to be sure of this. If so, the structure of the summer research program evaluation may have helped to compensate for such experience effects. The program's monitoring of student and mentor progress, combined with regular communications with mentors by one of us (Singer), seemed to have helped ensure that all mentors, regardless of their level of experience, were able to implement the evaluation instruments effectively and use them to provide meaningful feedback to their students. This observation is based on personal communications between Singer and the program mentors and on comments made by some mentors on their assessment forms.

The data for the period 2008 to 2010 showed little difference either in the scores or patterns of scoring among inexperienced, somewhat experienced, and very experienced mentors. The mean scores from all mentor experience levels increased over time.

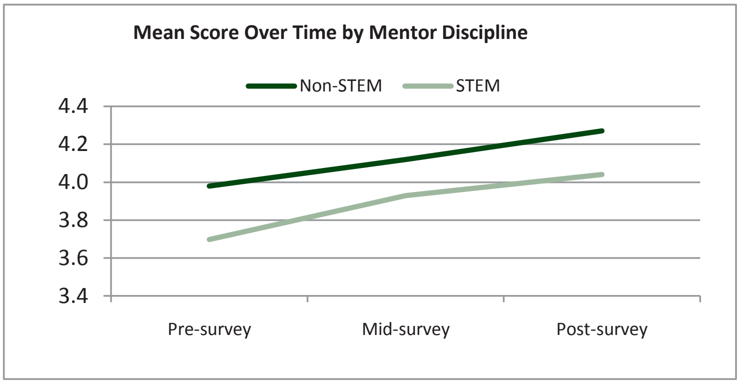
The finding (although not statistically significant) that very experienced mentors scored more conservatively (assigned lower scores) than did inexperienced or somewhat experienced mentors suggest that very experienced mentors may have questioned students more closely in order to learn more about them and make them more self-aware.

We also coded the mentors according to whether they were in a STEM or non-STEM academic discipline. When comparing STEM discipline mentors (n=23) to non-STEM discipline mentors (n=37), there was a modest difference in how mentors, regardless of experience, scored students. On 28 of the assessment's 34 items, non-STEM mentors gave their students slightly higher ratings than did the mentors in STEM disciplines. Ten items were statistically significant at $p < .05$ or better, and two other items were close to significance. STEM discipline mentors rated students higher on three items in the assessment, but the differences were not statistically significant. An overall mean score for all 34 items over time, by mentor's discipline, illustrates the trend. Note that mean differences are small. These findings suggest that the assessment instrument is, in fact, applicable to all disciplines. There were no statistically significant differences between student self-assessment scores for students in STEM disciplines and students in non-STEM disciplines.

Concluding Remarks

The evaluation of the SUNY-Buffalo State summer research program achieved its dual goals of providing a reliable assessment of program impact and helping to advance student

Table 5. Mean Score Over Time by Mentor Discipline



intellectual and professional growth. Our data—including comments from students and mentors on the assessment forms and participants' personal communications with one of the authors (Singer)—confirm that participating in the student research and evaluation processes fostered meaningful reflection by both students and mentors and encouraged frequent, constructive student-mentor dialogue. While the students were the primary beneficiaries of these activities, our data confirm that mentors also gained from the experience. Asking mentors to reflect on how much confidence they had in their assessment scores and asking them to explain why they raised or lowered their ratings from one assessment to the next appeared to help them become more effective mentors, no matter how much prior experience they brought to the program. It also appears that asking students about why they raised or lowered their self-assessments helped them gain a more realistic understanding of their abilities and identify areas they desired to improve.

We are continuing this evaluation and will look for ways to refine and enhance the assessment instruments themselves, while being mindful of the educational opportunities the evaluation and its assessment components provide. We intend to further explore the dynamics of mentoring and its impact on students. Alumni surveys also will be implemented to assess the long-term impact of the research experience on plans for graduate school, employment, and the undergraduate experience in general.

Acknowledgments

The authors gratefully acknowledge Daniel Weiler for the role he has played throughout the years. He has provided encouragement and much feedback as we progressed from the pilot evaluation effort to full-scale implementation. We thank Buffalo State for its continuing strong support for this program. We also thank the students and faculty who participated in Buffalo State's Undergraduate Summer Research Program and embraced our efforts to evaluate the impact of the program.

References

Alexander, Baine, Julie Foertsch, Susan Daffinrud, and Richard Tapia. 2000. "The Spend a Summer with a Scientist Program at Rice: A Study of Program Outcomes and Essential Elements 1991-1997." *CUR Quarterly* 20:127-133.

Bauer, Karen, and Joan Bennett. 2003. "Alumni Perceptions Used to Assess Undergraduate Research Experience." *The Journal of Higher Education* 74:210-230.

Gregerman, Sandra, Jennifer Lerner, William von Hippel, John Jonides, and Biren Nagda. 1998. "Undergraduate Student-Faculty Research Partnerships Affect Student Retention." *The Review of Higher Education* 22:55-72.

Hunter, Anne-Barrie, Sandra L. Laursen, and Elaine Seymour. 2007. "Becoming a Scientist: The Role of Undergraduate Research in Students' Cognitive, Personal, and Professional Development." *Science Education* 91:36-74.

Lopatto, David. 2004. "Survey of Undergraduate Research Experiences (SURE): First Findings." *Cell Biology Education* 3:270-277.

Merkel, Carolyn Ash. 2001. Undergraduate Research at Six Research Universities: A Pilot Study for the Association of American Universities. Pasadena, CA: California Institute of Technology.

Seymour, Elaine, Anne-Barrie Hunter, Sandra L. Laursen, and Tracee DeAntoni. 2004. "Establishing the Benefits of Research Experiences for Undergraduates: First Findings From a Three-year Study." *Science Education* 88:495-594.

Singer, Jill and Daniel Weiler. 2009. "A Longitudinal Student Outcomes Evaluation of the Buffalo State College Summer Undergraduate Research Program." *CUR Quarterly* 29:20-25.

Jill Singer

SUNY-Buffalo State, singerjk@buffalostate.edu

Jill Singer, a professor of geology in the Earth Sciences Department at Buffalo State, is director of Buffalo State's Office of Undergraduate Research. She was a principal investigator on a CUR NSF-Course, Curriculum, and Laboratory Improvement (CCLI) grant that supported eight regional workshops and related follow-up activities designed to help institutions implement or advance campus undergraduate research programs. Her research interests focus on sediment transport and the interactions between river flow and Lake Erie seiches and how these processes inform decisions about restoring the environmental integrity of the Buffalo River.

Bridget Zimmerman is associate for research and evaluation in the research foundation at Buffalo State, where she evaluates numerous state, federal, and local grants. She has more than 25 years of experience in research and program evaluation, including work in the areas of education, healthcare operations, and clinical outcomes.